# Genomic clues of chemosynthetic symbiosis from deep-sea seep-living *Lamellibrachia*

Yuanning Li1,2\*, Michael G. Tassia1, Damien S. Waits1, Viktoria E. Bogantes1, Kyle T. David1, Kenneth M. Halanych1

1 Department of Biological Sciences & Molette Biology Laboratory for Environmental and Climate Change Studies, Auburn University

2 Department of Ecology and Evolutionary Biology, Yale University

\* Corresponding author, [yuanning.li@yale.edu](mailto:yuanning.li@yale.edu)

## Abstract

Since the discovery of chemosynthetic symbiosis in hydrothermal vents and cold seeps, the genetic mechanism(s) of how organisms maintain the unique host-symbiont associations at the molecular level is still mostly unknown. The gutless tubeworm *Lamellibrachia* represents one of the longest-lived invertebrates and is widespread in many deep-sea cold seeps, providing an excellent model to study mechanisms like the association of chemosynthetic symbiosis, adaptation to extreme environments. However, the genomic and molecular evolutionary information of tubeworms is exceptionally scarce. Here, we present and characterize the genome of *Lamellibrachia* *luymesi*. We found genomic evidence that symbionts may compensate for the host’s deficiency in amino acid biosynthesis. Unexpectedly, we found a large expansion of hemoglobin B1 genes - many of which also possess a free cysteine residue which is hypothesized to function in sulfide-binding. Moreover, we found no evidence that sulfide-binding mediated by zinc ions is conserved across tubeworms, suggesting the hemoglobin structure and the sulfide-binding mechanism is potentially more complex than previously thought. Our comparative analyses suggest the Toll-like receptor pathway may be essential to host immunity and tolerance/sensitivity to symbionts and pathogens. Last, we identified and compiled several genes potentially play an important role in longevity. Thus, these results help elucidate previously unknown links and potential mechanisms related to the evolution of holobionts, adaptations to chemosynthetic environments, and likely extend to other chemosynthetic holobionts.

## Introduction



**Figure 1.** *Lamellibrachia* from seep localities in the Gulf of Mexico.

Since their discovery in the 1980s, siboglinid tubeworms at hydrothermal vents and cold seeps have fascinated evolutionary biologists with their unique symbioses between the worms and chemosynthetic endosymbionts. However, the genetic underpinnings of how organisms maintain the host-symbiont associations at the cellular level are still mostly unknown. The vestimentiferan tubeworm *Lamellibrachia* *luymesi* is distributed along the Gulf of Mexico and is one of the most dominant taxa living in deep-sea cold seeps. At some seeps the vestimentiferan tubeworms are so abundant that they form a unique habitat that is attractive for other species (Fig 1), maintaining alpha and beta biodiversity and facilitating adaptive radiations and evolutionary novelties. Siboglinids typically lack a digestive tract and rely on sulfide-oxidizing bacterial symbionts for nutrition. Hosts acquire their symbionts from the surrounding environment and store them in a specialized tissue called the trophosome (Nussbaumer, Fisher, and Bright 2006). Previous studies suggested that vestimentiferan chemoautotrophic symbionts can use both reverse TCA cycle (rTCA) and the Calvin cycle for carbon fixation as a nutrient source for the hosts (Markert et al. 2007, @li2018endosymbiont). However, previous metabolic and physiological studies primarily focus on symbionts associated with tubeworm host lineages, and studies on how exchanges of nutritions and metabolites between two partners from the host’s perspective are still lacking.

One of the key adaptations contributing to the remarkable ability of tubeworms to thrive in chemosynthetic habitats is the ability for siboglinid hemoglobins (Hbs) to bind oxygen and sulfide simultaneously and reversibly at two different sites (Zal et al. 1996). To avoid the toxicity of sulfide while transferring it to their chemoautotrophic symbionts, siboglinids possess a multi hemoglobin system with three different extracellular hemoglobins (Hbs; V1, V2, and C1): two dissolved in the vascular blood, V1 and V2, and one in the coelomic fluid, C1 (Arp and Childress 1981, @zal1996multi). Siboglinid Hbs consist of four heme-containing chains (A1, A2, B1, B2). Sulfur-binding capabilities are hypothesized to be due to free cysteine residues at key positions in Hbs, especially in the A2 and B2 chains (Zal et al. 1997). V1 Hb can form persulfide groups on its four linker chains (L1-L4), a mechanism that can account for the higher sulfide-binding potential of this Hb (Zal et al. 1997). However, another study found no evidence of the free cysteine hypothesis and suggested sulfide-binding affinity was mediated by the zinc moieties bound to amino acid residues at the interface between pairs of A2 chains in *Riftia* (Flores et al. 2005). Thus, it is still not clear which mechanism is primarily responsible for sulfide-binding in siboglinids.

In contrast to the rapidly growing vent-living vestimentiferans (Lutz et al. 1994), seep-living vestimentiferans are thought to be the most long-lived non-colonial marine invertebrates (up to 250 year to grow up to two meters) (Bergquist, Williams, and Fisher 2000). Immunity is a critical evolutionary driver of maintaining symbiosis (Chu and Mazmanian 2013), has important implications in aging (Quesada et al. 2018), and is also potentially important for organisms being held in high densities in chemosynthetic environments. However, very little is known about genetic mechanisms related to immunity in most chemosynthetic host species, and it is far more complex in some groups of invertebrates than traditionally recognized from recent studies (Halanych and Kocot 2014) (Tassia, Whelan, and Halanych 2017). Thus, *Lamellibrachia* provides an excellent model to study immunity function associated with host-symbiont relationships and adaptations to chemosynthetic habitats. However, the genomic and molecular evolution information on extremophile immunity and/or immune tolerance is exceptionally scarce.

Considerable debate persists concerning the evolutionary origins of siboglinids, owing to conflicting theories of its origins from fossil and molecular age estimates. Siboglinids have been claimed to be as old as 430 MYA based on fossil tubes found in the Silurian fossil vent communities (Little 2002), but molecular clock analyses suggest a more recent (50-126 MYA) origin based on mitochondrial *COI* or *16S* gene sequences (Little and Vrijenhoek 2003). Recently, Late Cretaceous *Osedax* fossil traces on reptile falls provided a solid calibration point for the molecular clock of the siboglinids (~100 MYA) (Danise and Higgs 2015). Moreover, a detailed chemical and morphological analyses of tubes from the Figueroa deposits suggested they were made by vestimentiferans, which significantly extends the age of this lineage to a Mesozoic origin than previous molecular clock analyses (Georgieva et al. 2017).

Here, we describe and investigate the genomic, transcriptomic and proteomic sequencing and comparative genomic analysis of tubeworm *Lamellibrachia* *luymesi* collected from the Gulf of Mexico. Our comparative genomic analyses of *Lamellibrachia* provide genomic evidence for genetic pathway and novel candidate genes which may underlie the extraordinary characteristics of tubeworms, including nutrition mode, hemoglobin evolution, immunity function and longevity. Moreover, we also conducted a detailed molecular clock analysis to discern among the hypotheses posed by previous debates concerning the evolutionary origins of siboglinids. Our findings provide hitherto previously unknown evidence of genetic machinery related to tubeworm success in deep-sea chemosynthetic communities, such as vents and seeps.

## Methods

Scripts and data for the analyses are available in a git repository at <https://github.com/yzl0084/Lamellibrachia-genome>.

## NCBI Accession

### Biological materials.

Adult *Lamellibrachia* *luymesi* specimens were collected from seep localities in the Mississippi Canyon at 754 m depth in the Gulf of Mexico (N 28°11.58’, W 89°47.94’), using the R/V Seward Johnson and Johnson Sea Link in October 2009. All samples were frozen at 80˚C following collection.

### Genome sequencing and assembly.

Vestimentum tissue was dissected from one individual, and high molecular weight genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer’s protocols. Sequencing of six paired-end or mate-pair genomic DNA libraries with insert sizes ranging from 180 bp to 7 kb were performed by The Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama on an Illumina HiSeq 2000 platform (see details in Table S1). Paired-end libraries (180 bp, 400 bp, 750 bp) were prepared using the 125 bp TrueSeq protocols, and mate-pair libraries (3-5 kbp, 5-7 kbp) were generated using the Illumina Nextera Mate Pair Library Kit followed by size selection. Moreover, a 10X sequencing library was constructed using the 10X Chromium protocol (10X genomics) at the Hudson Alpha Institute. The finished library was sequenced on an Illumina HiSeqX platform, using paired 151 bp reads with a single 8 bp index read.

Our genome assembly workflow is shown in Fig. S1. The paired-end and 10X raw reads were checked using FastQC v0.11.5 (Andrews and others 2010) and quality filtered (Q score >30) using Trimmomatic v0.36 (Bolger, Lohse, and Usadel 2014). The estimation of genome size, level of heterozygosity, and repeat contents of the *Lamellibrachia* genome was determined by analyzing the kmer histograms generated from the paired-end libraries using Jellyfish v2.2.3 (Marçais and Kingsford 2011) and GenomeScope (Vurture et al. 2017) (Fig. S2). The mate-pair reads were trimmed and sorted using NxTrim v0.3.1 (O’Connell et al. 2015) which can recognize and trim the artificial Nextera mate-pair circulation adapters. Only reads from category “mp” (true mate-pair reads) and “unknown” (mostly large insert size reads) were used for downstream scaffolding analysis. Reads from “pe” (paired-end reads) and “se” (single ends) categories were discarded.

Given the high heterozygosity of the *Lamellibrachia* genome, all reads were assembled using Platanus v1.2.4 (Kajitani et al. 2014) with a kmer size of 32. Scaffolding was conducted by mapping Illumina paired-end and mate-pair reads to contigs generated by Platanus using SSPACE v3.0 (Boetzer and Pirovano 2014). Gaps in the scaffolds were then filled with GapCloser v1.12 (Luo et al. 2012). Redundant allele scaffolds were further removed using Redundans v0.13c with default settings (Pryszcz and Gabaldón 2016). Genome assembly quality was assessed using QUAST v3.1 (Gurevich et al. 2013). Completeness of obtained genome was assessed using BUSCO v3(Waterhouse et al. 2017) with Metazoa\_odb9 database (978 Busco genes). We also assembled the genome using Supernova 1.2.0 (Weisenfeld et al. 2017) with 10X data alone by default settings, but the genome quality and completeness derived from Supernova is worse than the Platanus assembly (Fig. S1).

### Transcriptome assembly and analysis

Total RNA was extracted from the the plume, vestimentum and trophosome tissue from the same individual of *Lamellibrachia* using Trizol. RNA-Seq of adult tissues from plume, vestimentum, and trophosome was performed using Illumina HiSeq 2000 platform at Hudson Alpha as described above. After quality checking and trimming of raw sequencing reads, transcripts were assembled de novo with Trinity. Transcript isoforms with high similarity (≥ 95%) were removed with CD-HIT-EST v4.7 (Li and Godzik 2006). Transcript abundance was estimated with Bowtie v2.2.9 (Langmead and Salzberg 2012) and RSEM v1.2.26 (Li and Dewey 2011) by mapping reads back to the transcriptomic assembly based on transcripts per million (TPM). A tissue-specifically expressed gene was defined as a gene that had over 75% of the total transcripts in a particular tissue based on TPM (Albertin et al. 2015).

### Proteomics characterization.

Proteomic analyses of *Lamellibrachia* trophosome tissue were performed by Proteomics & Metabolomics Facility at Colorado State University. Briefly, samples were transferred to 5mL tubes appropriate for use in the Bullet Blender Storm 5 (Next Advanced). Freshly cleaned, 3.2mm stainless steel beads were added at an approximate 1:1 sample: bead volume ratio (1-5 beads). The tissue/bead mixture was then combined with 1X PBS (Hyclone) spiked with 1X Halt Protease Inhibitor cocktail (Thermo Scientific) for an approximate 2:1 liquid: solid ratio (150-600 μl). Tissue homogenization was achieved using speed setting 12 for 5 minutes. The stainless steel beads were removed and 0.1mm glass beads added at approximately 1:1 sample:bead ratio. Further homogenization was achieved at speed 8 for 3 minutes. Homogenate was then transferred to 1.5mL microcentrifuge tubes and subjected to cup horn sonication (Amplitude 70, 10s pulse, 20s off, 7 minutes total exposure time; Qsonica) while suspended in ice water. Insoluble material was pelleted by brief centrifugation at 3000xg. The resulting supernatant was subjected to protein quantification by the Pierce BCA Protein Assay Kit (ThermoFisher-Pierce) following manufacturer’s instructions. Absorbance of reacted samples was measured at 550nm and protein quantification stemmed from a quadratic fit of the Bovine Serum Albumin (BSA) standard curve.

50 μg total protein was aliquoted from each sample and processed for in-solution trypsin digestion as previously described (Schauer et al. 2013). A total of 0.5μg of peptides were then purified and concentrated using an on-line enrichment column (Waters Symmetry Trap C18 100Å, 5μm, 180 μm ID x 20mm column). Subsequent chromatographic separation was performed on a reverse phase nanospray column (Waters, Peptide BEH C18; 1.7μm, 75 μm ID x 150mm column, 45°C) using a 90 minute gradient: 5%-30% buffer B over 85 minutes followed by 30%-45%B over 5 minutes (0.1% formic acid in ACN) at a flow rate of 350 nanoliters/min. Peptides were eluted directly into the mass spectrometer (Orbitrap Velos Pro, Thermo Scientific) equipped with a Nanospray Flex ion source (Thermo Scientific) and spectra were collected over a m/z range of 400–2000 under positive mode ionization. Ions with charge state +2 or +3 were accepted for MS/MS using a dynamic exclusion limit of 2 MS/MS spectra of a given m/z value for 30 s (exclusion duration of 90 s). The instrument was operated in FT mode for MS detection (resolution of 60,000) and ion trap mode for MS/MS detection with a normalized collision energy set to 35%. Compound lists of the resulting spectra were generated using Xcalibur 3.0 software (Thermo Scientific) with a S/N threshold of 1.5 and 1 scan/group.

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by ProteoWizard MsConvert (version 3.0). Spectra from all samples were searched using Mascot (Matrix Science, London, UK; version 2.6.0) against gene models of *Lamellibrachia* host and symbiont genomes (derived from (Yuanning Li, Liles, and Halanych 2018)) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 20 PPM. Oxidation of methionine and carbamidomethyl of cysteine were specified in Mascot as variable modifications. Search results from all samples were imported and combined using the probabilistic protein identification algorithms (Keller et al. 2002) implemented in the Scaffold software (version Scaffold\_4.8.4, Proteome Software Inc., Portland, OR) (Searle, Turner, and Nesvizhskii 2008). Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al. 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

### Genome annotation.

Our genome annotation workflow is shown in Fig. S3. Gene models of the *Lamellibrachia* genome were constructed following the Funannotate pipeline 1.3.0 (<https://github.com/nextgenusfs/funannotate>). Briefly, repetitive regions in the *Lamellibrachia* genome assembly were identified using RepeatModeler v1.0.8 (Smit and Hubley 2008) and were subsequently soft-masked using RepeatMasker v4.0.6 (Chen 2004). RNA-Seq data from different tissue were leveraged to improve the accuracy of gene prediction. RNA-Seq data were assembled *de* *novo* into a single transcriptome using Trinity v2.4.0 (Haas et al. 2013) and HISAT 2.1.0 (Kim, Langmead, and Salzberg 2015) was used to align RNA-Seq reads to the *Lamellibrachia* genome assembly. the transcriptome assembly was then passed to the PASA pipeline v2.3.3 (Haas et al. 2003) to identify high-quality gene models. The aligned RNA-Seq data was used to train the *ab* *initio* gene predictor using AUGUSTUS v3.3 (Stanke et al. 2006). Protein alignments from the SwissProt database to *Lamellibrachia* assembly was generated using exonerate (Slater and Birney 2005) and Trinity/PASA transcripts were aligned to the genome using Minimap2 v2.1 (H. Li 2018). The tRNA genes were identified using tRNAscan-SE v1.3.1 (Lowe and Eddy 1997). Finally, EvidenceModeler 1.1.0 (Haas et al. 2008) was used to combine all the evidence of gene prediction from protein alignments, transcript alignments, and *ab* *initio* predictions to construct high-quality consensus gene models. Functional annotations of predicted gene models was performed using several curated databases: KEGG orthology was assigned using the KEGG Automatic Annotation server (Moriya et al. 2007). Gene models were further annotated with domain structure and protein identity by InterProScan (Zdobnov and Apweiler 2001) and SwissProt database, respectively. Secreted proteins were predicted using SignalP (Petersen et al. 2011) and Phobius (Käll, Krogh, and Sonnhammer 2007) using InterProScan.

### Gene family analysis

Following all-to-all Diamond v1.0 (Buchfink, Xie, and Huson 2014) BLASTP searches against 22 selected lophotrochozoan proteomes (Table S2), orthology groups (OGs) were identified using Orthofinder with a default inflation parameter (I=1.5). Gene ontology annotation was performed using PANTHER v13.1 (Mi et al. 2016) with the PANTHER HMM scoring tool (pantherScore2.pl). Gene family expansion and contraction was estimated using CAFÉ v2.1 (De Bie et al. 2006). For each gene family, CAFÉ generated a family-wide P value, with a significant *P* value indicating a possible gene-family expansion or contraction event as suggested in the manual. Significantly expanded gene families (*p* < 0.05)were then classified by InterProscan.

In addition to the above gene family expansion or contraction analysis, the protein-coding sequences were used for discovery of genes under positive selection in from *L*. *luymesi* in comparison with selected lophotrochozoan species (including *Capitella* *teleta*, *Lottia* *gigantea*, *Helobdella* *robusta*, *Octopus* *bimaculoides*, *Crassostrea* *virginica*, *Notospermus* *geniculatus* and *Phoronis* *australis*) using PosiGene, which is an automated pipeline for genome-wide detection of positively selected genes , with default settings (p < 0.05). In this positive selection analysis through the PosiGene pipeline, *L*. *luymesi* was used as the reference and anchored species for ortholog assignment of each OG.

### Manual annotation of gene families with potential interest.

In addition to the automatic annotation process mentioned above, we manually annotated genes that are general interest, such as hemoglobin gene families, genes related to amino acid synthesize, immunity function or longevity. These genes were specifically selected *a* *priori* based on our experience after a revision of the available publications in the field. For hemoglobin genes, datasets used for Hb gene family analyses, identified homologs, and their respective accession numbers are available in Fig. S4. Hbs and linker sequences of interest were obtained from *Lamellibrachia* genome and assembled siboglinid transcriptomes (see molecular clock analysis below) via diamond blastp (evalue cutoff 1E-5) by utilizing Hb and linker sequences acquired from GenBank of siboglinids as databases. Sequences with best hits to target proteins were annotated for protein domain architecture using the Pfam databases included in InterProscan. After manual removal of redundant and incorrect sequences (e.g., sequences are too short or lack of globin domain), we used MAFFT 7.2.15 (Katoh and Standley 2013) to align Hb amino acid sequences. Maximum likelihood analyses were performed in IQTree under the best-fitting models for associated partition schemes determined by Modelfinder implemented in IQTree with ultrafast bootstrapping of 1000 replicates.

Our analysis for immunity-related genes largely follows the same workflow as used for Hbs, but instead were processed through the Extract\_Homologs2 script used in (Tassia, Whelan, and Halanych 2017) (available at <https://github.com/mtassia/Homolog_identification>). We examined gene families of C-type lectins, Chitin binding, immunoglobulins, NACHT containing proteins and TLRs. We only included identification of TLR signaling components here since other immunity related genes did not reveal any notable evolutionary patterns across the lophotrochozoan datasets in our analysis (results from immunity related genes list above is available at <https://github.com/yzl0084/Lamellibrachia-genome/tree/master/Immunity_related/Orthofinder_Archive>). Target proteins of TLR signaling pathway are available in Table S3. Importantly, the Extract\_Homologs2 script identifies unique protein sequences within an animo acid dataset which fall within user-defined domain architecture criteria (Table S4). Due to this stringency, the pipeline identifies the complement of unique proteins for any target family, not the number of synonymous genes encoded in a genome. Major output files produced by this pipeline are available as Supplementary Data 1. Full amino acid sequences for TLRs were placed in a phylogenetic context using the bioinformatic workflow delineated above for Hbs.

Genes related to amino acids synthesis from *Lamellibrachia*, *Lamellibrachia* symbionts, and *Capitella* *teleta* genomes were performed by using the KEGG2 KAAS genome annotation web server and then visualized by the KEGG Mapper Reconstruct Pathway too. A BLASTp search of protein sequences from the genome annotation were queried against the Swiss-Prot database was used to search and supplement for proteins that were missing in the visualized KEGG pathway (Table S#).

### Phylogenomics and Molecular clock analysis

Analysis of siboglinid phylogeny was conducted utilizing from publically available transcriptomic datasets (n=16) in conjunction with our newly generated *Lamellibrachia* proteome (Table S5). Sequence assembly, annotation, homology evaluation, gene tree construction, parsing of genes trees to OGs, and supermatrix construction were conducted with Agalma (Dunn, Howison, and Zapata 2013). The final analyses presented here contained 13 siboglinids and four outgroups based on current understanding of annelid phylogeny. Maximum likelihood phylogenetic reconstructions were performed in IQTree under the best-fitting models for associated partition schemes determined by Modelfinder implemented in IQTree 1.6.3 (Nguyen et al. 2014) with ultrafast bootstrapping of 1000 replicates (Fig. S5).

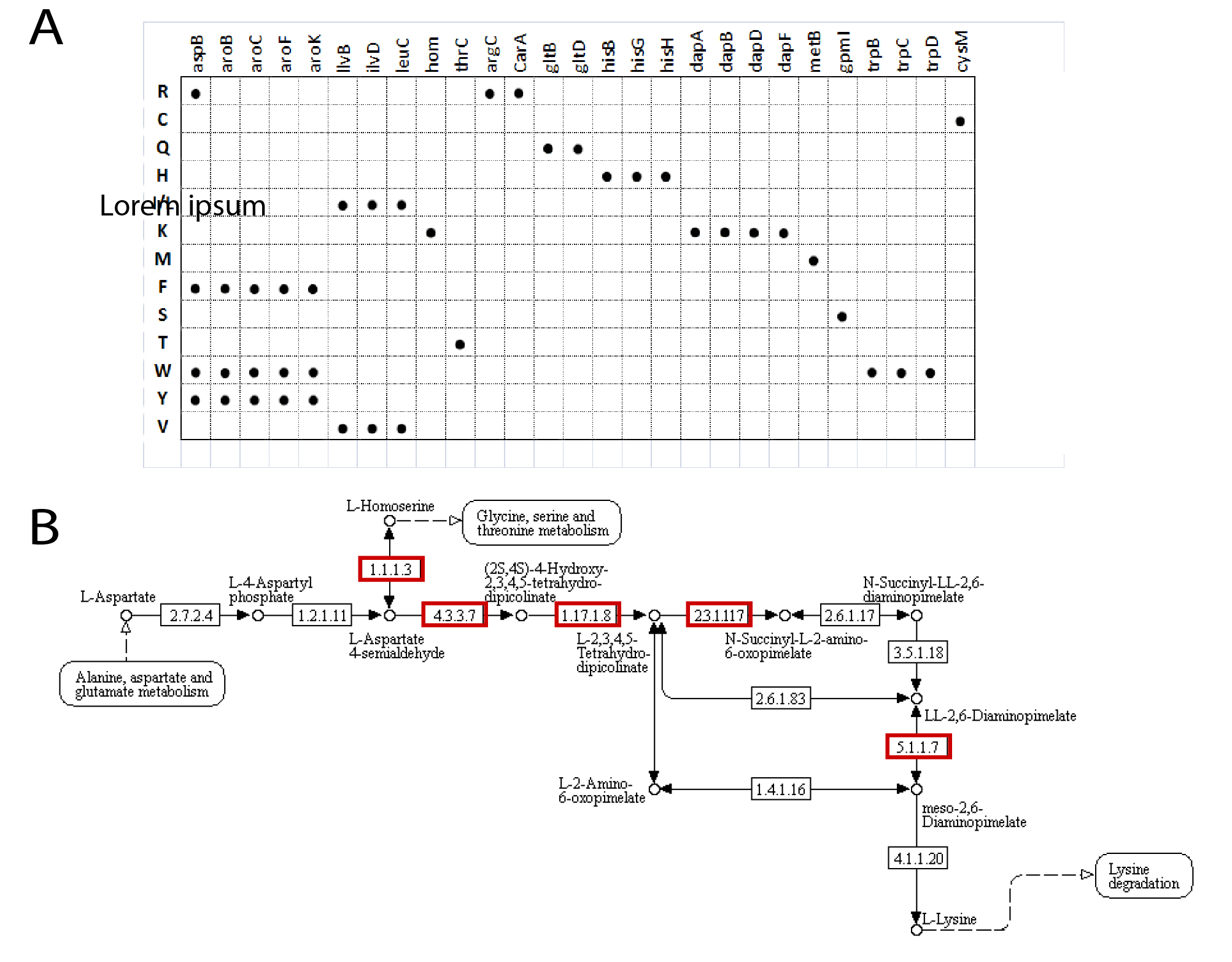
For the molecular clock analysis, a relaxed molecular clock with a lognormal distribution and a Yule tree model was used in BEAST 2 v2.5.1 (Bouckaert et al. 2014). A calibration was placed on the node representing the most recent common ancestor (MRCA) of four *Osedax* species using a normal distribution with a mean of 100 MYA and a standard deviation of 10 following the findings of (Danise and Higgs 2015). Another calibration was placed on the node of MRCA of Serpulida and Sabellida using a normal distribution with a mean of 267 MYA (Sanfilippo et al. 2017). Molecular clock analyses with BEAST 2 consisted of two independent runs with 1 million MCMC generations sampled every 1000 generations. Convergence was checked and confirmed by comparing trace plots in Tracer ensuring the effective sample size of each parameter was greater than 100 and that stationarity appeared to have been achieved. Log and tree files were combined using Logcombiner. A maximum clade credibility tree with mean heights was calculated using TreeAnnotater. The resulting time-calibrated tree was plotted using R packages, including phyloch, strap, and OutbreakTools. Bayesian inference using a molecular clock resulted in identical branching patterns as analysis with IQTree.

## Results and Discussion

### Genome features

The genome of *Lamellibrachia* was sequenced using a combination of Illumina paired-end, mate-pair, and 10x techniques. Results from high-throughput sequencing and genome assembly for *Lamellibrachia* are presented in Table S6 with at least 300 fold coverage using a combination of Illumina paired-end, mate-pair and 10X genomic sequencing with a variety of insert sizes. The haploid genome assembly size is ~687 Mb, similar to the result of genome size estimation using GenomeScope (~646 Mb) (Fig. S2). The N50 value of the assembled scaffolds and contigs is 373 Kb and 24 Kb, respectively. Although N50 lengths and assembly quality of *Lamellibrachia* are comparable to those of other annelids (e.g. *Capitella* *teleta*, *Helobdella* *robusta*) (Table S6), the overall genome completeness measured by BUSCO (~ 95%) is one of the highest among other lophotrochozoans (Table S7), indicating the completeness of the genome assembly. With the support of RNA-seq data from different tissues, we estimated *Lamellibrachia* genome contains 38,998 gene models. The genome also exhibits a similar level of heterozygosity (0.6%) and repetitive contents (36.92%), compared to other lophotrochozoan genomes. The repeat contents of *Lamellibrachia* genome is shown in Table S8. We found 94 OGs appear to have undergone a genomic expansion compared to other lophotrochozoan genomes (Table S9).

### Nutritional adaptations to a symbiotic life stage



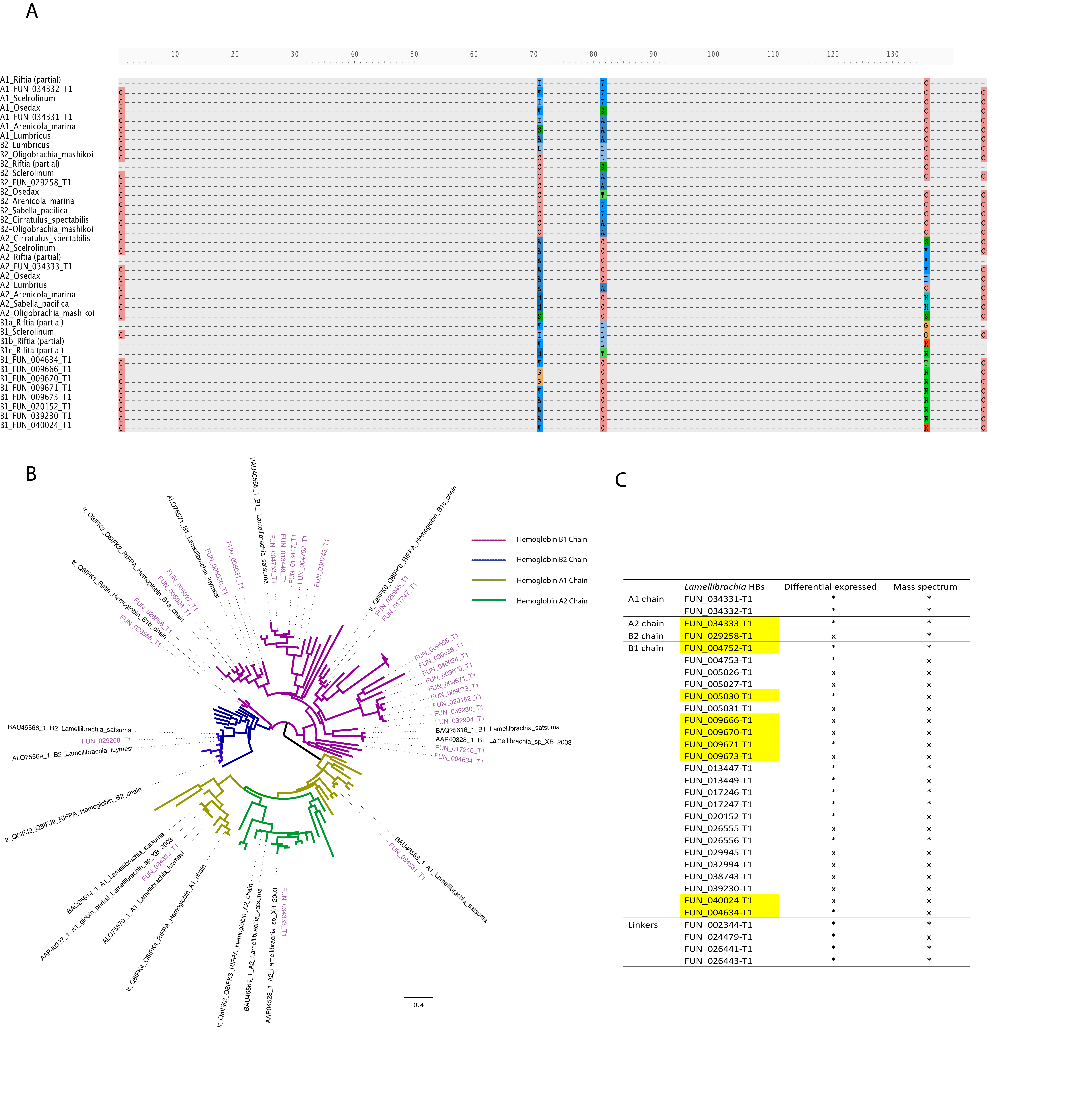
**Figure 2.** *Lamellibrachia* lacks many genes related to amino acid biosynthesis. (A) Genes present in *Lamellibrachia* symbiont and *Capitella* *teleta* genomes but absent in *Lamellibrachia* genome. (B) Example of Lysine biosynthesis pathway in *Lamellibrachia*. Red boxes represent missing genes in the pathway.

Adult *Lamellibrachia* worms lack a digestive tract and rely solely on endosymbionts for nutrition. Noticeably, only 57 genes associated with amino acid biosynthesis were found in the *Lamellibrachia* host genome, of which eight were identified as proteins (Supplementary Data 2), whereas 90 genes were identified in *Capitella* genome (Fig. 2A). In contrast, the genome of the *Lamellibrachia* symbiont (110 genes) contains an essentially complete gene sets for the biosynthesis of all 20 proteinogenic amino acids and of 11 vitamins/cofactors. The lack of 30 genes related amino acid synthesis in *Lamellibrachia* compared to *Capitella* genome are involved in the biosynthesis of 13 out of 20 amino acids (e.g. five key enzymes in the Lysine biosynthesis pathway Fig. 2B). As amino acids are indispensable for protein biosynthesis in the host, the sparse presence of amino acid synthesis-related genes in *Lamellibrachia* genome may indicate that the host depends on its symbionts for supply with amino acids and cofactors. Moreover, we found a large gene expansion of nutrient uptake ABC transport protein-coding genes in *Lamellibrachia* compared with other lophotrochozoans (Table S9). These findings are consistent with previous biochemical analyses which suggest that *Riftia* is also dependent on its bacterial symbiont for the biosynthesis of polyamines that are important for host metabolism and physiology (Minic and Hervé 2003). Interestingly, a similar mechanism of overcoming host-deficiency in amino acid synthesis has also been suggested in thiotrophic vesicomyid symbioses of *Calyptogena* *magnifia* (Newton, Girguis, and Cavanaugh 2008) and vent mussel symbioses of *Bathymodiolus* *azoricus* (Ponnudurai et al. 2017). These results emphasize the potential that host-deficiency in amino acid syntheses and exchanges of metabolites between two partners might to be a common strategy for chemosynthetic symbioses.

In addition to the immediate release of fixed carbon and provision of amino acids by the symbionts, we have found proteomic evidence of a second possible nutritional mode - digestion of symbionts directly by the host, as showed by the detection of abundant host-derived digestive enzymes in the trophosome tissue (Table S11). Although previous observations indicated that symbiont could be digested by *Riftia* (Bright, Keckeis, and Fisher 2000), the direct evidence and underlying genetic mechanisms related to symbiont digestion were not characterized. We identified fifteen the host proteins related to lysosomal proteases that were both highly expressed and detected as proteins in the trophosome tissue, such as *Saposin* and multi copies of *Cathepsin* (Table S11). Lysosomes contain an array of digestive enzymes used for protein degradation in eukaryotic cells and are thought to play an essential role in symbiont digestion of *Bathymodilus*. *azoricus* (Ponnudurai et al. 2017). We additionally identified 19 major proteasome components as proteins in the trophosome tissue, indicating a potential role in protein degradation of symbiont digestion. Thus, our results suggest that lysosomal host proteases and proteasome may facilitate the degradation of symbiont proteins during symbiont digestion, and may play a role maintaining population load of symbionts within trophosome, or possibly both.

Because the trophosome tissue of *Lamellibrachia* contains sulfur-oxidizing symbionts, we also characterized ~ 200 bacterial proteins present in the trophosome tissue to gain insights into the symbiotic relationship between the bacteria and their host. Key enzymatic genes, RubisCO, and ATP citrate lyase (ACL) type II associated with these carbon fixation cycles, were identified as proteins in *Lamellibrachia* symbionts (Table S10). Our results further corroborate the idea that the presence of rTCA in addition to CBB pathways for carbon fixation might be common in all vent-dwelling and seep-living vestimentiferan endosymbionts. We also identified several key components related to sulfide and nitrogen metabolic pathways, and the results are largely consistent with previous analyses (Markert et al. 2007, @li2018endosymbiont).

### Hemoglobin evolution



**Figure 3.** Hemoglobin gene family evolution in *Lamellibrachia* genome. (A) Partial alignment of sampled siboglinid Hb subunit A1, A2, B1, B2 sequences. Only positions contain free cysteines or cysteine residues are involved in disulfide bridge were shown in the alignment. (B) Gene tree of siboglinid Hb subunit A1, A2, B1 and B2. Only *Lamellibrachia* Hb sequences (from NCBI or this study) were labeled at the tips. (C) *Lamellibrachia* Hb sequences identified that are highly expressed in the trophosome tissue or as proteins from proteomic data.

Partial alignment and gene tree of siboglinid Hbs are shown in Figures 3A, 3B, S5. The expression of Hb subunit A1, A2, B1, B2, and linker L1-L4 were among the highest in the trophosome compared to the plume and vestimentum, and we were able to identify most of them in the mass spectrometry (Fig. 3C). Consistent with previous studies (Zal et al. 1996), a single copy of A2 and B2 Hb was identified in all siboglinids to possess conserved-free cysteines (i.e., cysteine residues not involved in disulfide bridges) at position 77 and 67, respectively. With exception of A2 and B2 Hb globins in *Lumbricus* *terrestris*, homologous cysteine residues were identified in *Cirratulus* *spectabilis*, *Sabella* *pacifica*, and *Sternapsis* sp. from sulfide-free environments and *Arenicola* *marina* living in sulfide-rich environments (Fig. 3A). These results support the hypothesis that free cysteine residues in A2 and B2 globins were present and conserved prior to the radiation of the Siboglinidae and potentially involved in H2S detoxification process (Bailly et al. 2002).

Surprisingly, we found a significant expansion of B1 globins in *Lamellibrachia* genome which possesses 25 copies (Fig. 3B). *Riftia* may also contain a comparable B1 globin expansion since three B1 globins were identified using a set of degenerate PCR primers in the previous study (Bailly et al. 2002). Noticeably, we found that eight copies of *Lamellibrachia* B1 sequences also contains a free cysteine at position 77, same position as free cysteine in A2 globins. Five out the eight copies were highly expressed in the trophosome, and one copy was identified at the protein level. Unlike A2 and B2 globins, it was long thought that B1 globins could only bind O2 and lacks the capacity to bind sulfide. Although the presence of free cysteine residues is not proof of sulfide binding also occurs in B1 globins, it is at least possible that the hemoglobin structure and sulfide binding function is more complex than previously thought due to unexpected a large B1 expansion and high expression of *Lamellibrachia* B1 sequences (Fig. 3C).

Instead of free cysteines, another hypothesis suggested that the H2S binding affinity has been mediated by the zinc moieties bound to amino acid residues at the interface between pairs of A2 chains in *Riftia* (Flores et al. 2005). Figure S6 shows an alignment of amino acid sequences of A2 globins from Siboglinidae and other annelids. Amino acids involved in the binding to of zinc ions in *Riftia* A2 globin pairs are highlighted (Fig. S6). Each A2 chain posesses one complete Zn2+-binding site and shares two others with adjoining A2 subunits (Flores et al. 2005). The Zn2+-binding site contained within A2 chain is composed of three His residues (B12, B16, and G9) (Flores et al. 2005). However, none of these sites are conserved across siboglinids, or even in vestimentiferans (Fig. S6). For example, B16 His site is replaced by a variety of amino acids in seep-living vestimentiferans. Thus, it is less likely that the zinc sulfide binding mechanism is a conserved function in siboglinids, and might only be a derived function just for vent-living vestimentiferans.

### Immunity function

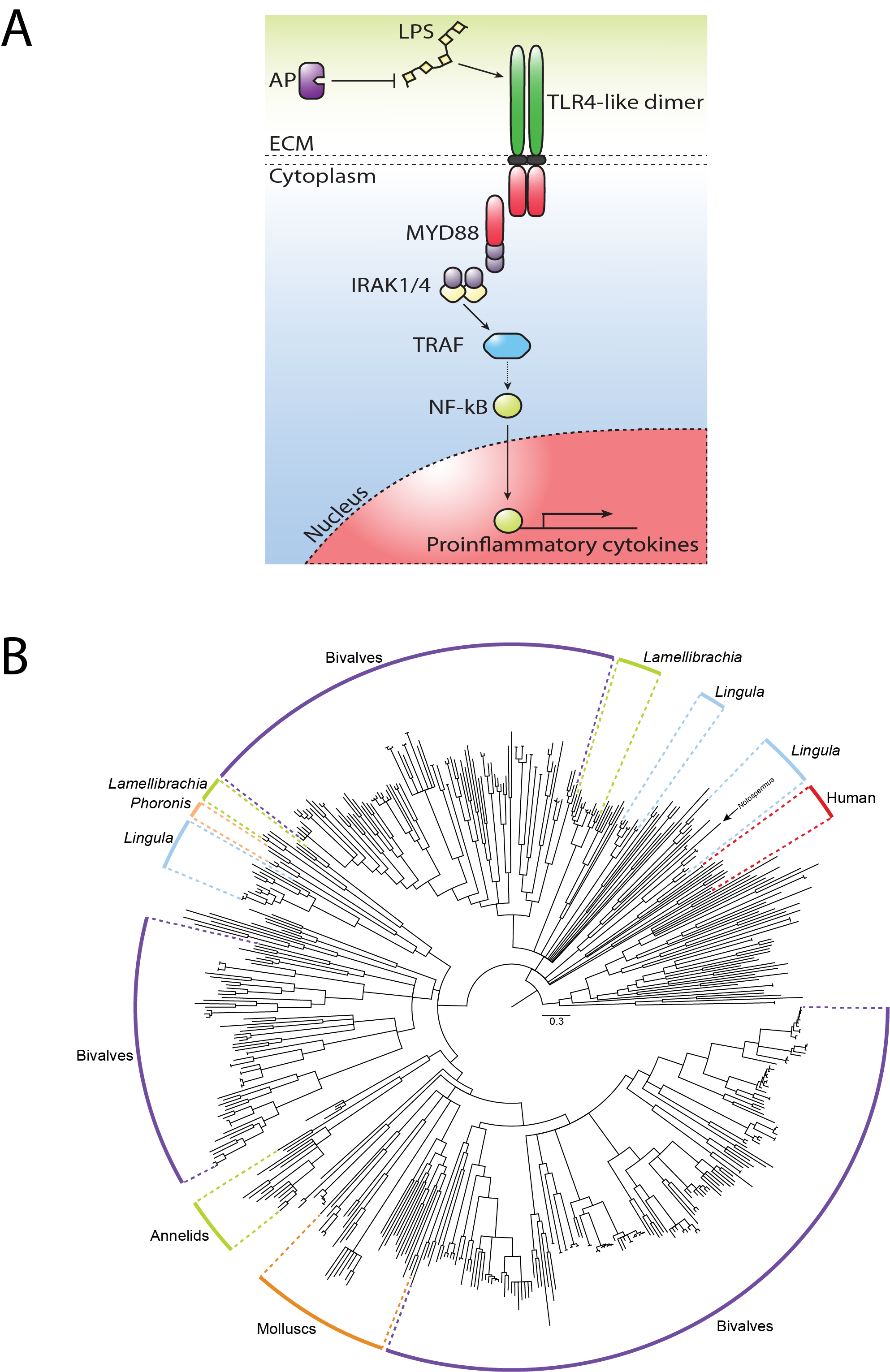


Figure 4. Putative Toll-like receptor pathway in *Lamellibrachia* genome. (A) Putative TLR4-like pathway in *Lamellibrachia* might be essential for immunity and response to symbionts and pathogens. (B) Toll-like receptor gene trees from selected lophotrochozoan genomes.

The immune function between hosts and symbionts is a key evolutionary driver that is under strong selective pressure and has potential implications in aging (REFERENCE). However, the genetic machinery and functionality of the immune system in most chemosynthetic symbioses has not been extensively characterized. We found several genes related to immune function are expanded in *Lamellibrachia* compared with other lophotrochozoans (Table S8). Toll-like receptor (TLRs) provides a core cellular and molecular interface between invading pathogens and recognition of host-microbial symbiosis (Chu and Mazmanian 2013) (Fig. 4A). Consistent with previous analyses (Luo et al. 2018), we found that TLR genes were individually expanded in among lophotrochozoan lineages, especially in bivalves (Fig. 4B). Noticeably, we found several lineage-specific expansion of putative TLR genes in *Lamellibrachia* genome with 33 copies, whereas only five copies were found in *Capitella* genome, suggesting TLR genes in *Lamellibrachia* may play some discernably important role in *Lamellibrachia* independent of their role(s) in *Capitella*. Whereas our phylogenetic reconstruction of TLR relationships resolves major orthology groups among lophotrochozoan TLRs, showing yet another example of lineage-driven expansions among invertebrate clades (Tassia et al. 2017), our taxon sampling does not provide the appropriate phylogenetic distribution to resolve relationships between lophotrochozoan TLRs and vertebrate TLRs. To resolve pan-metazoan TLR orthology groups, deep taxonomic sampling and time/resource-intensive phylogenetic inference protocols will be required.

Based on KEGG mapping, a substantial subset of all *Lamellibrachia* TLR proteins best identify as TLR4 by primary sequence identity (and among all other TLRs in the KEGG GENES database). In mammals, TLR4 recognizes and binds lipopolysaccharide (LPS; a major cell-membrane component of Gram-negative bacteria, including tubeworm symbionts). LPS-bound TLR4 then initiates a signal-transduction pathway that activates NF-kB, a transcription factor that promotes the expression of pro-inflammatory cytokines (Park and Lee 2013). Because orthology is uncertain (Fig. 4), the TLRs identified by KEGG will be referred to as TLR4-like. *Lamellibrachia* encodes seven TLR4-like proteins, which is in contrast to *Capitella* and *Helobdella*’s single TLR4-like protein – lending to the potential for increased sensitivity to Gram-negative bacteria in *Lamellibrachia* (Cavanaugh et al. 1981). Interestingly, we also found genomic expansions of tumor necrosis factor receptors (TNFRs) and TNFR-associated factors (TRAFs) (Table S7). TNFRs and TRAFs play vital roles in activation and the downstream responses of NF-kB, further supporting a specialized role for TLR4-like signaling in *Lamellibrachia*. We additionally investigated components of the innate immunity’s RIG-1-like receptor signaling pathway (which recognizes virus-derived nucleotide present in the cytoplasm) and found no indication of gene expansion.

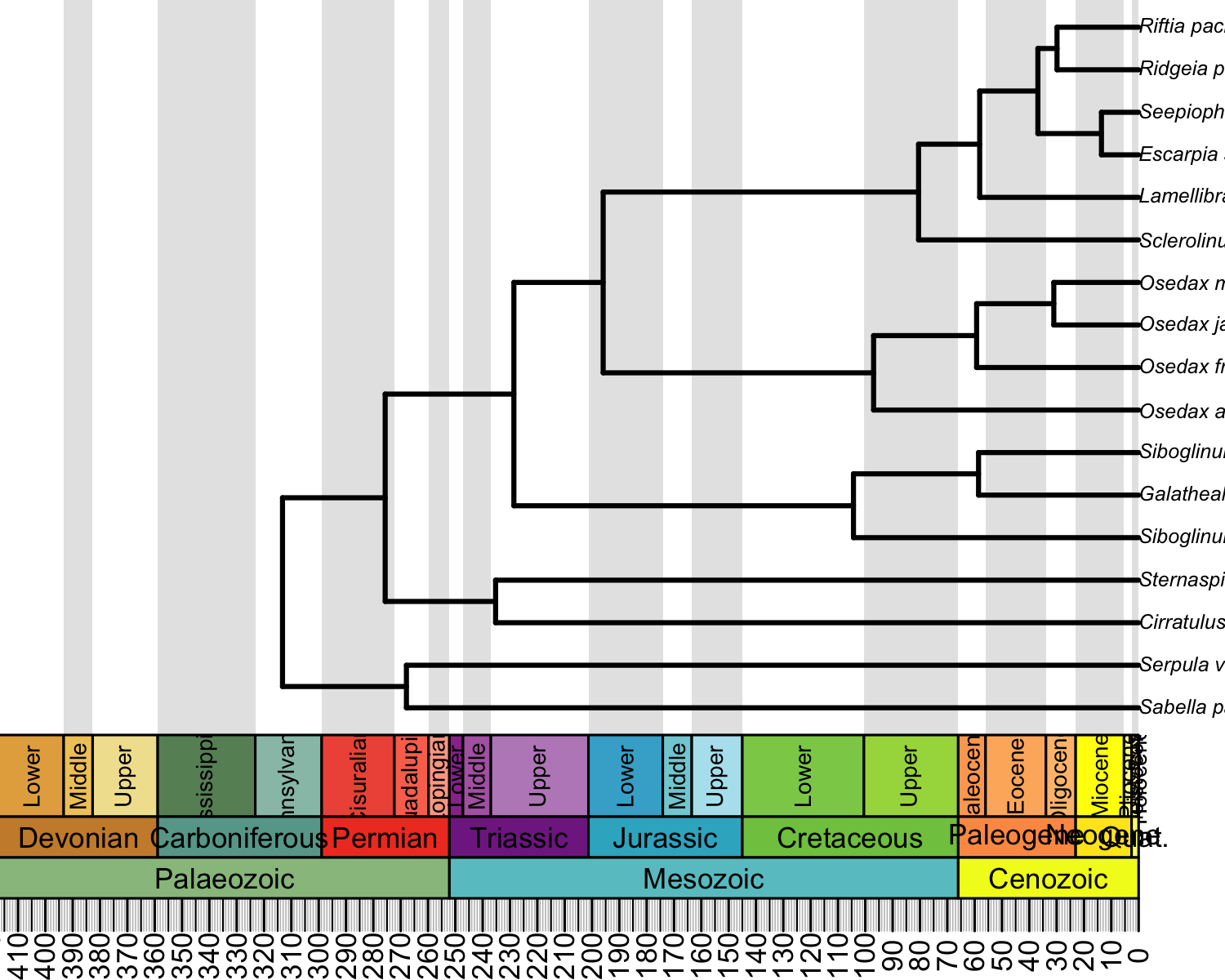
The initial physical encounter between tubeworms and symbionts occurs in an extracellular mucus secreted by pyriform glands by newly settled larvae (Nussbaumer, Fisher, and Bright 2006). Within these mucus matrices, symbionts can attach to the host using extracellular components secreted from symbionts, such as LPS. The symbiont’s colonization process terminates with massive apoptosis of skin tissue as symbionts travel from host epidermal cells into trophosome (Nussbaumer, Fisher, and Bright 2006). This result is consistent with that recognition of lipopolysaccharide (LPS) by TLR4 can result in the induction of signaling cascades that lead to activation of NF-kb and the production of proinflammatory cytokines (Chu and Mazmanian 2013). Although it is still not clear how the host distinguishes between symbionts and pathogens in most symbioses, alkaline phosphatase has been shown to be involved in the maintenance of homeostasis of commensal bacteria in the squids, mouse, and zebrafish (Bates et al. 2007). The commensal bacterially-derived LPS signaling via TLR4 yields an upregulation of intestinal alkaline phosphatase and prevents inflammatory responses to the resident microbiota. Importantly, we also identified eight copies of alkaline phosphatase in *Lamellibrachia*, whereas only one copy was found in *Capitella* and *Helobdella* genomes, further supporting a potential mechanism of tolerating Gram-negative bacteria and facilitating symbiotic colonization. Interestingly, alkaline phosphatase activities were significantly higher in the *Riftia* symbiont recombinants than the *Escherichia* *coli* controls (Hughes, Felbeck, and Stein 1997). Thus, although further analysis is warranted to investigate these hypotheses mechanistically, a TLR4-like signaling pathway may be central for host immunity and in distinguishing between symbionts and pathogens (Fig. 4A).

In addition to TLRs and their cohort of signaling intermediates, we also found evidence of the genomic expansion of NLRP family, which plays a key role in the inflammasome, an innate immunity protein complex that recognizes infectious pathogens and regulates the activation of inflammatory caspases. We found large expansions of several members of Sushi domain-containing proteins that could be involved recognition and adhesion between hosts and symbionts (REFERENCE).

### Aging

The extreme longevity of seep-living vestimentiferans is expected to involve genes affecting aging. However, genetic mechanisms involved in aging on *Lamellibrachia* are incredibly scarce. Besides genomic expansions of genes related to innate immunity mentioned above, our analysis also highlighted several genes in *Lamellibrachia* that may play a direct role in aging. Overcome of oxidative stress is thought to be a hallmark of aging (Liguori et al. 2018). We found genomic expansions of Mn-superoxide dismutase (Sod2) and CuZn-superoxide dismutase (Sod1) genes in *Lamellibrachia* genome compared to other lophotrochozoans (Fig. S7). Most lophotrochozoan genomes contain one or two copies of Sod1 and Sod2 , whereas we found five copies of each gene (Fig. S7). Importantly, we are be able to identify four of five Sod2 genes as proteins. The Sods have important function role in cells to protect against oxidative damage induced by metabolism, and Sod2 has drawn particular interest because it is located within mitochondria and appears to be involved in other processes, including tumor suppression and cellular differentiation in mammals (Li et al. 1995). Previous studies suggested that overexpression of Sod1 or Sod2 could significantly extend lifespan in mammals, fruit flies and *C*. *elegans* (Citations). Interestingly, Sod gene product was thought to help symbionts overcome host cellular immune responses (Bright and Bulgheresi, 2010), and Sod genes were identified in all vestimentiferan symbiont genomes in previous studies. We also be able to identify symbionts’ Sod gene as proteins in proteomic analysis (Table S11). Thus, Sods may play a central role in tubeworm symbiosis for overcoming oxidative damage and essential for extreme longevity for seep-living vestimentiferans. Moreover, a large expansion of interleukin 6 receptors (IL6R) was found in comparative analysis (Table S9). IL6R is the key component of IL-6 pathways which is the main signaling implicated in aging (Maggio et al. 2006).

### Phylogeny and molecular clock of tubeworm

 Figure 5. Time-calibrated phylogeny inferred with BEAST2 and 191 proteincoding genes in units of millions of years. All nodes have 1.00 posterior probability.

Two additional *Osedax* taxa were added in the study compared to the previous siboglinid phylogenomic analyses (Li et al. 2017) (Table S3). The final supermatrix dataset contains 191 genes single-copy orthologs. Bayesian inference with a relaxed molecular clock (Fig. 5) recovered the same topology as ML analysis in IQTree with strong nodal support (Fig. S5). Both analyses strongly support *Osedax* is most closely related to the Vestimentifera + *Sclerolinum* clade and Frenulata is the early diverging group, as recently reported (Li et al. 2015). Within Vestimentifera, *Lamellibrachia* is sister to the remaining sampled vestimentiferans. Molecular clock analyses based on phylogenomic dataset suggest modern siboglinid diversity originated in Mesozoic (223 MYA ± 80 MY), conflicting with previous hypotheses indicate a Late Mesozoic or Cenozoic approximately 50-126 MYA. The previous analyses are solely based on the estimation of nucleotide divergence of COI sequences on limited taxa sampling (mainly vestimentiferans). However, mitochondrial genes of vestimentiferans may have experienced a “slowdown” in the rate of nucleotide substitution relative to other siboglinid lineages (Li et al. 2015).

Moreover, molecular clock analysis suggests a younger split of vestimentiferans during the Cenozoic (60 MYA). Recent analyses indicated that Jurassic tubes from Figueroa deposits are likely to have been made by vestimentiferans (Georgieva et al. 2017), but this hypothesis is not supported in our analysis. Thus, we provide the molecular clock of the siboglinid phylogenetic tree, placing a common siboglinid ancestor as far back as the Jurassic. Siboglinid vestimentiferans might exploit and adapt to vents and seeps during Cenozoic (<65 MYA). This adds to the growing evidence that the Cenozoic was a key period for the radiation of many dominant invertebrate taxa now occupying in deep-sea chemosynthetic communities (Vrijenhoek 2013).

## Conclusion

In summary, we report the preliminary characterization of deep-sea seep-living *Lamellibrachia genome*. The comparative analysis offers candidate genes and pathways that may, in part, explain underlie the extraordinary characteristics of host–symbiont associations adapt to extreme chemosynthetic environments, including their nutrition mode, hemoglobin structures and functions, immunity function and longevity. A better understanding of these genetic mechanisms that we have studied may help to further extend to other taxa (e.g., deep-sea mussels, clams, snails and shrimp, or shallow water bivalves) that harbor chemosynthetic endosymbionts.

## Acknowledgments

This study was supported by awards from the U.S. National Science Foundation (NSF) (DEB-1036537 and IOS-0843473). Yuanning Li is supported by a scholarship from the China Scholarship Council (CSC) for studying and living abroad. We thank Chris Little, Maggie Georgieva and Luke Parry for the helpful discussion about molecular dating. We thank Jon Palmer for the help about trouble shooting Funannoate pipeline for genome annotation. We thank Jason Flores for the help of pointing out the zinc binding sites for sulfide binding in siboglinid A2 chains. We thank Kitty Brown for the help about interpretation of proteomic dataset. All bioinformatic analyses were conducted on the Auburn University Molette Laboratory SkyNet server and Auburn University Hopper HPC system. This is Molette Biology Laboratory contribution ### and Auburn University Marine Biology Program contribution ###.

## Author contribution

Albertin, Caroline B, Oleg Simakov, Therese Mitros, Z Yan Wang, Judit R Pungor, Eric Edsinger-Gonzales, Sydney Brenner, Clifton W Ragsdale, and Daniel S Rokhsar. 2015. “The Octopus Genome and the Evolution of Cephalopod Neural and Morphological Novelties.” *Nature* 524 (7564): 220.

Andrews, Simon, and others. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data.”

Arp, Alissa J, and James J Childress. 1981. “Blood Function in the Hydrothermal Vent Vestimentiferan Tube Worm.” *Science* 213 (4505): 342–44.

Bailly, Xavier, Didier Jollivet, Stephano Vanin, Jean Deutsch, Franck Zal, François Lallier, and André Toulmond. 2002. “Evolution of the Sulfide-Binding Function Within the Globin Multigenic Family of the Deep-Sea Hydrothermal Vent Tubeworm Riftia Pachyptila.” *Molecular Biology and Evolution* 19 (9): 1421–33.

Bates, Jennifer M, Janie Akerlund, Erika Mittge, and Karen Guillemin. 2007. “Intestinal Alkaline Phosphatase Detoxifies Lipopolysaccharide and Prevents Inflammation in Zebrafish in Response to the Gut Microbiota.” *Cell Host & Microbe* 2 (6): 371–82.

Bergquist, Derk C, Frederick M Williams, and Charles R Fisher. 2000. “Longevity Record for Deep-Sea Invertebrate.” *Nature* 403 (6769): 499.

Boetzer, Marten, and Walter Pirovano. 2014. “SSPACE-Longread: Scaffolding Bacterial Draft Genomes Using Long Read Sequence Information.” *BMC Bioinformatics* 15 (1): 211.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20.

Bouckaert, Remco, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A Suchard, Andrew Rambaut, and Alexei J Drummond. 2014. “BEAST 2: A Software Platform for Bayesian Evolutionary Analysis.” *PLoS Computational Biology* 10 (4): e1003537.

Bright, M, H Keckeis, and CR Fisher. 2000. “An Autoradiographic Examination of Carbon Fixation, Transfer and Utilization in the Riftia Pachyptila Symbiosis.” *Marine Biology* 136 (4): 621–32.

Buchfink, Benjamin, Chao Xie, and Daniel H Huson. 2014. “Fast and Sensitive Protein Alignment Using Diamond.” *Nature Methods* 12 (1): 59.

Cavanaugh, Colleen M, Stephen L Gardiner, Meredith L Jones, Holger W Jannasch, and John B Waterbury. 1981. “Prokaryotic Cells in the Hydrothermal Vent Tube Worm Riftia Pachyptila Jones: Possible Chemoautotrophic Symbionts.” *Science*, 340–42.

Chen, Nansheng. 2004. “Using Repeatmasker to Identify Repetitive Elements in Genomic Sequences.” *Current Protocols in Bioinformatics* 5 (1): 4–10.

Chu, Hiutung, and Sarkis K Mazmanian. 2013. “Innate Immune Recognition of the Microbiota Promotes Host-Microbial Symbiosis.” *Nature Immunology* 14 (7): 668.

Danise, Silvia, and Nicholas D Higgs. 2015. “Bone-Eating Osedax Worms Lived on Mesozoic Marine Reptile Deadfalls.” *Biology Letters* 11 (4): 20150072.

De Bie, Tijl, Nello Cristianini, Jeffery P Demuth, and Matthew W Hahn. 2006. “CAFE: A Computational Tool for the Study of Gene Family Evolution.” *Bioinformatics* 22 (10): 1269–71.

Dunn, Casey W, Mark Howison, and Felipe Zapata. 2013. “Agalma: An Automated Phylogenomics Workflow.” *BMC Bioinformatics* 14 (1): 330.

Flores, Jason F, Charles R Fisher, Susan L Carney, Brian N Green, John K Freytag, Stephen W Schaeffer, and William E Royer. 2005. “Sulfide Binding Is Mediated by Zinc Ions Discovered in the Crystal Structure of a Hydrothermal Vent Tubeworm Hemoglobin.” *Proceedings of the National Academy of Sciences* 102 (8): 2713–8.

Georgieva, Magdalena N, Crispin TS Little, Jonathan S Watson, Mark A Sephton, Alexander D Ball, and Adrian G Glover. 2017. “Identification of Fossil Worm Tubes from Phanerozoic Hydrothermal Vents and Cold Seeps.” *Journal of Systematic Palaeontology*, 1–43.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. “QUAST: Quality Assessment Tool for Genome Assemblies.” *Bioinformatics* 29 (8): 1072–5.

Haas, Brian J, Arthur L Delcher, Stephen M Mount, Jennifer R Wortman, Smith JrRoger K, Linda I Hannick, Rama Maiti, et al. 2003. “Improving the Arabidopsis Genome Annotation Using Maximal Transcript Alignment Assemblies.” *Nucleic Acids Research* 31 (19): 5654–66.

Haas, Brian J, Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D Blood, Joshua Bowden, Matthew Brian Couger, et al. 2013. “De Novo Transcript Sequence Reconstruction from Rna-Seq Using the Trinity Platform for Reference Generation and Analysis.” *Nature Protocols* 8 (8): 1494.

Haas, Brian J, Steven L Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E Allen, Joshua Orvis, Owen White, C Robin Buell, and Jennifer R Wortman. 2008. “Automated Eukaryotic Gene Structure Annotation Using Evidencemodeler and the Program to Assemble Spliced Alignments.” *Genome Biology* 9 (1): 1.

Halanych, Kenneth M, and Kevin M Kocot. 2014. “Repurposed Transcriptomic Data Facilitate Discovery of Innate Immunity Toll-Like Receptor (Tlr) Genes Across Lophotrochozoa.” *The Biological Bulletin* 227 (2): 201–9.

Hughes, Deborah S, Horst Felbeck, and Jeffrey L Stein. 1997. “A Histidine Protein Kinase Homolog from the Endosymbiont of the Hydrothermal Vent Tubeworm Riftia Pachyptila.” *Applied and Environmental Microbiology* 63 (9): 3494–8.

Kajitani, Rei, Kouta Toshimoto, Hideki Noguchi, Atsushi Toyoda, Yoshitoshi Ogura, Miki Okuno, Mitsuru Yabana, et al. 2014. “Efficient de Novo Assembly of Highly Heterozygous Genomes from Whole-Genome Shotgun Short Reads.” *Genome Research*, gr–170720.

Katoh, Kazutaka, and Daron M Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80.

Käll, Lukas, Anders Krogh, and Erik LL Sonnhammer. 2007. “Advantages of Combined Transmembrane Topology and Signal Peptide Prediction—the Phobius Web Server.” *Nucleic Acids Research* 35 (suppl\_2): W429–W432.

Keller, Andrew, Alexey I Nesvizhskii, Eugene Kolker, and Ruedi Aebersold. 2002. “Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by Ms/Ms and Database Search.” *Analytical Chemistry* 74 (20): 5383–92.

Kim, Daehwan, Ben Langmead, and Steven L Salzberg. 2015. “HISAT: A Fast Spliced Aligner with Low Memory Requirements.” *Nature Methods* 12 (4): 357.

Langmead, Ben, and Steven L Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357.

Li, Bo, and Colin N Dewey. 2011. “RSEM: Accurate Transcript Quantification from Rna-Seq Data with or Without a Reference Genome.” *BMC Bioinformatics* 12 (1): 323.

Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” *Bioinformatics* 1: 7.

Li, Weizhong, and Adam Godzik. 2006. “Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences.” *Bioinformatics* 22 (13): 1658–9.

Li, Yibing, Ting-Ting Huang, Elaine J Carlson, Simon Melov, Philip C Ursell, Jean L Olson, Linda J Noble, et al. 1995. “Dilated Cardiomyopathy and Neonatal Lethality in Mutant Mice Lacking Manganese Superoxide Dismutase.” *Nature Genetics* 11 (4): 376.

Li, Yuanning, Kevin M Kocot, Christoffer Schander, Scott R Santos, Daniel J Thornhill, and Kenneth M Halanych. 2015. “Mitogenomics Reveals Phylogeny and Repeated Motifs in Control Regions of the Deep-Sea Family Siboglinidae (Annelida).” *Molecular Phylogenetics and Evolution* 85: 221–29.

Li, Yuanning, Kevin M Kocot, Nathan V Whelan, Scott R Santos, Damien S Waits, Daniel J Thornhill, and Kenneth M Halanych. 2017. “Phylogenomics of Tubeworms (Siboglinidae, Annelida) and Comparative Performance of Different Reconstruction Methods.” *Zoologica Scripta* 46 (2): 200–213.

Li, Yuanning, Mark R Liles, and Kenneth M Halanych. 2018. “Endosymbiont Genomes Yield Clues of Tubeworm Success.” *The ISME Journal* 12 (11): 2785.

Liguori, Ilaria, Gennaro Russo, Francesco Curcio, Giulia Bulli, Luisa Aran, David Della-Morte, Gaetano Gargiulo, et al. 2018. “Oxidative Stress, Aging, and Diseases.” *Clinical Interventions in Aging* 13: 757.

Little, Crispin TS. 2002. “The Fossil Record of Hydrothermal Vent Communities.” *Cahiers de Biologie Marine* 43 (3/4): 313–16.

Little, Crispin TS, and Robert C Vrijenhoek. 2003. “Are Hydrothermal Vent Animals Living Fossils?” *Trends in Ecology & Evolution* 18 (11): 582–88.

Lowe, Todd M, and Sean R Eddy. 1997. “TRNAscan-Se: A Program for Improved Detection of Transfer Rna Genes in Genomic Sequence.” *Nucleic Acids Research* 25 (5): 955.

Luo, Ruibang, Binghang Liu, Yinlong Xie, Zhenyu Li, Weihua Huang, Jianying Yuan, Guangzhu He, et al. 2012. “SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler.” *Gigascience* 1 (1): 18.

Luo, Yi-Jyun, Miyuki Kanda, Ryo Koyanagi, Kanako Hisata, Tadashi Akiyama, Hirotaka Sakamoto, Tatsuya Sakamoto, and Noriyuki Satoh. 2018. “Nemertean and Phoronid Genomes Reveal Lophotrochozoan Evolution and the Origin of Bilaterian Heads.” *Nature Ecology & Evolution* 2 (1): 141.

Lutz, Richard A, Timothy M Shank, Daniel J Fornari, Rachel M Haymon, Marvin D Lilley, Von DammKaren L, and Daniel Desbruyeres. 1994. “Rapid Growth at Deep-Sea Vents.” *Nature* 371 (6499): 663.

Maggio, Marcello, Jack M Guralnik, Dan L Longo, and Luigi Ferrucci. 2006. “Interleukin-6 in Aging and Chronic Disease: A Magnificent Pathway.” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 61 (6): 575–84.

Marçais, Guillaume, and Carl Kingsford. 2011. “A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of K-Mers.” *Bioinformatics* 27 (6): 764–70.

Markert, Stephanie, Cordelia Arndt, Horst Felbeck, Dörte Becher, Stefan M Sievert, Michael Hügler, Dirk Albrecht, et al. 2007. “Physiological Proteomics of the Uncultured Endosymbiont of Riftia Pachyptila.” *Science* 315 (5809): 247–50.

Mi, Huaiyu, Xiaosong Huang, Anushya Muruganujan, Haiming Tang, Caitlin Mills, Diane Kang, and Paul D Thomas. 2016. “PANTHER Version 11: Expanded Annotation Data from Gene Ontology and Reactome Pathways, and Data Analysis Tool Enhancements.” *Nucleic Acids Research* 45 (D1): D183–D189.

Minic, Zoran, and Guy Hervé. 2003. “Arginine Metabolism in the Deep Sea Tube Worm Riftia Pachyptila and Its Bacterial Endosymbiont.” *Journal of Biological Chemistry*.

Moriya, Yuki, Masumi Itoh, Shujiro Okuda, Akiyasu C Yoshizawa, and Minoru Kanehisa. 2007. “KAAS: An Automatic Genome Annotation and Pathway Reconstruction Server.” *Nucleic Acids Research* 35 (suppl\_2): W182–W185.

Nesvizhskii, Alexey I, Andrew Keller, Eugene Kolker, and Ruedi Aebersold. 2003. “A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry.” *Analytical Chemistry* 75 (17): 4646–58.

Newton, Irene LG, Peter R Girguis, and Colleen M Cavanaugh. 2008. “Comparative Genomics of Vesicomyid Clam (Bivalvia: Mollusca) Chemosynthetic Symbionts.” *BMC Genomics* 9 (1): 585.

Nguyen, Lam-Tung, Heiko A Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2014. “IQ-Tree: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” *Molecular Biology and Evolution* 32 (1): 268–74.

Nussbaumer, Andrea D, Charles R Fisher, and Monika Bright. 2006. “Horizontal Endosymbiont Transmission in Hydrothermal Vent Tubeworms.” *Nature* 441 (7091): 345.

O’Connell, Jared, Ole Schulz-Trieglaff, Emma Carlson, Matthew M Hims, Niall A Gormley, and Anthony J Cox. 2015. “NxTrim: Optimized Trimming of Illumina Mate Pair Reads.” *Bioinformatics* 31 (12): 2035–7.

Park, Beom Seok, and Jie-Oh Lee. 2013. “Recognition of Lipopolysaccharide Pattern by Tlr4 Complexes.” *Experimental & Molecular Medicine* 45 (12): e66.

Petersen, Thomas Nordahl, Søren Brunak, Gunnar von Heijne, and Henrik Nielsen. 2011. “SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions.” *Nature Methods* 8 (10): 785.

Ponnudurai, Ruby, Manuel Kleiner, Lizbeth Sayavedra, Jillian M Petersen, Martin Moche, Andreas Otto, Dörte Becher, et al. 2017. “Metabolic and Physiological Interdependencies in the Bathymodiolus Azoricus Symbiosis.” *The ISME Journal* 11 (2): 463.

Pryszcz, Leszek P, and Toni Gabaldón. 2016. “Redundans: An Assembly Pipeline for Highly Heterozygous Genomes.” *Nucleic Acids Research* 44 (12): e113–e113.

Quesada, Vı'ctor, Freitas-Rodrı'guezSandra, Joshua Miller, José G Pérez-Silva, Zi-Feng Jiang, Washington Tapia, Olaya Santiago-Fernández, et al. 2018. “Giant Tortoise Genomes Provide Insights into Longevity and Age-Related Disease.” *Nature Ecology & Evolution*, 1.

Sanfilippo, Rossana, Antonietta Rosso, Agatino Reitano, and Gianni Insacco. 2017. “First Record of Sabellid and Serpulid Polychaetes from the Permian of Sicily.” *Acta Palaeontologica Polonica* 62 (1): 25–38.

Searle, Brian C, Mark Turner, and Alexey I Nesvizhskii. 2008. “Improving Sensitivity by Probabilistically Combining Results from Multiple Ms/Ms Search Methodologies.” *The Journal of Proteome Research* 7 (1): 245–53.

Slater, Guy St C, and Ewan Birney. 2005. “Automated Generation of Heuristics for Biological Sequence Comparison.” *BMC Bioinformatics* 6 (1): 31.

Smit, AFA, and R Hubley. 2008. “RepeatModeler Open-1.0.” *Available Fom Http://Www. Repeatmasker. Org*.

Stanke, Mario, Oliver Keller, Irfan Gunduz, Alec Hayes, Stephan Waack, and Burkhard Morgenstern. 2006. “AUGUSTUS: Ab Initio Prediction of Alternative Transcripts.” *Nucleic Acids Research* 34 (suppl\_2): W435–W439.

Tassia, Michael G, Nathan V Whelan, and Kenneth M Halanych. 2017. “Toll-Like Receptor Pathway Evolution in Deuterostomes.” *Proceedings of the National Academy of Sciences* 114 (27): 7055–60.

Vrijenhoek, Robert C. 2013. “On the Instability and Evolutionary Age of Deep-Sea Chemosynthetic Communities.” *Deep Sea Research Part II: Topical Studies in Oceanography* 92: 189–200.

Vurture, Gregory W, Fritz J Sedlazeck, Maria Nattestad, Charles J Underwood, Han Fang, James Gurtowski, and Michael C Schatz. 2017. “GenomeScope: Fast Reference-Free Genome Profiling from Short Reads.” *Bioinformatics* 33 (14): 2202–4.

Waterhouse, Robert M, Mathieu Seppey, Felipe A Simão, Mosè Manni, Panagiotis Ioannidis, Guennadi Klioutchnikov, Evgenia V Kriventseva, and Evgeny M Zdobnov. 2017. “BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics.” *Molecular Biology and Evolution* 35 (3): 543–48.

Weisenfeld, Neil I, Vijay Kumar, Preyas Shah, Deanna M Church, and David B Jaffe. 2017. “Direct Determination of Diploid Genome Sequences.” *Genome Research*.

Zal, Franck, François H Lallier, Brian N Green, Serge N Vinogradov, and André Toulmond. 1996. “The Multi-Hemoglobin System of the Hydrothermal Vent Tube Worm Riftia Pachyptila Ii. Complete Polypeptide Chain Composition Investigated by Maximum Entropy Analysis of Mass Spectra.” *Journal of Biological Chemistry* 271 (15): 8875–81.

Zal, F, T Suzuki, Y Kawasaki, JJ Childress, FH Lallier, and A Toulmond. 1997. “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: Structure, Function, and Bioinformatics* 29 (4): 562–74.

Zdobnov, Evgeni M, and Rolf Apweiler. 2001. “InterProScan–an Integration Platform for the Signature-Recognition Methods in Interpro.” *Bioinformatics* 17 (9): 847–48.