Marine Genomics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

First transcriptome assembly of a newly discovered vent mussel, *Gigantidas vrijenhoeki*, at Onnuri Vent Field on the northern Central Indian Ridge

Taewoo Ryu [a,1], Jong Guk Kim [b,1], Jimin Lee [b], Ok Hwan Yu [b], Seungshic Yum [c], Dongsung Kim [b], Seonock Woo [d,*]

[a] Marine Climate Change Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan
[b] Marine Ecosystem and Biological Research Center, Korea Institute of Ocean Science and Technology, Busan 49111, Republic of Korea
[c] Ecological Risk Research Division, Korea Institute of Ocean Science and Technology, Geoje 53201, South Korea
[d] Marine Biotechnology Research Center, Korea Institute of Ocean Science and Technology, Busan 49111, South Korea

ARTICLE INFO

Keywords:
Gigantidas vrijenhoeki
Mussel
Transcriptome
Hydrothermal vent
Onnuri Vent Field

ABSTRACT

This is the first report of a transcriptome assembly of a newly discovered hydrothermal vent mussel, *Gigantidas vrijenhoeki* (Bivalvia: Mytilidae), on the Central Indian Ridge. *Gigantidas vrijenhoeki* was identified from material collected at the newly discovered Onnuri Vent Field (OVF) on the Central Indian Ridge in 2018, and was reported as a new species, distinct from another dominant hydrothermal vent mussel, *Bathymodiolus marisindicus*, in 2020. We sequenced the transcriptome of *G. vrijenhoeki* using the Illumina HiSeq X System. *De novo* assembly and analysis of the coding regions predicted 25,405 genes, 84.76% of which was annotated by public databases. The transcriptome of *G. vrijenhoeki* will be a valuable resource in studying the ecological and biological characteristics of this new species, which is distinct from other deep-sea mussels. These data should also support the investigation of the relationship between the environmental conditions of hydrothermal vents and the unique distribution of *G. vrijenhoeki* in the OVF of the Central Indian Ridge.

1. Introduction

Deep-sea hydrothermal vents on the seafloor are dynamic environments with steep gradients of nutrient and physicochemical conditions resulting from both volcanic and tectonic events (Childress and Fisher 1992). Despite the extreme conditions, hydrothermal vent communities are characterized by high endemism and a large biomass (Rogers et al. 2012), predominantly composed of vestimentiferan tube worms, bathymodiolin mussels, vesicomyid clams, and shrimps. These organisms are maintained in this habitat by chemosynthetic energy sources from hydrothermal mineral deposits, such as hydrogen sulfide and methane (Dubilier et al. 2008; Nakamura and Takai 2015). The composition of the vent fauna is affected by the type of hydrothermal vent, which can be a basalt- or ultramafic-supported system (Rogers et al. 2012; Copley et al. 2016).

After the discovery of the Kairei hydrothermal vent field, an active hydrothermal vent, in 2000, only five hydrothermal vent communities (Dodo, Edmond, Kairei, Longqu, and Solitaire) were known in the Indian Ocean (Gamo et al. 2001). However, the new hydrothermal vent, the Onnuri Vent Field (OVF), located on the northern Central Indian Ridge, was recently discovered by the Korea Institute of Ocean Science and Technology (KIOST) during a 2017–2018 research expedition. In this chemosynthesis-based ecosystem, the commonest species are the scaly-foot gastropod *Chrysomallon squamiferum*, the shrimps *Rimicaris kairei* and *Mirocaris indica*, the stalked barnacle *Neolepas marisindica*, and the mussel *Bathymodiolus marisindicus*. Species of bathymodiolin mussels are conspicuous and dominant taxa in the Indian hydrothermal vent fields and play an important role in providing energy and habitats for various animals (Taylor and Glover 2010; Xu et al. 2018, 2019). Although mussels of the genus *Bathymodiolus* are commonly found in other vent fields in the Indian Ocean, a new mussel species, *Gigantidas vrijenhoeki* Jang et al. 2020, was found for the first time in the bathymodiolin mussel habitats in the OVF. *Gigantidas vrijenhoeki* is the first member of the genus *Gigantidas* recorded in the Indian Ocean. Most members of *Gigantidas* occur in the western Pacific, and some are found in the Atlantic Ocean. This new mussel species was found together with *B. marisindicus*, but has a completely different shell morphology and symbionts, and is also uniquely found in the OVF of the Central Indian

* Corresponding author.
E-mail address: cwoo@kiost.ac.kr (S. Woo).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.margen.2020.100819
Received 21 April 2020; Received in revised form 27 August 2020; Accepted 31 August 2020
Available online 12 September 2020

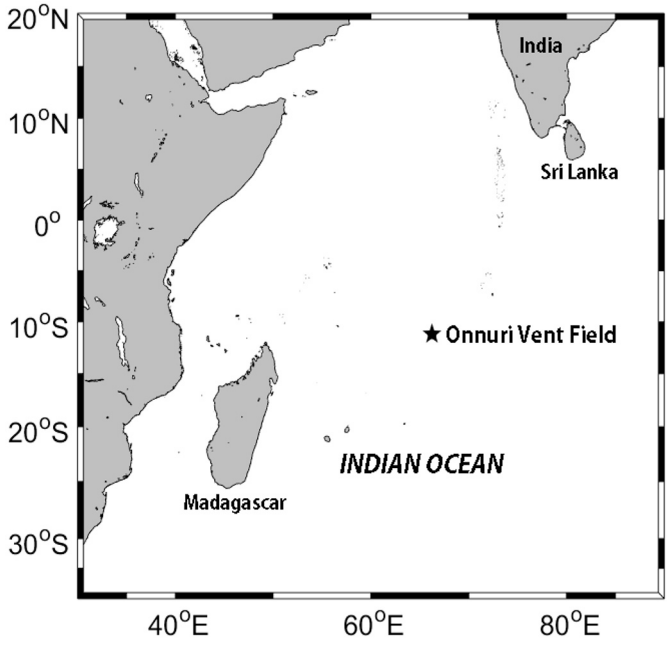

Fig. 1. Sampling location (star) of *Gigantidas vrijenhoeki*.

Ridge.

Several studies have examined the mitochondrial genes of bathymodiolin mussels, which has allowed the reconstruction of their phylogeny at the genus level (Jones and Vrijenhoek 2006; Jones et al. 2006; Lorion et al. 2010; Thubaut et al. 2013). A recent population genetic analysis focused on their large-scale distributions across geographic barriers, such as the world ocean and ocean currents (Breusing et al. 2015; Xu et al. 2018). Using DNA sequence analyses, Breusing et al. (2015) showed that the evolutionary history of a Central Indian Ridge mussel, *B. marisindicus*, was independent of that of the western Pacific mussels. Studies of the molecular phylogeny and the symbiotic bacteria in the gills of bathymodiolin mussels have also been undertaken (Duperron et al. 2009; Fontanez and Cavanaugh 2013; Jang et al. 2020). However, there has been no large-scale study of this taxon, limiting the further analysis of topics such as its evolution in and adaption to its harsh environment. Therefore, we have constructed the first transcriptome assembly for *G. vrijenhoeki* to provide baseline data for future studies of its survival mechanisms in deep-sea hydrothermal environments.

2. Data description

2.1. Sample collection

A *G. vrijenhoeki* specimen was collected with a video-guided hydraulic grab (Oktopus, Germany) around the OVF (11°14′55.92”S, 66°15′15.10″E, at 2014.5 m depth) in June 2019 (Figs. 1 and 2). The collected sample was immediately rinsed with seawater, directly frozen in a deep-freeze, and stored at −80 °C until the extraction of its total RNA in the laboratory.

2.2. Data production and de novo assembly

The total RNA was extracted from the whole body of *G. vrijenhoeki* with TRIzol® Reagent (Thermo Fisher Scientific Inc., Waltham, MA, USA). In brief, the whole deep-frozen *G. vrijenhoeki* body was pulverized in a mortar with liquid nitrogen. This mussel powder was homogenized in 1 mL of TRIzol® Reagent, and the RNA was extracted with the 1/5 volume of chloroform. After vigorous shaking and incubation at 4 °C for 1 h, the phase separation was done by centrifugation at 16,000 ×*g* at 4 °C for 20 min and the aqueous phase was transferred to a fresh tube. The RNA was precipitated by centrifugation at 16,000 ×*g* for 30 min and the precipitate was resuspended in 300 μL of diethyl pyrocarbonate (DEPC)-treated water. The RNA was precipitated with a 1/10 volume of 3 M sodium acetate (pH 5.2) and the same volume of isopropanol. The precipitated RNA was rinsed with 70% ethanol (diluted with DEPC-treated water), and dissolved in an appropriate volume of DEPC-treated water (30–40 μL). Contaminating DNA was removed with 1 μL of DNase. The yield and purity of the RNA were measured with a NanoDrop ND-1000 spectrophotometer (Thermo Scientific Inc.), and the RNA integrity was checked as the RNA integrity number (RIN) with the Bioanalyzer 2100 (Agilent Technology, Santa Clara, CA, USA). The final concentration of RNA was 2.156 μg/μL and the RIN value was 9.1.

Library construction and transcriptome sequencing were conducted with one sample of *G. vrijenhoeki* using the Illumina TruSeq Stranded Total RNA LT Sample Prep Kit and Illumina HiSeq X Sequencing System (Macrogen, Seoul, South Korea), according to the manufacturer's instructions. Briefly, the purified RNA was randomly fragmented and reverse-transcribed into cDNA. The cDNA fragments were ligated to adapter sequences at both ends and then amplified with PCR. Fragments with insert sizes of 200–400 base pairs (bp) were selected and subjected to paired-end sequencing, generating 86,863,648 reads of 151 bp (~13.1 Gbp in total). The overall GC content was 41.74% and base call accuracy represented by Q30 (ratio of bases with Phred quality score ≥30) was 94.52%. The RNA-Seq dataset has been deposited in the National Center for Biotechnology Information (NCBI) GenBank database (BioProject ID PRJNA597854).

Before transcriptome assembly, the adaptor sequences and low-quality sequences (window size: 4, mean quality threshold: 15, and minimum read length: 36) were removed with Trimmomatic v0.38 (Bolger et al. 2014), leaving 85,300,970 reads containing 12,374,473,378 total bases. The quality-checked reads were assembled *de novo* into contigs without a reference genome sequence using Trinity version trinityrnaseq_r20140717, with the default settings (Grabherr et al. 2011). The assembled contigs were deemed to be the reconstructed

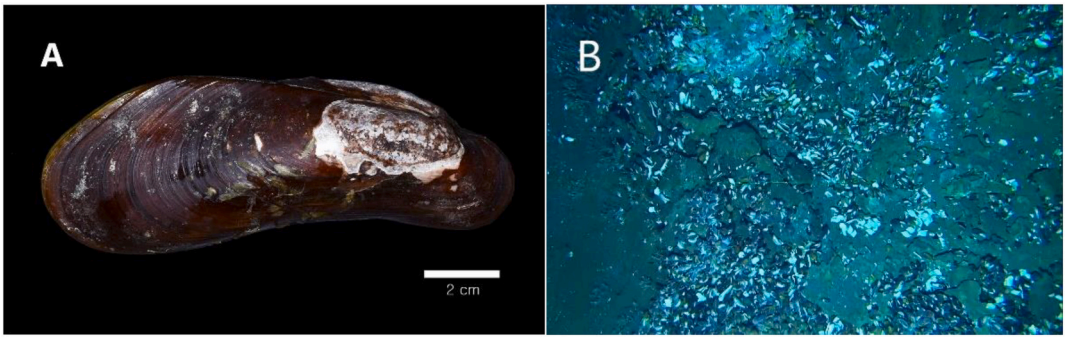

Fig. 2. *Gigantidas vrijenhoeki* specimen (A) and the sampling site view of the Onnuri Vent Field (B, depth 2014.5 m).

Table 1
MIxS description of the *Gigantidas vrijenhoeki* transcriptome.

Item	Description
Investigation_type	Eukaryote
Project_name	Transcriptome assembly of *Gigantidas vrijenhoeki*
Organism	*Gigantidas vrijenhoeki*
Classification	Metazoa (kingdom); Mollusca (phylum); Bivalvia (class); Pteriomorphia (subclass); Mytilida (order); Mytiloidea (Superfamily); Mytilidae (family); Bathymodiolinae (Subfamily); *Gigantidas* (genus)
Lat_lon	11°14′55.92″S, 66°15′15.10″E
Geo_loc_name	Indian Ocean, Onnuri Vent Field
Collection_date	2019-06-28
Collector	Dongsung Kim, Ok Hwan Yu, Jong Guk Kim
Environment (biome)	marine benthic biome (ENVO:01000024)
Environment (feature)	marine hydrothermal vent (ENVO:01000122)
Environment (material)	sea water (ENVO:00002149)
Env_package	Water
Seq_meth	Illumina
Transcriptome_platform	HiSeq X System
Assembly_method	trinityrnaseq_r20140717
Submitted_to_INSDC	Bioproject (PRJNA597854)
	Biosample (SAMN13685211)
	SRA (SRR10802050)
	GenBank (GIIM00000000)

universal set of transcribed mRNAs in *G. vrijenhoeki*. The total number of contigs was 249,529 with a GC content of 34.17%. The minimum contig length required to cover 50% of the transcriptome assembly (N50) was 1014 bp and the longest contig had 64,578 bp. We further calculated ExN50, an alternative statistic of N50, which account for transcriptome expression level. Abundance of transcripts were quantified by kallisto (Bray et al. 2016) and the largest expression-dependent N50 was obtained by 'contig_exn50_statistic.pl' module provided by Trinity (Grabherr et al. 2011), resulted in E85N50 of 1901 bp.

The minimum information about any (x) sequence (MIxS) specifications, a standard way of reporting genetic material (Yilmaz et al. 2011), for the transcriptome assembly of *G. vrijenhoeki* is provided in Table 1.

2.3. Gene prediction and functional analysis

Unigenes were identified by clustering the contigs into the nonredundant and longest transcripts using cdhit-est v4.6 (Li and Godzik 2006). This process produced 194,770 unigenes with N50 of 712 bp and a maximum contig length of 64,578 bp. The open reading frames (ORFs) within the unigenes were then predicted with TransDecoder v3.0.1 (Haas et al. 2013) to identify candidate protein-coding sequences, with a length threshold of 100 amino acids. The number of ORFs within the *G. vrijenhoeki* unigenes was 25,405.

To annotate the identified transcripts, the ORF sequences were used as queries in public databases, including EggNog v4, Gene Ontology (GO) v20180319, KEGG pathway v20190104, NCBI nonredundant protein (nr) v20180503, NCBI nucleotide (nt) v20180116, Pfam v20160316, and UniProtKB v20180116, using BLASTN of NCBI BLAST v2.4.0+ (Altschul et al. 1990) or BLASTX of DIAMOND v0.9.21 (Buchfink et al. 2015) with an *E*-value default cutoff of 10^{-5}. Of the *G. vrijenhoeki* ORFs, 14.9% (nt) ~ 83.9% (nr) were annotated using the database search (Fig. 3). In total, 84.76% of ORFs were annotated by at least one database.

We cataloged *G. vrijenhoeki* ORFs using the GO database (Supplementary Fig. 1). For the Biological Process category, the highest number

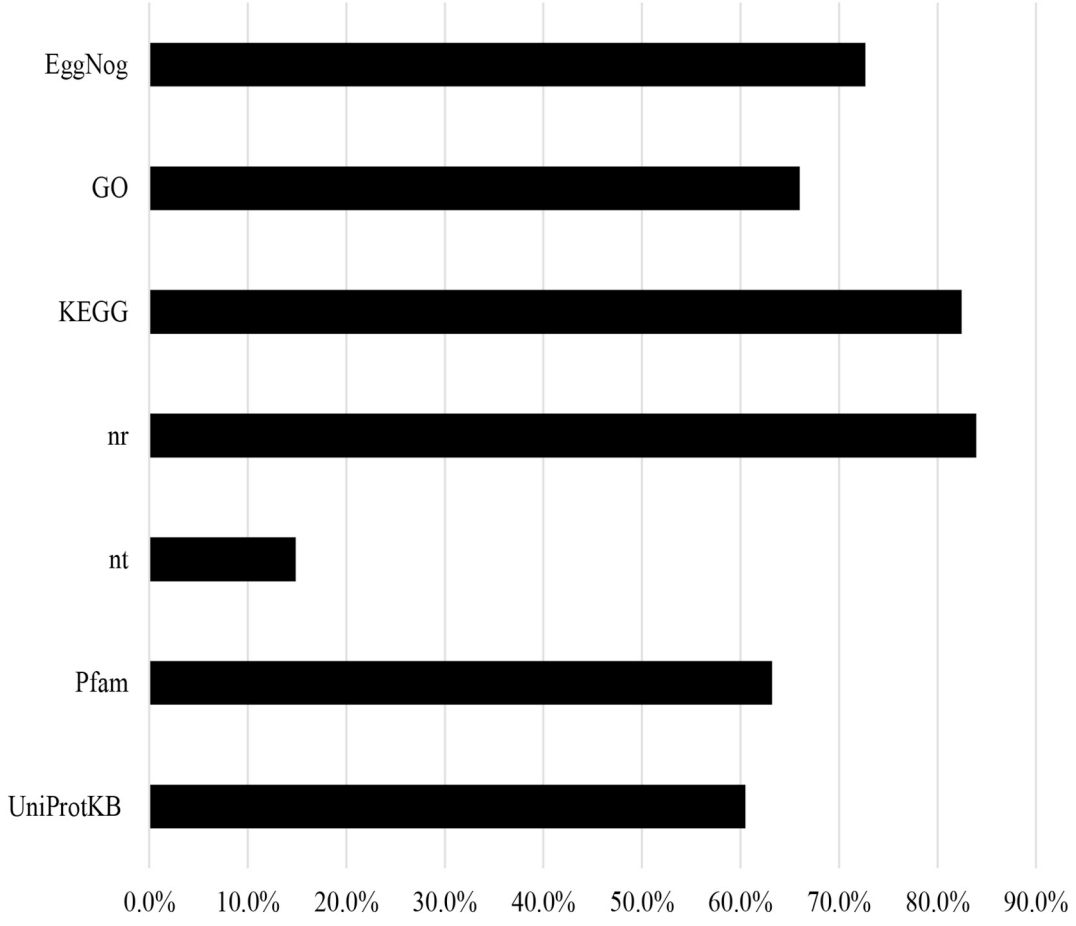

Fig. 3. Proportions of annotated *Gigantidas vrijenhoeki* proteins in each database.

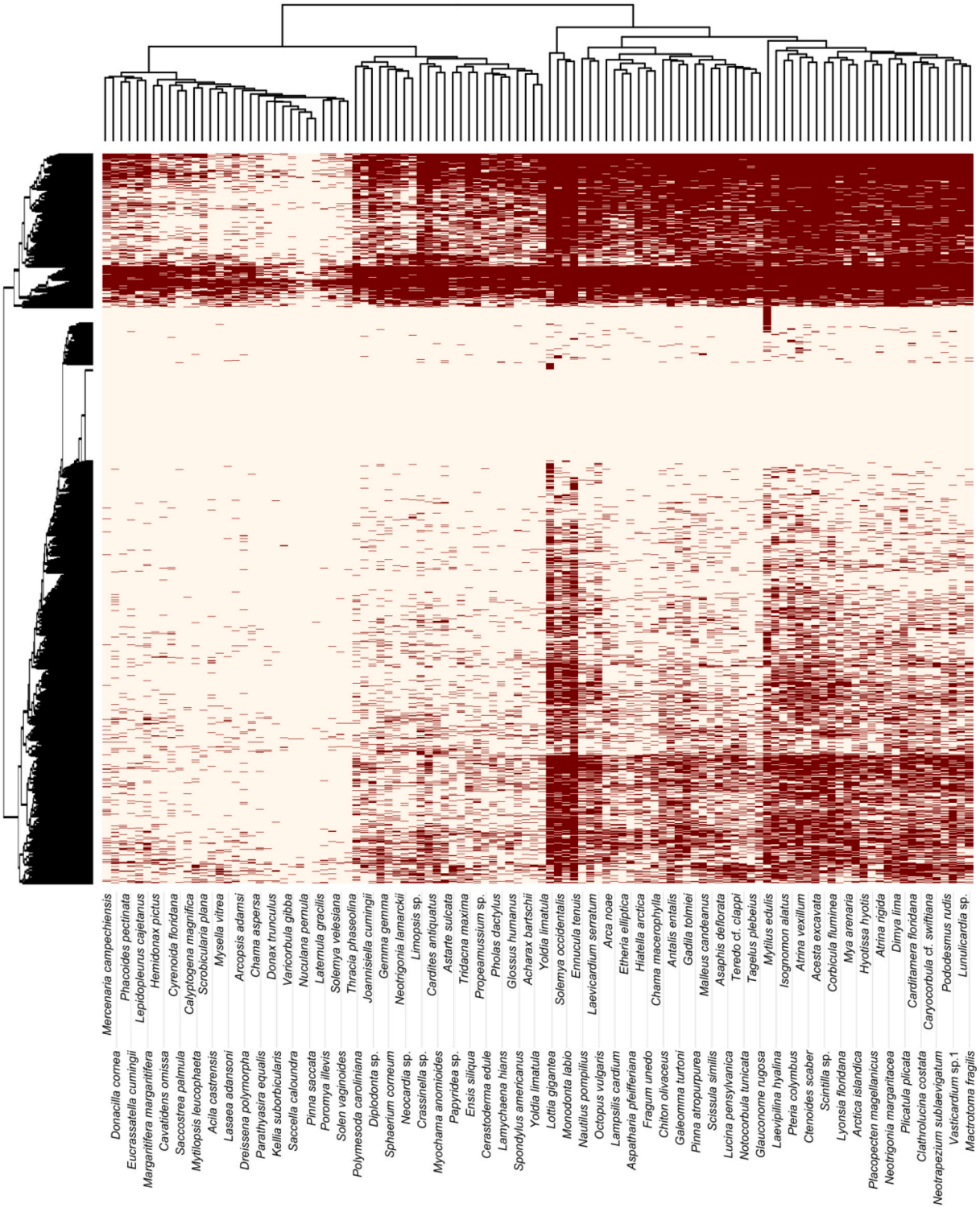

Fig. 4. Heatmap representation of the *Gigantidas vrijenhoeki* transcriptome by similarity to other molluscs transcriptomes. Proteins identified from the *Gigantidas vrijenhoeki* transcriptome were queried to those from 108 mollusc transcriptomes cataloged by Lemer et al. (2019) using BLASTP. Presence or absence of homologues (criteria: e-value $<10^{-4}$ and alignment length >50%) in other mollusc transcriptomes are represented by red or white color, respectively. Rows and columns correspond to *Gigantidas vrijenhoeki* proteins and 108 transcriptomes, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of annotations were 'metabolic process' (27.8%) followed by 'biological regulation' (20.9%). In the Cellular Component category, most unigenes were related to 'cell part' (36.8%), 'membrane part' (16.9%), and 'organelle' (12.5%). 'Binding' (40.3%) and 'catalytic activity' (39%) were dominant terms in the Molecular Function category.

Protein domain analysis identified that 16,059 ORFs have at least one Pfam or SuperFamily domains (Supplementary Fig. 2). The 'zinc finger C2H2 superfamily' domain was the most dominant domain class. The next abundant domains were P-loop containing nucleoside triphosphate hydrolase domain, immunoglobulin-like domain

superfamily, cadherin-like superfamily, and ankyrin repeat-containing domain superfamily. Approximately 21% of the identified protein domains belonged to those top 5 domain classes which are mainly related to protein structure and folding and critical for rigid scaffold for protein interactions and stability.

The assembly completeness was quantitatively assessed with BUSCO v3.0.2 (Waterhouse et al. 2017). Among the 978 near-universal single-copy orthologues in the metazoan lineage, 701 single-copy and 266 duplicated genes were detected as complete forms in the *G. vrijenhoeki* transcriptome assembly, indicating 98.9% assembly completeness. Nine partial orthologues were found and only two orthologues were missing from the assembly, indicating the high gene content of our results.

Additionally, the ORFs of *G. vrijenhoeki* were also compared with the transcriptome datasets of 108 molluscs (99 bivalves and 9 non-bivalve molluscs), including diverse taxa such as monoplacophorans, polyplacophorans, scaphopods, bivalves, gastropods, and cephalopods (Lemer et al. 2019). Homologues in other molluscs were identified with a BLASTP analysis, with two criteria: e-value $<10^{-4}$ and alignment length $>$ 50% with *G. vrijenhoeki* proteins. The overall transcriptomic similarity of each mollusc species to *G. vrijenhoeki* was visualized with the heatmap.2 package (Fig. 4) (Warnes et al. 2012). Of the 25,405 *G. vrijenhoeki* ORFs, the highest number of homologues was found in a protobranch bivalve (*Ennucula tenuis*, 62.6%) and the next highest in an owl limpet (*Lottia gigantea*, 60.6%), a chemosymbiotic bivalve (*Solemya occidentalis*, 54.4%), a lipped periwinkle (*Monodonta labio*, 53.7%), and a mussel (*Mytilus edulis*, 53.0%). *Ennucula tenuis* and *S. occidentalis* belong to the subclass Protobranchia, which is a cosmopolitan group at abyssal depths (Sharma et al. 2013), and *M. edulis* belongs to the family Mytilidae in the subclass Pteriomorphia, a morphologically diverse group that includes mussels, scallops, oysters, and ark clams, including *G. vrijenhoeki*. The 10 most-distant species were morphologically very different from *G. vrijenhoeki* and belonged to the subclass Heterodonta, a group that generally includes heteromyarian species, distributed from shallow waters to deep seas (Canapa et al. 2001; Taylor et al. 2007). *Laevipilina hyaline* (46.6%), in the class Monoplacophora, also shared high homology with *G. vrijenhoeki*. Monoplacophorans are known as 'living fossil' molluscs, have gills on both sides of the oval foot and no eyes or head, and occur at abyssal and hadal depths (Sigwart et al. 2019).

Of the *G. vrijenhoeki* genes, 3180 shared no homology with the 108 molluscs transcriptomes examined. We assigned biological functions of these orphan genes with GO, Pfam, and SuperFamily entries (Supplementary Figs. 3 and 4). 'Cellular process' and 'metabolic process' were the most frequent terms in the Biological Process category, and were assigned to 11.4% and 8.8% of these genes, respectively. The genes related to the terms 'cell part' (10.4%) and 'organelle' (5.2%) were observed frequently in the Cellular Component category. In the Molecular Function category, 'catalytic activity' (8.5%) and 'binding' (8.5%) were frequently assigned to these lineage-specific genes (Supplementary Fig. 3).

The *G. vrijenhoeki* orphan genes contained 291 Pfam and SuperFamily domains. The 12 most abundant superfamily domains in the *G. vrijenhoeki* orphan genes are shown in Supplementary Fig. 4. The most frequent domain was the alpha/beta hydrolase fold domain, which is related to the hydrolytic enzyme superfamily, in which members with widely differing phylogenetic origins and catalytic functions occur. The next most frequent superfamily domains were the 'cytochrome c-like domain superfamily', involved in cellular stress, and 'P-loop containing nucleoside triphosphate hydrolase', which is the most prevalent domain among several distinct nucleotide-binding protein folds.

In the deep-sea environment, hydrostatic pressure strongly inhibits protein functions, affecting their folding and activities. Consequently, organisms living at immense depths must maintain intracellular conditions that preserve the properties and pressure resistance of their proteins (Jamieson 2015). Mechanisms based on structural adaptations have been proposed to explain the preservation of protein functions in deep-sea organisms. The results of our protein domain analysis of both *G. vrijenhoeki* ORFs and orphan genes demonstrate that a large proportion of protein domains have functions that involve protein structures, interactions, and stability.

The deep-sea vent habitat is one of the most productive ecosystems on earth because hydrothermal mineral deposits provide a source of energy for chemosynthetic organisms that is independent of solar energy (Nakamura and Takai 2015). Bathymodiolin mussels are particularly conspicuous animals in deep-sea chemosynthesis-based ecosystems, including hydrothermal vents and cold seeps, and these mussels often dominate the biomass of chemosynthesis-based communities, forming dense mussel beds that provide critical habitats for many other deep-sea creatures (Taylor and Glover 2010; Xu et al. 2018, 2019). In this study, we generated a reference transcriptome with 194,770 unigenes for *G. vrijenhoeki* Jang et al. 2020, which was found for the first time in the OVF in the Indian Ocean Central Ridge and reported functional classification of unigenes, the represented protein domains, and the expression distribution of transcripts and also demonstrated the homology of translated peptides comparing with 108 mollusc transcriptomes. This analysis provides baseline data and should facilitate the study of the functional attributes that permit this mussel to live in the OVF, uniquely among the six Indian Ocean hydrothermal vents. Future studies should assess the functional expression profiles related to specific adaptations to the OVF environment and compare the transcriptomic strategies of *G. vrijenhoeki* with those of other *Gigantidas* species in the Pacific and Atlantic oceans.

Data availability

The RNA-Seq reads have been deposited in GenBank (BioProject ID PRJNA597854). The transcriptome shotgun assembly project has been deposited in the DDBJ/EMBL/GenBank in the FASTA format under accession number GIIM00000000. The version described in this paper is the first version, GIIM01000000.

Author contributions

JK, OY, SY, JL, and DK sampled the specimen and investigated the experiments. TR performed the computational analyses. TR, JK, and SW wrote the original draft and revised it.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was part of the project titled 'Understanding the deep-sea biosphere on seafloor hydrothermal vents in the Indian Ridge (no. 20170411)' funded by the Ministry of Oceans and Fisheries, Korea and Korea Institute of Ocean Science and Technology (PM61740) and National Research Foundation of Korea funded by the Korea government (MSIT) (2020R1A2B5B02001619).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2020.100819.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bray, N.L., Pimentel, H., Melsted, P., Patcher, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. https://doi.org/10.1038/nbt.3519Buchfink.

Breusing, C., Johnson, S.B., Tunnicliffe, V., Vrijenhoek, R.C., 2015. Population structure and connectivity in indo-Pacific deep-sea mussels of the *Bathymodiolus septemdierum* complex. Conserv. Genet. 16, 1415–1430. https://doi.org/10.1007/s10592-015-0750-0.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176.

Canapa, A., Barucca, M., Marinelli, A., Olmo, E., 2001. A molecular phylogeny of Heterodonta (Bivalvia) based on small ribosomal subunit RNA sequences. Mol. Phylogenet. Evol. 21, 156–161. https://doi.org/10.1006/mpev.2001.0992.

Childress, J.J., Fisher, C.R., 1992. The biology of hydrothermal vent animals: physiology, biochemistry, and autotrophic symbioses. Oceanogr. Mar. Biol. Annu. Rev. 30, 337–441.

Copley, J.T., Marsh, L., Glover, A.G., Hühnerbach, V., Nye, V.E., Reid, W.D.K., Sweeting, C.J., Wigham, B.D., Wiklund, H., 2016. Ecology and biogeography of megafauna and macrofauna at the first known deep-sea hydrothermal vents on the ultraslow-spreading southwest Indian ridge. Sci. Rep. 6, 39158. https://doi.org/10.1038/srep39158.

Dubilier, N., Bergin, C., Lott, C., 2008. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat. Rev. Microbiol. 6, 725–740. https://doi.org/10.1038/nrmicro1992.

Duperron, S., Lorion, J., Samadi, S., Gros, O., Gaill, F., 2009. Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. C. R. Biol. 332, 298–310. https://doi.org/10.1016/j.crvi.2008.08.003.

Fontanez, K.M., Cavanaugh, C.M., 2013. Phylogenetic relationships of hydrothermal vent mussels (Bathymodiolinae) and their symbionts. Mar. Ecol. Prog. Ser. 474, 147–154. https://doi.org/10.3354/meps10086.

Gamo, T., Chiba, H., Yamanaka, T., Okudaira, T., Hashimoto, J., Tsuchida, S., Ishibashi, J., Kataoka, S., Tsunogai, U., Okamura, K., Sano, Y., Shinjo, R., 2001. Chemical characteristics of newly discovered black smoker fluids and associated hydrothermal plumes at the Rodriguez triple junction, Central Indian Ridge. Earth Planet. Sci. Lett. 193, 371–379. https://doi.org/10.1016/S0012-821x(01)00511-8.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Jamieson, A.J., 2015. The Hadal Zone: Life in the Deepest Oceans. Cambridge Univ. Press.

Jang, S.-J., Ho, P.-T., Jun, S.-Y., Kim, D., Won, Y.-J., 2020. A newly discovered *Gigantidas* bivalve mussel from the Onnuri Vent Field on the northern Central Indian Ridge. Deep-Sea Res. I Oceanogr. Res. Pap. https://doi.org/10.1016/j.dsr.2020.103299.

Jones, W.J., Vrijenhoek, R.C., 2006. Evolutionary relationships within the "*Bathymodiolus*" *childressi* group. Cah. Biol. Mar. 47, 403–407.

Jones, W.J., Won, Y.J., Maas, P.A.Y., Smith, P.J., Lutz, R.A., Vrijenhoek, R.C., 2006. Evolution of habitat use by deep-sea mussels. Mar. Biol. 148, 841–851. https://doi.org/10.1007/s00227-005-0115-1.

Lemer, S., Bieler, R., Giribet, G., 2019. Resolving the relationships of clams and cockles: dense transcriptome sampling drastically improves the bivalve tree of life. Proc. R. Soc. B Biol. Sci. 286, 20182684. https://doi.org/10.1098/rspb.2018.2684.

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

Lorion, J., Buge, B., Cruaud, C., Samadi, S., 2010. New insights into diversity and evolution of deep-sea Mytilidae (Mollusca: Bivalvia). Mol. Phylogenet. Evol. 57, 71–83. https://doi.org/10.1016/j.ympev.2010.05.027.

Nakamura, K., Takai, K., 2015. Geochemical constraints on potential biomass sustained by subseafloor water–rock interactions. In: Ishibashi, J., Okino, K., Sunamura, M. (Eds.), Subseafloor Biosphere Linked to Hydrothermal Systems. Springer, Tokyo, pp. 11–30. https://doi.org/10.1007/978-4-431-54865-2_2.

Rogers, A.D., Tyler, P.A., Connelly, D.P., Copley, J.T., James, R., Larter, R.D., Linse, K., Mills, R.A., Garabato, A.N., Pancost, R.D., Pearce, D.A., Polunin, N.V., German, C.R., Shank, T., Boersch-Supan, P.H., Alker, B.J., Aquilina, A., Bennett, S.A., Clarke, A., Dinley, R.J., Graham, A.G., Green, D.R., Hawkes, J.A., Hepburn, L., Hilario, A., Huvenne, V.A., Marsh, L., Ramirez-Llodra, E., Reid, W.D., Roterman, C.N., Sweeting, C.J., Thatje, S., Zwirglmaier, K., 2012. The discovery of new deep-sea hydrothermal vent communities in the Southern Ocean and implications for biogeography. PLoS Biol. 10, e1001234 https://doi.org/10.1371/journal.pbio.1001234.

Sharma, P.P., Zardus, J.D., Boyle, E.E., González, V.L., Jennings, R.M., McIntyre, E., Wheeler, W.C., Etter, R.J., Giribet, G., 2013. Into the deep: a phylogenetic approach to the bivalve subclass Protobranchia. Mol. Phylogenet. Evol. 69, 188–204. https://doi.org/10.1016/j.ympev.2013.05.018.

Sigwart, J.D., Wicksten, M.K., Jackson, M.G., Herrera, S., 2019. Deep-sea video technology tracks a monoplacophoran to the end of its trail (Mollusca, Tryblidia). Mar. Biodivers. 49, 825–832. https://doi.org/10.1007/s12526-018-0860-2.

Taylor, J.D., Glover, E.A., 2010. Chemosymbiotic bivalves. In: Kiel, S. (Ed.), The Vent and Seep Biota: Aspects from Microbes to Ecosystems. Springer, Dordrecht, pp. 107–135. https://doi.org/10.1007/978-90-481-9572-5_5.

Taylor, J.D., Williams, S.T., Glover, E.A., Dyal, P., 2007. A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes. Zool. Scr. 36, 587–606. https://doi.org/10.1111/j.1463-6409.2007.00299.x.

Thubaut, J., Puillandre, N., Faure, B., Cruaud, C., Samadi, S., 2013. The contrasted evolutionary fates of deep-sea chemosynthetic mussels (Bivalvia, Bathymodiolinae). Ecol. Evol. 3, 4748–4766. https://doi.org/10.1002/ece3.749.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B., 2012. Gplots: various R programming tools for plotting data. http://CRAN.R-project.org/package=gplots.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35, 543–548. https://doi.org/10.1093/molbev/msx319.

Xu, T., Sun, J., Watanabe, H.K., Chen, C., Nakamura, M., Ji, R., Feng, D., Lv, J., Wang, S., Bao, Z., Qian, P.Y., Qiu, J.W., 2018. Population genetic structure of the deep-sea mussel *Bathymodiolus platifrons* (Bivalvia: Mytilidae) in the Northwest Pacific. Evol. Appl. 11, 1915–1930. https://doi.org/10.1111/eva.12696.

Xu, T., Feng, D., Tao, J., Qiu, J.W., 2019. A new species of deep-sea mussel (Bivalvia: Mytilidae: *Gigantidas*) from the South China Sea: morphology, phylogenetic position, and gill-associated microbes. Deep-Sea Res. I Oceanogr. Res. Pap. 146, 79–90. https://doi.org/10.1016/j.dsr.2019.03.001.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., Gilbert, J.A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B.W., Blaser, M.J., Bonazzi, V., Booth, T., Bork, P., Bushman, F.D., Buttigieg, P.L., Chain, P.S.G., Charlson, E., Costello, E.K., Huot-Creasy, H., Dawyndt, P., DeSantis, T., Fierer, N., Fuhrman, J.A., Gallery, R.E., Gevers, D., Gibbs, R.A., Gil, I.S., Gonzalez, A., Gordon, J.I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A.L., Kelley, S.T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C.L., Legg, T., Ley, R.E., Lozupone, C.A., Ludwig, W., Lyons, D., Maguire, E., Methé, B.A., Meyer, F., Muegge, B., Nakielny, S., Nelson, K.E., Nemergut, D., Neufeld, J.D., Newbold, L.K., Oliver, A.E., Pace, N.R., Palanisamy, G., Peplies, J., Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D.A., Assunta-Sansone, S., Schloss, P.D., Schriml, L., Sinha, R., Smith, M.I., Sodergren, E., Spor, A., Stombaugh, J., Tiedje, J.M., Ward, D.V., Weinstock, G.M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J.R., Yatsunenko, T., Glöckner, F.O., 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. 29, 415–420. https://doi.org/10.1038/nbt.1823.