ZR Zoological Research

Supplementary Materials

Phylogenetic context of a deep-sea clam (Bivalvia: Vesicomyidae) revealed by DNA from 1 500-years-old shells

Yi-Xuan Li[1,2], Yanjie Zhang[3], Jack Chi-Ho Ip[1], Jing Liu[4], Chong Chen[5], Crispin T. S. Little[6,7], Yusuke Yokoyama[8], Moriaki Yasuhara[9,10], Jian-Wen Qiu[1,2,*]

[1] Department of Biology, Hong Kong Baptist University, Hong Kong, China
[2] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, Guangdong, China
[3] School of Life Sciences, Hainan University, Haikou, Hainan 570228, China
[4] Laboratory of Marine Organism Taxonomy and Phylogeny, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, Shandong 266071, China
[5] X-STAR, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan
[6] School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK
[7] Life Sciences Department, Natural History Museum, London SW7 5BD, UK
[8] Department of Earth and Planetary Science, The University of Tokyo, Hongo, 113-0033, Japan
[9] School of Biological Sciences, Area of Ecology and Biodiversity,-Swire Institute of Marine Science, Institute for Climate and Carbon Neutrality, and Musketeers Foundation Institute of Data Science, The University of Hong Kong, Hong Kong, China
[10] State Key Laboratory of Marine Pollution, City University of Hong Kong, Hong Kong SAR, China

[*]Corresponding author, E-mail: qiujw@hkbu.edu.hk

Supplementary Note

Supplementary Figure S1. Shell photos of *Archivesica nanshaensis*: holotype collected from southern South China Sea (A, shell length = 6.5 cm) and shells collected from northeastern South China Sea (B-D) used in this study.

Supplementary Figure S2. Length distribution of reads (A) and Mapdamage fragment misincorporation (B) plot of *Archivesica nanshaensis* collapsed nuclear reads mapped against the *A. nanshaensis* assembly of its nuclear reads.

Supplementary Figure S3. *In situ* photograph showing an extinct population of *A. nanshaensis* on the sediment surface of the sampling site.

Supplementary Figure S4. The ML and BI phylogenetic tree of 58 vesicomyid species constructed using the concatenated dataset of *COI*, *H3*, *18S* and *28S*.

Supplementary Table S1. Results of carbon isotope analysis of *Archivesica nanshaensis* shells.

Supplementary Table S2. Results of *Archivesica nanshaensis* shell DNA sequencing and mapping against the nuclear genome and mitochondrial genome of *A. marissinica* using bwa-mem2[1] and BBmap[2] after removing duplicates.

Supplementary Table S3. The blast results of *A. nanshaensis* mitochondrial DNA contigs against the mitochondrial genome of *A. marissinica*.

Supplementary Table S4. The blast results of *A. nanshaensis* nuclear DNA contigs against the *18S* rRNA and *28S* rRNA sequences of *A. marissinica*.

Supplementary Table S5. The annotation description of 859 *A. marissinica* (Ama) annotated contigs mapped against nuclear reads of *A. nanshaensis*.

Supplementary Table S6. Predicted coding genes of *A. nanshaensis* annotated by eggNOG database and nt database.

Supplementary Table S7. Pairwise K2P (%) distances among selected vesicomyid species based on the *COⅠ* dataset. Table S14 contains data from more species of vesicomyids.

Supplementary Table S8. K2P distance (%) between and within (Italic value) vesicomyid groups.

Supplementary Table S9. Results of mapping of *A. nanshaensis* shell DNA contigs against the symbiont genome of *A. marissinica* and NCBI's nt database.

Supplementary Table S10. De novo assembly results of each *Archivesica nanshaensis* shell DNA sample.

Supplementary Table S11. GenBank accession numbers for Vesicomyidae species included in the phylogenetic analysis.

Supplementary Table S12. Best-fit models identified using ModelFinder for each partition used for phylogenetic analyses.

Supplementary Table S13. Information of Illumina sequencing data using in this study.

Supplementary Table S14. Pairwise K2P (%) distances among all vesicomyid species based on the *COⅠ* dataset.

Supplementary Table S15. Mitochondrial sequences recovered from *Archivesica nanshaensis* shells.

Supplementary Tables S14-S15 are listed as a separate excel file due to their large size.

Supplementary Note

Shell sample collection and age determination

Three pairs of articulated *Archivesica nanshaensis* shells (samples Anan_1, 2, 3) were collected from the South China Sea (N21°18.5114′, E119°11.9134′) at 3,003 m water depth on 8 May 2018 (Figure 1 and S1). They were sampled using a scoop net from *ROV* ROPOS on board *R/V* Tan Kah Kee. The studied shell samples do not originate from protected species. The specimens were identified by morphological comparison with the holotype of *A. nanshaensis* (Figure S1). Morphological comparison between the *A. nanshaensis* and congener *A. marissinica* has been included in Chen et al. (2018), which revealed the differences in the maximum shell size, hinge structure, dorsal margin curvature, and acuteness of the posterior end. Besides, *A. marissinica* is distributed in the Haima cold seep in northwestern South China Sea (SCS) at a shallower depth (~ 1400 m depth), whereas *A. nanshaensis* has been found in deeper waters of northeastern SCS (2 626 m) and southwestern SCS (3 003 m) shown at Figure 1.

We conducted AMS (Accelerator Mass Spectrometry) radiocarbon dating on fragments of two valves (samples Anan_1 and 2) at the Atmosphere and Ocean Research Institute, The University of Tokyo (Yokoyama et al., 2019). A small piece (~500 μg) was taken from each specimen and used for dating. The radiocarbon dates were calibrated to calendar ages by OxCal 4.4. (Ramsey, 2009) using the Marine20 dataset (Heaton et al., 2020) and the global standard reservoir age ($\Delta R = 0$). The $\delta^{13}C_{PDB}$ value of *Archivesica nanshaensis* shell samples Anan_1 and Anan_2 were -2.8 ± 0.5‰ and 0.5 ± 0.7‰, respectively (Table S1). The close to zero $\delta^{13}C_{PDB}$ values indicate that the source of carbon for shell formation in this species was ultimately atmospheric CO_2, therefore age determination of the shells should not be affected by methane-derived carbon that might otherwise confound the age determination using the ^{14}C method (Adkins et al., 2002; Ambrose et al., 2015).

Shell DNA extraction, library preparation and sequencing

Three valves of *A. nanshaensis* (one from each of the three pairs available) were used for shell DNA extraction using a procedure adopted from Kemp et al. (2007). The shell DNA extractions were performed in a UV light sterilized, HEPA filtered, dusk free laminar air flow hood to reduce the risk of contamination. The shells were rinsed using Milli-Q water to remove sediment and other potential contaminants, immersed in 1% sodium hypochlorite solution for 10 min, rinsed again using Milli-Q water, then air-dried in a fume hood. The dried shells were pounded to powder with a mortar and pestle and filtered with a 200-mesh Nylon cloth. Approximately 0.6 g of shell powder from each sample was transferred to a 50 mL Falcon tube containing 30 mL 0.5 M EDTA buffer (pH = 8.0). The mixture was gently shaken for 24 h at room temperature on a rotating mixer, then the Falcon tubes were centrifuged at 5,000 ×g for 10 min using a centrifuge (Eppendorf Model 5804, Hamburg, Germany). The supernatant was discarded and the pellet was redissolved in 2 mL 2× CTAB buffer (100 mM Tris-HCl in pH 8.0, 1.4 M NaCl, 20 mM EDTA in pH 8.0, 2% CTAB, 2% PVP40). The dissolved shell sample was transferred into two sterilized 1.5 mL Eppendorf tubes and digested by overnight incubation at 60°C with 50 μL Proteinase K (20 mg/mL), the final digestion mixture was centrifuged at 10 000 rpm for 5 min, and the supernatant was transferred to four new sterilized 1.5 mL tubes. A 500 μL solution of phenol:chloroform:isoamyl alcohol (25:24:1) was added to each digested shell solution (500 μL), and the tubes were mixed gently on a rotating mixer for 10 min, then centrifuged at 12 000 rpm for 10 min. The supernatant was transferred to new sterilized tubes with 400 μL of chloroform:isoamyl alcohol (24:1), then mixed and centrifuged as in the previous step. After centrifugation, the supernatant was transferred to new sterilized tubes containing 300 μL absolute isopropanol and 3 M NaAc (10% volume as the total volume of supernatant), vortexed, and incubated overnight at -20°C.

The DNA was precipitated by centrifugation at 12,000 rpm for 10 min. The solution was discarded, 1 mL 80% cold ethanol was added, and the tubes were centrifuged at 12,000 rpm for 15 min. The supernatant was then discarded, and the tube was air-dried in a fume hood for 10 min to remove all remaining ethanol. Shell DNA in each tube was resuspended with 20 μL sterilized Milli-Q water. For each shell sample, DNA solution with a final volume of 80 μL was stored at -80°C until use. The shell DNA was qualified using a Bioanalyzer (Agilent Technologies, California, USA) and quantified with a Qubit fluorometer v.4.0 (Thermo Fisher Scientific, Massachusetts, USA). Qualified DNA was used to prepare 350 bp Illumina libraries by NEB Next® Ultra™ DNA Library Prep Kit (New England Biolabs, MA, USA) according to the manufacturer's protocol without special treatment. The PCR-amplified libraries were sequenced on an Illumina Novaseq 6000 sequencer to produce paired-end 150 bp reads at Novogene (Tianjin, China).

Data trimming, quality control, and reference genome assembly

The raw data were trimmed using Trimmomatic v.0.38 (Bolger et al., 2014) with the following settings: "Illuminaclip:TruSeq3-PE; fa:2:30:10; Leading:3; Trailing:3; Slidingwindow: 4:15; Minlen:36". The filtered data were qualified using FastQC v0.11.9 (Brown et al., 2017). The clean reads were assembled using three methods: (1) *de novo* assembly: clean reads of each sample were assembled using SPAdes v.3.14 (Prjibelski et al., 2020) with setting of "--careful"(Table S10); (2) reference-guided assembly: clean reads of each sample were mapped to the nuclear genome (Ip et al., 2021) and the mitochondrial genome (assembled using raw Illumina sequencing data from Ip et al., 2021) of the congeneric vesicomyid *Archivesica marissinica* using bbsplit in BBMap v. 38 (Bushnell, 2014), and the mapped reads were assembled using SPAdes with the "--only-assembler" and "--cov-cutoff off" settings; (3) reference-guided assembly of the combined sequencing data: reads from the three shells were combined and then filtered using bbduk in BBMap to remove the viral, bacterial, and vector contaminants (databases downloaded from GenBank). The filtered reads were mapped against both a reference nuclear genome (Ip et al., 2021) and a mitochondrial genome of *A. marissinica* mentioned above, as well as a transcriptome of *Phreagena okutanii* (Lan et al., 2019) using "bbsplit.sh" and assembled using SPAdes, the same way as for the three individual samples. To obtain better assembly results by reference-guide assembly methods, we tested two mapping methods, bwa-mem2 (Vasimuddin et al., 2019) and BBMap, in order to select the more sensitive and effective one. The assembly processed by BBMap was chosen to continue to generate contigs, as it was faster and more sensitive.

The sequencing of DNA extracted from the three shells of *A. nanshaensis* produced 75.2–89.0 million reads, among them 98.9–99.2% of reads passed quality check for downstream analyses (Table S2). Apart from *de novo* assembly, we also applied two methods to filter our reads before conducting the assembly. First, applying bwa-mem2 resulted in only 2.06% to 5.98% mapping rates against the reference genome of *A. marissinica* (Table S2) and decontamination filtered a few reads, when compared to without filtration. Second, applying BBMap resulted in only 2,980 (0.004%) to 29,670 (0.035%) mapped reads. Through comparing the three different assembly methods of shell DNA, we found that two steps should be implemented before assembly, including filtering reads from contaminants and mapping reads to genomic data of a closely related species. Data filtering and sensitive mapping using BBMap also reduced redundant reads, allowing for better recovery of authentic endogenous DNA.

To extract mitochondrial and nuclear genes for phylogenetic analyses, the assembled *A. nanshaensis* contigs were used for BLASTn analysis under an e-value threshold of "1e-10" in BLAST v.2.10 against the mitochondrial genome (OM397546), *18S* rRNA gene (OM914736) and *28S* rRNA gene (OM914741) of *A. marissinica*. The result of mapping against the *A. marissinica* mitochondrial genome was visualized using Chloroplot (Zheng et al., 2020). To identify more nuclear genes from the *A. nanshaensis* assemblies, the reads were mapped to the available vesicomyid transcriptomes of *A. marissinica* (Ip et al., 2021) and *Phreagena okutanii* (Lan et al., 2019) using bbsplit, and the results were searched using diamond v.2.0.10 (Buchfink et al., 2021) "blastx" under the "fast" mode, and annotated by comparison with the *A. marissinica* genome.

DNA damage analysis

Post-mortem DNA damage was assessed using the Paleomix pipeline v.1.6.1 (Schubert et al., 2014). The filtered nuclear reads were mapped to the reference-guided assembly of *A. nanshaensis* using bwa v.0.5.9 (Li & Durbin, 2009) with the "backtrack", "no seed" and mapping quality over 30 (Q30) settings, after collapsing the overlapping reads as suggested by the manual of Paleomix. The reads were further filtered to remove PCR duplicates using Picard v.2.26.3 (http://broadinstitute.github.io/picard/). The sequencing coverage and depth of each sample mapped to the *A. nanshaensis* assembly generated by assembly method 3 (reference-guided assembly of the combined sequencing data) mentioned in the previous section were calculated using "flagstat" and "depth" in samtools v.1.14 (Danecek et al., 2021). The DNA fragmentation, base misincorporation and fragment size distributions in shell DNA were assessed using MapDamage v.2 (Jónsson et al., 2013) with 100 000 downstream sampling.

Genetic distances, phylogenetic relationships, and dating of divergence times

A 1,653-bp *COⅠ* dataset was used to calculate the pairwise genetic distances between vesicomyid species. The Kimura 2-parameter distance (K2P) model was run with 1,000 bootstrap using MEGA v.11 (Tamura et al., 2021).

To determine the phylogenetic position of *A. nanshaensis* within the Pliocardiinae, mitochondrial *CO I* and nuclear genes (*H3*, *18S* and *28S*) recovered from the shells were aligned separately using "AUTO" setting of MAFFT v.7.490 (Katoh et al., 2002) with the corresponding pliocardiine and outgroup sequences downloaded from GenBank (Table S11). The single-gene datasets were each trimmed using trimAl v.1.3 (Capella-Gutiérrez et al., 2009) with the setting of "gappyout" to retain the conserved segments. The retained segments were then combined using SeqKit v. 0.11.0 (Shen et al., 2016) with the function of "concat". The 4 525 bp concatenated dataset was exported to ModelFinder v.1.5.4 (Kalyaanamoorthy et al., 2017) to obtain best-fit models (Table S3) for Maximum Likelihood (ML) and Bayesian Inference (BI) phylogenetic analyses. The ML analysis was conducted using IQ-TREE v.2 (Minh et al., 2020) with 10 000 ultrabootstrap and visualised using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The BI analysis was conducted by Mrbayes v.3.2.7 (Ronquist et al., 2012) under partition model with 1,000,000 generations,100 sample frequency and two parallel runs. The initial 25% of sampled data were discarded as burn-in and the convergence diagnostic of Estimated Sample Size (ESS) were over 200. The phylogenetic tree of vesicomyids was rooted with the clade of *Vesicomya* and Johnson et al. (2017).

To estimate the divergence times among various vesicomyids, two fossils were used for calibration as in Johnson et al. (2017): the oldest known pliocardiine based on the fossil species *Archivesica* cf. *tschudi* (47 ± 2 Ma) and the first emergence of the *Calyptogena*, based on the fossil species *Calyptogena katallaensis* (30 ± 7 Ma) (Kiel & Amano, 2010). BEAST v.2.6.6 (Bouckaert et al., 2019) was used to estimate the divergence time was estimated and tree topology of 4 525 bp concatenated dataset examined under the following settings: (1) partition model of the three genes as the site model (Table S12); (2) "Relaxed Clock Log Normal" as the clock model with estimated clock rate; (3) "Yule model" as the tree prior and fossil calibration dates were set in the *Calyptogena* clade ("Log normal" distribution, M parameter = 3.4, S parameter = 0.108) and the pliocardiine clade ("Log normal" distribution, M parameter = 3.85, S parameter = 0.021); (4) running chain length as 10 000 000 MCMC generations with 1 000 sample frequencies. The default settings were adopted for other parameters. The convergence of the likelihood parameters in the log file was determined by Tracer v.1.7.2 (Rambaut et al., 2018) with the ESS over 200. The final tree file was generated using TreeAnnotator (Rambaut & Drummond, 2015) with 25% burn-in and plotted with FigTree.

Annotation of protein coding genes, and discovery of nuclear gene markers

We annotated the assembly of *A. nanshaensis* using Maker pipeline v.2.32 (Holt & Yandell, 2011) following Ip et al. (2021). The assembly was "soft-masked" using RepeatMasker v.2.1 (http://www.repeatmasker.org/) with the repeat libraries derived from Ip et al. (2021). Augustus v3.1 (Stanke & Morgenstern, 2005) was used to predict genes in the repeat-masked genome sequences, the genome of *A. marissinica* was inputted as the protein homology evidence and the minimum contig length was set as 300 bp. Results from different gene predictors were integrated into a consensus weighted annotation by EVidenceModeler v1.1.1 (Haas et al., 2008) with "augustus = 5, evmprot = 10" and default settings for other parameters in the software. The predicted genes of *A. nanshaensis* were functionally annotated by eggNOG-MAPPER v2 (Cantalapiedra et al., 2021) and contigs belong to Eukaryota and Metazoa are shown in Table S6.

We collected and assembled available Illumina sequencing data from vesicomyids (Table S13), including RNA-seq and metagenomic data, to obtain more nuclear genes. Those sequencing data were filtered using "bbduck.sh" to remove contaminates, especially symbionts and other bacteria, and then assembled using SPAdes v.3.14 (Prjibelski et al., 2020) for metagenomic data or Trinity v.2.13.2 (Haas et al., 2013) for RNA-seq data.

Furthermore, for these nuclear contigs, the predicted genes of *A. nanshaensis* were used to do blastn analysis under a e-value threshold of "1e-10" against the assembly of 12 vesicomyids species (Table S13). The blast results were filtered by removing records in low identity, short sequence length and repeated. The filtered genes were further identified by nt database and four nuclear markers (*ITS1*, *H3*, *ACTB*, *C482*). *ITS1* and *H3* are common nuclear markers for phylogenetic studies of bivalves and aDNA. *ACTB* is a *β-Actin* gene and *C482* is a hypothetical coding gene of *A. nanshaensis*. *C482* is a function-unknown contig but this repeat was found can be mapped to repeats in the chromosomes of *A. marissinica* and *Mercenaria mercenaria*. We finally identified four nuclear genes from each assembly of 12 vesicomyids, including *ITS1* for seven species, *H3* for 10 species, *ACTB* for seven species, and *C482* for 10 species (Table S2). These gene sequences were not available for other vesicomyid species on the databases, except for a few records of *ITS1* and *H3*

(Goffredi et al., 2003; Johnson et al., 2017; Kojima et al., 2006; Sharma et al., 2013; Wood et al., 2007). Nevertheless, these sequences represent a potential resource for other applications, including a broader phylogenetic analysis with other molluscs.

References

Adkins JF, Griffin S, Kashgarian M, et al. 2002. Radiocarbon dating of deep-sea corals. *Radiocarbon*, **44**(2): 567–580.

Ambrose WG, Panieri G, Schneider A, et al. 2015. Bivalve shell horizons in seafloor pockmarks of the last glacial-interglacial transition: a thousand years of methane emissions in the Arctic Ocean. *Geochemistry*, *Geophysics*, *Geosystems*, **16**(12): 4108–4129.

Audzijonyte A, Krylova EM, Sahling H, et al. 2012. Molecular taxonomy reveals broad trans-oceanic distributions and high species diversity of deep-sea clams (Bivalvia: Vesicomyidae: Pliocardiinae) in chemosynthetic environments. *Systematics and Biodiversity*, **10**(4): 403–415.

Bieler R, Mikkelsen PM, Collins TM, et al. 2014. Investigating the Bivalve Tree of Life - an exemplar-based approach combining molecular and novel morphological characters. *Invertebrate Systematics*, **28**(1): 32–115.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15): 2114–2120.

Bouckaert R, Vaughan TG, Barido-Sottani J, et al. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **15**(4): e1006650.

Brown J, Pirrung M, McCue LA. 2017. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, **33**(19): 3137–3139.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, **18**(4): 366–368.

Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. https://sourceforge.net/projects/bbmap/.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, et al. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, **38**(12): 5825–5829.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15): 1972–1973.

Chen C, Okutani T, Liang QY, et al. 2018. A noteworthy new species of the family Vesicomyidae from the South China Sea (Bivalvia: Glossoidea). *Venus* (*Journal of the Malacological Society of Japan*), **76**(1–4): 29–37.

Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2): giab008.

Decker C, Olu K, Cunha RL, et al. 2012. Phylogeny and diversification patterns among vesicomyid bivalves. *PLoS One*, **7**(4): e33359.

Goffredi S, Hurtado L, Hallam S, et al. 2003. Evolutionary relationships of deep-sea vent and cold seep clams (Mollusca: Vesicomyidae) of the "*pacifica*/*lepta*" species complex. *Marine Biology*, **142**(2): 311–320.

Haas BJ, Papanicolaou A, Yassour M, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**(8): 1494–1512.

Haas BJ, Salzberg SL, Zhu W, et al. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**(1): R7.

Heaton TJ, Köhler P, Butzin M, et al. 2020. Marine20—the marine radiocarbon age calibration curve (0–55,000 cal BP). *Radiocarbon*, **62**(4): 779–820.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, **12**: 491.

Ip JCH, Xu T, Sun J, et al. 2021. Host–Endosymbiont Genome Integration in a Deep-Sea Chemosymbiotic Clam. *Molecular Biology and Evolution*, **38**(2): 502–518.

Jónsson H, Ginolhac A, Schubert M, et al. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, **29**(13): 1682–1684.

Johnson SB, Krylova EM, Audzijonyte A, et al. 2017. Phylogeny and origins of chemosynthetic vesicomyid clams. *Systematics and Biodiversity*, **15**(4): 346–360.

Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**(6): 587–589.

Katoh K, Misawa K, Kuma KI, et al. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**(14): 3059–3066.
Kiel S, Amano K. 2010. Oligocene and Miocene vesicomyid bivalves from the Katalla district, Southern Alaska. *The Veliger*, **51**(1): 76–84.
Kemp BM, Malhi RS, McDonough J, et al. 2007. Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *American Journal of Physical Anthropology*, **132**(4): 605–621.
Kojima S, Fujikura K, Okutani T. 2004. Multiple trans-Pacific migrations of deep-sea vent/seep-endemic bivalves in the family Vesicomyidae. *Molecular Phylogenetics and Evolution*, **32**(1): 396–406.
Kojima S, Tsuchida E, Numanami H, et al. 2006. Synonymy of *Calyptogena solidissima* with *Calyptogena kawamurai* (Bivalvia: Vesicomyidae) and its population structure revealed by mitochondrial DNA sequences. *Zoological Science*, **23**(10): 835–842.
Krylova EM, Kamenev GM, Vladychenskaya IP, et al. 2015. Vesicomyinae (Bivalvia: Vesicomyidae) of the Kuril–Kamchatka Trench and adjacent abyssal regions. *Deep Sea Research Part II: Topical Studies in Oceanography*, **111**: 198–209.
Lan Y, Sun J, Zhang W, et al. 2019. Host–symbiont interactions in deep-sea chemosymbiotic vesicomyid clams: Insights from transcriptome sequencing. *Frontiers in Marine Science*, **6**: 1–13.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.
Liu HL, Cai SY, Liu J, et al. 2018. Comparative mitochondrial genomic analyses of three chemosynthetic vesicomyid clams from deep-sea habitats. *Ecology and Evolution*, **8**(15): 7261–7272.
Liu HL, Cai SY, Zhang HB, et al. 2016. Complete mitochondrial genome of hydrothermal vent clam *Calyptogena magnifica*. *Mitochondrial DNA Part A*, **27**(6): 4333–4335.
Martin AM, Goffredi SK. 2012. *'Pliocardia' krylovata*, a new species of vesicomyid clam from cold seeps along the Costa Rica Margin. *Journal of the Marine Biological Association of the United Kingdom*, **92**(5): 1127–1137.
Minh BQ, Schmidt HA, Chernomor O, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**(5): 1530–1534.
Ozawa G, Shimamura S, Takaki Y, et al. 2017. Ancient occasional host switching of maternally transmitted bacterial symbionts of chemosynthetic vesicomyid clams. *Genome Biology and Evolution*, **9**(9): 2226–2236.
Perez M, Breusing C, Angers B, et al. 2022. Divergent paths in the evolutionary history of maternally transmitted clam symbionts. *Proceedings of the Royal Society B*, **289**(1970): 20212137.
Prjibelski A, Antipov D, Meleshko D, et al. 2020. Using SPAdes de novo assembler. *Current Protocols in Bioinformatics*, **70**: e102.
Rambaut A, Drummond AJ. 2015. TreeAnnotator v1.8.2. MCMC Output Anal.
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**(5): 901–904.
Ramsey CB. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon*, **51**(1): 337–360.
Rodrigues CF, Cunha MR, Olu K, et al. 2012. The smaller vesicomyid bivalves in the genus *Isorropodon* (Bivalvia, Vesicomyidae, Pliocardiinae) also harbour chemoautotrophic symbionts. *Symbiosis*, **56**(3): 129–137.
Ronquist F, Teslenko M, van der Mark P, et al. 2012. MRBAYES 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**(3): 539–542.
Schubert M, Ermini L, Der Sarkissian C, et al. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, **9**(5): 1056–1082.
Sharma PP, Zardus JD, Boyle EE, et al. 2013. Into the deep: a phylogenetic approach to the bivalve subclass Protobranchia. *Molecular Phylogenetics and Evolution*, **69**(1): 188–204.
Shen W, Le S, Li Y, et al. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file Manipulation. *PLoS One*, **11**(10): e0163962.
Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, **33**(S2): W465–W467.
Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, **38**(7): 3022–3027.
Taylor JD, Williams ST, Glover EA, et al. 2007. A molecular phylogeny of heterodont bivalves (Mollusca: Bivalvia: Heterodonta): new analyses of 18S and 28S rRNA genes. *Zoologica Scripta*, **36**(6): 587–606

Valdés F, Sellanes J, D'Elía G, 2012. Phylogenetic position of vesicomyid clams from a methane seep off central chile (~36 °S) with a molecular timescale for the diversification of the vesicomyidae. *Zoological Studies*, **51**(7): 1154–1164.

Vasimuddin M, Misra S, Li H, et al. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *In*: IEEE International Parallel and Distributed Processing Symposium (IPDPS). Rio de Janeiro: IEEE, 314–324.

Wiklund H, Taylor JD, Dahlgren TG, et al. 2017. Abyssal fauna of the UK-1 polymetallic nodule exploration area, Clarion-Clipperton Zone, central Pacific Ocean: Mollusca. *ZooKeys*, **707**: 1–46.

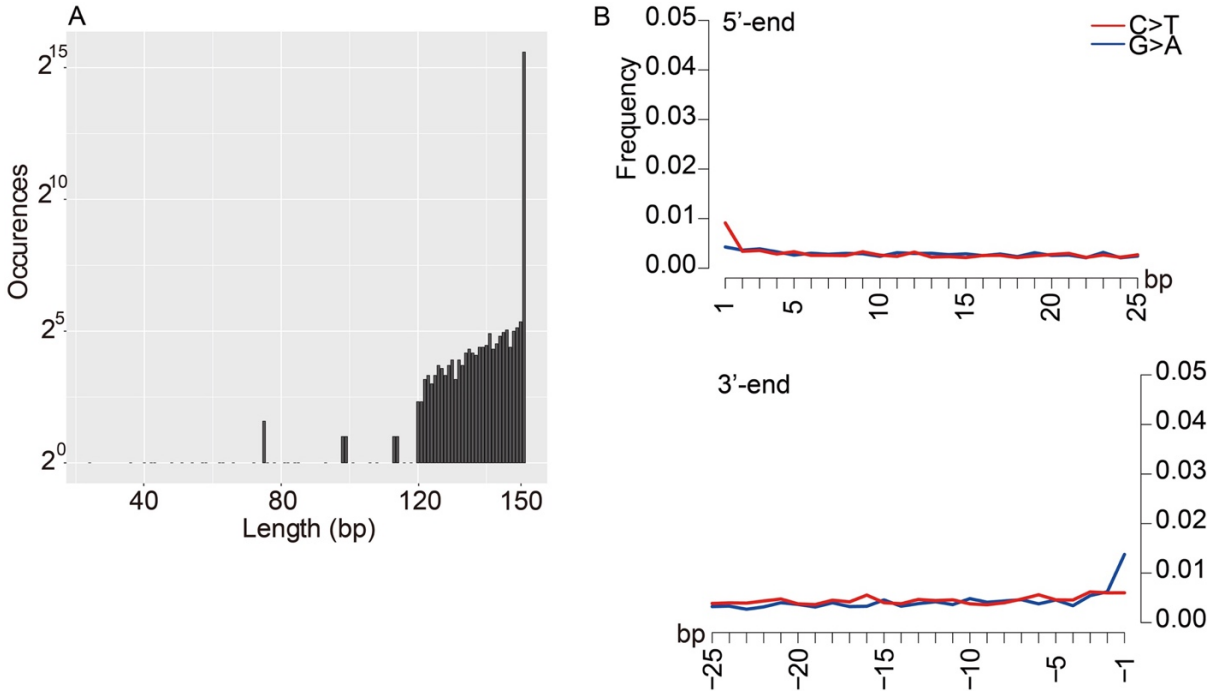
Wood AR, Apte S, MacAvoy ES, et al. 2007. A molecular phylogeny of the marine mussel genus *Perna* (Bivalvia: Mytilidae) based on nuclear (ITS1&2) and mitochondrial (COI) DNA sequences. *Molecular Phylogenetics and Evolution*, **44**(2): 685–698.

Yokoyama Y, Miyairi Y, Aze T, et al. 2019. A single stage accelerator mass spectrometry at the atmosphere and ocean research institute, the university of Tokyo. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, **455**: 311–316.

Zheng SY, Poczai P, Hyvönen J, et al. 2020. Chloroplot: an online program for the versatile plotting of organelle genomes. *Frontiers in Genetics*, **11**: 576124.

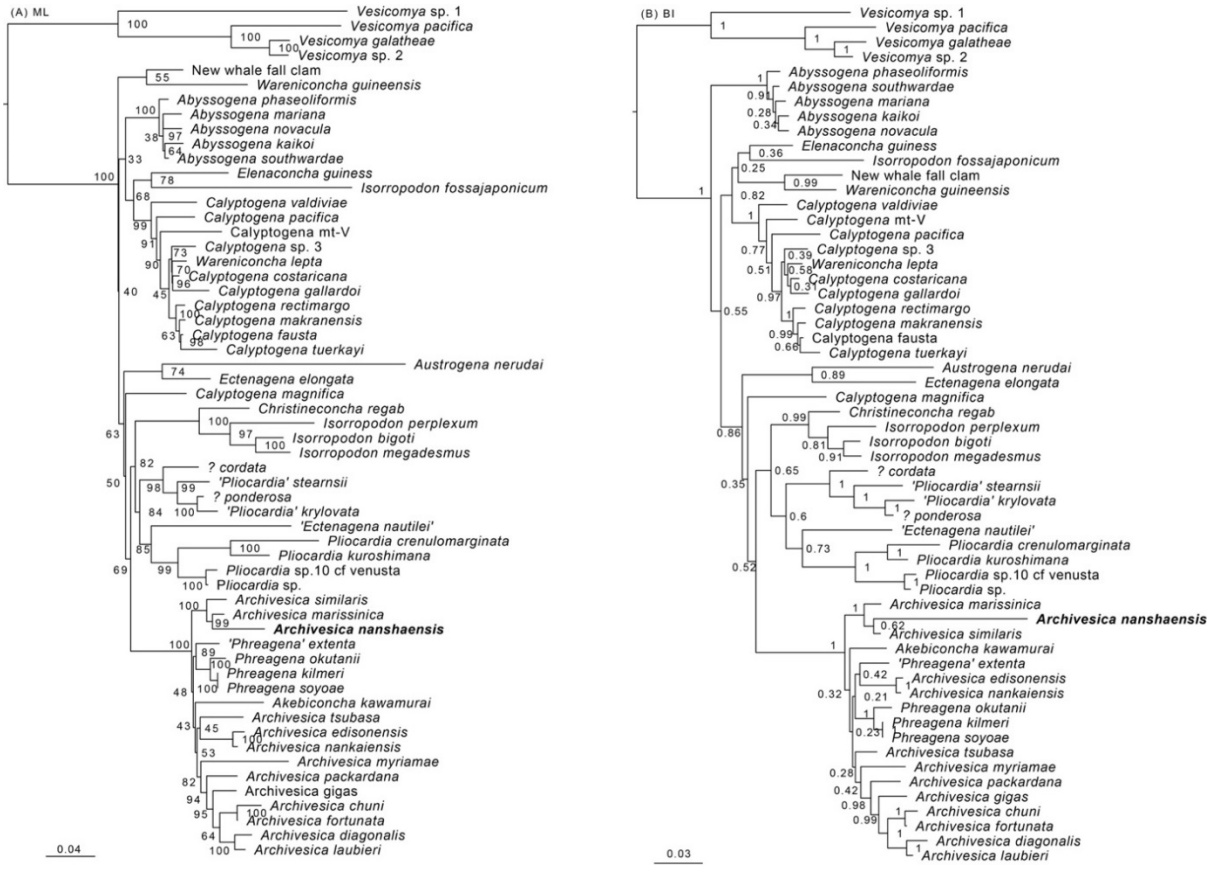

A
B Sample: Anan_1 (shell length: 6.4 cm)

Supplementary Figure S1. Shell photos of *Archivesica nanshaensis*: holotype collected from southern South China Sea (A, shell length = 6.5 cm) and shells collected from northeastern South China Sea (B-D) used in this study.



Supplementary Figure S2. Length distribution of reads (A) and Mapdamage fragment misincorporation (B) plot of *Archivesica nanshaensis* collapsed nuclear reads mapped against the *A. nanshaensis* assembly of its nuclear reads.

Supplementary Figure S3. *In situ* photograph showing an extinct population of *A. nanshaensis* on the sediment surface of the sampling site.



Supplementary Figure S4. The ML and BI phylogenetic tree of 58 vesicomyids constructed using the concatenated dataset of *COⅠ*, *H3*, *18S* and *28S*.

Value at node: BPP/Bootstrap value by BI and ML analysis

Supplementary Table S1. Results of carbon isotope analysis of *Archivesica nanshaensis* shells.

Sample	$\delta^{13}C_{PDB}$ ± Err. 1δ (‰)	^{14}C age ± Err. 1δ (yr)	pMC ± Err. 1δ (%)	$\Delta^{14}C$ ± Err. 1δ (‰)	Calibrated age (yr)	
					Mean	Range (95%)
Anan_1	-2.8 ± 0.5	2,080 ± 21	77.19 ± 0.2	-234.69 ± 1.95	1475	1616 – 1340
Anan_2	0.5 ± 0.7	2,082 ± 21	77.17 ± 0.2	-234.93 ± 2.02	1478	1618 – 1341

$\delta^{13}C_{PDB}$: carbon isotope ratio of carbonates, compared with standards used limestone called Pee Dee Belemnite; pMC: percent modern carbon, 100% pMC was defined at 1950; $\Delta^{14}C$ indicates the carbon source.

Supplementary Table S2. Results of *Archivesica nanshaensis* shell DNA sequencing and mapping against the nuclear genome and mitochondrial genome of *A. marissinica* using bwa-mem2[1] and BBMap[2] after removing duplicates.

Sample ID	Raw reads	Clean reads	Reads mapped to genome[1]	Filtered reads mapped to genome[1]	Mapping rate (%)[1]	Reads mapped to mtgenome[1]	Coverage (%)[1]	Mean depth[1]	Reads mapped to genome[2]	Reads mapped to mtgenome[2]
Anan_1	84,099,202	83,364,168	5,074,346	4,989,168	5.98	37	17.39	0.16	29,670	20
Anan_2	75,262,274	74,397,368	2,007,918	1,562,045	2.10	20	15.64	0.25	2,980	18
Anan_3	89,010,760	88,277,246	1,839,143	1,822,817	2.06	9	7.25	0.07	12,211	7

Supplementary Table S3. The blast results of *A. nanshaensis* mitochondrial DNA contigs against the mitochondrial genome of *A. marissinica*.

Sample ID	Number of contigs	Gene name	Length (bp)	Corresponding positions in the *A. marissinica* mtgenome
Anan_1	2	*CO Ⅰ*	585	4,795–5,144 5,268–5,504
Anan_2	4	*CYTB*	240	8,334–8,571
		ND4	151	11,069–11,220
		trnE-trnS-ATP6	150	11,678–11,828
		COIII	147	3,792–3,939
Anan_3	4	*CYTB*	293	7,889–8,036 8,101–8,247
		trnE-trnS-ATP6	150	11,695–11,844
		ND1-trnV-trnN	275	14,006–14,279
Combined data	7	*ND2*	290	1,663–1,953
		CO Ⅰ	566	4,814–5,162 5,248–5,464
		CYTB	267	8,334–8,600
		ATP6	178	12,320–12,497
		ND3	156	12,510–12,766
		ND1-trnV-trnN	275	14,006–14,279

Supplementary Table S4. The blast results of *A. nanshaensis* nuclear DNA contigs against the *18S* rRNA and *28S* rRNA sequences of *A. marissinica*.

Gene	Sample	Contig ID	Identity (%)	Length	Range	E value
28S	Amar	(reference)		2,136		
	Anan_1	NODE_16_length_885_cov_2.353960	96.912	421	1–421	0
	Anan_3	NODE_20_length_539_cov_2.370130	99.629	539	358–896	0
18S	Amar	(reference)		1,777		
	Anan_1	NODE_79_length_447_cov_1.456757	100	447	508–954	0
	Anan_3	NODE_42_length_410_cov_1.174174	96.35	410	233–642	0
H3	Amar	(reference)		309		
	Anan_combined	NODE_352_length_517_cov_1.920455	83	309	1–309	1e-77

#The low identity of *H3* (nucleotide) of *A. nanshaensis* might be effected by sequencing error.

Supplementary Table S5. The annotation description of 859 *A. marissinica* (Ama) annotated contigs mapped against nuclear reads of *A. nanshaensis*.

ID of mapped Ama contigs	COG category	Description
Ama38352	A	WD domain, G-beta repeat
Ama36046	A	RNA recognition motif
Ama13693	A	inductive cell-cell signaling
Ama16822	A	Belongs to the DEAD box helicase family
Ama18282	A	nucleotide binding. It is involved in the biological process described with mRNA processing
Ama16949	A	RecQ protein-like 4
Ama31938	A	helicase activity
Ama08493	A	DEAD (Asp-Glu-Ala-Asp) box polypeptide 41
Ama06338	A	double-strand break repair via alternative nonhomologous end joining
Ama27655	A	Helicase POLQ-like
Ama34273	A	RNA 3'-terminal phosphate cyclase
Ama10720	A	ATP-dependent RNA helicase activity
Ama08858	A	negative regulation of mRNA polyadenylation
Ama08787	A	cellular response to selenite ion
Ama38727	A	piRNA biosynthetic process
Ama05345	A	Interferon induced with helicase C domain 1
Ama05524	A	zinc ion binding
Ama19042	A	protein transport
Ama30140	A	Ribonuclease
Ama30317	A	Catalytic subunit of the queuine tRNA-ribosyltransferase (TGT) that catalyzes the base-exchange of a guanine (G) residue with queuine (Q) at position 34 (anticodon wobble position) in tRNAs with GU(N) anticodons (tRNA-Asp, -Asn, -His and -Tyr), resulting in the hypermodified nucleoside queuosine (7-(((4,5-cis- dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine). Catalysis occurs through a double-displacement mechanism. The nucleophile active site attacks the C1' of nucleotide 34 to detach the guanine base from the RNA, forming a covalent enzyme-RNA intermediate. The proton acceptor active site deprotonates the incoming queuine, allowing a nucleophilic attack on the C1' of the ribose to form the product
Ama00681	A	pri-miRNA transcription by RNA polymerase II
Ama38910	A	RNA secondary structure unwinding
Ama33198	A	Belongs to the DEAD box helicase family
Ama32770	A	helicase
Ama33199	A	regulation of RNA-directed 5'-3' RNA polymerase activity
Ama38245	AJ	iron-responsive element binding
Ama18111	AT	mRNA methylation
Ama37844	B	Herpesvirus alkaline exonuclease
Ama35968	B	protein heterodimerization activity
Ama36960	B	Herpesvirus alkaline exonuclease

Ama40030	B	Herpesvirus alkaline exonuclease
Ama25990	B	Core histone H2A/H2B/H3/H4
Ama26003	B	protein heterodimerization activity
Ama25989	B	Histone H2A
Ama26002	B	protein heterodimerization activity
Ama25986	B	protein heterodimerization activity
Ama18099	B	protein heterodimerization activity
Ama10675	B	Belongs to the MYST (SAS MOZ) family
Ama22274	B	protein heterodimerization activity
Ama22275	B	protein heterodimerization activity
Ama20230	B	methylated histone binding
Ama00311	B	protein heterodimerization activity
Ama00941	B	NACHT and WD repeat
Ama26191	BD	pogo transposable element with KRAB domain
Ama09434	BD	pogo transposable element with KRAB domain
Ama11643	BD	Structural maintenance of chromosomes
Ama31385	BDLT	Transformation transcription domain-associated protein
Ama27658	BK	Poly (ADP-ribose) polymerase
Ama19813	BT	Cupin-like domain
Ama36815	C	Flavoprotein (FP) subunit of succinate dehydrogenase (SDH) that is involved in complex II of the mitochondrial electron transport chain and is responsible for transferring electrons from succinate to ubiquinone (coenzyme Q)
Ama36462	C	NADP transhydrogenase
Ama25146	C	Produces ATP from ADP in the presence of a proton gradient across the membrane
Ama26057	C	Belongs to the complex I 20 kDa subunit family
Ama25981	C	L-aminoadipate-semialdehyde dehydrogenase activity
Ama13982	C	malate dehydrogenase 1B
Ama14097	C	Belongs to the mitochondrial carrier (TC 2.A.29) family
Ama32087	C	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism
Ama22883	C	Isocitrate dehydrogenase
Ama07789	C	Scaffold protein for the de novo synthesis of iron- sulfur (Fe-S) clusters within mitochondria, which is required for maturation of both mitochondrial and cytoplasmic 2Fe-2S and 4Fe-4S proteins
Ama05907	C	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (non-phosphorylating) activity
Ama05986	C	Aldehyde dehydrogenase 16 family member A1
Ama05989	C	malate dehydrogenase (decarboxylating) (NAD+) activity
Ama27897	C	Succinyl-CoA synthetase functions in the citric acid cycle (TCA), coupling the hydrolysis of succinyl-CoA to the synthesis of either ATP or GTP and thus represents the only step of substrate-level phosphorylation in the TCA. The alpha subunit of the enzyme binds the substrates coenzyme A and phosphate, while succinate binding and specificity for either ATP or GTP is provided by different beta subunits
Ama35724	C	isocitrate dehydrogenase NAD subunit

Ama35501	C	3 iron, 4 sulfur cluster binding
Ama34538	C	Acetyl-CoA hydrolase/transferase C-terminal domain
Ama35082	C	Fumarase C-terminus
Ama34666	C	cAMP biosynthetic process
Ama10194	C	dehydrogenase E1 and transketolase domain containing 1
Ama09358	C	Catalyzes a 2-step reaction, involving the ATP-dependent carboxylation of the covalently attached biotin in the first step and the transfer of the carboxyl group to pyruvate in the second
Ama09356	C	pyruvate carboxylase activity
Ama21258	C	4 iron, 4 sulfur cluster binding
Ama21464	C	mitochondrial electron transport, cytochrome c to oxygen
Ama13115	C	ATP synthase, H transporting, mitochondrial F1 complex, delta subunit
Ama12733	C	deoxyhypusine monooxygenase activity
Ama19749	C	proton-transporting ATPase activity, rotational mechanism. It is involved in the biological process described with ATP synthesis coupled proton transport
Ama30152	C	prohibitin homologues
Ama02340	C	aconitate hydratase activity
Ama33249	C	methylcrotonoyl-CoA carboxylase activity
Ama33918	C	NADH dehydrogenase (quinone) activity
Ama14471	D	Poly(A) RNA polymerase
Ama17110	D	structural constituent of cuticle
Ama23471	D	autophagy of peroxisome
Ama29913	D	Nucleotide binding
Ama40003	D	mitochondrion morphogenesis
Ama23499	DKT	aminophospholipid transmembrane transporter activity
Ama11425	DO	Cell division cycle
Ama29891	DO	protein K11-linked ubiquitination
Ama26187	DUZ	Belongs to the TRAFAC class TrmE-Era-EngA-EngB-Septin- like GTPase superfamily. Septin GTPase family
Ama06822	DZ	protein tyrosine kinase inhibitor activity
Ama36602	E	Pyrroline-5-carboxylate reductase
Ama24880	E	Ribonuclease H protein
Ama14138	E	K02A2.6-like
Ama13963	E	glycine biosynthetic process from serine
Ama17752	E	K02A2.6-like
Ama18317	E	cysteine desulfurase activity
Ama16287	E	Ribonuclease H protein
Ama32432	E	K02A2.6-like
Ama39664	E	dipeptidase activity
Ama22400	E	Ribonuclease H protein
Ama23748	E	Ribonuclease H protein
Ama22555	E	Ribonuclease H protein
Ama06857	E	homogentisate 1,2-dioxygenase activity
Ama06362	E	choline dehydrogenase activity

Ama26718	E	aminopeptidase activity
Ama09191	E	5-methyltetrahydrofolate-dependent methyltransferase activity
Ama09801	E	K02A2.6-like
Ama39522	E	K02A2.6-like
Ama21538	E	argininosuccinate lyase activity
Ama21682	E	Ribonuclease H protein
Ama29198	E	K02A2.6-like
Ama28994	E	homocysteine catabolic process
Ama28126	E	Ribonuclease H protein
Ama29274	E	Copper type II ascorbate-dependent monooxygenase, N-terminal domain
Ama05260	E	Ribonuclease H protein
Ama03615	E	cobalamin biosynthetic process
Ama05259	E	K02A2.6-like
Ama11468	E	glycine dehydrogenase (decarboxylating) activity
Ama19892	E	Nitrilase
Ama30286	E	Ribonuclease H protein
Ama30074	E	argininosuccinate synthase
Ama01497	E	K02A2.6-like
Ama01752	E	glycine biosynthetic process from serine
Ama33576	E	Ribonuclease H protein
Ama04812	EF	5-phosphoribose 1-diphosphate metabolic process
Ama31958	EH	thiamine pyrophosphate binding
Ama13075	EI	methylcrotonoyl-CoA carboxylase beta chain, mitochondrial-like
Ama32921	EI	propionyl-CoA carboxylase activity
Ama38406	EO	Peptidase family M1 domain
Ama06518	EPT	extracellularly glutamate-gated ion channel activity
Ama18943	ET	serine racemase
Ama36152	F	GMP synthase (glutamine-hydrolyzing) activity
Ama37078	F	Catalyzes the reversible transfer of the terminal phosphate group between ATP and AMP. Plays an important role in cellular energy homeostasis and in adenine nucleotide metabolism. Adenylate kinase activity is critical for regulation of the phosphate utilization and the AMP de novo biosynthesis pathways
Ama13890	F	Si dkey-6n6.2
Ama17836	F	Catalyzes the irreversible NADPH-dependent deamination of GMP to IMP. It functions in the conversion of nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and in maintaining the intracellular balance of A and G nucleotides
Ama23817	F	5,10-methylenetetrahydrofolate-dependent methyltransferase activity
Ama35535	F	phosphoribosylamine-glycine ligase activity
Ama34973	F	carbamoyl-phosphate synthase (glutamine-hydrolyzing) activity
Ama10844	F	Xanthine dehydrogenase
Ama08844	F	nucleoside transmembrane transporter activity
Ama29294	F	GTP binding

Ama05755	F	phosphoglycerate kinase
Ama03428	F	phosphoribosylformylglycinamidine synthase activity
Ama04279	F	phosphoribosylaminoimidazole carboxylase activity
Ama12560	F	dihydropyrimidine dehydrogenase (NADP+) activity
Ama19496	F	IMP dehydrogenase activity
Ama01122	F	negative regulation of oxidative stress-induced intrinsic apoptotic signaling pathway
Ama33816	G	cellulose 1,4-beta-cellobiosidase activity
Ama38035	G	pyruvate kinase activity
Ama14078	G	Glucoside xylosyltransferase
Ama15099	G	Belongs to the ribulose-phosphate 3-epimerase family
Ama14407	G	positive regulation of TOR signaling
Ama17916	G	GDP-mannose 4,6-dehydratase activity
Ama17949	G	Glucose-6-phosphate
Ama17649	G	positive regulation of TOR signaling
Ama23361	G	positive regulation of TOR signaling
Ama26781	G	positive regulation of TOR signaling
Ama27707	G	Iduronidase, alpha-L
Ama35031	G	enolase 2
Ama10760	G	Glycosyl hydrolase family 9
Ama09004	G	magnesium ion binding
Ama09947	G	aldose 1-epimerase activity
Ama05200	G	N-terminal region of glycosyl transferase group 7
Ama04643	G	N-terminal region of glycosyl transferase group 7
Ama04537	G	N-terminal region of glycosyl transferase group 7
Ama12929	G	Transferase activity, transferring glycosyl groups. It is involved in the biological process described with carbohydrate metabolic process
Ama19118	G	pyrimidine nucleotide-sugar transmembrane transporter activity
Ama18343	G	glucose-6-phosphate dehydrogenase activity
Ama29821	G	trimming of terminal mannose on B branch
Ama30489	G	major facilitator superfamily
Ama02860	G	Glycosyl hydrolase family 20, catalytic domain
Ama26817	GO	nucleoside-diphosphatase activity
Ama17942	H	S-adenosylhomocysteine catabolic process
Ama23619	H	porphobilinogen synthase activity
Ama24258	H	Integrase core domain
Ama21111	H	Phenylalanine ammonia-lyase
Ama12665	H	Catalytic histone acetyltransferase subunit of the RNA polymerase II elongator complex, which is a component of the RNA polymerase II (Pol II) holoenzyme and is involved in transcriptional elongation
Ama02586	H	methionine adenosyltransferase activity
Ama02587	H	methionine adenosyltransferase activity
Ama37298	I	sphingomyelin phosphodiesterase
Ama36893	I	GPI anchor biosynthetic process
Ama36449	I	3-hydroxy-lignoceroyl-CoA dehydratase activity

Ama24974	I	queuine tRNA-ribosyltransferase activity
Ama25280	I	3-hydroxy-3-methylglutaryl-coenzyme A reductase
Ama14073	I	queuine tRNA-ribosyltransferase activity
Ama15696	I	MHC class II protein binding, via antigen binding groove
Ama16563	I	CRAL/TRIO domain
Ama16590	I	ATPase activity, coupled to transmembrane movement of substances
Ama22869	I	Phosphatidic acid phosphatase type 2B
Ama34855	I	queuine tRNA-ribosyltransferase activity
Ama10854	I	queuine tRNA-ribosyltransferase activity
Ama22043	I	3-oxoacyl- acyl-carrier-protein synthase
Ama22012	I	lathosterol oxidase activity
Ama28434	I	very-long-chain-acyl-CoA dehydrogenase activity
Ama05507	I	Phytanoyl-CoA dioxygenase (PhyH)
Ama05389	I	Putative methyltransferase
Ama11837	I	It is involved in the biological process described with acetyl-CoA biosynthetic process from acetate
Ama12226	I	Belongs to the thiolase family
Ama11836	I	acetyl-CoA biosynthetic process from acetate
Ama19615	I	ATPase activity, coupled to transmembrane movement of substances
Ama20293	I	Carboxylesterase family
Ama29898	I	Phytanoyl-CoA dioxygenase (PhyH)
Ama33657	I	acyl-CoA synthetase long-chain family member 3b
Ama14301	IQ	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama16758	IQ	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama22281	IQ	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama38274	J	protein phosphatase regulator activity
Ama36078	J	protein phosphatase regulator activity
Ama37114	J	regulation of response to stimulus
Ama37010	J	translation initiation factor activity
Ama24960	J	cytoplasmic translation
Ama25290	J	elongation factor Tu
Ama24643	J	threonyl-tRNA aminoacylation
Ama25012	J	protein phosphatase regulator activity
Ama15665	J	protein phosphatase regulator activity
Ama15443	J	Ribosomal protein L14p/L23e
Ama15800	J	protein phosphatase regulator activity
Ama15594	J	protein phosphatase regulator activity
Ama17935	J	Belongs to the class-I aminoacyl-tRNA synthetase family
Ama17141	J	60S ribosomal protein L13
Ama16614	J	Catalyzes the attachment of alanine to tRNA(Ala) in a two-step reaction alanine is first activated by ATP to form Ala- AMP and then transferred to the acceptor end of tRNA(Ala). Also edits incorrectly charged tRNA(Ala) via its editing domain
Ama31593	J	protein phosphatase regulator activity
Ama23157	J	Translation initiation factor

Ama22965	J	release factor 1-like
Ama39335	J	protein phosphatase regulator activity
Ama07268	J	Tryptophanyl tRNA synthetase 2, mitochondrial
Ama08232	J	Elongation factor
Ama07565	J	maturation of LSU-rRNA
Ama07490	J	protein phosphatase regulator activity
Ama27927	J	protein phosphatase regulator activity
Ama27811	J	protein phosphatase regulator activity
Ama35381	J	positive regulation of translational termination
Ama35503	J	exosome component 10
Ama35007	J	Carboxypeptidase
Ama34969	J	protein phosphatase regulator activity
Ama34484	J	protein phosphatase regulator activity
Ama09176	J	Mitochondrial GTPase that mediates the disassembly of ribosomes from messenger RNA at the termination of mitochondrial protein biosynthesis. Acts in collaboration with MRRF. GTP hydrolysis follows the ribosome disassembly and probably occurs on the ribosome large subunit. Not involved in the GTP-dependent ribosomal translocation step during translation elongation
Ama10817	J	polyribonucleotide nucleotidyltransferase activity. It is involved in the biological process described with mRNA catabolic process
Ama10480	J	Ribosomal protein S12/S23
Ama20832	J	aspartyl-tRNA synthetase 2, mitochondrial
Ama21135	J	protein phosphatase regulator activity
Ama21973	J	mitochondrial tRNA wobble uridine modification
Ama29116	J	protein phosphatase regulator activity
Ama28801	J	Ribosomal protein P0 is the functional equivalent of E.coli protein L10
Ama29529	J	methionyl-tRNA formyltransferase activity. It is involved in the biological process described with
Ama28217	J	Methionyl-tRNA formyltransferase
Ama03065	J	rRNA binding
Ama05374	J	GTP binding
Ama04782	J	Promotes mitochondrial protein synthesis. May act as a fidelity factor of the translation reaction, by catalyzing a one- codon backward translocation of tRNAs on improperly translocated ribosomes. Binds to mitochondrial ribosomes in a GTP-dependent manner
Ama11969	J	Belongs to the class-I aminoacyl-tRNA synthetase family
Ama12434	J	protein phosphatase regulator activity
Ama19242	J	Eukaryotic translation initiation factor 5A
Ama19315	J	translation elongation factor activity
Ama19014	J	ribosomal protein S7
Ama19130	J	glutamine-tRNA ligase activity
Ama30534	J	It is involved in the biological process described with seryl-tRNA aminoacylation
Ama30818	J	ribosomal protein
Ama29883	J	ATP-binding cassette, subfamily F, member

Ama30508	J	RNA-binding component of the eukaryotic translation initiation factor 3 (eIF-3) complex, which is involved in protein synthesis of a specialized repertoire of mRNAs and, together with other initiation factors, stimulates binding of mRNA and methionyl-tRNAi to the 40S ribosome. The eIF-3 complex specifically targets and initiates translation of a subset of mRNAs involved in cell proliferation. This subunit can bind 18S rRNA
Ama30119	J	Belongs to the class-I aminoacyl-tRNA synthetase family
Ama01597	J	rRNA export from nucleus
Ama01210	J	Ribosomal protein L3
Ama33433	J	ribosomal protein
Ama33894	J	translation elongation factor activity
Ama36024	K	Ligand binding domain of hormone receptors
Ama24417	K	zinc ion binding
Ama25079	K	binding. It is involved in the biological process described with steroid hormone mediated signaling pathway
Ama14798	K	homeobox
Ama17820	K	DNA binding. It is involved in the biological process described with regulation of transcription, DNA-templated
Ama17609	K	CCR4-NOT transcription complex subunit 1
Ama31639	K	RNA polymerase I CORE element sequence-specific DNA binding
Ama07172	K	helix loop helix domain
Ama06300	K	carbohydrate response element binding
Ama07654	K	DNA-templated transcription, initiation
Ama06523	K	negative regulation of protein localization to nucleolus
Ama26851	K	regulation of transcription from RNA polymerase II promoter involved in spinal cord motor neuron fate specification
Ama26278	K	high mobility group
Ama26724	K	SRY (sex determining region Y)-box 30
Ama27767	K	Domain first found in the mice T locus (Brachyury) protein
Ama26512	K	DNA binding
Ama10602	K	90 kDa subunit
Ama10672	K	DNA- binding
Ama05282	K	negative regulation of type B pancreatic cell apoptotic process
Ama05273	K	nuclear factor related to
Ama12201	K	maintenance of DNA methylation
Ama12168	K	SAM domain binding
Ama18835	K	canonical Wnt signaling pathway involved in negative regulation of apoptotic process
Ama39137	K	transcription corepressor activity
Ama30202	K	DNA binding
Ama29919	K	DNA-templated transcriptional preinitiation complex assembly
Ama02344	K	regulation of nucleic acid-templated transcription
Ama33206	K	It is involved in the biological process described with regulation of sequence-specific DNA binding transcription factor activity
Ama31497	KL	E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA damage. It is unclear

		whether it also mediates the formation of other types of polyubiquitin chains. The E3 ubiquitin-protein ligase activity is required for its tumor suppressor function. The BRCA1-BARD1 heterodimer coordinates a diverse range of cellular pathways such as DNA damage repair, ubiquitination and transcriptional regulation to maintain genomic stability. Regulates centrosomal microtubule nucleation. Required for normal cell cycle progression from G2 to mitosis. Required for appropriate cell cycle arrests after ionizing irradiation in both the S-phase and the G2 phase of the cell cycle. Involved in transcriptional regulation of P21 in response to DNA damage. Required for FANCD2 targeting to sites of DNA damage. May function as a transcriptional regulator
Ama08544	KL	Helicase-like transcription factor
Ama08755	KLO	ankyrin repeats
Ama35790	KLT	MyD88-dependent toll-like receptor signaling pathway
Ama37894	L	transposition, RNA-mediated
Ama37994	L	Integrase core domain
Ama36209	L	It is involved in the biological process described with nucleobase-containing compound metabolic process
Ama36753	L	transcription, DNA-templated
Ama36303	L	Integrase core domain
Ama24948	L	Helix-turn-helix of DDE superfamily endonuclease
Ama25745	L	transposition, RNA-mediated
Ama15061	L	double-stranded DNA 3'-5' exodeoxyribonuclease activity
Ama15728	L	Integrase core domain
Ama13747	L	ATP-dependent 5'-3' DNA helicase activity
Ama14380	L	protein heterodimerization activity
Ama15480	L	retrotransposable element Tf2 155 kDa protein type 1-like
Ama15535	L	DNA replication initiation
Ama14059	L	Integrase core domain
Ama14075	L	K02A2.6-like
Ama16335	L	transposition
Ama17025	L	transposition, RNA-mediated
Ama32173	L	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)
Ama31279	L	transposition, RNA-mediated
Ama32002	L	transcription, DNA-templated
Ama23520	L	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama22452	L	protein heterodimerization activity
Ama23961	L	transposition, RNA-mediated
Ama08408	L	PNMA
Ama06287	L	transposition, RNA-mediated
Ama06858	L	DDE superfamily endonuclease
Ama07037	L	protein heterodimerization activity
Ama35322	L	nuclease activity
Ama10786	L	Integrase core domain
Ama09181	L	Integrase core domain

Ama09052	L	transcription, DNA-templated
Ama10835	L	transposition, RNA-mediated
Ama09367	L	Helix-turn-helix of DDE superfamily endonuclease
Ama39509	L	hydrolase activity, acting on ester bonds
Ama20976	L	THAP
Ama22165	L	Integrase core domain
Ama29074	L	DDE superfamily endonuclease
Ama28491	L	Integrase core domain
Ama28937	L	DNA replication initiation
Ama29075	L	nuclease activity
Ama04215	L	protein heterodimerization activity
Ama05157	L	Integrase core domain
Ama03015	L	PNMA
Ama03016	L	Integrase core domain
Ama04039	L	Conserved hypothetical protein
Ama11692	L	Integrase core domain
Ama20037	L	protein heterodimerization activity
Ama19995	L	DNA replication initiation
Ama30953	L	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama29833	L	protein heterodimerization activity
Ama30909	L	Integrase core domain
Ama00673	L	Serine threonine-protein kinase SMG1
Ama00972	L	transposition, RNA-mediated
Ama33366	L	Integrase core domain
Ama33539	L	NFX1-type zinc finger-containing protein
Ama33331	L	Integrase core domain
Ama28513	M	Transferase activity, transferring hexosyl groups
Ama12972	M	UDP-N-acetylglucosamine 4-epimerase activity
Ama02284	M	It is involved in the biological process described with carbohydrate biosynthetic process
Ama38096	O	calcium ion binding
Ama38492	O	DnaJ heat shock protein family (Hsp40) member A1
Ama37558	O	Plasmin dissolves the fibrin of blood clots and acts as a proteolytic factor in a variety of other processes including embryonic development, tissue remodeling, tumor invasion, and inflammation. In ovulation, weakens the walls of the Graafian follicle. It activates the urokinase-type plasminogen activator, collagenases and several complement zymogens, such as C1 and C5. Cleavage of fibronectin and laminin leads to cell detachment and apoptosis. Also cleaves fibrin, thrombospondin and von Willebrand factor. Its role in tissue remodeling and tumor invasion may be modulated by CSPG4. Binds to cells
Ama36656	O	Cathepsin C exclusion domain
Ama24416	O	Caseinolytic mitochondrial matrix peptidase chaperone subunit
Ama24897	O	Zinc ion binding
Ama25835	O	Removes the N-terminal methionine from nascent proteins. The N-

		terminal methionine is often cleaved when the second residue in the primary sequence is small and uncharged (Met-Ala-, Cys, Gly, Pro, Ser, Thr, or Val)
Ama24980	O	peptidylamidoglycolate lyase activity
Ama25180	O	FKBP-type peptidyl-prolyl cis-trans isomerase
Ama25874	O	ATP-dependent serine protease that mediates the selective degradation of misfolded, unassembled or oxidatively damaged polypeptides as well as certain short-lived regulatory proteins in the mitochondrial matrix. May also have a chaperone function in the assembly of inner membrane protein complexes. Participates in the regulation of mitochondrial gene expression and in the maintenance of the integrity of the mitochondrial genome. Binds to mitochondrial DNA in a site-specific manner
Ama26166	O	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama15332	O	MreB/Mbl protein
Ama13974	O	ATP-dependent zinc metalloprotease
Ama16810	O	mitochondrial protein processing
Ama16624	O	lytic vacuole organization
Ama16992	O	ATP-dependent peptidase activity
Ama17139	O	zinc ion binding
Ama16348	O	endopeptidase inhibitor activity
Ama32069	O	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama32080	O	protein import into mitochondrial intermembrane space
Ama23325	O	regulation of vesicle fusion
Ama24353	O	regulation of protein catabolic process
Ama06283	O	organic anion transmembrane transporter activity
Ama08273	O	microtubule-severing ATPase activity
Ama07095	O	Peroxiredoxin 4
Ama08020	O	Small ubiquitinrelated modifier
Ama26940	O	E3 ubiquitin-protein ligase DTX3L
Ama34251	O	zinc ion binding
Ama34781	O	E3 ubiquitin-protein ligase TRIM56
Ama35473	O	proteasome-activating ATPase activity
Ama34947	O	G-protein coupled receptor activity
Ama35470	O	proteasome-activating ATPase activity
Ama09070	O	biogenesis factor 6
Ama10587	O	Belongs to the peptidase S8 family
Ama08903	O	Dienelactone hydrolase family
Ama10619	O	regulation of defense response to virus by host
Ama09028	O	MreB/Mbl protein
Ama10277	O	respiratory chain complex III assembly
Ama10452	O	stress-induced-phosphoprotein 1
Ama08855	O	Heat shock protein
Ama08646	O	Belongs to the AAA ATPase family
Ama08797	O	ubiquitin
Ama20951	O	Matrixin

Ama22252	O	protein K6-linked deubiquitination
Ama29240	O	Calcium ion binding. It is involved in the biological process described with protein folding
Ama28804	O	ATP binding
Ama29019	O	homolog subfamily C member 7
Ama28288	O	Tripartite motif-containing protein
Ama28082	O	zinc ion binding
Ama04178	O	SUMO binding
Ama05548	O	ATP binding
Ama02973	O	heat shock protein
Ama05140	O	ubiquitinyl hydrolase activity
Ama03635	O	regulation of axon guidance
Ama04775	O	carbohydrate derivative binding
Ama04809	O	Ubiquitin-conjugating enzyme
Ama03362	O	translational attenuation
Ama04869	O	aryl hydrocarbon receptor interacting
Ama05544	O	ATP binding
Ama11628	O	Belongs to the AAA ATPase family
Ama13536	O	peptidyl-prolyl cis-trans isomerase activity
Ama11571	O	protein modification by small protein conjugation
Ama11720	O	ATP-dependent Clp protease proteolytic subunit
Ama12018	O	peptidyl-prolyl cis-trans isomerase activity
Ama18749	O	negative regulation of IRE1-mediated unfolded protein response
Ama19179	O	Belongs to the peptidase M16 family
Ama19910	O	heat shock protein 90kDa beta (Grp94), member 1
Ama19617	O	Mib_herc2
Ama19317	O	peptidyl-prolyl cis-trans isomerase activity
Ama30531	O	negative regulation of IRE1-mediated unfolded protein response
Ama30532	O	negative regulation of IRE1-mediated unfolded protein response
Ama01645	O	A-macroglobulin receptor
Ama00843	O	serine-type endopeptidase activity. It is involved in the biological process described with proteolysis
Ama02187	O	protein modification by small protein conjugation
Ama32668	O	Caseinolytic mitochondrial matrix peptidase chaperone subunit
Ama29902	OT	blood vessel endothelial cell proliferation involved in sprouting angiogenesis
Ama32815	OT	Belongs to the peptidase C2 family
Ama08173	OU	K02A2.6-like
Ama13913	P	sodium:potassium-exchanging ATPase activity
Ama17612	P	Transporter
Ama31690	P	solute carrier family 4, sodium bicarbonate
Ama27885	P	iduronate 2-sulfatase
Ama21840	P	Solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 2
Ama29219	P	Structural component of the gap junctions

Ama04889	P	secondary active sulfate transmembrane transporter activity
Ama20038	P	delayed rectifier potassium channel activity
Ama39937	P	Cation-transporting atpase
Ama14150	PQ	cuticle development
Ama14658	PT	Receptor family ligand binding region
Ama02486	PT	Lipoxygenase homology
Ama01324	PT	calcium channel activity
Ama37996	Q	ATPase activity, coupled to transmembrane movement of substances
Ama36380	Q	Hephaestin-like protein 1
Ama17771	Q	N,N-dimethylaniline monooxygenase activity
Ama31192	Q	It is involved in the biological process described with transmembrane transport
Ama32539	Q	ATPase activity, coupled to transmembrane movement of substances
Ama32177	Q	ATPase activity, coupled to transmembrane movement of substances
Ama08330	Q	ATP-binding cassette subfamily B
Ama08329	Q	ATP-binding cassette subfamily B
Ama28197	Q	ATPase activity, coupled to transmembrane movement of substances
Ama28225	Q	ATPase activity, coupled to transmembrane movement of substances
Ama11270	Q	oxidoreductase activity
Ama19204	Q	ATPase activity, coupled to transmembrane movement of substances
Ama20297	Q	ATPase activity, coupled to transmembrane movement of substances
Ama02570	Q	Belongs to the zinc-containing alcohol dehydrogenase family. Class-III subfamily
Ama01590	Q	Pyridine nucleotide-disulphide oxidoreductase, dimerisation domain
Ama36615	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama24571	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama16028	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama17089	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama31416	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama23393	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama34472	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama28124	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama05717	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama11822	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama18511	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama02039	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama34205	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama37845	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama37526	S	Domain of unknown function (DUF4371)
Ama38220	S	Kelch domain-containing protein 4
Ama37916	S	DDE superfamily endonuclease
Ama37722	S	DDE superfamily endonuclease
Ama37977	S	DDE superfamily endonuclease
Ama38362	S	SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain

Ama38482	S	DDE superfamily endonuclease
Ama36812	S	regulation of DNA damage checkpoint
Ama36230	S	Helix-turn-helix of DDE superfamily endonuclease
Ama36840	S	positive regulation of microtubule motor activity
Ama35909	S	transcription by RNA polymerase III
Ama36958	S	Helix-turn-helix of DDE superfamily endonuclease
Ama36034	S	Transposase IS4
Ama36698	S	RNase H
Ama36307	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama24582	S	Ribonuclease H protein
Ama25367	S	Chitin-binding domain type 2
Ama25315	S	DDE superfamily endonuclease
Ama26071	S	DDE superfamily endonuclease
Ama25567	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama24987	S	DDE superfamily endonuclease
Ama24581	S	Ribonuclease H protein
Ama24477	S	ZnF_TTF
Ama26149	S	Family with sequence similarity 172 member A
Ama26108	S	Ribonuclease H protein
Ama24840	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama24808	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama15919	S	52 kDa repressor of the inhibitor of the protein
Ama14086	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama15807	S	DDE superfamily endonuclease
Ama15255	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama15803	S	Transposase IS4
Ama14182	S	G8
Ama15608	S	MULE transposase domain
Ama14215	S	Thyroglobulin type-1 repeat
Ama14011	S	DDE superfamily endonuclease
Ama14302	S	DDE superfamily endonuclease
Ama14674	S	oxidation-reduction process
Ama14702	S	negative regulation of canonical Wnt signaling pathway involved in cardiac muscle cell fate commitment
Ama15055	S	gamma-glutamylaminecyclotransferase activity
Ama17191	S	zinc finger
Ama17510	S	HEPN domain
Ama16408	S	Metal ion binding
Ama16315	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama16663	S	carbohydrate binding
Ama16658	S	signal transduction involved in G2 DNA damage checkpoint
Ama16290	S	Zinc finger, AN1-type domain 1
Ama17386	S	Helix-turn-helix of DDE superfamily endonuclease
Ama16284	S	Transcriptional repressor TCF25

Ama18030	S	BtpA family
Ama16578	S	positive regulation of protein localization to cilium
Ama16296	S	axonemal central apparatus assembly
Ama31280	S	Transposase
Ama31730	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama31399	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama32046	S	partitioning defective 3 homolog
Ama32042	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama32119	S	positive regulation of rRNA processing
Ama32302	S	Transposase
Ama31538	S	regulation of alternative mRNA splicing, via spliceosome
Ama31485	S	Ribonuclease H protein
Ama31424	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama23282	S	Jerky protein homolog-like
Ama23266	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama23208	S	Tumour protein p53-inducible protein 11
Ama23163	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama23906	S	family with sequence similarity 78, member
Ama24168	S	52 kDa repressor of the inhibitor of the protein
Ama23212	S	activity. It is involved in the biological process described with oxidation-reduction process
Ama23492	S	Fanconi Anaemia group E protein FANCE
Ama24117	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama23562	S	Ddb1 and cul4 associated factor 7
Ama07562	S	DDE superfamily endonuclease
Ama06222	S	SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain
Ama05987	S	PIH1 domain-containing protein 1
Ama06234	S	Ribonuclease H protein
Ama06780	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama08270	S	zinc finger
Ama07115	S	Peptidase family C78
Ama07652	S	oxidation-reduction process
Ama06141	S	Zinc finger MYM-type protein
Ama06579	S	hyaluronic acid binding
Ama06792	S	Ribonuclease H protein
Ama06388	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama07563	S	Helix-turn-helix of DDE superfamily endonuclease
Ama27784	S	ubiquitin-protein transferase activity
Ama26401	S	CREB3 regulatory factor
Ama26889	S	Ribonuclease H protein
Ama26985	S	Ribonuclease H protein
Ama27646	S	Retinoic acid induced 16-like protein
Ama26234	S	YqaJ-like viral recombinase domain
Ama27826	S	Reverse transcriptase (RNA-dependent DNA polymerase)

Ama27596	S	CD40 signaling pathway
Ama27332	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama26825	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama27775	S	Stabilization of polarity axis
Ama35057	S	Endonuclease/Exonuclease/phosphatase family
Ama35684	S	HELP motif
Ama34866	S	IQ calmodulin-binding motif
Ama34585	S	DDE superfamily endonuclease
Ama35184	S	trypsinogen activation
Ama35402	S	fibroblast migration
Ama35425	S	YqaJ-like viral recombinase domain
Ama35607	S	Ribonuclease H protein
Ama35339	S	DDE superfamily endonuclease
Ama35359	S	peroxisome matrix targeting signal-1 binding
Ama34645	S	oxidation-reduction process
Ama35093	S	WWE domain
Ama34871	S	DDE superfamily endonuclease
Ama40062	S	DDE superfamily endonuclease
Ama09102	S	DDE superfamily endonuclease
Ama10056	S	Ral GTPase-activating protein subunit
Ama09851	S	carboxyl-O-methyltransferase activity
Ama10004	S	UPF0764 protein C16orf89 homolog
Ama10138	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama10982	S	oxidation-reduction process
Ama11021	S	positive regulation of actin filament depolymerization
Ama09940	S	52 kDa repressor of the inhibitor of the protein
Ama10347	S	Coiled-coil domain containing 151
Ama10706	S	7 transmembrane receptor (rhodopsin family)
Ama09079	S	Helix-turn-helix of DDE superfamily endonuclease
Ama09197	S	response to other organism
Ama10051	S	Domain of unknown function (DUF4371)
Ama09493	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama38869	S	DDE superfamily endonuclease
Ama21401	S	Inter-alpha-trypsin inhibitor heavy chain
Ama22219	S	YqaJ-like viral recombinase domain
Ama21107	S	Inhibitor of Apoptosis domain
Ama21547	S	Domain of unknown function (DUF4371)
Ama20645	S	calcium ion binding
Ama21300	S	DDE superfamily endonuclease
Ama20644	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama21232	S	DDE superfamily endonuclease
Ama21494	S	oxidation-reduction process
Ama21238	S	Endonuclease/Exonuclease/phosphatase family

Ama22089	S	PHR domain
Ama20800	S	scavenger receptor activity
Ama21857	S	DDE superfamily endonuclease
Ama29485	S	negative regulation of canonical Wnt signaling pathway involved in cardiac muscle cell fate commitment
Ama29174	S	DDE superfamily endonuclease
Ama28797	S	carbohydrate binding
Ama28924	S	Helix-turn-helix of DDE superfamily endonuclease
Ama28554	S	ZnF_TTF
Ama28125	S	zinc finger
Ama04269	S	Inhibitor of Apoptosis domain
Ama05127	S	Ribonuclease H protein
Ama04336	S	methylated histone binding
Ama04864	S	Transposase DDE domain
Ama03993	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama04017	S	eel-Fucolectin Tachylectin-4 Pentaxrin-1 Domain
Ama04392	S	Neurexophilin and PC-esterase domain family, member 3
Ama04033	S	oxidation-reduction process
Ama05306	S	Protein of unknown function (DUF229)
Ama03123	S	52 kDa repressor of the inhibitor of the protein
Ama04028	S	MULE transposase domain
Ama05450	S	YqaJ-like viral recombinase domain
Ama03641	S	Transposase
Ama04598	S	DDE superfamily endonuclease
Ama03053	S	C2 calcium-dependent domain containing 3
Ama05268	S	protein localization to perinuclear region of cytoplasm
Ama05522	S	Ribonuclease H protein
Ama05067	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama05391	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama05474	S	Transmembrane protein 115
Ama03516	S	zinc finger
Ama04247	S	Protein of unknown function (DUF229)
Ama11770	S	zinc finger
Ama11588	S	7 transmembrane receptor (rhodopsin family)
Ama11702	S	NAD/NADP octopine/nopaline dehydrogenase, alpha-helical domain
Ama13087	S	Enhancer of
Ama12363	S	Potential calcium-dependent cell-adhesion protein. May be involved in the establishment and maintenance of specific neuronal connections in the brain
Ama12839	S	Ribonuclease H protein
Ama12896	S	DDE superfamily endonuclease
Ama13496	S	PiggyBac transposable element-derived protein
Ama12769	S	ZnF_TTF
Ama11677	S	metal ion transport

Ama19977	S	Endonuclease/Exonuclease/phosphatase family
Ama19391	S	Domain of unknown function (DUF4062)
Ama18641	S	nuclear membrane organization
Ama18570	S	Ribonuclease H protein
Ama19567	S	Pre-rRNA-processing protein TSR2 homolog
Ama19059	S	protocadherin
Ama18522	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama20117	S	Helix-turn-helix of DDE superfamily endonuclease
Ama20258	S	Ectodermal ciliogenesis protein
Ama38808	S	52 kDa repressor of the inhibitor of the protein
Ama39143	S	Transposase
Ama30355	S	Protein of unknown function (DUF2475)
Ama30270	S	DDE superfamily endonuclease
Ama30265	S	hAT family C-terminal dimerisation region
Ama30402	S	Guanine nucleotide binding protein-like 1
Ama30847	S	Ribonuclease H protein
Ama39542	S	ZnF_TTF
Ama02356	S	Thyroglobulin type-1 repeat
Ama00654	S	regulation of otic vesicle morphogenesis
Ama02146	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama01980	S	DDE superfamily endonuclease
Ama00773	S	DNA-binding transcription factor activity, RNA polymerase II-specific
Ama00897	S	Helix-turn-helix of DDE superfamily endonuclease
Ama02495	S	Transposase
Ama00557	S	YqaJ-like viral recombinase domain
Ama01574	S	Von Willebrand factor A
Ama02673	S	regulation of ligase activity
Ama02603	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama01920	S	Leucine Rich repeat
Ama01249	S	protein C4orf22 homolog
Ama01943	S	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
Ama00809	S	GTP binding
Ama33103	S	Transposase
Ama32678	S	Helix-turn-helix of DDE superfamily endonuclease
Ama33118	S	Transposase protein
Ama33925	S	suppressors of cytokine signalling
Ama34112	S	Endonuclease/Exonuclease/phosphatase family
Ama33388	S	Transposase IS4
Ama33235	S	zinc-binding in reverse transcriptase
Ama33523	S	Reverse transcriptase (RNA-dependent DNA polymerase)
Ama37126	T	transmembrane receptor protein tyrosine kinase activity
Ama36978	T	Citron rho-interacting serine threonine kinase

Ama36858	T	netrin receptor activity
Ama25110	T	heparin binding
Ama24813	T	Neuronal acetylcholine receptor subunit
Ama24988	T	Adenylyl- / guanylyl cyclase, catalytic domain
Ama15635	T	Mucin 4, cell surface associated
Ama15016	T	Camp-dependent protein kinase type
Ama13939	T	Deleted in malignant brain tumors 1
Ama15639	T	sequestering of TGFbeta in extracellular matrix
Ama15770	T	carbohydrate binding
Ama13750	T	Domain found in Dishevelled, Egl-10, and Pleckstrin (DEP)
Ama38601	T	semaphorin receptor activity
Ama16553	T	Forkhead associated domain
Ama18109	T	protein serine/threonine phosphatase activity
Ama17016	T	inositol 1,4,5-trisphosphate-sensitive calcium-release channel activity
Ama16858	T	Serine threonine-protein kinase PRP4 homolog
Ama31662	T	Hemicentin 1
Ama32427	T	7 transmembrane receptor (Secretin family)
Ama31261	T	Protein tyrosine kinase
Ama31682	T	DBD domain binding
Ama22915	T	Ferric-chelate reductase 1
Ama08119	T	G-protein coupled receptor activity
Ama07614	T	carbohydrate binding
Ama06073	T	protein tyrosine phosphatase activity
Ama07482	T	Patched family
Ama06084	T	Protein tyrosine phosphatase, catalytic domain
Ama08071	T	Complement receptor type
Ama27145	T	tetratricopeptide
Ama27194	T	fibroblast growth factor-activated receptor activity
Ama27482	T	Belongs to the G-protein coupled receptor 1 family
Ama35714	T	neurotransmitter:sodium symporter activity
Ama34408	T	von Willebrand factor D and EGF
Ama35165	T	Coagulation factor 5/8 C-terminal domain, discoidin domain
Ama09153	T	Guanyl-nucleotide exchange factor activity. It is involved in the biological process described with small GTPase mediated signal transduction
Ama09817	T	Si dkey-234i14.3
Ama10865	T	suppression of tumorigenicity 5
Ama10019	T	myosin-light-chain-phosphatase activity
Ama10575	T	motor activity
Ama10696	T	cell division
Ama21325	T	regulation of replicative cell aging
Ama20992	T	Modified RING finger domain
Ama20341	T	mitogen-activated protein kinase

Ama20799	T	scavenger receptor activity
Ama22026	T	serine threonine-protein phosphatase
Ama22191	T	C-terminal of Roc, COR, domain
Ama22080	T	nucleosome binding
Ama20868	T	serine threonine-protein phosphatase
Ama28196	T	glycolipid transporter activity
Ama28283	T	glycolipid transporter activity
Ama29510	T	sequestering of TGFbeta in extracellular matrix
Ama29489	T	sequestering of TGFbeta in extracellular matrix
Ama04503	T	protein tyrosine phosphatase activity
Ama04667	T	protein tyrosine phosphatase activity
Ama03548	T	negative regulation of canonical Wnt signaling pathway involved in cardiac muscle cell fate commitment
Ama04574	T	protein tyrosine phosphatase activity
Ama04442	T	extracellularly glycine-gated chloride channel activity
Ama04430	T	protein tyrosine phosphatase activity
Ama03544	T	negative regulation of canonical Wnt signaling pathway involved in cardiac muscle cell fate commitment
Ama03377	T	cAMP biosynthetic process
Ama13386	T	Protein kinase domain
Ama12145	T	calcium ion binding
Ama11906	T	prostaglandin E receptor 4 (subtype EP4)
Ama11792	T	Protein tyrosine phosphatase activity. It is involved in the biological process described with protein dephosphorylation
Ama12964	T	calcium ion binding
Ama11601	T	stromal membrane-associated protein
Ama38733	T	C-terminal of Roc, COR, domain
Ama18693	T	Serine/Threonine protein kinases, catalytic domain
Ama19187	T	regulation of replicative cell aging
Ama18668	T	positive regulation of endothelial cell chemotaxis
Ama19647	T	GTPase-activator protein for Rho-like GTPases
Ama19466	T	neurotransmitter:sodium symporter activity
Ama18355	T	caveola assembly
Ama18991	T	negative regulation of canonical Wnt signaling pathway involved in cardiac muscle cell fate commitment
Ama19215	T	calcium-independent protein kinase C activity
Ama30182	T	bundle of His cell to Purkinje myocyte communication
Ama30682	T	neurogenic locus notch homolog protein
Ama30190	T	perineurial glial growth
Ama30460	T	Phosphotyrosine-binding domain
Ama30081	T	containing protein, X-linked 2
Ama02126	T	7 transmembrane receptor (rhodopsin family)
Ama02802	T	cellular response to sorbitol
Ama01694	T	neural precursor cell proliferation
Ama01253	T	Bone morphogenetic protein 3

Ama00425	T	Serine threonine-protein phosphatase 2A catalytic subunit
Ama02751	T	receptor
Ama33939	T	It is involved in the biological process described with protein phosphorylation
Ama10350	TU	1-phosphatidylinositol binding
Ama28684	TU	protein domain specific binding
Ama38091	TV	carbohydrate binding
Ama07521	TV	biological adhesion
Ama09392	TW	Cadherin repeats.
Ama29060	TZ	Catenin (cadherin-associated protein), beta 1
Ama37465	U	Golgi vesicle fusion to target membrane
Ama36068	U	protein transport
Ama36054	U	transmembrane 9 superfamily member
Ama39106	U	calcium ion transmembrane transport
Ama26114	U	extracellular ligand-gated ion channel activity
Ama16612	U	hematopoietic progenitor cell differentiation
Ama10213	U	positive regulation of natural killer cell degranulation
Ama21543	U	Syntaxin-like protein
Ama05204	U	alpha-N-acetylglucosaminidase activity
Ama03719	U	Golgi organization
Ama12277	U	endosome to melanosome transport
Ama12026	U	Sar guanyl-nucleotide exchange factor activity
Ama12280	U	protein complex 3, delta 1 subunit
Ama13223	U	N-acetylglucosaminidase, alpha
Ama29904	U	ADP-ribosylation factor-like
Ama02241	U	Belongs to the small GTPase superfamily. Rho family
Ama02222	U	Belongs to the small GTPase superfamily. Rho family
Ama02204	U	clathrin adaptor activity
Ama22464	UY	structural constituent of nuclear pore
Ama27315	UY	nuclear export signal receptor activity
Ama07304	V	Beta-lactamase
Ama38347	W	X11-like protein binding
Ama37087	W	Bystin
Ama17415	W	negative regulation of dendritic cell antigen processing and presentation
Ama27850	W	Complement component 1, q subcomponent-like 2
Ama11370	W	laminin subunit
Ama13481	W	regulation of embryonic development
Ama30429	W	collagen
Ama37238	Z	Myosin VB
Ama13734	Z	translation initiation factor activity
Ama15452	Z	dynein light chain binding
Ama31999	Z	structural constituent of cytoskeleton
Ama32468	Z	structural constituent of cytoskeleton

Ama31199	Z	ATP-dependent microtubule motor activity, minus-end-directed
Ama23410	Z	structural constituent of cytoskeleton
Ama22961	Z	microtubule plus-end binding
Ama23411	Z	structural constituent of cytoskeleton
Ama07598	Z	inner dynein arm assembly
Ama07594	Z	heavy chain 1
Ama06171	Z	ATP-dependent microtubule motor activity, minus-end-directed
Ama07088	Z	structural constituent of cytoskeleton
Ama26730	Z	structural constituent of cytoskeleton
Ama27626	Z	cytoskeleton organization
Ama35084	Z	structural constituent of cytoskeleton
Ama09788	Z	actinin, alpha
Ama10474	Z	Pecanex protein (C-terminus)
Ama10228	Z	cytoskeletal anchoring at nuclear membrane
Ama10552	Z	Hydrolytic ATP binding site of dynein motor region D1
Ama10374	Z	response to drug
Ama21393	Z	translation initiation factor activity
Ama22228	Z	ATP-dependent microtubule motor activity, minus-end-directed
Ama39523	Z	microtubule motor activity
Ama28485	Z	heavy chain
Ama05626	Z	Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha chain
Ama05625	Z	structural constituent of cytoskeleton
Ama05611	Z	structural constituent of cytoskeleton
Ama12215	Z	Actin
Ama12313	Z	Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha chain
Ama12213	Z	cytoskeleton organization
Ama12288	Z	cytoskeleton organization
Ama12214	Z	cytoskeleton organization
Ama38895	Z	dynein light chain binding
Ama19703	Z	structural constituent of cytoskeleton
Ama19267	Z	P-loop containing dynein motor region D3
Ama19177	Z	Dynein heavy chain 12, axonemal
Ama30595	Z	structural constituent of cytoskeleton
Ama30704	Z	structural constituent of cytoskeleton
Ama30594	Z	structural constituent of cytoskeleton
Ama30950	Z	cytoskeleton organization
Ama30549	Z	structural constituent of cytoskeleton
Ama00207	Z	Functions as component of the Arp2 3 complex which is involved in regulation of actin polymerization and together with an activating nucleation-promoting factor (NPF) mediates the formation of branched actin networks
Ama32886	Z	sperm axoneme assembly

Ama33902	Z	Filamin-type immunoglobulin domains
Ama33991	Z	structural constituent of cytoskeleton
Ama33735	Z	structural constituent of cytoskeleton

Supplementary Table S6. Predicted coding genes of *A. nanshaensis* annotated by eggNOG database and nt database.

Bold contigs: gene markers

#query	max_annot_lvl	COG_category	Description	Preferred_name	PFAMs	Notes
augustus_masked-Anan_contig_351-abinit-gene-0.0-mRNA-1	2759|Eukaryota	B	Core component of nucleosome. Nucleosomes wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template. Histones thereby play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. DNA accessibility is regulated via a complex set of post-translational modifications of histones, also called histone code, and nucleosome remodeling	-	CENP-T_C	Histone H3
augustus_masked-Anan_contig_714-abinit-gene-0.0-mRNA-1	33208|Metazoa	Z	ATP binding	ACTB	Actin	Actin beta
augustus_masked-Anan_contig_783-abinit-gene-0.0-mRNA-1	33208|Metazoa	T	Low affinity immunoglobulin epsilon Fc	FCER2	Lectin_C	-
augustus_masked-Anan_contig_685-abinit-gene-0.0-mRNA-1	33208|Metazoa	Z	Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha chain	-	Tubulin,Tubulin_C	Tubulin
augustus_masked-Anan_contig_615-abinit-gene-0.0-mRNA-1	33208|Metazoa	Z	ATP binding	ACTBL2	Actin	Actin beta
augustus_masked-	33208|Metazoa	B	Centromere kinetochore	-	CENP-T_C,Histone	Histone H4

Anan_contig_122-abinit-gene-0.0-mRNA-1			component CENP-T histone fold			
augustus_masked-Anan_contig_242-abinit-gene-0.0-mRNA-1	-	-	ITS1	ITS1	-	Identified by nt database
augustus_masked-Anan_contig_482-abinit-gene-0.0-mRNA-1	-	-	Hypothetical protein	-	-	*C482*; This gene is found repeated in multi-chromosome of *A. marissinica* and *Mercenaria mercenaria*

Supplementary Table S7. Pairwise K2P (%) distances among selected vesicomyid species based on the *CO I* dataset. Table S14 contains data from more species of vesicomyids.

	Anan	Amar	Asim	Achu	Adia	Afor	Agig	Alau	Apac	Poku	Pext	Pkil	Psoy	Atsu	Anan
Anan															
Amar	3.4														
Asim	5.8	4.2													
Achu	9.1	10.9	9.5												
Adia	10.4	11.3	8.7	5.8											
Afor	10.0	10.9	8.7	1.9	5.0										
Agig	10.0	10.1	7.1	7.0	4.2	5.4									
Alau	10.8	11.7	9.1	5.8	2.3	4.2	4.2								
Apac	11.4	10.6	8.4	8.7	8.3	7.8	7.1	7.8							
Poku	9.6	8.7	6.2	8.7	8.3	8.7	6.6	8.7	6.3						
Pext	9.6	8.7	6.2	8.7	8.3	8.7	7.5	8.7	7.5	4.6					
Pkil	10.1	9.6	6.2	11.3	10.4	10.4	8.7	10.8	8.0	3.4	7.1				
Psoy	10.1	9.6	6.2	11.3	10.4	10.4	8.7	10.8	8.0	3.4	7.1	0.0			
Atsu	11.4	10.1	6.2	9.1	7.8	8.3	7.1	7.8	6.7	6.2	6.7	6.2	6.2		
Anan	13.3	11.9	9.2	13.1	12.7	12.2	9.7	12.2	8.4	9.3	9.7	8.8	8.8	7.1	

Abbreviations: Anan, *Archivesica nanshaensis*; Amar, *Archivesica marissinica*; Psim, *Archivesica similaris*; Achu, *Archivesica chuni*; Adia, *Archivesica diagonalis*; Afor, *Archivesica fortunata*; Agig, *Archivesica gigas*; Alau, *Archivesica laubieri*; Apac, *Archivesica packardana*; Poku, *Phreagena okutanii*; Pext, *'Phreagena' extenta*; Pkil, *Phreagena kilmeri*; Psoy, *Phreagena soyoae*; Atsu, *Archivesica tsubasa*; Anan, *Archivesica nankaiensis*. Table S11 contains the sources these sequences.

Supplementary Table S8. K2P distance (%) between and within (Italic value) vesicomyid groups.

	"g1a"	"g1b"	"g1c"	"g1d"	Akka	*Abyssogena*	*Calyptogena*	*Pliocardia*	*Vesicomya*
"g1a"	*5.8*								
"g1b"	9.0	*4.5*							
"g1c"	10.0	8.5	*4.6*						
"g1d"	9.9	8.4	10.1	*7.5*					
Akka	9.8	7.9	11.2	9.2	-				
Abyssogena	12.5	11.1	12.6	12.5	12.3	*2.3*			
Calyptogena	13.3	11.4	12.4	12.2	13.0	10.3	*5.9*		
Pliocardia	14.6	13.0	16.2	14.1	15.0	11.6	13.1	*8.8*	
Vesicomya	26.0	26.1	28.6	26.6	27.2	26.4	26.9	27.4	*18.0*

Abbreviation: Akka, *Akebiconcha kawamurai*. The name of groups follows the Figure 1.

Supplementary Table S9. Results of mapping of *A. nanshaensis* shell DNA contigs against the symbiont genome of *A. marissinica* and NCBI's nt database.

Sample	Contig ID	Identity (%) against *A. marissinica*	Length	Range	E value	Reference sequence in nt database	Identity (%)	Range	E value
Anan_1	NODE_5_length_2389_cov_519.511678	86.25	1,564	958,471-956,816	0	Uncultured G-proteobacterium 16S, FJ497543.1	99.79	1-1,451	0
	NODE_7_length_2094_cov_784.355974	87.02	2,094	953,376-955,474	0	Candidatus *Reidiella endopervernicosa*, CP054491.1	91.01	1,219,830-1,221,919	0
Anan_2	NODE_10_length_2094_cov_995.046604	87.02	2,094	953,376-955,474	0	Candidatus *Reidiella endopervernicosa*, CP054491.1	91.01	1,219,830-1,221,919	0
	NODE_6_length_2572_cov_622.465731	86.25	1,564	956,816-958,271	0	*Acidihalobacter yilgarnensis*, CP017415.1	89.84	772,505-775,435	0
	NODE_1_length_3655_cov_200.631917	84.25	2,939	953,070-955,989	0	Uncultured G-proteobacterium 16S, FJ497543.1	99.79	1-1,451	0
Anan_3	NODE_7_length_2584_cov_745.101715	86.25	1,564	956,816-958,371	0	Uncultured G-proteobacterium 16S, GU302419.1	99.47	1-1,500	0
	NODE_2_length_3420_cov_419.710739	85.62	2,942	955,990-953,070	0	*Sedimenticola thiotaurini*, CP011412.1	90.24	3,180,723-3,183,643	0
	NODE_74_length_941_cov_813.344907	82.41	941	99,836-100,785	0	Candidatus *Vesicomyosocius* sp., CP072535.1	83.12	101,360-102,309	0

Supplementary Table S10. *De novo* assembly results of each *Archivesica nanshaensis* shell DNA sample.

Sample	No. assembled contigs	Total length	Contig N50
Anan_1	839,188	439,233,241	2,296
Anan_2	374,882	216,389,232	1,921
Anan_3	413,335	267,064,958	2,519

Supplementary Table S11. GenBank accession numbers for Vesicomyidae species included in the phylogenetic analysis (Mtgenome, *H3*, *18S* and *28S*) and other new discovered markers.

-: data unavailable; bold numbers: assembled sequence in this study.

Taxon	Previous taxon name in Johnson et al. (2017)	*CO I*	Mtgen-ome	*18S*	*28S*	*ITS1*	*H3*	*ACTB*	*C482*	Reference
Archivesica nanshaensis	-	**OM291911**; **OM291912**	**Table S15**	**OM914734**; **OM914735**	**OM914739**; **OM914740**	**ON817205**	**ON542508**	**ON525714**	**ON525718**	This study
Archivesica marissinica	-	**OM397546**	**OM397546**	**OM914736**	**OM914741**	-	**ON542507**	**ON525713**	**ON525717**	This study
Archivesica gigas	*Archivesica gigas*	MT947382	MT947382	KT345534	KX010162	**ON817204**	**ON542506**	-	**ON525725**	Johnson et al., 2017; Liu et al., 2018; Perez et al., 2022
Archivesica diagonalis	*Archivesica diagonalis*	MT947381	MT947381	-	KX010161	-	-	-	-	Perez et al., 2022; Johnson et al., 2017
Archivesica packardana	*Archivesica packardana*	KT345577	-	**ON532911**	**ON532896**	**ON817206**	-	**ON525727**	-	Liu et al., 2016; Johnson et al., 2017
Archivesica fortunata	*'Archivesica' fortunata*	JX196977	-	KX010204	KX010183	-	KX010154	-	-	Audzijonyte et al., 2012; Johnson et al., 2017
Archivesica laubieri	*'Archivesica' laubieri*	AP014553	AP014553	-	-	-	-	-	-	Ozawa et al., 2017
Archivesica chuni	*'Archivesica' chuni*	JN563828	-	-	-	-	-	-	-	Decker et al., 2012
Archivesica myriamae	*Laubiericoncha myriamae*	JN563829	-	-	-	-	-	-	-	Decker et al., 2012
'*Phreagena*' *extenta*	'*Phreagena*' *extenta*	MF981085	MT947388	**ON532913**	**ON532898**	**ON817207**	KX010153	**ON525728**	**ON525720**	Johnson et al., 2017; Perez et al., 2022
Archivesica nankaiensis	*'Phreagena' nankaiensis*	AB197913	-	KX010188	KX010164	-	KX010139	-	-	Kojima et al., 2004; Johnson et al., 2017
Archivesica edisonensis	*'Phreagena' edisonensis*	AB110776	-	-	-	-	-	-	-	Kojima et al., 2004
Phreagena kilmeri	*Phreagena kilmeri*	AP014552	AP014552	-	-	-	-	-	-	Ozawa et al., 2017
Phreagena okutanii	*Phreagena okutanii*	AP014555	AP014555	-	-	-	**ON542512**	**ON525716**	**ON525726**	Ozawa et al., 2017
Phreagena soyoae	*Phreagena soyoae*	AP014558	AP014558	KT345536	**ON532900**	**ON817209**	**ON542511**	-	**ON525724**	Liu et al., 2016; Ozawa et al., 2017
Archivesica tsubasa	*'Phreagena' tsubasa*	AB110753	-	-	-	-	-	-	-	Kojima et al., 2004
Archivesica similaris	*'Phreagena' similaris*	AB110751	-	-	-	-	-	-	-	Kojima et al., 2004
Akebiconcha kawamurai	*Akebiconcha kawamurai*	AP014551	AP014551	-	-	-	-	-	-	Ozawa et al., 2017

Isorropodon fossajaponicum	*Isorropodon fossajaponicum*	AP014550	AP014550	-	-	-	-	-	-	Ozawa et al., 2017
Isorropodon bigoti	*Isorropodon bigoti*	HE680074	-	-	-	-	-	-	-	Rodrigues et al., 2012
Isorropodon perplexum	*Isorropodon perplexum*	JN563832	-	-	-	-	-	-	-	Decker et al., 2012
Isorropodon megadesmus	*Isorropodon megadesmus*	HE680075	-	-	-	-	-	-	-	Rodrigues et al., 2012
'Ectenagena nautilei'	*'Ectenagena' nautilei*	AP014554	AP014554	-	-	-	-	-	-	Ozawa et al., 2017
Ectenagena elongata	*Ectenagena elongata*	MH290879	-	KT345533	KX010174	-	KX010149	-	-	Liu et al., 2016; Johnson et al., 2017; NCBI records
Elenaconcha guiness	*Elenaconcha guiness*	JN563835	-	-	-	-	-	-	-	Decker et al., 2012
Wareniconcha guineensis	*Wareniconcha guineensis*	JN563830	-	-	-	-	-	-	-	Decker et al., 2012
Calyptogena fausta	*Calyptogena fausta*	AP014549	AP014549	**ON532914**	**ON532899**	-	**ON669135**	**ON525715**	**ON525721**	Ozawa et al., 2017
Turnerocon cha magnifica	*Turnerocon cha magnifica*	KR862368	KR862368	KC429381	KC429487	-	KX010150	-	-	Sharma et al., 2013; Liu et al., 2016
Calyptogena pacifica	*Calyptogena pacifica*	AP014556	AP014556	AM774564	AM779738	**ON817208**	**ON542509**	-	**ON525722**	Taylor et al., 2007; Ozawa et al., 2017
Calyptogena rectimargo	*Calyptogena rectimargo*	MT947387	MT947387	**ON532915**	KX010165	-	**ON542510**	**ON525729**	**ON525723**	Johnson et al., 2017; Perez et al., 2022
Calyptogena costaricana	*Calyptogena costaricana*	JX196980	-	KX010192	KX010168	-	KX010143	-	-	Audzijonyte et al., 2012; Johnson et al., 2017
Wareniconcha lepta	*Wareniconcha lepta*	OK174411	-	KX010191	KX010167	-	KX010142	-	-	Audzijonyte et al., 2012; Johnson et al., 2017
Calyptogena makaranensis	*Calyptogena makaranensis*	JX196976	-	-	KX010166	-	-	-	-	Audzijonyte et al., 2012; Johnson et al., 2017
Calyptogena valdiviae	*Calyptogena valdiviae*	JN563831	-	-	-	-	-	-	-	Decker et al., 2012
Calyptogena gallardoi	*Calyptogena gallardoi*	KC164255	-	-	-	-	-	-	-	Valdés et al., 2012
Calyptogena tuerkayi	*Calyptogena tuerkayi*	FN999998	-	-	-	-	-	-	-	Krylova et al., 2015
Caplyptogena sp. 3	*Caplyptogena* sp. 3	JX196975	-	KX010193	KX010169	-	KX010144	-	-	Audzijonyte et al., 2012; Johnson et al., 2017
Calyptogena mt-V	*Calyptogena* mt-V	AF114395	-	-	-	-	-	-	-	Goffredi et al.,

										2003
Austrogena nerudai	*Austrogena nerudai*	KC87658	-	-	-	-	-	-	-	Krylova et al., 2015
Abyssogena kalkoi	*Abyssogena kalkoi*	AB110761	-	-	-	-	-	-	-	Kojima et al., 2004
Abyssogena novacula	*Abyssogena novacula*	JX196970	-	-	-	-	-	-	-	Audzijonyte et al., 2012
Abyssogena mariana	*Abyssogena mariana*	LC126311	LC126311	-	-	-	-	-	-	Ozawa et al., 2017
Abyssogena southwardae	*Abyssogena southwardae*	MT947385	MT947385	**ON532912**	**ON532897**	-	**ON542514**	-	-	Johnson et al., 2017; Perez et al., 2022
Abyssogena phaseoliformis	*Abyssogena phaseoliformis*	AP014557	AP014557	KX010196	KX010172	-	KX010151	-	-	Johnson et al., 2017; Ozawa et al., 2017; Perez et al., 2022
Pliocardia sp.	-	MT947391	MT947391	**ON532916**	**ON532901**	**ON817210**	**ON542513**	-	**ON525719**	Perez et al., 2022; This study
Pliocardia sp. 10 aff. venusta	*Pliocardia* sp. 10 aff. venusta	JX196971	-	KX010189	-	-	KX010140	-	-	Audzijonyte et al., 2012; Johnson et al., 2017
? ponderosa	*? ponderosa*	MF981084	MF981084	KT345541	KT345561	-	KX010159	-	-	NCBI records; Johnson et al, 2017
'Pilocardia' stearnsii	*'Pilocardia' stearnsii*	AP014559	AP014559	KX010206	KX010184	-	-	-	-	Johnson et al., 2017; Ozawa et al., 2017
'Pliocardia' krylovata	*'Pliocardia' krylovata*	JF784422	-	KX087165	KX010178	-	KX010160	-	-	Martin & Goffredi, 2012; Johnson et al., 2017
? cordata	*? cordata*	AF114397	-	KT345530	KX010173	-	KX010148	-	-	Liu et al., 2016; Johnson et al., 2017;
Pliocardia crenulomarginata	*Pliocardia crenulomarginata*	AB110741	-	-	-	-	-	-	-	Kojima et al., 2004
Vesicomya galatheae	*Vesicomya galatheae*	MF157509	-	MF157474	-	-	-	-	-	Wiklund et al., 2017
Vesicomya pacifica	*Vesicomya pacifica*	KM388731	-	-	-	-	-	-	-	Krylova et al., 2015
Vesicomya sp. 1	*Vesicomya* sp. 1 (as *Vesicomya atlantica*)	JX196991	-	KX010197	-	-	KX010137	-	-	Audzijonyte et al., 2012; Johnson et al., 2017
Vesicomya sp. 2	*Vesicomya* sp. 2 (as *Kelliella* sp.)	KC429129	-	KC429380	KC429485	-	KC429213	-	-	Sharma et al., 2013; Bieler et al., 2014; Johnson et al., 2017
New whale-fall clam	New whale-fall clam	KX010210	-	-	KX010185	-	KX010158	-	-	Johnson et al., 2017

#*'Pliocardia' stearnsii*, *'Pliocardia' krylovata*, we keep the quota markers as possibly new genera that have not been described yet (Johnson et al.,2017). We keep the '*Phreagena*' *extenta* because it doesn't have confirmed generic name and need additional information. *'Ectenagena' nautilei*, we change it to *'Ectenagena nautilei'*, it lacks morphological information that we cannot confirm.

Supplementary Table S12. Best-fit models identified using ModelFinder for each partition used for phylogenetic analyses.

Dataset	Total length (bp)	Partition	Length (bp)	Best-fit model (IQTREE and Mrbayes)	Best-fit model (BEAST)
CO I -H3-18S-28S	4,525	*CO I*	1,653	GTR+F+I+G4	TN+F+G4
		H3	343	K2P+G4	HKY+G4
		18S	1,783	SYM+I	TNe+G4
		28S	746	HKY+F+I	TN+F+G4

Supplementary Table S13. Information of Illumina sequencing data using in this study.

Species	Number of contigs	Total length	Minium length of contigs	Average length of contigs	Maximum length of contigs	Data source	SRA number
Abyssogena southwardae	39,922	18,970,149	128	475.2	11,576	metagenomic data	SRR12077202
Calyptogena extenta	569,233	134,654,232	78	236.6	200,407	metagenomic data	SRR12077196
Calyptogena fausta	815,451	580,537,012	78	711.9	216,241	genome skimming	SRR12008158
Calyptogena pacifica	61,649	32,318,021	128	524.2	15,219	metagenomic data	SRR12077193
Calyptogena rectimargo	60,242	28,417,206	128	471.7	27,840	metagenomic data	SRR12077201
Calyptogena soyoae	49,944	22,812,861	128	456.8	19,372	metagenomic data	SRR12077194
Pliocardia sp.	69,904	34,426,138	128	492.5	17,211	metagenomic data	SRR12077197
Pheagena okutanii	25,249	28,184,488	297	1,116.3	17,323	transcriptome	SRR7156763
Archivesica gigas	60,781	28,822,343	128	474.2	25,143	metagenomic data	SRR12077198
Archivesica marissinica	28,949	41,107,808	156	1,420	21,610	transcriptome	Figshare: https://doi.org/10.6084/m9.figshare.12198987.v1
Archivesica nanshaensis	2,913	968,596	81	332.5	3,938	metagenomic data	SRR19046124, SRR19046125, SRR19046126 (This study)
Archivesica packardana	44,525	45,691,011	255	1,026.2	28,584	transcriptome	SRR8820491