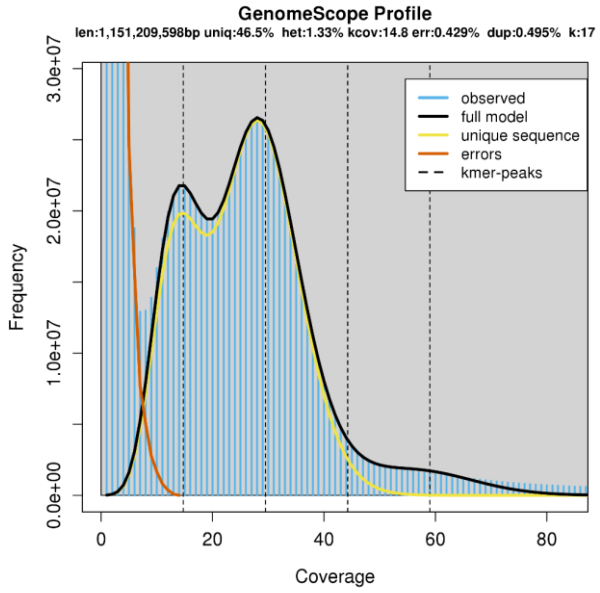


Fig. S1 Estimation of genome size, repetitive content and level of heterozygosity from Illumina reads of *Bathymodiolus marisindicus*.

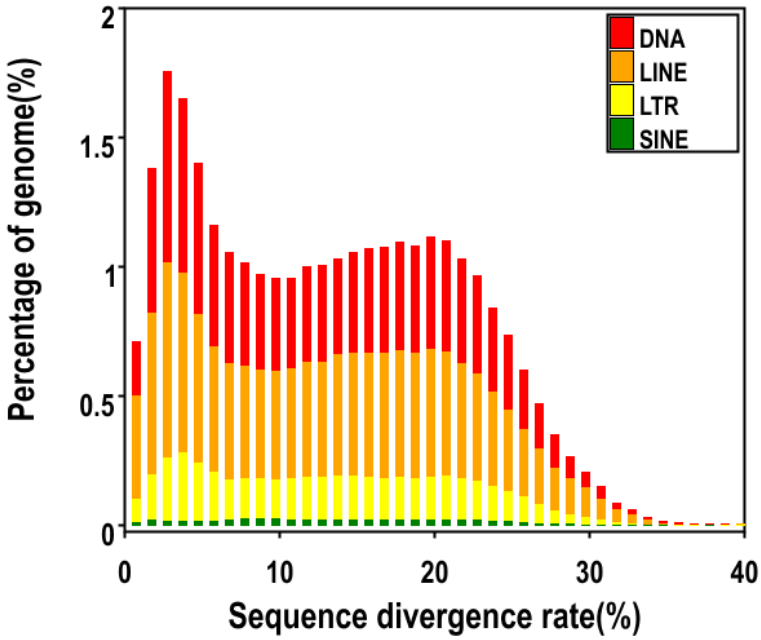
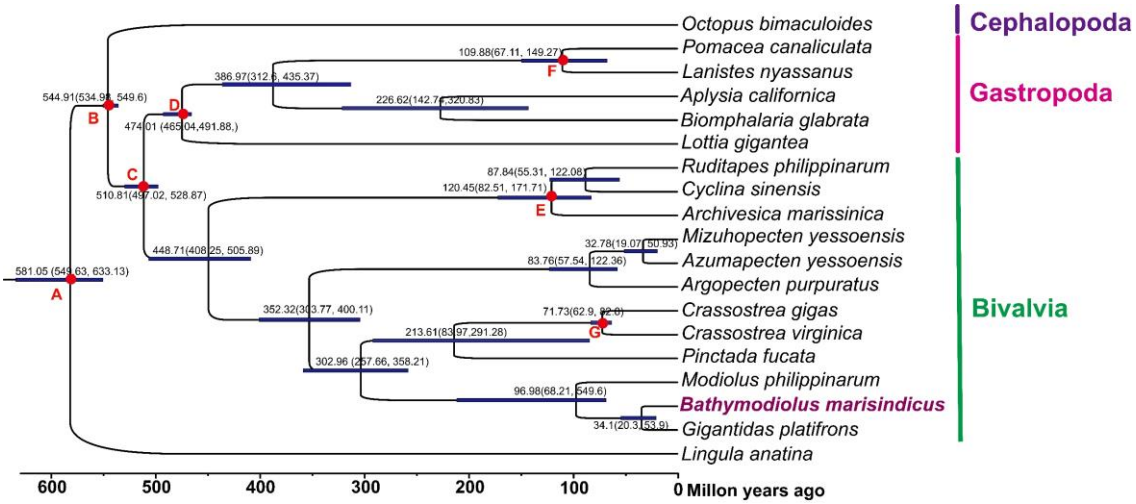


Fig. S2 History from major superfamilies of transposable elements in the *Bathymodiolus marisindicus*.

A



B

	Fossil records/geographic events	Time
A	The most recent common ancestor of lophotrochozoa	550.25-636.1 Ma
B	The most recent common ancestor of molluscs	532-549 Ma
C	The split of *Aplysia californica* and *Lottia gigantea*	470.2-531.5 Ma
D	The first appearance of Pteriomorpha	>465 Ma
E	The fossil record of pliocardiine	>470 Ma
F	The split of South America and Africa	<150 Ma
G	The split of *Crassostrea gigas* and *Crassostrea virginica*	63-83 Ma

Fig. S3 **A** Maximum-likelihood tree constructed based on 388 shared single-copy gene families of 19 lophotrochozoans with seven calibrated divergence time points indicating by red dots on nodes. At each node the posterior mean divergence time and upper and lower 95% confidence intervals are shown. **B** Details of calibration points based on fossil records and geological events.

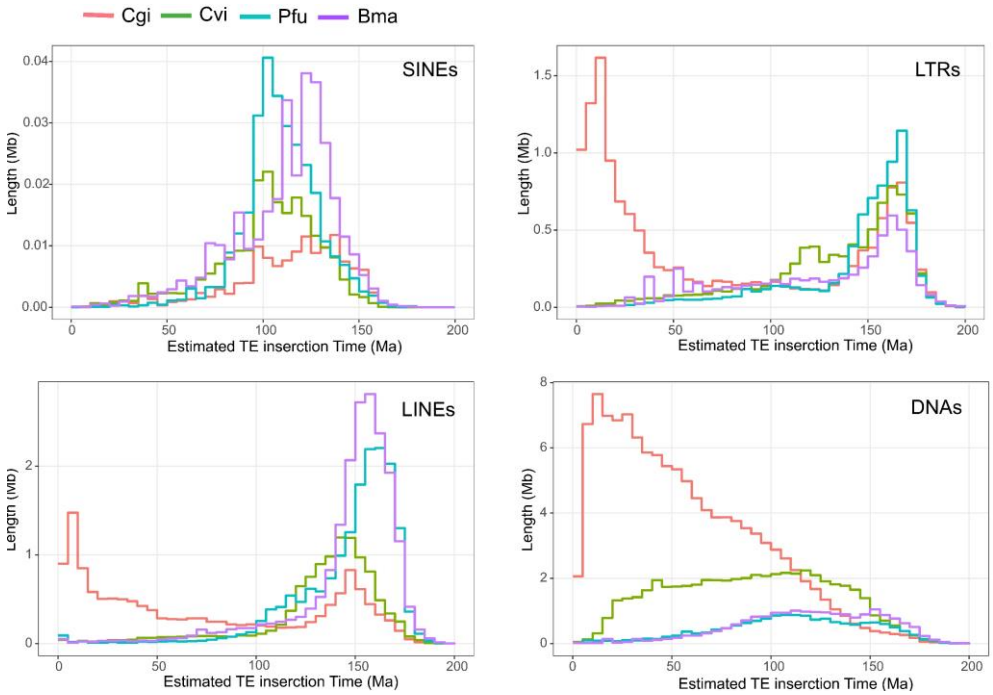


Fig. S4 Comparison of estimated insertion times of four TE classes among bivalves. Cgi, *Crassostrea gigas* (oyster); Cvi, *Crassostrea virginica*; Bma, *Bathymodiolus marisindicus*; Pfu, *Pinctada fucata*;

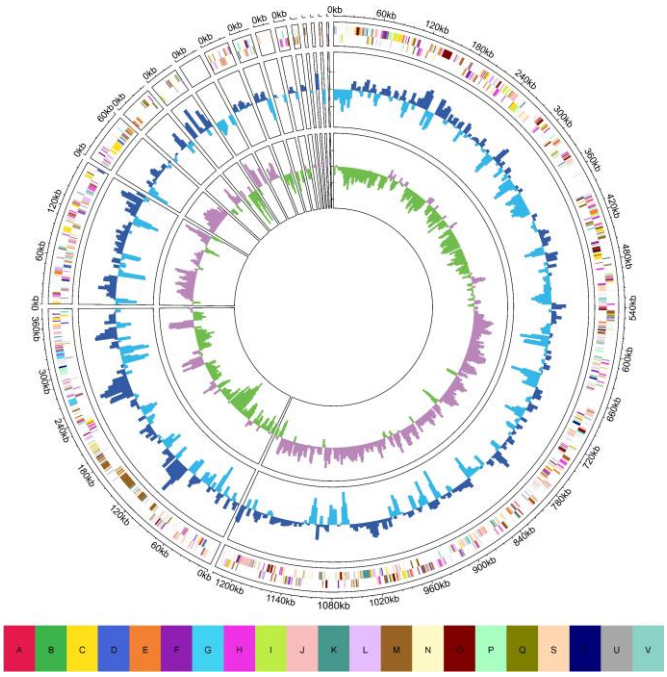


Fig. S5 Key genome features of both the sulfur-oxidizing endosymbiont of *B. marisindicus* with genome sizes in the centre, the inner circle showing the GC skew, the second circle showing the GC content of the deviation from the average, and the outer circle showing the COG annotation categories with different colors.

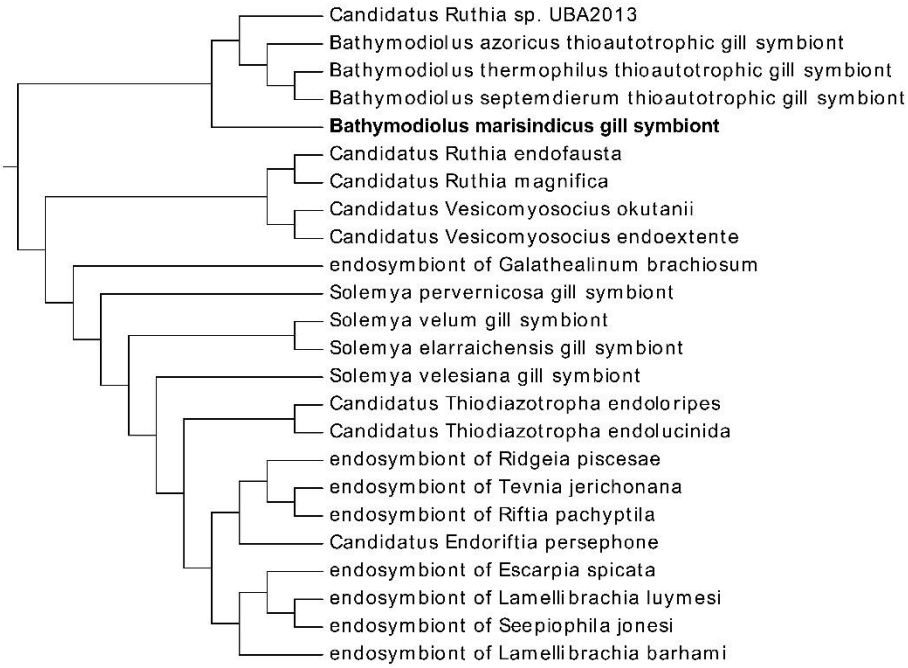


Fig. S6. A phylogenetic tree of *B. marisindicus* sulphur-oxidising endosymbiont and its closest relatives based on 16S rRNA gene.

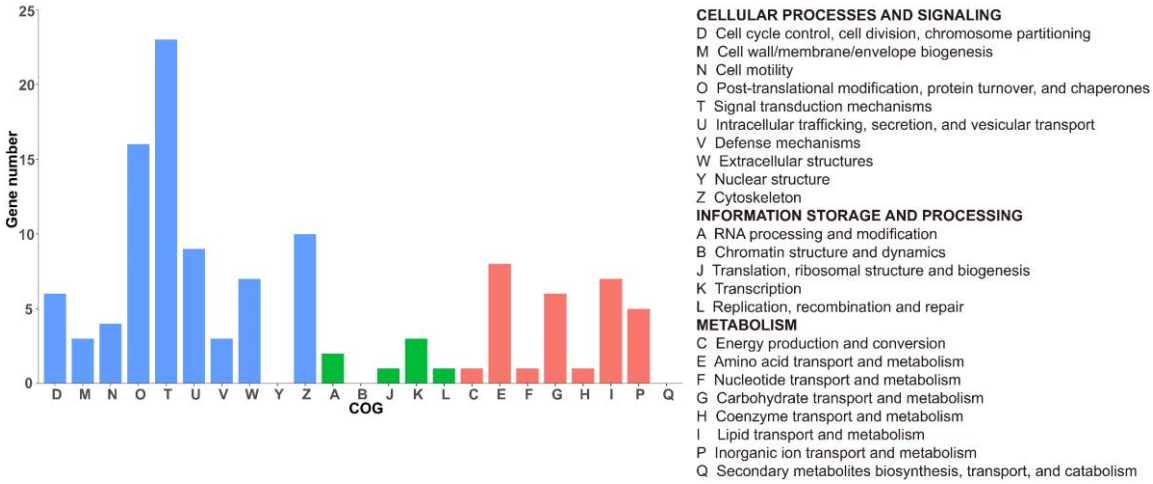


Fig. S7 Functional distribution of gill-specific highly expressed pseudogenes in *Bathymodiolus marisindicus*.

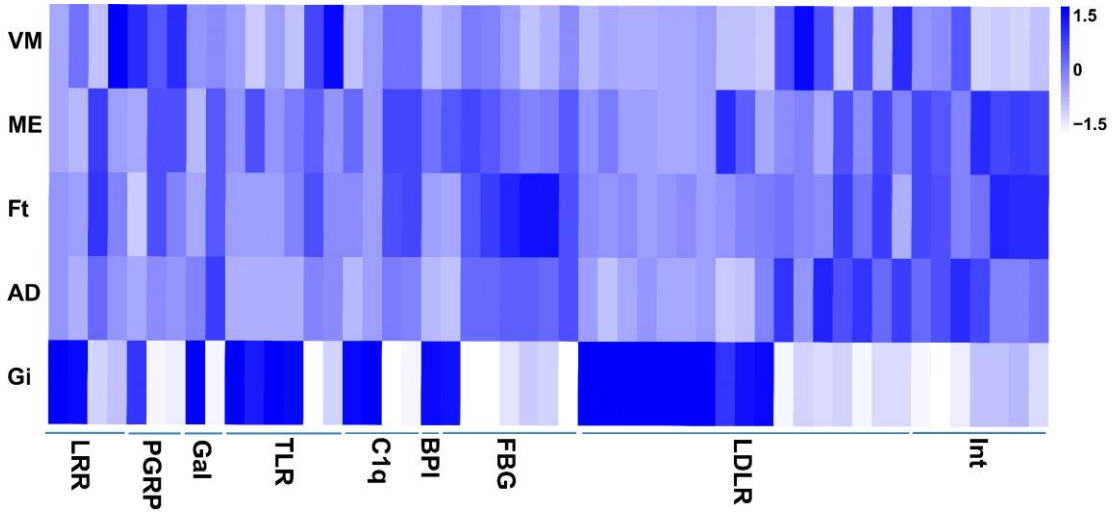


Fig. S8 Heat maps showing expression levels of pattern recognition receptors in *Bathymodiolus marisindicus*. Ad, adductor muscle; Mt, mantle; VM, visceral mass; Ft, foot; Gi, gill.

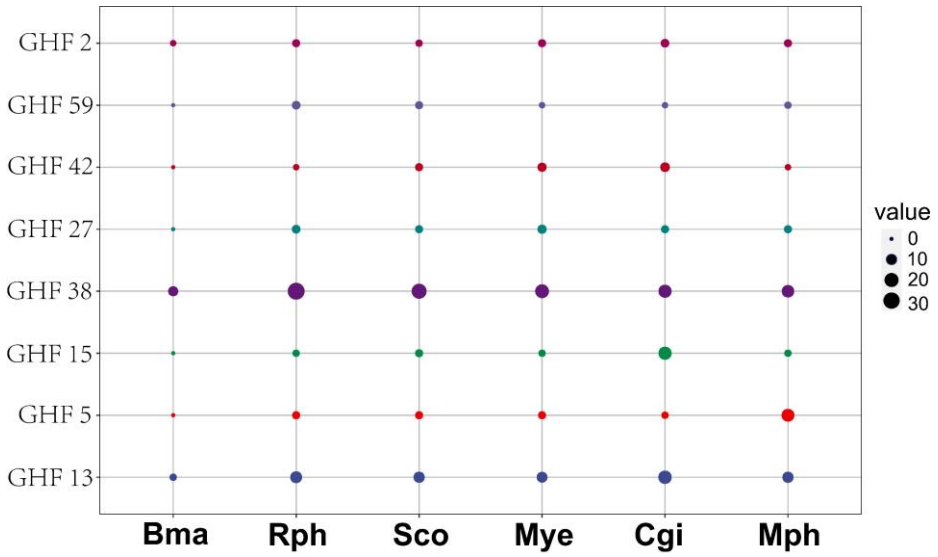


Fig. S9 The numbers of contracted GHF genes among 6 bivalve species. Mye, Mizuhopecten yessoensis (scallop); Cgi, Crassostrea gigas (oyster); Gpl, Gigantidas platifrons (mussel); Mph, Modiolus philippinarum (shallow imparidentian); Rph, Ruditapes philippinarum (shallow imparidentian); Sco, Sinonovacula constricta (shallow imparidentian).

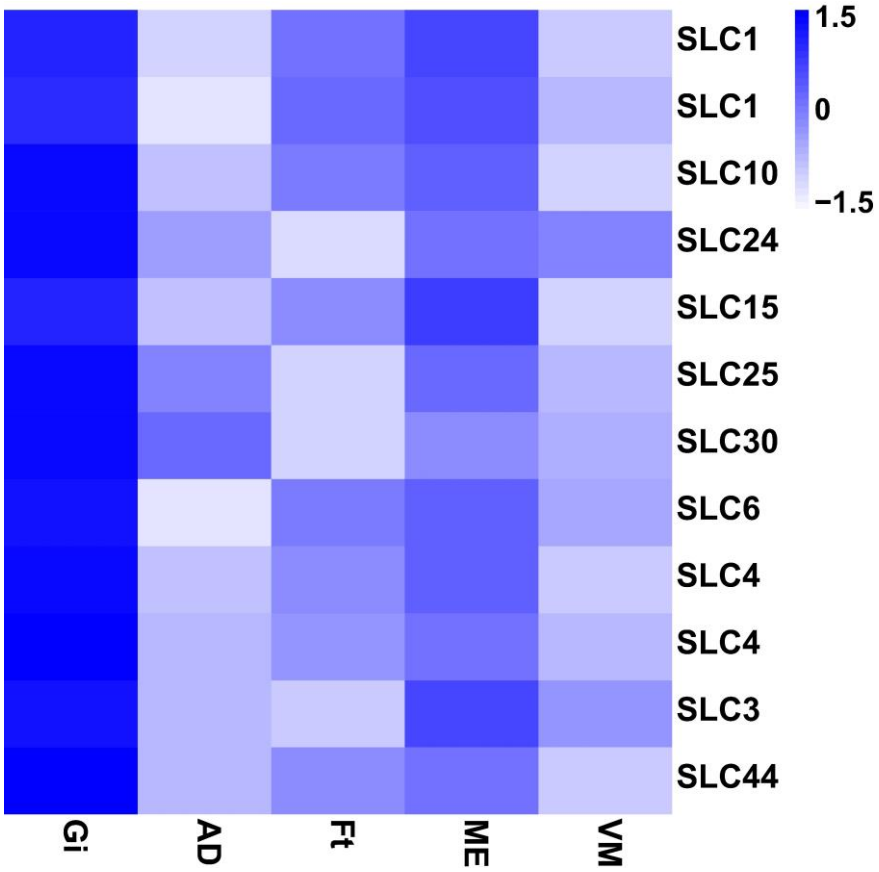


Fig. S10 The heat map showing gill-specific highly expressed solute carrier family related genes in *Bathymodiolus marisindicus.*

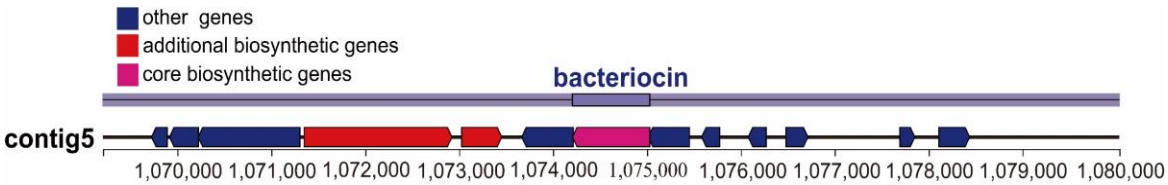


Fig. S11 Schematic representation of the bacteriocin gene cluster.

Table S1. Statistics of the assembly of the *Bathymodiolus marisindicus* genome

Genome assembly parameter	
Total size (Gb)	1.03
No of scaffold	10,561
Scaffold number, > 1Kb	10,549
Scaffold N50	301.96 kb
Scaffold L50	5,534
Longest scaffold	2.28 Mb
Mean size	98.21 kb
BUSCO	C:92.66% [S:91.8%,D:0.8%], F:4.0%,M:3.4%

Table S4. Categories of TEs predicted in *Bathymodiolus marisindicus* genome.

Type	Repbase TEs		TE protiens		De novo		Combined TEs	
	Length (Mb)	% in genome	Length (Mb)	% in genome	Length (Mb)	% in genome	Length (Mb)	% in genome
DNA	35.525	3.425	5.464	0.526	299.973	28.921	310.443	29.930
LINE	19.748	1.904	22.488	2.168	134.151	12.934	144.633	13.944
SINE	0.372	0.036	0	0	6.310	0.608	6.672	0.643
LTR	6.551	0.632	3.992	0.385	49.660	4.788	55.076	5.310
Other	0.006	0.001	0	0	0	0	0.006	0.001
Unknown	0	0	0	0	32.957	3.177	32.957	3.177
Total	61.102	5.891	31.926	3.078	460.538	44.401	477.808	46.066

Table S8. Statistics of the assembly of the *Bathymodiolus marisindicus* endosymbiont genome.

Genome assembly parameter	
Genome size (M)	2.1
Number of scaffolds	36
Scaffold N50 (bp)	1,225,115
CheckM completeness (%)	97.88
Contamination (%)	3.97
Strain heterogeneity (%)	66.67
Genes	2,164