Title: Tripartite holobiont system in a vent snail broadens the concept of chemosymbiosis

Authors:

Yi Yang[1], Jin Sun[1], Chong Chen[2], Yadong Zhou[3], Yi Lan[1], Cindy Lee Van Dover[4], Chunsheng Wang[3,5], Jian-Wen Qiu[6], Pei-Yuan Qian[1*]

Affiliations:

[1] Department of Ocean Science, Division of Life Science and Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), The Hong Kong University of Science and Technology, Hong Kong, China

[2] X-STAR, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

[3] Key Laboratory of Marine Ecosystem Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China

[4] Division of Marine Science and Conservation, Nicholas School of the Environment, Duke University, Beaufort, NC, United States

[5] School of Oceanography, Shanghai Jiao Tong University, Shanghai, China

[6] Department of Biology, Hong Kong Baptist University, Hong Kong, China

Correspondence to:

Pei-Yuan Qian: boqianpy@ust.hk

Abstract

Many animals inhabiting deep-sea vents are energetically dependent on chemosynthetic endosymbionts, but how such symbiont community interacts with host, and whether other nutritional sources are available to such animals remain unclear. To reveal the genomic basis of symbiosis in the vent snail *Alviniconcha marisindica*, we sequenced high-quality genomes of the host and gill campylobacterial endosymbionts, as well as metagenome of the gut microbiome. The gill endosymbiont has a streamlined genome for efficient chemoautotrophy, but also shows metabolic heterogeneity among populations. Inter- and intra-host variabilities among endosymbiont populations indicate the host poses low selection on gill endosymbionts. Virulence factors and genomic plasticity of the endosymbiont provide advantages for cooperating with host immunity to maintain mutualism and thriving in changing environments. In addition to endosymbiosis, the gut and its microbiome expand the holobiont's utilisation of energy sources. Host-microbiota mutualism contributes to a highly flexible holobiont that can excel in various extreme environments.

Introduction

Since the discovery of deep-sea hydrothermal vents in 1977, many intricate symbioses have been reported between vent-endemic animals and chemoautotrophic microbes. While some crustaceans such as the shrimp *Rimicaris exoculata* (Durand et al., 2015; Petersen et al., 2010) and the squat lobster *Shinkaia crosnieri* (Watsuji et al., 2015) rely on ectosymbionts living on their gills or their chaetae for nutrients, annelids in the family Siboglinidae (Dubilier et al., 2008) and molluscs in several families such as Vesicomyidae and Mytilidae (Dubilier et al., 2008) host endosymbiotic microbes in their bacteriocytes. Due to their intimate relationships with endosymbionts, it is generally accepted that in endosymbiosis the host relies entirely on symbionts for nutrition (Childress and Girguis, 2011; Dubilier et al., 2008).

Alviniconcha is a genus of chemosymbiotic provannid vent snails, with five species distributed in the Pacific Ocean and one in the Indian Ocean (Johnson et al., 2015). Among genera in the superfamily Abyssochrysoidea and those currently assigned to family Provannidae, only *Alviniconcha* and its sister genus *Ifremeria* that live in hydrothermal vents harbour endosymbionts in the gill epithelia (Beinart et al., 2019). Five species of *Alviniconcha* have been reported – four from the Pacific Ocean and one from the Indian Ocean (Johnson et al., 2015). The *Alviniconcha* species from the South Pacific are known to harbour both chemoautotrophic Gammaproteobacteria and Campylobacteria in its gills (Beinart et al., 2019), with the endosymbiont type and relative abundance varying with differences in vent fluid geochemistry (Beinart et al., 2012; Sanders et al., 2013). In contrast, *A. marisindica* from vent fields on the Central Indian Ridge (CIR) hosts a single ribotype of Campylobacterota endosymbionts (Miyazaki et al., 2020). Campylobacterota are abundant in vent habitats and also live as ectosymbionts on polychaete worms, molluscs, and crustaceans (Assié et al., 2016; Campbell et al., 2006; Goffredi, 2010; Watsuji et al., 2015), but do not commonly assume the role of intracellular symbionts. Campylobacterota are capable of oxidising sulfur, formate, and hydrogen to produce energy (Beinart et al., 2019; Miyazaki et al., 2020; Takai et al., 2005), and mostly rely on the reductive tricarboxylic acid cycle (rTCA) for carbon fixation with the exception of a bathymodiolin mussel epibiont possessing a complete Calvin–Benson–Bassham (CBB) cycle (Assié et al., 2020). Depending on the abundance of hydrogen and hydrogen sulfide,

the Campylobacterota endosymbiont of *A. marisindica* is capable of shifting between these reduced compounds as its main energy source (Miyazaki et al., 2020).

Unlike siboglinid tubeworms and clam *Solemya reidi* that have lost their gut (Dubilier et al., 2008), *Alviniconcha* species retained theirs, albeit a much reduced one (Warèn and Bouchet, 1993). Dissection revealed soft biogenic substances and mineral grains inside the snail gut, indicating it is functionally active (Suzuki et al., 2005). As *Alviniconcha* hosts endosymbionts in the gill and has a previously overlooked functional gut, it serves as a good model system to tease out the complex host-microbiota interactions that are key to our understanding of the adaptations of these animals to the extreme environments in the deep ocean (McFall-Ngai et al., 2013). Here, we report comprehensive analyses of the holobiont of *Alviniconcha marisindica* from a newly discovered northern Indian Ocean population (Zhou et al., 2019). Through analysing the symbiont genome and transcriptome, we aim to unravel the chemoautotrophic metabolism of the symbionts and their machinery for interaction with host, whether such symbiont populations contain streamlined and heterogeneous genomes that may enable them to utilise diverse substrates effectively, and how genomic plasticity of such populations provide advantages for thriving in their deep-sea habitat and interacting with host to establish symbiosis. Through analysing the host genome and transcriptome, we aim to understand how the host cooperates with symbionts to maintain mutualism and how the host's innate immunity has been remodelled to support the symbiosis. We also test the hypothesis that the gut and its microbiome are likely to provide nutrients that supplement the nutrition provided by the endosymbionts. Through dissecting the complex relations among the host, gill endosymbiont and gut microbiome, our study refine the holobiont concept in chemosymbiotic ecosystems which have enabled many animals to thrive in the extreme hydrothermal vent environments.

Results

Hologenome assembly and characterisation

The genome of the snail *Alviniconcha marisindica*, sequenced using a hybrid approach, is 829.61 Mb in length (N50 = 727.6 kb, genome completeness 96.5%) (Supplementary Table S1 and S2) with 21,456 predicted gene models (79.5% comparatively annotated) (Supplementary Figure S1). Comparative analyses among available lophotrochozoan genomes (n = 26; Figure 1A)

reveal a lack of obvious gene family expansion in the *A. marisindica* genome. The *A. marisindica* genome encodes 737 unique gene families (6.8%; Figure 1B) when compared with the genomes of a freshwater snail, a vent-endemic chemosymbiont hosting snail in a distant clade, and a scallop. Analyses of these unique genes reveal an enrichment of genes related to oxidoreductases, hydrolases, endocytosis, transporters, and signal transduction (Supplementary Figure S2 A and B). Among the annotated genes, those involved in immune response, substrate transportation, macromolecular digestion, and absorption are highly expressed in the intestinal tissue (Supplementary Figure S2C), indicating active functioning of the *A. marisindica* gut.

Sequencing a bacterial 16S rRNA gene clone library from the gill tissue revealed over 99% sequence similarity among the clones, confirming the presence of a single endosymbiont phylotype in the bacteriocytes. The campylobacterotal endosymbiont genome is 1.47 Mb in length, located in two scaffolds (98.16% completeness, 0.82% contamination) with 1,429 predicted genes, among which 92.65% were successfully annotated (Figure 2A, Supplementary Figure S3A). The campylobacterotal endosymbiont genome, here named *Sulfurovum alviniconcha* CR, possesses fewer coding sequences than other available whole genomes within the phylum (Figure 3), but has the highest coding density (97.0%, Table 1) and minimal loss-of-function mutations (Figure 4A), and similar average lengths in coding regions (Supplementary Figure S4A). There are almost no flagellar or chemotaxis genes in this endosymbiont genome. When compared with its four Campylobacterota close relatives (Table 1), *Sulfurovum alviniconcha* CR lacks many cell envelope biogenesis and non-essential metabolic genes. For example, genes involved in capsular polysaccharides biogenesis (*cps*), and genes involved in partial Citrate cycle (*ace* and *DLAT*) which is one of the optional from pyruvate to aceyl-coA are missing. *Sulfurovum alviniconcha* CR and its pathogenic relatives lack many DNA-repair genes (Supplementary Figure S5) that will lead to frequent gene loss, mutation, and recombination (Kang and Blaser, 2006; Monack et al., 2004). Nevertheless, *Sulfurovum alviniconcha* CR genome contains 180 unique orthologues when compared with its four Campylobacterota close relatives (Table 1, Figure 4B), including those involved in cell wall/membrane/envelope biogenesis that modify the bacterial surface for immune evasion (e.g. *eptA*), enzymes related to oxidoreductases and translocases that promote energy production and conversion in the endosymbiont (e.g. *putA*), and extracellular proteases secretion enhancing bacterial virulence

125 factors that associated with symbiotic interactions (e.g. *aprE* and *pulD*) (Supplementary Figure
126 S4 B and C). In addition, *Sulfurovum alviniconcha* CR shares many virulence genes with its non-
127 pathogenic (*Sulfurovum* species) and pathogenic (*Helicobacter* and *Campylobacter* species)
128 Campylobacterota relatives, such as bacterial virulence factors haemolysin and MviN/MurJ,
129 intracellular invasion CiaB, and N-linked glycosylation (NLG) with vital roles in infectivity.
130
131 In contrast to harbouring only one dominant phylotype of endosymbiont in the gill, the *A.*
132 *marisindica* gut contains diverse microbiota. Analyses of guts from three snail individuals reveal
133 169 microbial genera from 38 phyla, with a different composition and relative abundance
134 compared to those in gastropods that do not rely energetically on endosymbionts (Aronson et al.,
135 2016; Li et al., 2019). For example, the dominant genus in gut microbes of *A. marisindica* is
136 *Sulfurovum* (Supplementary Figure S3B) – a genus of chemoautotrophic Campylobacterota.
137 *Sulfurovum* is a minor community in the gut of deep-sea bone-eating snail *Rubyspira osteovora*
138 (Aronson et al., 2016) and it is rare in the gut of fresh-water polyphagous snail *Pomacea*
139 *canaliculata* (Li et al., 2019). The multi-taxa associations of gut microbiome in *A. marisindica*
140 exhibit a significant non-random co-occurrence pattern (Figure 2B), indicating the effects of the
141 intestinal microenvironment in shaping microbial community composition. Especially, lactic
142 acid bacteria, vital for maintaining the gut ecological balance (Koleva et al., 2014), account for at
143 least ~2.7% of gut microbes in *A. marisindica*, also shows that the gut microbiome is not
144 contaminants even if they are in low density (2.61–5.57%) in *A. marisindica* (Supplementary
145 Figure S6).
146
147 ***Diversity and metabolism of the gill endosymbiont***
148 Sequencing 23 isolates from 13 host snails (Table 2) reveals a *Sulfurovum alviniconcha* CR core-
149 campylobacterotal genome with 1,001 shared gene families (51.8–73.0% of the predicted
150 orthologues per isolate genome). Thiosulfate oxidation, oxygen reduction, and key reverse TCA
151 cycle genes are present in the core-genome. Each isolate represents a subpopulation, the pan-
152 genome of 23 subpopulations contains 2,783 orthologues, of which 1,475 are isolate-specific,
153 and exhibit high metabolic flexibility, especially along the chemoautotrophic pathways of sulfur
154 metabolism, hydrogen oxidation, and carbon fixation (Supplementary Figure S7). For example,
155 the pan-genome contains hydrogen oxidation genes (*hydABCDE* gene cluster), nitrate reduction

genes (*napA* and *napD*), and genes involved in hydrogen sulfide utilisation (*metZ*, catalyses the formation of L-homocysteine from O-succinyl-L-homoserine and hydrogen sulfide). Principal component and phylogenomic analyses on 941 shared single-copy orthologues of 23 isolates show that *Sulfurovum alviniconcha* CR are not clustered by their host individuals (Supplementary Figure S8). Among 23 isolates, selecting 20 isolates from the anterior and posterior gills of 10 host snails (Table 2), 20 *Sulfurovum alviniconcha* CR genomes are obtained with a total of 190 genomic average nucleotide identity (ANI) values ranging from 98.5% to 99.7% (Figure 4C). Nevertheless, these isolates belong to the same phylotype but have 28,448 single-nucleotide polymorphisms (SNPs) among them, indicating a high genetic diversity. Based on the results of genomic ANI and phylogenomic analysis of SNPs, the 20 endosymbiont isolates are classified into five types (Figure 4C) in a panmictic state among the 10 snails, showing that each snail hosts genetically diverse endosymbionts and with different types.

Analysis of the core metabolic genes of *Sulfurovum alviniconcha* CR genome reveals its chemolithoautotroph capability, especially in anaerobic oxidation of thiosulfate (*sox* genes), but it lacks the sulfide oxidation pathway as indicated by the lack of genes in the *dsrAB* complex. The sox multi-enzyme system allows generation of energy from thiosulfate oxidation, and *soxX-soxY-soxZ-soxA-soxB* genes are highly expressed (among the top 150) in *Sulfurovum alviniconcha* CR (Supplementary Figure S9A). The absence of a sulfate/thiosulfate transporter in *Sulfurovum alviniconcha* CR genome indicates that it can only use thiosulfate from endogenous organic sulfur compounds (Figure 5). A previous study shows the gill tissue of *A. marisindica* from the Kairei hydrothermal site actively consumed environmental sulfide (Miyazaki et al., 2020), which is consistent with *sqr* (25th in transcriptome, Supplementary Figure S9A) and *cysK* genes in the *Sulfurovum alviniconcha* CR genome involving in the conversion of sulfide to polysulfides. In addition, *Sulfurovum alviniconcha* CR lacks the sulfur globule protein genes (*sgp*) for intracellular sulfur storage, indicating this endosymbiont might dependent on intracellular polysulfides for sulfur storage.

Host-microbe syntrophic interactions

The tripartite *A. marisindica* holobiont is supported by their tight metabolic complementarity (Figure 5). Both *Sulfurovum alviniconcha* CR and the gut microbiome of the Wocan *A.*

marisindica possess typical metabolic pathways for synthesising carbohydrates, amino acids, and vitamins/cofactors and transporters for supplying these to the host. *Sulfurovum alviniconcha* CR uses the rTCA cycle to fix carbon and synthesises 20 amino acids and 4 vitamins/cofactors (Supplementary Figure S10A). The gut microbiome as a whole possess biosynthetic pathways for 10 amino acids and 4 vitamins (Supplementary Figure S10A), among them all of the 10 amino acids and 2 of the vitamins are shared with *Sulfurovum alviniconcha* CR, but the vitamins thiamine and nicotinate (and its derivative nicotinamide) are unique to the gut microbiome. Four amino acids and eight vitamins/cofactors cannot be synthesised *de novo* by either the symbiont system alone (Supplementary Figure S10B) and their production requires the complementary metabolic pathways of the host and symbionts to collaborate. For example, only *Sulfurovum alviniconcha* CR is capable of synthesising tryptophan yet it lacks genes for tryptophan metabolism, whereas the host genome contains the full tryptophan metabolic pathway from tryptophan to quinolinate. The host is further able to use quinolinate as a principal precursor to synthesise nicotinate and nicotinamide (vitamin B3) (Figure 5). Neither the host nor *Sulfurovum alviniconcha* CR alone can synthesise thiamine (vitamin B1), and the host lacks the thiamine transporter (*THTR*) for absorbing thiamine extracellularly. However, *Sulfurovum alviniconcha* CR can produce the thiamine phosphate precursor and pass it to the host. Thiamine is then synthesised as indicated by the highly expressed *PHO* that catalysing the conversion of thiamine phosphate to thiamine in the gill (Figure 5 and Supplementary Figure S9B). Similarly, the host cannot synthesise pantothenate but can obtain it from *Sulfurovum alviniconcha* CR in order to synthesise coenzyme A (Figure 5). Fatty acids (FAs) are essential nutrients required by most animals (Pranal et al., 1996). Holo-[carboxylase] serves as a biotin carrier protein and is essential in the biosynthesis of fatty acids in *A. marisindica*. Since only *Sulfurovum alviniconcha* CR can synthesise biotin (Supplementary Figure S10A), *A. marisindica* likely uses biotin derived from its endosymbionts *Sulfurovum alviniconcha* CR. Although *Sulfurovum alviniconcha* CR only possess biosynthetic pathways for saturated FA precursors (Figure 5), they may provide these precursors to the host, which can continue the FA biosynthesis by using the genes *MCH* and *fasN*, both of which are highly expressed in the gills (Figure 5 and Supplementary Figure S9B).

Neither *Sulfurovum alviniconcha* CR nor gut microbiome alone are able to generate all nutrients needed by the host (Supplementary Figure S10B). For example, the host expresses highly active

pathways of pancreatic secretion and bile secretion in addition to metabolic pathways of folate and octadecanoic acid (Figure 5 and Supplementary Figure S9B), among other nutrients that cannot be synthesised by the host or endosymbiont. Numerous genes responsible for key hydrolases that are responsible for breaking down macromolecules, and specialised transport proteins are highly expressed and enriched in the intestine (Supplementary Figure S2C, S11A and S12A). In addition, the gut microbial enzymes include hydrolases (30.6–34.7%) (Dataset S1), transferases (26.7–28.2%), and oxidoreductases (9.6–17.4%). Large amounts of multi-exohydrolase complexes in gut microbiome may promote the host's intestinal nutrient digestion (Supplementary Figure S9C, Table S3). For example, lactic acid bacteria (LAB) in the gut are found to possess a major facilitator, sugar transporters and enzymes for utilising large carbohydrate molecules. The gut microbiome even encodes additional enzymes such as oligoendopeptidase F (*pepF1*) and alginate lyase (*algL*) that can enhance digestion. Importantly, Campylobacterota in the gut are chemoautotrophic and found to encode the Sox system and [NiFe]-hydrogenases, and fix carbon with a complete rTCA cycle (Figure 5).

Strategies of symbiosis maintenance

Sulfurovum alviniconcha CR lacks genes to assemble surface layer proteins (SLPs) or capsular polysaccharides (CPs). Nevertheless, *Sulfurovum alviniconcha* CR encodes and actively expresses transmembrane signalling receptors, lipid A and its modification (Supplementary Figure S11B). *Sulfurovum alviniconcha* CR does not encode putative virulence-related proteins (*pag*) for Cationic antimicrobial peptides (CAMPs) resistance, but its genome harbours the *eptA* gene (Supplementary Figure S11B) involving in bacterial surface charge modification. In addition, genes encoding various proteases (e.g. subtilisin-like serine proteases) and the type II secretion system (T2SS) are highly expressed in *Sulfurovum alviniconcha* CR (Supplementary Figure S11B), along with Sec and Tat secretory pathways. On the other hand, genes involved in the assembly of bacterial cloaks (CP, SLP), lipopolysaccharide (LPS), and other surface-associated antigens responsible for bacterial adhesion to the intestinal epithelium and activating the complement system (Sára and Sleytr, 2000; Futoma-Koloch, 2016) are found in the gut microbiome. Surface-layer glycoprotein variation in the gut microbiome is evident from the differential expression of S-layer genes, a type of antigenic variation responding to the lytic activity of the host immune system (Supplementary Figure S11B). In the gut microbiome of *A.*

marisindica, genes encoding for sialate O-acetylesterase (SIAE) are highly expressed (Supplementary Figure S9C), which indicate their active sialic acid degradation in the gut.

The host immune system responds differently to *Sulfurovum alviniconcha* CR and the gut microbiome. The gills harbour a much higher abundance of bacteria than the gut (Supplementary Figure S6), but with weaker host immune responses (Figure 6A). Pattern recognition receptors (PRRs) are essential in the host's innate immune system. They can be divided into membrane-bound PRRs and cytoplasmic PRRs. Genes encoding membrane-bound C-type lectin receptors (CLRs) and cytoplasmic RIG-I-like receptors (RLRs) are more active in the intestine than in the gills of host invertebrates (Figure 6A). Toll-like receptors (TLRs) recognise structurally conserved molecules derived from microbes and activate immune responses. Genes encoding an endosomal TLR13 are highly expressed in the gill tissue, similar to the finding in the symbiont-hosting gills of the vent mussel *Bathymodiolus platifrons* (Sun et al., 2017). In the gut, however, membrane-bound TLR2 and TLR6 are more active (Figure 6A). Once the host recognises the symbionts, the gut and gills take different approaches to deal with the invading symbionts. In the gut tissue of *A. marisindica*, the component cascade is activated as indicated by the highly expressed complement component 1 complex (*C1*) and complement C3 (*C3*) (Figure 6A and 6B). In the gill tissue, genes encoding the signal repressor NF-κB1 (*NFKB1*), negative regulators (NF-κB inhibitor zeta (*NFKBIZ*)), and TNFAIP3-interacting protein 1 (*TNIP1*) of NF-κB response, all of which involved in attenuation of NF-κB, are highly expressed (Figure 6A). Genes encoding four members of the GiMAP gene family are also highly expressed in the gill tissue (Figure 6A). In addition, genes involved in TNF signalling, the MyD88-independent TLR signalling pathway, and leukocyte differentiation, which related to antimicrobial activity, are enriched in the gills (Supplementary Figure S12B).

Discussion

Sulfurovum alviniconcha CR has a relatively compact (1.47-Mbp) and streamlined genome. As maintaining the symbionts involves costs (Douglas and Smith, 1983; Meyer and Weis, 2012), the host may prefer a cellularly economised symbiont genome for energetic efficiency (Nicks and Rahn-Lee, 2017). A small endosymbiont genome may also enhance growth efficiency and intracellular competitiveness (Moran, 2002). *Sulfurovum alviniconcha* CR lacks most flagellar or

chemotaxis genes like its Campylobacterota close relatives from deep-sea sediments (Inagaki et al., 2004) and mounds (Nakagawa et al., 2007), whereas the Campylobacterota endosymbionts of *Alviniconcha boucheti* from Kilo Moana vent field at the Eastern Lau Spreading Centre has complete flagellar genes but no genes for the chemotactic signaling system (Beinart et al., 2019). The Campylobacterota endosymbionts of *Alviniconcha boucheti* are thought to be motile at free-living stage and such motility machinery could be used for finding a host. In this case, the non-motile *Sulfurovum alviniconcha* CR has a different machinery for adhesion to or interaction with *Alviniconcha marisindica*. Several symbioses have shown that motility and chemotaxis are not indispensable for recruiting symbionts from the environment, for example, some non-motile sulfate-reducing bacteria and methane-producing archaea in marine sediments use adhesins to colonise their host (Orphan et al., 2001; Raina et al., 2019). In addition, low-fidelity repair in the *Sulfurovum alviniconcha* CR genome increase its mutagenic potential, and such genomic plasticity has been found in human/animal pathogenic Campylobacterota (Kang and Blaser, 2006; Monack et al., 2004) and deep-sea vent Campylobacterota (Nakagawa et al., 2007), leading to micro-diversity increasement that confers a competitive advantages enabling bacteria persist in infections (Kang and Blaser, 2006; Monack et al., 2004) or thriving in ever-changing environments such as deep-sea vents (Nakagawa et al., 2005; Nakagawa et al., 2007). The *Sulfurovum alviniconcha* CR genome has the core of virulence for important animal pathogens, indicating its infectivity. Even if the *Sulfurovum alviniconcha* CR genome lacks many genes, it shows the ability to face with a changing environment, infect the animal host and survive intracellularly.

The host selectivity of endosymbionts in *Alviniconcha* snails is low when compared to other chemosymbiotic animals such as tubeworms (Beinart et al., 2012; Beinart et al., 2019; Yang et al., 2020) which may harbour a high diversity (even multiple classes) of endosymbionts with different types of metabolism within a single host (Beinart et al., 2012). This probably reflects the combined effect of environment selectivity on the available phylotypes and differences in vent fluid chemistry (Wang et al., 2017). A rarely discussed anatomical characteristic of the gill endosymbionts in the *Alviniconcha* species is that these endosymbionts residing inside bacteriocytes are present in a state between truly intracellular and extracellular (Endow and Ohta, 1989). Electron microscopy revealed that the vacuoles in bacteriocytes housing the

311 endosymbionts are exposed to the ambient seawater through duct-like openings (Endow and
312 Ohta, 1989). Considering the aggregated distribution of symbionts near the more exposed, outer
313 surface of the bacteriocytes, the *Alviniconcha* gill symbionts are in a 'semi-endosymbiotic'
314 condition (Windoffer and Giere, 1997), which likely provides *Alviniconcha* snails with the
315 ability to exchange or reacquire gill symbionts according to the local habitat and environment
316 selectivity through the endocytosis of free-living bacteria. Such 'semi-endosymbiotic' condition
317 provides gill endosymbiont populations with heterogeneous genomes regarding metabolic genes
318 along the chemoautotrophic pathways that may enable the utilisation of diverse substrates.
319 *Sulfurovum alviniconcha* CR is capable of anaerobically oxidising thiosulfate and hydrogen. The
320 environmental sulfide is conversed to polysulfides in *Sulfurovum alviniconcha* CR and then
321 bacterial organic polysulfides such as sulfur-containing amino acids are degraded to produce
322 intracellular thiosulfate for oxidation to produce cellular energy (Figure 5). This method of
323 sulfide utilisation and storage is different from those seen in many deep-sea holobionts such as in
324 siboglinid tubeworms, where the host haemoglobin binds to and transports the sulfides to
325 endosymbionts for direct oxidation or storage in bacterial sulfur globule proteins (Yang et al.,
326 2020), and in *Bathymodiolus* mussels, where the host oxidise sulfides and provide a reservoir of
327 thiosulfate for the endosymbionts' oxidation (Ponnudurai et al., 2020). We supposed that the
328 storage of environmental sulfides in *Sulfurovum alviniconcha* CR's polysulfides and the
329 utilisation of thiosulfate degraded from these intracellular sulfur compounds, is more efficient
330 than those symbionts which use thiosulfate provided by extracellular host tissues.
331
332 The synergistic biosynthesis of nutrients in *A. marisindica* gives the holobiont a capability of
333 nutrient production that is controlled by mutual supply of intermediates between the host and the
334 endosymbionts. Although the semi-endosymbiotic mode of housing the gill endosymbiont
335 provides *Alviniconcha* with the ability to utilise a rather wide array of bacteria as gill
336 endosymbiont (Beinart et al., 2019), it comes at a cost in that some symbiont phylotypes may
337 lack genes for certain syntrophic interactions. Although the digestive tract is substantially
338 reduced in the adult snail (Warèn and Bouchet, 1993), *A. marisindica* has a functioning gut
339 which contains faecal-like black substances suggesting that this snail ingests food by either
340 grazing or filter-feeding like *A. hessleri* from the Mariana Back Arc Basin (Warèn and Bouchet,
341 1993) and *A. marisindica* from the Central Indian Ridge (Suzuki et al., 2005). By supplying

genes and functions that are missing in *Sulfurovum alviniconcha* CR and the host, the gut microbiome help ensures the nutritional viability of the holobiont as a whole. For example, *pepF1* and *algL* genes in the gut microbiome that enhance digestion are lacking in the host snail and thus provide direct evidence that the gut microbiota has the potential to fulfil the nutritional demands of the holobiont. Overall, the results show that the gut microbiome has the potential to provide nutrition benefits to the *Alviniconcha* snail, an aspect of symbiosis that has been neglected in previous studies of many deep-sea endosymbiotic holobionts. As the *Alviniconcha* species has a semi-endosymbiotic model of gill symbiosis, the endosymbionts have flexible symbiotic associations with the host, which may at times impose nutritional limits on the holobiont system. In such situations, the gut likely contributes to keep the holobiont functionally versatile, ensuring its thriving in vent fields featuring different geochemical environments and available energy sources.

Sulfurovum alviniconcha CR lacks two common bacterial physical "cloaks" – SLPs and CPs that protect intracellular bacteria from host defences but also being recognised by the host as immunodominant antigens (Zamze et al., 2002; Sára and Sleytr, 2000). The absence of CPs likely helps *Sulfurovum alviniconcha* CR enter host cells (Deghmane et al., 2002) and reduces the risk of polysaccharide recognition by the host immune system (Zamze et al., 2002). In addition, lipid A modification enzymes and surface signal receptors help bacterial pathogens to avoid detection by TLRs (Thakur et al., 2019), this may also apply to *Sulfurovum alviniconcha* CR. CAMPs are key components of the host's innate immune response (Le et al., 2017; Noore et al., 2013). The presence of the *eptA* gene (Supplementary Figure S11B) involved in surface charge modification implies that *Sulfurovum alviniconcha* CR increases its surface positive charge to repel CAMPs. T2SS enables the transport of various cytoplasmic proteins into extracellular milieu, including bacterial toxins and degradative enzymes such as proteases and lipases. Previous study of tubeworm endosymbionts shows that the endosymbiont may use serine proteases to modulate the host's immune response by diminishing the function of host signal proteins (Yang et al., 2020). *Sulfurovum alviniconcha* CR may use a similar strategy. Surface antigenic molecules of *Sulfurovum alviniconcha* CR are distinct from the gut microbiome, indicating its host-specific immune-evasion mechanisms. In gut tissues, the mucus layer is the interface between the gut flora and the host, and sialic acids are prominent carbohydrates of the

intestinal mucus layer (Schroeder, 2019). Thus sialic acid breakdown of the gut microbiome indicates the way of intestinal bacterial encroachment and survival. Such differences in interaction with the host can lead to the establishment of different animal-microbe associations (Koropatnick et al., 2004).

Accordingly, the *A. marisindica* host has distinct recognition profiles for *Sulfurovum alviniconcha* CR and the gut microbiome (Figure 6A). After being recognised by the host, the invading symbionts will be controlled by the host's different corresponding strategies. The component cascade is activated in the gut which attack the microbe's cell membrane and eliminate microbes to control bacterial infections (Janeway et al., 2001) (Figure 6B). In the gills, potential attenuation of NF-κB are important as they grant the invaded cells additional protection (Burns et al., 2017; Best et al., 2019), and four members of the GiMAP gene family play critical roles in constraining and compartmentalising pathogens within cells (Weiss et al., 2013; Hunn et al., 2011). In addition, genes responsible for the majority of antimicrobial activity are enriched in the gills (Supplementary Figure S12B). In this case, the gill tissue shows a strong potential to constrain the intracellular symbionts and resist environmental invasion, which also indicate the ability of *Sulfurovum alviniconcha* CR to evade these host antimicrobial activities at the free-living stage. The weak host immune responses in the gills (Figure 6A) indicate *Sulfurovum alviniconcha* CR are more adept at evading recognition by the host immune system or inhibiting activation of the host immune system. Overall, the results show that *Sulfurovum alviniconcha* CR may have evolved an immunomodulation mechanism that they modulate the cell's outermost layer and release proteins enabling them to effectively evade recognition by the host immune system. In addition, the semi-intracellular position of *Sulfurovum alviniconcha* CR may allow it to avoid areas of high lysosomal activity in host cells that are part of the host self-defence mechanism (Endow and Ohta, 1989). The balance between the host's immune activity and bacterial counter-defence contributes to the complexity of the persistent symbioses.

We show, through hologenomic and holotranscriptomic analyses, that the *Alviniconcha marisindica* holobiont is more complex than previously recognised, being a tripartite system with the host snail and gill endosymbiont additionally supported by functional gut microbiome. The relative importance of each partner in *A. marisindica* may fluctuate depending on the immediate

availability of resources impacting the interplay downstream, as has been shown for other invertebrate symbioses (Morris et al., 2019; Belda et al., 1993). We unravel complex interactions among symbiotic parties in the *A. marisindica* holobiont, which deepen our understanding of the adaptation of many dominant chemosymbiotic holobionts that rely on gill endosymbionts for nutrition and also retain a functional gut, such as *Alviniconcha*'s sister genus *Ifremeria* (Windoffer and Giere, 1997), peltospirid snails (Chen et al., 2018), and *Bathymodiolus* mussels (Page et al., 1991).

Materials and Methods

Sample collection and nucleic acid preparation

Alviniconcha marisindica individuals were collected from a water depth of 2,919 m at the Wocan vent site on the Carlsberg Ridge (CR) of the northwestern Indian Ocean (60.53°E, 6.36°N) (Supplementary Figure S13). Sampling was conducted using the human occupied vehicle (HOV) *Jiaolong* onboard the research vehicle *Xiangyanghong 9* on March 19, 2017. Snails were placed into an insulated bio-box with a closed lid using a manipulator to minimise changes in water temperature. *Jiaolong* took approximately 2.5 hours to return to deck. Once the snails were onboard the research vessel, all specimens were immediately flash-frozen in liquid nitrogen and then transferred to a -80^{o}C freezer for storage. The morphological observation and molecular taxonomy of snail samples were shown in Supplementary Note 1.

The frozen snails were thawed in RNAlater® (Invitrogen, USA) on ice, dissected with different tissues fixed separately in RNAlater®, and then prepared for nucleic acid extraction. A single specimen of *Alviniconcha marisindica* from Wocan was used for the holobiont genome assembly. The foot and neck muscles were used for host genomic DNA extraction, and the endosymbiont-harbouring gills were used for endosymbiont DNA extraction. A total of three snail individuals, including the one used for identifying the host genome, were dissected into 7–10 tissue types each with RNA extraction performed on the different tissues. The gills of 10 other individuals were divided into anterior and posterior parts, and the DNA of these 20 parts were extracted separately for metagenome sequencing. The intestines of the three individuals were dissected for total DNA and RNA extraction for the gut microbiome, and the gills of these individuals were also dissected for total RNA extraction (Supplementary Note 1). Genomic DNA (gDNA) was

435 extracted using the E.Z.N.A.® Mollusc DNA Kit (Omega Bio-tek, Georgia, USA) and then
436 purified using Genomic DNA Clean & Concentrator™-10 Kit (Zymo Research, CA, USA)
437 according to the manufacturer's protocol. Total DNA of the gills and that of the intestines were
438 extracted using the same protocol. Total RNA was extracted using Trizol (Invitrogen, USA)
439 from different tissues following the manufacturer's protocol and prepared for RNA-Seq. Nucleic
440 acid quality was evaluated using agarose gel electrophoresis and a BioDrop μLITE (BioDrop,
441 Holliston, MA, US), and nucleic acid concentrations were quantified using a Qubit fluorometer
442 v3.0 (Thermo Fisher Scientific, Singapore).
443
444 *Library construction and sequencing*
445 Genomic DNA was aliquoted and submitted to three sequencing platforms: Illumina, PacBio
446 Sequel, and Oxford Nanopore Technologies (ONT). A library with a 350 bp insert size was
447 constructed from gDNA following the standard protocol provided by Illumina (San Diego, CA,
448 USA). After paired-end sequencing of the library at Novogene (Beijing, China), approximately
449 50 Gb of Illumina NovaSeq reads with a read length of 150 bp were generated. Illumina
450 sequencing of total DNA from the gills and that of total DNA from the intestines were conducted
451 similarly, with approximately 50 Gb of reads generated from each gill sample for endosymbiont
452 genome assembly, approximately 6–8 Gb of reads generated from each of 20 gill filaments for
453 symbiont genetic diversity analysis, and approximately 12 Gb of reads generated from each of
454 three intestine specimens for metagenome analysis (see overview of sequencing data in
455 Supplementary Note 2).
456
457 For preparation of the single-molecule real-time (SMRT) DNA template for PacBio sequencing,
458 the gDNA was sheared into large fragments (10 kb on average) using a Covaris® g-TUBE®
459 device and then concentrated using AMPure® PB beads. DNA repair and purification were
460 carried out according to the manufacturer's instructions (Pacific Biosciences). The blunt adapter
461 ligation reaction was conducted on purified end-repair DNA, and after purification DNA
462 sequencing polymerases became bound to SMRTbell templates. Finally, the library was
463 quantified using a Qubit fluorometer v3.0. After sequencing with the PacBio Sequel System at
464 the Hong Kong University of Science and Technology (HKUST) and Novogene, approximately
465 72 Gb of long reads were generated, with reads less than 4 kb in length discarded.

For ONT sequencing, a total of 3–5 μg of gDNA were used for the construction of each library following the '1D gDNA selecting for long reads (SQK-LSK109)' protocol from ONT. Briefly, gDNA was repaired and end-prepped as per standard protocol, before it was cleaned up with a 0.4× volume of AMPure XP beads. Adapter ligation and clean-up of the cleaned-repaired DNA were performed as per the standard protocol and the purified-ligated DNA was eluted using elution buffer. The DNA library was mixed with sequencing buffer and loading beads before it was loaded onto the SpotON sample port. Finally, sequencing was performed following the manufacturer's guidelines using the FLO-MIN106 R9.4 flow cell coupled to the MinION™ platform (ONT, Oxford, UK). Raw reads were base-called according to the protocol in MinKNOW and written into fastq files, and 9.8 Gb of long reads were generated with reads less than 4Kb discarded. MinION sequencing of total DNA from one gill specimen was conducted using the same procedures, generating 3.5 Gb of reads for endosymbiont genome scaffolding (see details of PacBio and ONT library construction in Supplementary Note 2). Illumina reads from gDNA were used for the genome survey of the Wocan *Alviniconcha marisindica*, and PacBio and ONT reads were used for the genome assembly (Supplementary Note 3).

For eukaryotic transcriptome sequencing of different tissues, a 250–300 bp insert cDNA library of each tissue was constructed after removing the prokaryotic RNA and sequenced on the Illumina NovaSeq platform at Novogene to produce 150 bp paired-end reads. Since the RNA of gills includes the sequences from both the host and the symbiont, another 250–300 bp insert strand-specific library of each gill specimen was constructed using Ribo-Zero™ Magnetic Kit to sequence both eukaryotic and microbial RNA. Therefore, two sets of transcript sequencing data were produced for the gills, one for both the host and the symbiont, and the other for only the host. The meta-transcriptome sequencing of the intestine was conducted using the same methods. Approximately 5–10 Gb of reads were generated from each tissue.

de novo *hybrid assembly of the host genome*

Trimmomatic v0.39 (Bolger et al., 2014) was used to trim the Illumina adapters and low-quality bases (base quality ≤ 20). The genome size of *A. marisindica* was estimated to be 809.1 Mb using the 17-mer histogram generated (Supplementary Note 3) and the genome heterozygosity

was evaluated as 0.88% using GenomeScope (Vurture et al., 2017). Several genome assembly pipelines were applied to assemble the genome with PacBio and ONT reads, including PacBio-only approaches (e.g. minimp2+miniasm (Li, 2016) and wtdbg2 (Ruan and Li, 2019)) and PacBio-ONT hybrid approaches (e.g. MaSuRCA version 3.2.8 (Zimin et al., 2013), FMLRC (Wang et al., 2018) + smartdenovo (Ruan, 2018) and FMLRC (Wang et al., 2018) +wtdbg2 (Ruan and Li, 2019)). The detailed settings of each assembly pipeline are shown in Supplementary Note 3.

A comparison of the assembly statistics of different pipelines (Supplementary Note 3) showed that the FMLRC+wtdbg2 assembly was the best and therefore this assembly was used in the downstream analyses. The assembly was carried out as follows: the ONT reads were concatenated with PacBio reads and error corrected with Illumina reads using FMLRC (Wang et al., 2018). This hybrid error correction method was selected based on previous benchmarking analysis on the available tools using Illumina reads for correction of PacBio/ONT long reads (Fu et al., 2019). The corrected long reads were then assembled using wtdbg2 using the setting "-x preset2" (Ruan and Li, 2019). Bacterial contamination was removed from the assembly using a genome binning method in MetaBAT 2 (Kang et al., 2015) and MaxBin 2.0 (Wu et al., 2016) (Supplementary Note 3). The Illumina reads were mapped to the clean assembly with Bowtie2 (Langmead and Salzberg, 2012), and only uniquely mapped reads were retained. The resultant sorted .bam file was used to correct errors in the assembly using Pilon v1.13 (Walker et al., 2014). Two rounds of error correction were performed. Redundant genomic assembled contigs from highly heterozygous regions were then removed using Redundans (Pryszcz and Gabaldón, 2016) with the settings of "--identity 0.8 --minLength 5000".

Quality check of the assembled host genome

We monitored the genome assembly completeness and redundancy using the metazoan Benchmarking Universal Single-Copy Orthologs (BUSCOs) v4.0.6 pipeline against the Metazoan dataset (Simão et al., 2015). A total of 921 out of the 954 searched BUSCO groups (96.5%) were complete in the assembled genome, and only 2.3% BUSCOs were missing, suggesting a high level of completeness of the *de novo* assembly (Supplementary Table S1).

QUAST v5.0.2 (Gurevich et al., 2013) was used to check genome assembly quality with PacBio and ONT reads (Supplementary Table S2).

Annotation of the host genome

The Wocan *Alviniconcha* host genome annotation pipeline generally followed a previously published procedure (Sun et al., 2017). Briefly, the repeat content and the transposable elements were predicted and classified using the RepeatMasker pipeline (Smit and Hubley, 2010) which searched against the known repeat library in Repbase and also the species-specific repeat library constructed by RepeatModeler (Supplementary Note 4).

Two versions of transcriptome assembly, i.e. the *de novo* assembled version and the genome-guided version, were independently assembled using Trinity v2.8.5 (Grabherr et al., 2011) and concatenated. Sequences with similarity over 0.97 were clustered with cd-hit-est (Li and Godzik, 2006). Maker v3.0 (Cantarel et al., 2008) was used to annotate the genome. In the first round of Maker annotation, only the transcriptomic evidence was used, and only a gene model with an annotation edit distance (AED) score less than 0.01 (Supplementary Figure S1) and predicted protein length over 200 amino acids was reported. The resultant genome annotation *.gff* file was used to train another *de novo* gene predictor, Augustus v3.3 (Stanke and Morgenstern, 2005). The gene model with only one exon with an incomplete open reading frame and inter-genic sequences less than 3 Kb was removed. The rest of the *bona-fide* gene models were used to train Augustus. In the second round of Maker, evidence from three different sources, i.e. the transcriptome, proteins from the Swiss-Prot database, and Augustus, were merged using EvidenceModer (Haas et al., 2008). The merged data was also integrated using Maker.

Gene functions were determined by using BLASTp to align the candidate sequences with NCBI non-redundant (NR) and Swiss-Prot protein databases with the settings of "-evalue 1e-5 -word_size 3 -num_alignments 20 -max_hsps 20". Blast2GO® (Götz et al., 2008) together with EggNOG mapper (Huerta-Cepas et al., 2017) was applied to assign Gene Ontology (GO) terms and clusters of orthologous groups (COGs) to the protein sequences via GO and EggNOG databases. The Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) (Kanehisa and Goto, 2000) was used to conduct the KEGG pathway annotation

558 analysis via the bidirectional best hit method. The Pfam database was searched using profile
559 hidden Markov models (profile HMMs) (with an e-value of 0.001) to classify the gene families
560 (El-Gebali et al., 2019).
561
562 *Host gene family identification and phylogenomic analysis*
563 A total of 26 lophotrochozoan genomes were analysed for clues to the gene family evolution
564 (Supplementary Note 4). Orthologs among all species were deduced via the OrthoMCL pipeline
565 (Li et al., 2003) with the BLASTp threshold set to 1e-5. Only single-copy genes in at least two-
566 thirds of the taxon sampled (i.e. in at least 18 species) were used in the phylogenetic tree analysis,
567 resulting in 492 orthologous groups. Protein sequences within each orthologue was aligned using
568 MUSCLE with the default settings; spurious sequences and poorly aligned sequences were
569 trimmed using TrimAL v1.4 (Capella-Gutiérrez et al., 2009). The final alignment of each
570 orthologue was concatenated with partition information for the phylogenetic analysis using
571 RaxML v8.2.11 (Stamatakis et al., 2005) with the GTR + Γ model. The MCMCTree v4.7 (Reis
572 and Yang, 2011) was used for tree dating. The root calibration point was set to 590 Ma in
573 MCMCTree, and the LG+Γ model of evolution was selected. Time frame constraints imposed to
574 calibrate the topology tree generated from RAxML are shown in Supplementary Note 4. The
575 MCMCTree was run for 1.0×10^7 generations, sampling every 1.0×10^3 and discarding 20% of
576 the samples as burn-in. Gene family expansion/contraction analysis was performed using CAFÉ
577 v3.1 (Han et al., 2013). Only a family level with $P<0.01$ and $P<0.01$ deduced by the Viterbi
578 method was considered to be expanded or contracted.
579
580 *Microbial metagenome assembly, annotation, and functional analysis*
581 For microbial metagenome assembly of the gill, Trimmomatic v0.39 (Bolger et al., 2014) and
582 FastUniq (Xu et al., 2012) were used to trim the Illumina reads and remove duplicates. The
583 bacterial abundance of gill metagenomic sequences was deduced using Kaiju (Menzel et al.,
584 2016) based on the subset of the NCBI BLAST *nr* database containing all proteins belonging to
585 Archaea, Bacteria, and Viruses. The clean reads were assembled using metaSPAdes v3.13.1
586 (Bankevich et al., 2012) with k-mer sizes of 21, 33, 55, 77, 99, and 127 bp, and the products
587 were pooled. Contigs potentially belonging to the campylobacterotal endosymbiont genome were
588 separated from its host genome using a genome binning method as described in previous studies

589 (Albertsen et al., 2013; Yang et al., 2020) (Supplementary Note 5). A genome of presumably
590 parasitic Mollicutes was removed (Supplementary Figure S3A). Contigs of the endosymbiont
591 genome were further determined using MetaBAT 2 (Kang et al., 2015) and MaxBin 2.0 (Wu et
592 al., 2016), assessed using CheckM v1.1.2 (Parks et al., 2015), and further scaffolded using
593 SSPACE-LongRead v1.1 (Boetzer and Pirovano, 2014) and *npScarf* (Cao et al., 2017) by adding
594 ONT long reads. The newly assembled scaffolds were binned again using the above pipeline.
595 GapFiller v1.10 (Boetzer and Pirovano, 2012) and Gap2Seq v3.1 (Salmela et al., 2016) were
596 used to fill the gaps in the binned endosymbiont genome. CheckM v1.1.2 (Parks et al., 2015) was
597 used to estimate the completeness and potential contamination of the binned genome. Coding
598 sequences (CDS) in the genome of the *Alviniconcha* endosymbiont were predicted and translated
599 using Prodigal v2.6.3 (Hyatt et al., 2010). Gene function annotation of the predicted protein
600 sequences followed the same pipeline as that described above for the host snail (Supplementary
601 Note 5). The protein sequences were annotated based on GO, EggNOG, KEGG and Pfam
602 databases.
603
604 For the gut metagenome assembly, reads were trimmed and duplicates removed as described
605 above. The host's interference in the analysis of intestinal content was minimised by removing
606 reads that were aligned with the host genome using Bowtie2 (Langmead and Salzberg, 2012)
607 before the assembly. The remaining reads were assembled using metaSPAdes v3.13.1
608 (Bankevich et al., 2012) with the same settings as above. The abundance and systematic
609 classification of intestinal metagenomic microbial sequences were carried out using Kaiju
610 (Menzel et al., 2016) (Supplementary Figure 3B and 6B). Network analysis of intestinal
611 microbes was conducted based on their relative abundance. To reduce the complexity of the
612 datasets, relative abundances higher than 0.01% were retained for the construction of the
613 network. All pairwise Spearman's rank correlations were calculated in the R package "picante".
614 Only robust (r>0.8 or r<-0.8) and statistically significant correlations (P<0.01) are shown in the
615 network. Network visualisation and modular analysis were conducted in Gephi v0.9.2. Prodigal
616 v2.6.3 (Hyatt et al., 2010) was used to predict and translate the coding sequences in the intestinal
617 metagenome, and BLASTp was then used to align the candidate sequences with the NCBI NR
618 protein database. The systematic assignment of each protein was imported to MEGAN v5.7.0
619 (Huson et al., 2011) using the lowest common ancestor (LCA) method with the parameters of

Min Score 50, Max Expected 0.01, Top Percent 5, and LCA Percent 100. Based on the systematic results, the microbial protein sequences were selected for further gene functional analysis, following the gene annotation pipeline described above. Blast2GO® (Götz et al., 2008) and EggNOG mapper (Huerta-Cepas et al., 2017) were applied to assign GO and COG terms to the intestinal prokaryotic protein sequences. KAAS (Kanehisa and Goto, 2000) was used to annotate the KEGG meta-pathway of intestinal flora using the single-directional best hit (SBH) method. All the annotated information of intestinal flora was in Dataset S2; the potential function and interaction of gut microbiome were shown in Supplementary Note 6.

Phylogenomic analysis and genomic comparison of the endosymbiont

A total of 120 single-copy orthologous genes found in all genomes of nine Deltaproteobacteria (outgroup) and 111 campylobacterotal representatives by Proteinortho6 (Lechner et al., 2011) (BLAST threshold $E = 1 \times 10^{-10}$) were retained for phylogenomic analysis. Sequences of each orthologue were aligned using MUSCLE and trimmed using TrimAL (Capella-Gutiérrez et al., 2009). The final alignment of each orthologue was concatenated with partition information for the phylogenetic analysis using RaxML v8.2.11 (Stamatakis et al., 2005) with the GTR + Γ model. The gill endosymbiont of the Wocan *Alviniconcha marisindica* was compared with the endosymbiont of *Lamellibrachia* tubeworm (Patra et al., 2016), the epibiont of the giant tubeworm *Riftia pachyptila* (Giovannelli et al., 2016), and two free-living Campylobacterota (Giovannelli et al., 2016; Nakagawa et al., 2007) from deep-sea hot vents (Table 1), which were clustered within the same clade (Figure 3). Whole-genome ANI of orthologous gene pairs shared between two microbial genomes was calculated using fastANI (Jain et al., 2018) with the default settings. A Venn web tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to illustrate the shared and unique orthologous genes among the five Campylobacterota representatives (Figure 4b). Orthologous genes only present in the Wocan *A. marisindica* endosymbiont were classified as its unique genes. Orthologous genes that were present in all the other four reference genomes but not in the endosymbiont of *Alviniconcha* were classified as reduced genes (Supplementary Note 7). In addition, an HMM-based approach delta-bitscore (Wheeler et al., 2016) was used to identify loss-of-function mutations in shared orthologous genes of the five Campylobacterota (Dataset S3, Supplementary Note 7).

In addition to the above genomic comparisons, a total of 23 metagenome sequences were obtained from 13 Wocan *A. marisindica* snails. A genome binning method was used following the pipeline described in previous sections to assemble and extract another 22 endosymbiont genomes and the ANI among these 23 endosymbiont genomes was calculated (Jain et al., 2018) (Table 2 and Supplementary Note 8). The core genome shared across all 23 endosymbiont genomes was obtained using Proteinortho6 (Lechner et al., 2011) (BLAST threshold $E = 1 \times 10^{-10}$). The pan genome including isolate-specific genes were also detected. In addition, we also captured genome-wide variation of endosymbionts by comparing variations present in two parts of the gills (anterior and posterior) in each snail individual, across multiple snails (Supplementary Note 8). Selecting from the above 23 genomes, SNPs among 20 endosymbiotic isolates of the anterior and posterior gills from 10 snail individuals were called by aligning clean high-quality Illumina reads from each gill sample with a complete reference genome using the novel high-accuracy pipeline BactSNP (Yoshimura et al., 2019), in a single step. Pseudo genomes of input isolates were obtained. For each isolate, all contigs in the pseudo genome were concatenated into one sequence and submitted to phylogeny analysis using RaxML v8.2.11 (Stamatakis et al., 2005) under the GTR + CAT model.

Quantification of gene expression level

For host transcriptome sequencing data, the raw reads of each tissue were trimmed with Trimmomatic v0.39 (Bolger et al., 2014), the gene expression level in each tissue was expressed in transcripts per million (TPM) with Salmon (Patro et al., 2017), and the number of read counts for genes was also included in the quantification results. For meta-transcriptome sequencing data of the symbionts, the same pipeline was followed, with a Salmon index built for the transcripts of symbionts obtained and translated from their genome data. The trimmed reads were then quantified directly against this index and expressed in TPM using Salmon (Patro et al., 2017). In addition, using this quantification method, the gene expression levels of the gills were produced from two sets of RNA sequencing data of the gills (one is a meta-transcriptome dataset including both the host and symbionts, and the other is a eukaryotic transcriptome including only host transcripts). The consistency of gene expression levels for the gills from these two sets of sequencing data also confirmed the accuracy of our transcript-level quantification in the host and its symbionts.

Differentially expressed genes were determined by DESeq2 using the normalisation method of Loess, a minimum read count of 10, and a paired test (n = 5). A gene was considered specifically expressed in a particular tissue based on its expression levels compared across all other tissue types (paired-test method). Genes overexpressed with over twofold changes and false discovery rate (FDR) < 0.05 when compared with other tissue types were considered to be highly expressed (Dataset S4). WEGO (http://wego.genomics.org.cn/cgi-bin/wego/index.pl) was used to plot GO annotations of highly expressed genes in the different selected tissues. Statistically overrepresented GO terms in the different tissues were identified through topGO package in R session (Alexa and Rahnenführer, 2009). The GO enrichment network is visualised using the Cytoscape application (Shannon et al., 2003). Differentially expressed genes of different tissues were shown in Supplementary Note 9.

Data availability

All raw sequencing data generated in the present study are available from NCBI via the accession numbers SRR11781614–SRR11781681, and BioSample accessions SAMN14907812–SAMN14907827. The data generated in the present study have been deposited in the NCBI database as BioProject PRJNA632343. All software commands used in the host genome assembly are given in the Supplementary Information. The assembled transcriptome, predicted transcripts, and proteins are openly available from Dryad (DOI: XXX).

References

1. Alexa A, Rahnenführer J (2009). Gene set enrichment analysis with topGO. *Bioconductor Improvements* **27**. 10.18129/B9.bioc.topGO

2. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, & Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature biotechnology* **31**: 533. https://doi.org/10.1038/nbt.2579

711 3. Aronson HS, Zellmer AJ, Goffredi SK (2016). The specific and exclusive microbiome of the
712 deep-sea bone-eating snail, *Rubyspira osteovora*. *FEMS Microbiology Ecology* **93**: fiw250.
713 https://doi.org/10.1093/femsec/fiw250
714
715 4. Assié A, Borowski C, van der Heijden K, Raggi L, Geier B, Leisch N, Schimak MP, Dubilier
716 N, Petersen JM (2016). A specific and widespread association between deep□sea *Bathymodiolus*
717 mussels and a novel family of Epsilonproteobacteria. *Environmental Microbiology Reports* **8**:
718 805–813. https://doi.org/10.1111/1758-2229.12442
719
720 5. Assié A, Leisch N, Meier DV, Gruber-Vodicka H, Tegetmeyer HE, Meyerdierks A, Kleiner M,
721 Hinzke T, Joye S, Saxton M, Dubilier N (2020). Horizontal acquisition of a patchwork Calvin
722 cycle by symbiotic and free-living Campylobacterota (formerly Epsilonproteobacteria). *The*
723 *ISME Journal* **14**: 104–122. https://doi.org/10.1038/s41396-019-0508-7
724
725 6. Bankevich A. Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
726 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV (2012). SPAdes: a new genome assembly
727 algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**:
728 455–477. https://doi.org/10.1089/cmb.2012.0021
729
730 7. Beinart RA, Luo C, Konstantinidis KT, Stewart FJ, Girguis PR (2019). The bacterial
731 symbionts of closely related hydrothermal vent snails with distinct geochemical habitats show
732 broad similarity in chemoautotrophic gene content. *Frontiers in Microbiology* **10**: 1818.
733 https://doi.org/10.3389/fmicb.2019.01818
734
735 8. Beinart RA, Sanders JG, Faure B, Sylva SP, Lee RW, Becker EL, Gartman A, Luther GW,
736 Seewald JS, Fisher CR, Girguis PR (2012). Evidence for the role of endosymbionts in regional-
737 scale habitat partitioning by hydrothermal vent symbioses. *Proceedings of the National Academy*
738 *of Science* **109**: E3241–E3250. https://doi.org/10.1073/pnas.1202690109
739

9. Belda C, Lucas J, Yellowlees D (1993). Nutrient limitation in the giant clam-zooxanthellae symbiosis: effects of nutrient supplements on growth of the symbiotic partners. *Marine Biology* **117**: 655–664. https://doi.org/10.1007/BF00349778

10. Best KT, Lee FK, Knapp E, Awad HA, Loiselle AE (2019). Deletion of NFKB1 enhances canonical NF-κB signaling and increases macrophage and myofibroblast content during tendon healing. *Scientific reports* **9**: 1–11. https://doi.org/10.1038/s41598-019-47461-5

11. Boetzer M & Pirovano W (2012). Toward almost closed genomes with GapFiller. *Genome Biology* **13**, R56. https://doi.org/10.1186/gb-2012-13-6-r56

12. Boetzer M & Pirovano W (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC bioinformatics* **15**: 211. https://doi.org/10.1186/1471-2105-15-211

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

14. Burns JA, Zhang H, Hill E, Kim E, Kerney R (2017). Transcriptome analysis illuminates the nature of the intracellular interaction in a vertebrate-algal symbiosis. *Elife* **6**: e22054. https://doi.org/10.7554/eLife.22054.001

15. Campbell BJ, Engel AS, Porter ML, Takai K (2006). The versatile ε-proteobacteria: key players in sulphidic habitats. *Nature Reviews Microbiology* **4**: 458. https://doi.org/10.1038/nrmicro1414

16. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research* **18**: 188–196. http://www.genome.org/cgi/doi/10.1101/gr.6743907

770 17. Cao MD, Nguyen SH, Ganesamoorthy D, Elliott AG, Cooper MA, Coin LJ (2017).
771 Scaffolding and completing genome assemblies in real-time with nanopore sequencing. *Nature*
772 *Communications* **8**: 1–10. https://doi.org/10.1038/ncomms14515
773
774 18. Capella-Gutiérrez S, Silla-Martínez JM & Gabaldón T (2009). trimAl: a tool for automated
775 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973.
776 https://doi.org/10.1093/bioinformatics/btp348
777
778 19. Chen C, Linse K, Uematsu K, Sigwart JD (2018). Cryptic niche switching in a
779 chemosymbiotic gastropod. *Proceedings of the Royal Society B: Biological Sciences* **285**:
780 20181099. https://doi.org/10.1098/rspb.2018.1099
781
782 20. Childress J, Girguis PR (2011). The metabolic demands of endosymbiotic chemoautotrophic
783 metabolism on host physiological capacities. *Journal of Experimental Biology* **214**: 312–325.
784 https://doi.org/10.1242/jeb.049023
785
786 21. Deghmane AE, Giorgini D, Larribe M, Alonso JM, Taha MK (2002). Down□regulation of
787 pili and capsule of *Neisseria meningitidis* upon contact with epithelial cells is mediated by CrgA
788 regulatory protein. *Molecular microbiology* 43: 1555–1564. https://doi.org/10.1046/j.1365-
789 2958.2002.02838.x
790
791 22. Douglas A, Smith D (1983). The cost of symbionts to their host in green hydra.
792 *Endocytobiology II* 631–648. https://doi.org/10.1515/9783110841237
793
794 23. Dubilier N, Bergin C, Lott C (2008). Symbiotic diversity in marine animals: the art of
795 harnessing chemosynthesis. *Nature Reviews Microbiology* **6**: 725.
796 https://doi.org/10.1038/nrmicro1992
797
798 24. Durand L, Roumagnac M, Cueff-Gauchard V, Jan C, Guri M, Tessier C, Haond M, Crassous
799 P, Zbinden M, Arnaud-Haond S, Cambon-Bonavita MA (2015). Biogeographical distribution of
800 *Rimicaris exoculata* resident gut epibiont communities along the Mid-Atlantic Ridge

hydrothermal vent sites. *FEMS Microbiology Ecology* **91**: fiv101. https://doi.org/10.1093/femsec/fiv101

25. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer EL (2019). The Pfam protein families database in 2019. *Nucleic acids research* **47**: D427–D432. https://doi.org/10.1093/nar/gky995

26. Endow K, Ohta S (1989). The symbiotic relationship between bacteria and a mesogastropod snail, *Alviniconcha hessleri*, collected from hydrothermal vents of the Mariana Back-Arc Basin. *Bulletin of Japanese Society of Microbial Ecology* **3**: 73–82. https://doi.org/10.1264/microbes1986.3.73

27. Fu S, Wang A, Au KF (2019). A comparative evaluation of hybrid error correction methods for error-prone long reads. *Genome biology* **20**, 26. https://doi.org/10.1186/s13059-018-1605-z

28. Futoma-Kołoch B (2016). Immune response against bacterial lipopolysaccharide. *Journal of Molecular Immunology* **2**: 105. https://dx.doi.org/10.4172/jmi.1000e106

29. Giovannelli D, Chung M, Staley J, Starovoytov V, Le Bris N, Vetriani C (2016). *Sulfurovum riftiae* sp. nov., a mesophilic, thiosulfate-oxidizing, nitrate-reducing chemolithoautotrophic epsilonproteobacterium isolated from the tube of the deep-sea hydrothermal vent polychaete *Riftia pachyptila*. *International journal of systematic and evolutionary microbiology* **66**, 2697–2701. https://doi.org/10.1099/ijsem.0.001106

30. Goffredi SK (2010). Indigenous ectosymbiotic bacteria associated with diverse hydrothermal vent invertebrates. *Environmental Microbiology Reports* **2**: 479–488. https://doi.org/10.1111/j.1758-2229.2010.00136.x

31. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420–3435. https://doi.org/10.1093/nar/gkn176

32. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Palma Fd, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N & Regev A (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644. https://doi.org/10.1038/nbt.1883

33. Gurevich A, Saveliev V, Vyahhi N & Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013). https://doi.org/10.1093/bioinformatics/btt086

34. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* **9**, R7. https://doi.org/10.1186/gb-2008-9-1-r7

35. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular biology and evolution* **34**, 2115-2122. https://doi.org/10.1093/molbev/msx148

36. Hunn JP, Feng CG, Sher A, Howard JC (2011). The immunity-related GTPases in mammals: a fast-evolving cell-autonomous resistance system against intracellular pathogens. *Mammalian genome* **22**: 43–54. https://doi.org/10.1007/s00335-010-9293-3

37. Huson DH, Mitra S, Ruscheweyh HJ, Weber N & Schuster SC (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Research* **21**, 1552–1560. http://www.genome.org/cgi/doi/10.1101/gr.120618.111

38. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119. https://doi.org/10.1186/1471-2105-11-119

39. Inagaki F, Takai K, Nealson KH & Horikoshi K (2004). *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the epsilon-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *International journal of systematic and evolutionary microbiology* **54**, 1477–1482. https://doi.org/10.1099/ijs.0.03042-0

40. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT & Aluru S (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications* **9**, 1–8. https://doi.org/10.1038/s41467-018-07641-9

41. Janeway CA Jr, Travers P, Walport M, Shlomchik MJ (2001). "The complement system and innate immunity." In *Immunobiology: The Immune System in Health and Disease. 5th edition.* Garland Science. https://www.ncbi.nlm.nih.gov/books/NBK27100/

42. Johnson SB, Warén A, Tunnicliffe V, Dover CV, Wheat CG, Schultz TF, Vrijenhoek RC (2015). Molecular taxonomy and naming of five cryptic species of *Alviniconcha* snails (Gastropoda: Abyssochrysoidea) from hydrothermal vents. *Systematics and Biodiversity* **13**: 278–295. https://doi.org/10.1080/14772000.2014.970673

43. Kanehisa M & Goto S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27

44. Kang DD, Froula J, Egan R & Wang Z (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165. https://doi.org/10.7717/peerj.1165

45. Kang J, Blaser MJ (2006). Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. *Nature Reviews Microbiology* **4**:826-836. https://doi.org/10.1038/nrmicro1528

893 46. Koleva Z, Dedov I, Kizheva J, Lipovanska R, Moncheva P, Hristova P (2014). Lactic acid
894 microflora of the gut of snail *Cornu aspersum*. *Biotechnology & Biotechnological Equipment* **28**:
895 627–634. https://doi.org/10.1080/13102818.2014.947071
896
897 47. Koropatnick TA, Engle JT, Apicella MA, Stabb EV, Goldman WE, McFall-Ngai MJ (2004).
898 Microbial factor-mediated development in a host-bacterial mutualism. *Science* **306**: 1186–1188.
899 https://doi.org/10.1126/science.1102218
900
901 48. Langmead B & Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
902 *methods* **9**: 357. https://doi.org/10.1038/nmeth.1923
903
904 49. Le CF, Fang CM, Sekaran SD (2017). Intracellular targeting mechanisms by antimicrobial
905 peptides. *Antimicrobial agents and chemotherapy* **61**: e02340–02316.
906 https://doi.org/10.1128/AAC.02340-16
907
908 50. Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ (2011). Proteinortho:
909 detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**: 124.
910 https://doi.org/10.1186/1471-2105-12-124
911
912 51. Li H (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long
913 sequences. *Bioinformatics* **32**: 2103–2110. https://doi.org/10.1093/bioinformatics/btw152
914
915 52. Li L, Stoeckert CJ, Roos DS (2003). OrthoMCL: identification of ortholog groups for
916 eukaryotic genomes. *Genome research* **13**: 2178–2189.
917 http://www.genome.org/cgi/doi/10.1101/gr.1224503
918
919 53. Li LH, Lv S, Lu Y, Bi DQ, Guo YH, Wu JT, Yue ZY, Mao GY, Guo ZX, Zhang Y, Tang YF
920 (2019). Spatial structure of the microbiome in the gut of *Pomacea canaliculata*. *BMC*
921 *Microbiology* **19**: 273. https://doi.org/10.1186/s12866-019-1661-x
922

923 54. Li W, Godzik A (2006). Cd-hit: a fast program for clustering and comparing large sets of
924 protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.
925 https://doi.org/10.1093/bioinformatics/btl158
926
927 55. Menzel P, Ng KL, Krogh A (2016). Fast and sensitive taxonomic classification for
928 metagenomics with Kaiju. *Nature Communications* **7**: 11257.
929 https://doi.org/10.1038/ncomms11257
930
931 56. Meyer E, Weis VM (2012). Study of cnidarian-algal symbiosis in the "omics" age. *The*
932 *Biological Bulletin* **223**: 44–65. https://doi.org/10.1086/692718
933
934 57. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Lošo T, Douglas AE,
935 Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH,
936 Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG,
937 Rumpho M, Sanders JG, Tautz D, Wernegreen JJ (2013). Animals in a bacterial world, a new
938 imperative for the life sciences. *Proceedings of the National Academy of Sciences* **110**: 3229–
939 3236. https://doi.org/10.1073/pnas.1218525110
940
941 58. Miyazaki J, Ikuta T, Watsuji TO, Abe M, Yamamoto M, Nakagawa S, Takaki Y, Nakamura
942 K, Takai K (2020). Dual energy metabolism of the Campylobacterota endosymbiont in the
943 chemosynthetic snail *Alviniconcha marisindica*. *The ISME Journal* **14**: 1273–1289.
944 https://doi.org/10.1038/s41396-020-0605-7
945
946 59. Monack D, Mueller A, Falkow S (2004). Persistent bacterial infections: the interface of the
947 pathogen and the host immune system. *Nature Reviews Microbiology* **2**: 747–765.
948 https://doi.org/10.1038/nrmicro955
949
950 60. Moran NA (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**:
951 583–586. https://doi.org/10.1016/S0092-8674(02)00665-7
952

953 61. Morris LA, Voolstra CR, Quigley KM, Bourne DG, Bay LK (2019). Nutrient availability and
954 metabolism affect the stability of coral–symbiodiniaceae symbioses. *Trends in microbiology* **27**:
955 678–689. https://doi.org/10.1016/j.tim.2019.03.004
956
957 62. Nakagawa S, Takai K, Inagaki F, Hirayama H, Nunoura T, Horikoshi K, Sako Y (2005).
958 Distribution, phylogenetic diversity and physiological characteristics of epsilon□Proteobacteria
959 in a deep□sea hydrothermal field. *Environmental Microbiology* **7**:1619–1632.
960 https://doi.org/10.1111/j.1462-2920.2005.00856.x
961
962 63. Nakagawa S, Takaki Y, Shimamura S, Reysenbach AL, Takai K, Horikoshi K (2007). Deep-
963 sea vent ε-proteobacterial genomes provide insights into emergence of pathogens. *Proceedings*
964 *of the National Academy of Sciences* **104**: 12146–12150.
965 https://doi.org/10.1073/pnas.0700687104
966
967 64. Nicks T, Rahn-Lee L (2017). Inside out: Archaeal ectosymbionts suggest a second model of
968 reduced-genome evolution. *Frontiers in Microbiology* **8**: 384.
969 https://doi.org/10.3389/fmicb.2017.00384
970
971 65. Noore J, Noore A, Li B (2013). Cationic antimicrobial peptide LL-37 is effective against
972 both extra-and intracellular *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy* **57**:
973 1283–1290. https://doi.org/10.1128/AAC.01650-12
974
975 66. Orphan VJ, House CH, Hinrichs KU, McKeegan KD, DeLong EF (2001). Methane-
976 consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science*
977 **293**:484–487. https://doi.org/10.1126/science.1061338
978
979 67. Page HM, Fiala-Medioni A, Fisher CR, Childress JJ (1991). Experimental evidence for filter-
980 feeding by the hydrothermal vent mussel, *Bathymodiolus thermophilus*. *Deep Sea Research Part*
981 *A. Oceanographic Research Papers* **38**: 1455–1461. https://doi.org/10.1016/0198-
982 0149(91)90084-S
983

68. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25**: 1043–1055. http://www.genome.org/cgi/doi/10.1101/gr.186072.114

69. Patra AK, Cho HH, Kwon YM, Kwon KK, Sato T, Kato C, Kang SG, Kim SJ (2016). Phylogenetic relationship between symbionts of tubeworm *Lamellibrachia satsuma* and the sediment microbial community in Kagoshima Bay. *Ocean Science Journal* **51**: 317–332. https://doi.org/10.1007/s12601-016-0028-6

70. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**: 417–419. https://doi.org/10.1038/nmeth.4197

71. Petersen JM, Ramette A, Lott C, Cambon□Bonavita MA, Zbinden M, Dubilier N *et al*. (2010). Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma□ and epsilonproteobacteria at four Mid□Atlantic Ridge hydrothermal vent fields. *Environmental Microbiology* **12**: 2204–2218. https://doi.org/10.1111/j.1462-2920.2009.02129.x

72. Ponnudurai R, Heiden SE, Sayavedra L, Hinzke T, Kleiner M, Hentschker C, Felbeck H, Sievert SM, Schlüter R, Becher D, Schweder T. Comparative proteomics of related symbiotic mussel species reveals high variability of host–symbiont interactions. *The ISME Journal* 2020; **14**: 649–656. https://doi.org/10.1038/s41396-019-0517-6

73. Pranal V, Fiala-Médioni A, Guezennec J (1996). Fatty acid characteristics in two symbiotic gastropods from a deep hydrothermal vent of the West Pacific. *Marine ecology progress series* **142**: 175–184. https://doi.org/10.3354/meps142175

74. Pryszcz LP, Gabaldón T (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic acids research* **44**: e113–e113. https://doi.org/10.1093/nar/gkw294

1014 75. Raina JB, Fernandez V, Lambert B, Stocker R, Seymour JR (2019). The role of microbial
1015 motility and chemotaxis in symbiosis. *Nature Reviews Microbiology* **17**: 284–294.
1016 https://doi.org/10.1038/s41579-019-0182-9
1017
1018 76. Reis Md, Yang Z (2011). Approximate likelihood calculation on a phylogeny for Bayesian
1019 estimation of divergence times. *Molecular Biology and Evolution* **28**: 2161–2172.
1020 https://doi.org/10.1093/molbev/msr045
1021
1022 77. Ruan J (2018). SMARTdenovo: Ultra-fast de novo assembler using long noisy reads. Github
1023 Available at: https://github. com/ruanjue/smartdenovo [Accessed January 10, 2019].
1024
1025 78. Ruan J, Li H (2019). Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, 1–
1026 4. https://doi.org/10.1038/s41592-019-0669-3
1027
1028 79. Salmela L, Sahlin K, Mäkinen V, Tomescu AI (2016). Gap filling as exact path length
1029 problem. *Journal of Computational Biology* **23**: 347–361.
1030 https://doi.org/10.1089/cmb.2015.0197
1031
1032 80. Sanders JG, Beinart RA, Stewart FJ, Delong EF, Girguis PR (2013). Metatranscriptomics
1033 reveal differences in in situ energy and nitrogen metabolism among hydrothermal vent snail
1034 symbionts. *The ISME Journal* **7**: 1556–1567. https://doi.org/10.1038/ismej.2013.45
1035
1036 81. Sára M, Sleytr UB (2000). S-layer proteins. *Journal of bacteriology* **182**: 859–868.
1037 https://doi.org/10.1128/JB.182.4.859-868.2000
1038
1039 82. Schroeder BO (2019). Fight them or feed them: how the intestinal mucus layer manages the
1040 gut microbiota. *Gastroenterology Report (Oxford)* **7**: 3–12.
1041 https://doi.org/10.1093/gastro/goy052
1042

1043 83. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov, EM (2015). BUSCO:
1044 assessing genome assembly and annotation completeness with single-copy orthologs.
1045 *Bioinformatics* **31**: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351
1046
1047 84. Smit AFA, Hubley R (2008). RepeatModeler/RepeatModeler. http://www.repeatmasker.org/
1048
1049 85. Stamatakis A, Ludwig T, Meier H (2005). RAxML-III: a fast program for maximum
1050 likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**: 456–463.
1051 https://doi.org/10.1093/bioinformatics/bti191
1052
1053 86. Stanke M, Morgenstern B (2005). AUGUSTUS: a web server for gene prediction in
1054 eukaryotes that allows user-defined constraints. *Nucleic acids research* **33**, W465–W467.
1055 https://doi.org/10.1093/nar/gki458
1056
1057 87. Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y *et al*. (2017). Adaptation to deep-sea
1058 chemosynthetic environments as revealed by mussel genomes. *Nature Ecology & Evolution* **1**:
1059 1–7. https://doi.org/10.1038/s41559-017-0121
1060
1061 88. Suzuki Y, Sasaki T, Suzuki M, Nogi Y, Miwa T, Takai K, Nealson KH, Horikoshi K (2005).
1062 Novel chemoautotrophic endosymbiosis between a member of the Epsilonproteobacteria and the
1063 hydrothermal-vent gastropod *Alviniconcha* aff. *hessleri* (Gastropoda: Provannidae) from the
1064 Indian Ocean. *Applied and Environmental Microbiology* **71**: 5440–5450.
1065 https://doi.org/10.1128/AEM.71.9.5440-5450.2005
1066
1067 89. Takai K, Campbell BJ, Cary SC, Suzuki M, Oida H, Nunoura T, Hirayama H, Nakagawa S,
1068 Suzuki Y, Inagaki F, Horikoshi K (2005). Enzymatic and genetic characterization of carbon and
1069 energy metabolisms by deep-sea hydrothermal chemolithoautotrophic isolates of
1070 Epsilonproteobacteria. *Applied and Environmental Microbiology* **71**: 7310–7320.
1071 https://doi.org/10.1128/AEM.71.11.7310-7320.2005
1072

1073 90. Thakur A, Mikkelsen H, Jungersen G (2019). Intracellular pathogens: host immunity and
1074 microbial persistence strategies. *Journal of immunology research* **2019**: 1356540.
1075 https://doi.org/10.1155/2019/1356540
1076
1077 91. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC
1078 (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**:
1079 2202–2204. https://doi.org/10.1093/bioinformatics/btx153
1080
1081 92. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
1082 Wortman J, Young SK, Earl AM (2014). Pilon: an integrated tool for comprehensive microbial
1083 variant detection and genome assembly improvement. *PloS One* **9**.
1084 https://doi.org/10.1371/journal.pone.0112963
1085
1086 93. Wang JR, Holt J, McMillan L, Jones CD (2018). FMLRC: Hybrid long read error correction
1087 using an FM-index. *BMC bioinformatics* **19**: 50. https://doi.org/10.1186/s12859-018-2051-3
1088
1089 94. Wang Y, Han X, Petersen S, Frische M, Qiu Z, Li H, Li H, Wu Z, Cui R (2017). Mineralogy
1090 and trace element geochemistry of sulfide minerals from the Wocan Hydrothermal Field on the
1091 slow-spreading Carlsberg Ridge, Indian Ocean. *Ore Geology Reviews* **84**: 1–19.
1092 https://doi.org/10.1016/j.oregeorev.2016.12.020
1093
1094 95. Warèn A, Bouchet P (1993). New records, species, genera, and a new family of gastropods
1095 from hydrothermal vents and hydrocarbon seeps. *Zoologica Scripta* **22**: 1–90.
1096 https://doi.org/10.1111/j.1463-6409.1993.tb00342.x
1097
1098 96. Watsuji TO, Yamamoto A, Motoki K, Ueda K, Hada E, Takaki Y, Kawagucci S, Takai K
1099 (2015). Molecular evidence of digestion and absorption of epibiotic bacterial community by
1100 deep-sea crab *Shinkaia crosnieri*. *ISME J* **9**: 821–831. https://doi.org/10.1038/ismej.2014.178
1101
1102 97. Weiss Y, Forêt S, Hayward DC, Ainsworth T, King R, Ball EE, Miller DJ (2013). The acute
1103 transcriptional response of the coral *Acropora millepora* to immune challenge: expression of

GiMAP/IAN genes links the innate immune responses of corals with those of mammals and plants. *BMC genomics* **14**: 400. https://doi.org/10.1186/1471-2164-14-400

98. Wheeler NE, Barquist L, Kingsley RA, Gardner PP (2016). A profile-based method for identifying functional divergence of orthologous genes in bacterial genomes. *Bioinformatics* **32**: 3566–3574. https://doi.org/10.1093/bioinformatics/btw518

99. Windoffer R, Giere O (1997). Symbiosis of the hydrothermal vent gastropod *Ifremeria nautilei* (Provannidae) with endobacteria-structural analyses and ecological considerations. *Biological Bulletin* **193**: 381–392. https://doi.org/10.2307/3593152

100. Wu YW, Simmons BA, Singer SW (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605–607. https://doi.org/10.1093/bioinformatics/btv638

101. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S (2012). FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One* **7**: e52249. https://doi.org/10.1371/journal.pone.0052249

102. Yang Y, Sun J, Sun Y, Kwan YH, Wong WC, Zhang Y, Xu T, Feng D, Zhang Y, Qiu JW, Qian PY (2020). Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts. *The ISME journal* **14**: 135–150. https://doi.org/10.1038/s41396-019-0520-y

103. Yoshimura D, Kajitani R, Gotoh Y, Katahira K, Okuno M, Ogura Y, Hayashi T, Itoh T (2019). Evaluation of SNP calling methods for closely related bacterial isolates and a novel high-accuracy pipeline: BactSNP. *Microbial genomics* **5**: e000261. https://doi.org/10.1099/mgen.0.000261

104. Zamze S, Martinez-Pomares L, Jones H, Taylor PR, Stillion RJ, Gordon S, Wong SY (2002). Recognition of bacterial capsular polysaccharides and lipopolysaccharides by the

macrophage mannose receptor. *Journal of Biological Chemistry* **277**: 41613–41623. https://doi.org/10.1074/jbc.M207057200

105. Zhou Y, Chen C, Sun Y, Watanabe HK, Zhang R, Wang C (2019). *Amphisamytha* (Annelida: Ampharetidae) from Indian Ocean hydrothermal vents: Biogeographic implications. *Deep-Sea Research Part I: Oceanographic Research Papers* **154**: 103148. https://doi.org/10.1016/j.dsr.2019.103148

106. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. *Bioinformatics* **29**: 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Acknowledgements

This work was supported by grants from the China Ocean Mineral Resources Research and Development Association (DY135-E2-1-03), the Hong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC01), the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0409), and the Major Basic and Applied Basic Research Projects of Guangdong Province (2019B030302004-04) awarded to P.Y.Q., and the National Natural Science Foundation of China (NSFC) (grant no. 91951201). Thank Ms Pui Shuen (Joyce) Wong at BioCRF, HKUST for her assistance in PacBio sequencing. Alice Cheung edited the final version of the paper.

Competing interests

The authors declare no competing interests.

Author Contributions

PYQ conceived the project. YY and JS designed the experiments. YZ and CW collected the *Alviniconcha* snails. CC dissected specimens. YY performed DNA extraction, RNA extraction, Nanopore and PacBio sequencing, gene expression and metabolic pathway analyses of the symbionts and the host *Alviniconcha* snail. YY performed genome assemblies, phylogenetic and genomic comparative analyses of the symbionts. JS performed genome assembly, phylogenetic and gene family analyses of the host *Alviniconcha* snail. JS and YY performed genome

annotation of the host *Alviniconcha* snail. YY checked bacteria contamination, performed genomic comparative and the remaining analyses of the host *Alviniconcha* snail. YY prepared the figures and tables and drafted the manuscript. CC, JS, LY, CVD, JWQ and PYQ contributed to manuscript editing.

Figure Legends

Figure 1. Genomic comparisons and gene family analyses across Lophotrochozoa. (**A**) Genome-based phylogeny of selected taxa showing the position of *Alviniconcha marisindica* among lophotrochozoans and divergence times among molluscan lineages. Error bars indicate 95% confidence levels. (**B**) Venn diagram depicting unique and shared gene families among four lophotrochozoan genomes.

Figure 2. Gill endosymbionts and intestinal microbiome of *Alviniconcha marisindica* from the Wocan vent field. (**A**) Circle diagram showing an overview of genome information of the binned endosymbiont based on COG annotation. (**B**) Correlation-based network of intestinal bacteria genera (relative abundance ≥ 0.5%) from three *A. marisindica* individuals. The network analysis displays the intra-associations within each sub-community and inter-associations between sub-communities. Node size is proportional to the number of connections (i.e. degree of connectivity). Connection between nodes represents strong (Spearman correlation efficiency >0.8 (yellow) or <-0.8 (blue)) and significant (p-value <0.01) correlation. The same colour of nodes shows their highly modularised (clustered) property within the network.

Figure 3. Genome-based phylogeny of campylobacterotal representatives. The position of the *Alviniconcha marisindica* endosymbiont among Campylobacterota belongs to the family Sulfurovaceae and marked in red. Nine deltaproteobacterial species are used to root the tree. Different lifestyles of the selected taxa are indicated by squares of different colours (purple: free-living, red: endosymbiont, yellow: epibiont, and green: pathogen/other). The right histogram indicates the size of each genome. The colour of a column represents the size range (grey: >2.0 Mb, dark blue: <2.0 and ≥1.5 Mb, red: <1.5 Mb). The line chart in orange indicates the number of coding sequences (CDS) of each genome. Circles of different colours are used to indicate different ranges of GC content in % (red: ≥38.0, yellow: ≥33.0 and <38.0, green: ≥28.0 and <33.0, black: <28.0). The genome size of the *A. marisindica* endosymbiont is the smallest among whole genomes within the phylum, and its GC content is slightly lower than those of the four closest relatives in the same clade.

Figure 4. **Genomic comparisons and gene family analyses of the Wocan *Alviniconcha marisindica* endosymbiont and four closely related members of Campylobacterota.** (**A**) The loss-of-function genes of the *A. marisindica* endosymbiont shown in different COG categories obtained via pairwise comparison with genomes of the other four Campylobacterota members. The histogram on the left presents the result of comparing the *A. marisindica* endosymbiont with the other four campylobacterotal bacteria. The *A. marisindica* endosymbiont has significantly fewer mutated genes than do the references. (**B**) Venn diagram depicting unique and shared gene families among the five campylobacterotal genomes. (**C**) SNP-based phylogeny on the whole-genome level of 20 endosymbiotic isolates from 10 *A. marisindica* individuals showing the inter- and intra-individual relationships of *A. marisindica* endosymbionts. The genomic ANIs among these 20 isolates obtained via pairwise comparisons are shown in the heat map. The number in the name of each isolate represents the host individual, and the capital A and B represent the anterior and posterior parts of the gills, respectively.

Figure 5. **Overview of metabolic pathways of the *Alviniconcha marisindica* holobiont from the Wocan vent field.** Metabolic pathways of different organisms including the gill endosymbiont, intestinal microbiome, and the *A. marisindica* host are presented in different colours (blue: gill endosymbiont, red: *A. marisindica* host, orange: intestinal microbiome, and grey: missing genes/pathways). Similarly, metabolites from different sources are also shown in different colours (blue: from gill endosymbiont, red: from *A. marisindica* itself, and orange: from intestinal food or microbiome). The compensation mechanism is revealed by the host's collaborating with its symbionts to synthesise nutrients or their mutually using important metabolic intermediates. The combination of endogenous and exogenous energy sources is shown here to explain the adaptive mechanism of the entire *A. marisindica* holobiont.

Figure 6. **Symbiosis constraints of the *Alviniconcha marisindica* holobiont.** (**A**) Heat map of the transcriptional activity of genes involved in host innate immunity in the foot, neck, mantle, intestine, and gill tissues showing distinct immune-expression profiles regulated by the two types of symbionts in the *A. marisindica* snail. Each grid in the heat map represents an identified gene. The colour represents the gene expression level (based on normalised TPM values of the selected tissues). The annotated gene names and their functional classifications are listed on the top side.

(**B**) Symbiosis model of the *A. marisindica* holobiont with two different symbiotic constraints and interactions with the external environment. All pattern recognition receptors (PRRs) and pathogen-associated molecular patterns (PAMPs) shown here are identified from the genome and transcriptome data. SLPs, surface layer proteins; LPS, lipopolysaccharide; CPS, capsular polysaccharides; SIAE, sialate O-acetylesterase; PGRPs, peptidoglycan recognition proteins; TLRs, toll-like receptors; C1q, complement component 1q; T2SS, type II secretion system.

1240 **Table 1.** Comparison of general genomic features of *Sulfurovum alviniconcha* CR and references.

Name	***Lamellibrachia satsuma***	***Sulfurovum lithotrophicum***	***Sulfurovum riftiae***	***Sulfurovum* sp. NBC37-1**	***Sulfurovum alviniconcha* CR**
INSDC	JQIX00000000.1	CP011308.1	LNKT00000000.1	AP009179.1	—
Size (Mb)	2.00	2.22	2.37	2.56	1.47
GC (%)	39.7	44.3	45.6	43.9	37.1
Protein	1,852	2,148	2,317	2,481	1,386
tRNA	37	44	45	44	40
Gene	2,019	2,227	2,432	2,583	1,429
Percentage coding (%)	91.7	96.5	95.3	96.1	97.0
Pseudogene	126	23	62	46	—
Habitat	Trophosome	Sediments	Tube	Sulfide mound	Gills
Depth (m)	~112	1,033	2,500	1,000	2,919

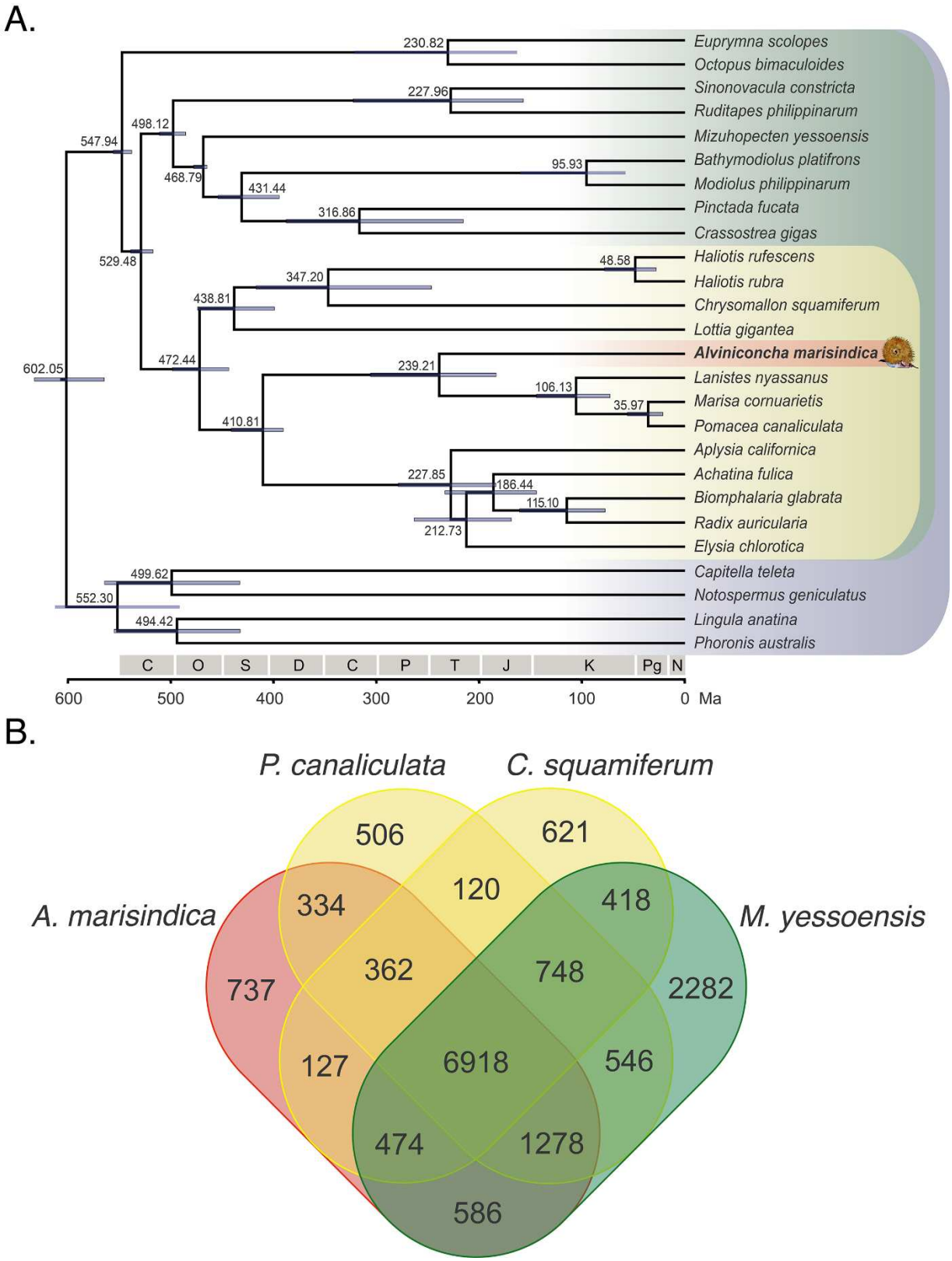
1241 Genome data — *Sulfurovum alviniconcha* CR: this study; the endosymbiont of *Lamellibrachia*
1242 *satsuma*: Patra AK et al., 2016; *Sulfurovum lithotrophicum*: Inagaki F et al., 2004; *Sulfurovum*
1243 *riftiae*: Giovanneli et al., 2016; *Sulfurovum* sp. NBC37-1: Nakagawa et al., 2007.
1244

1245 **Table 2.** General genomic features of the binned endosymbionts of *Alviniconcha marisindica*

1246 extracted from 23 metagenome datasets of gill filaments.

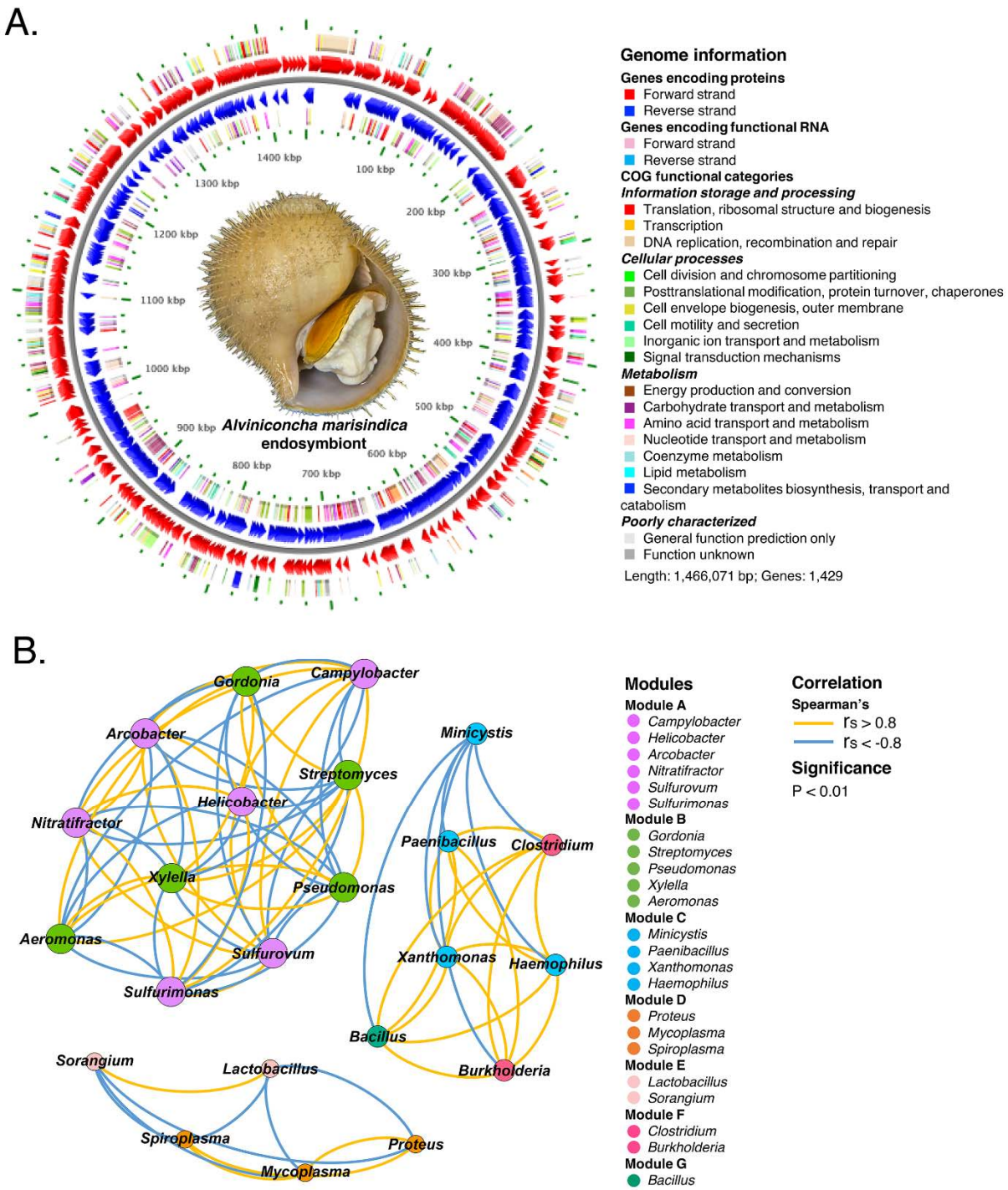
Individual	Sample No.	Gill part	Genome size (Mb)	Contig No.	GC (%)	CDS	Completeness (%)	Contamination (%)
38I-DV129-2	1	random	1.47	2	37.09	1,386	98.16	0.82
38I-DV129-3	2	random	1.31	153	37.49	1,396	96.11	3.89
38I-DV129-16	3	random	1.33	77	37.41	1,384	96.31	1.43
38I-DV129-14-1	4-B	posterior	1.76	298	37.77	1,734	99.18	2.25
	4-A	anterior	1.75	291	37.49	1,745	99.18	2.66
38I-DV129-14-2	5-B	posterior	1.55	126	37.71	1,524	98.36	2.05
	5-A	anterior	1.85	257	37.44	1,816	99.59	3.69
38I-DV129-19-1	6-B	posterior	1.75	306	37.82	1,722	99.18	1.30
	6-A	anterior	1.39	62	37.25	1,394	98.77	1.43
38I-DV129-19-2	7-B	posterior	1.72	149	37.66	1,676	99.18	1.23
	7-A	anterior	1.76	288	37.67	1,759	98.77	0.82
38I-DV129-20	8-B	posterior	1.55	172	37.92	1,525	97.95	2.46
	8-A	anterior	1.60	207	37.64	1,612	98.57	2.87
38I-DV131-1	9-B	posterior	1.46	176	37.79	1,493	98.36	1.23
	9-A	anterior	1.58	143	37.76	1,540	97.95	2.05
38I-DV131-3	10-B	posterior	1.43	134	37.59	1,445	97.95	2.05
	10-A	anterior	1.55	138	37.77	1,528	98.77	2.05
38I-DV131-8	11-B	posterior	1.45	109	37.45	1,541	97.54	2.46
	11-A	anterior	1.53	117	37.83	1,484	98.77	1.64
38I-DV131-9-1	12-B	posterior	1.38	59	37.25	1,409	97.95	0.61
	12-A	anterior	1.55	143	38.00	1,499	97.13	1.23
38I-DV131-9-2	13-B	posterior	1.39	68	37.81	1,430	98.16	1.02
	13-A	anterior	1.43	154	37.31	1,438	97.54	2.25

1247



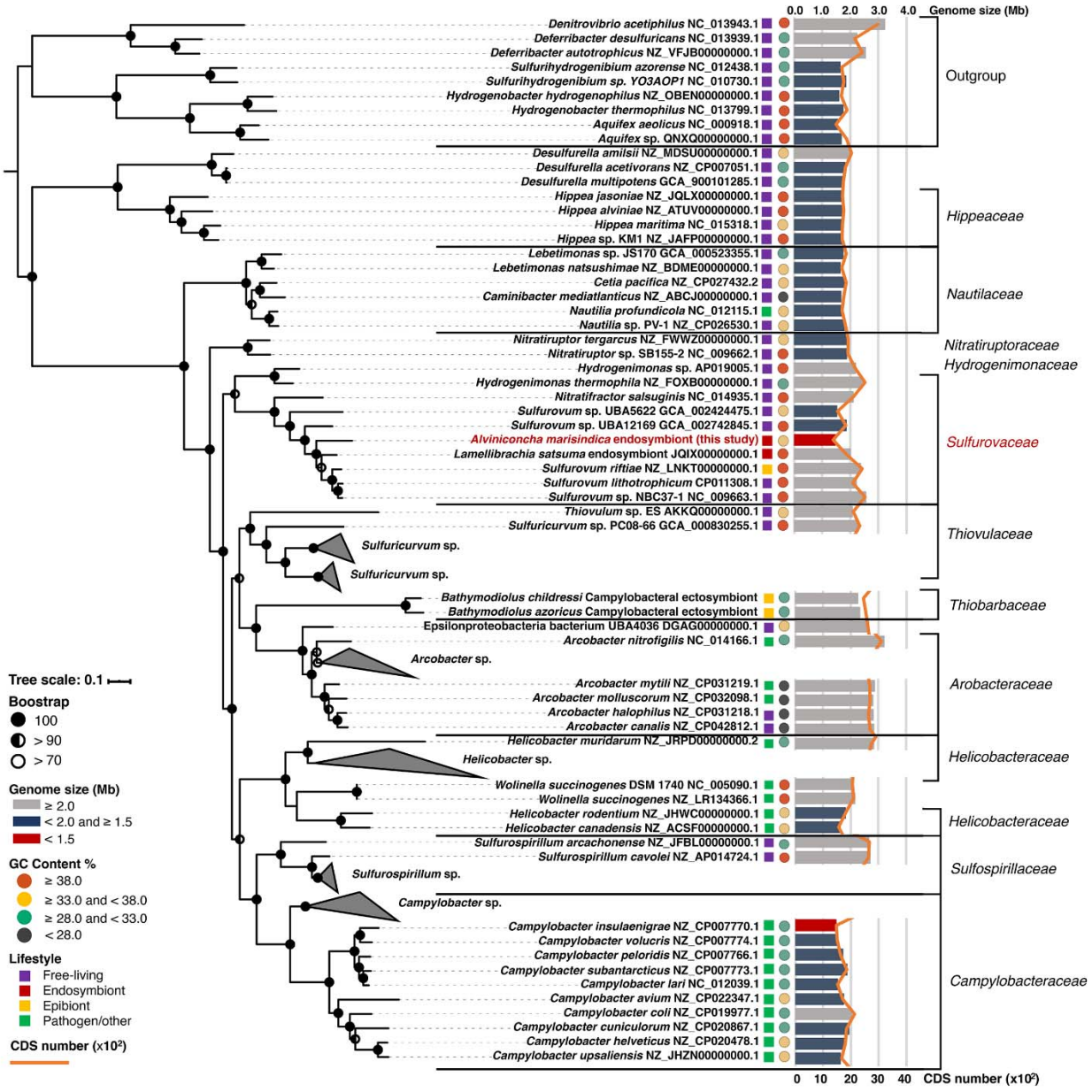
1248

1249 Figure 1.



1250

1251 Figure 2.



1252

1253 Figure 3.

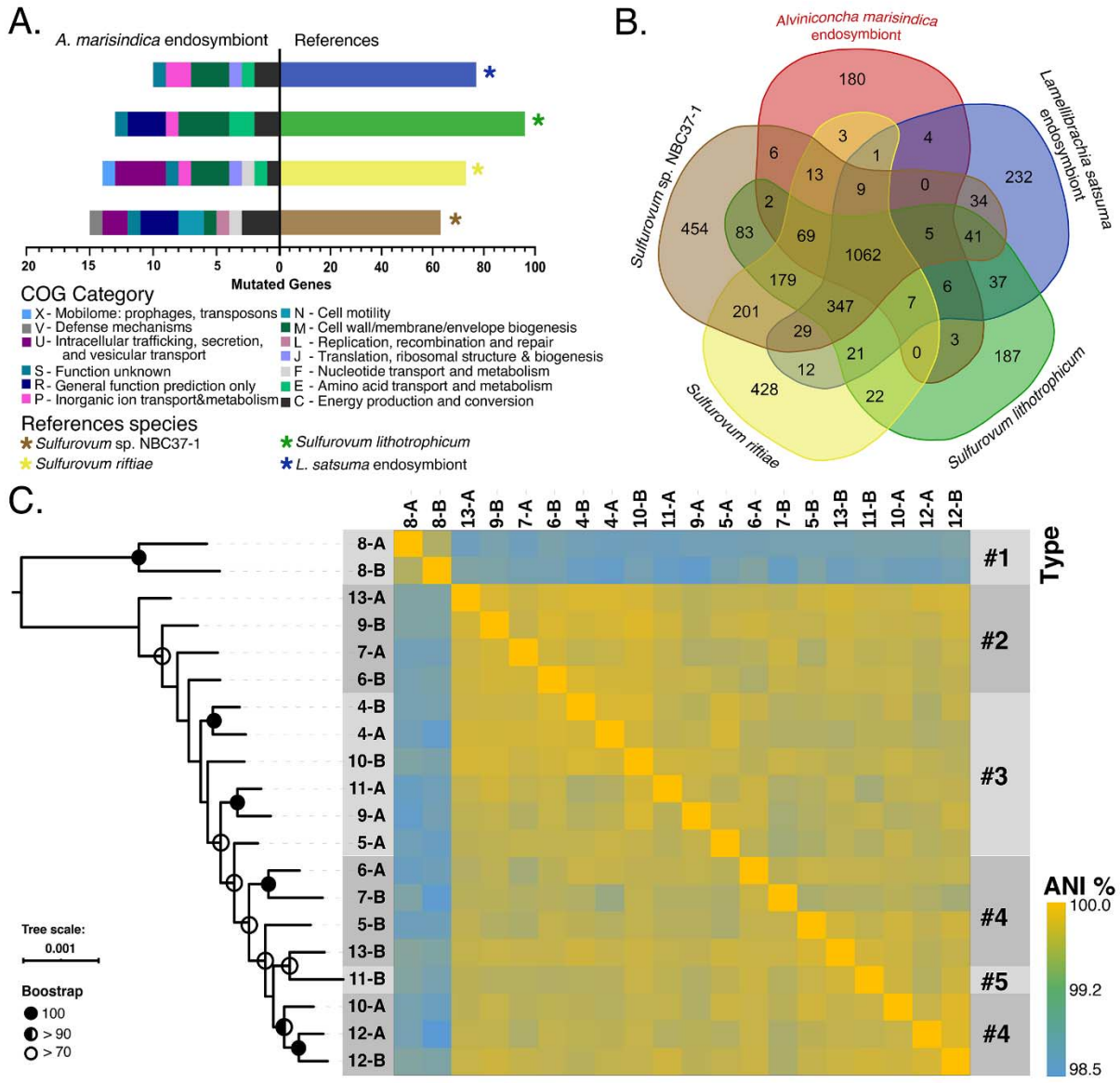
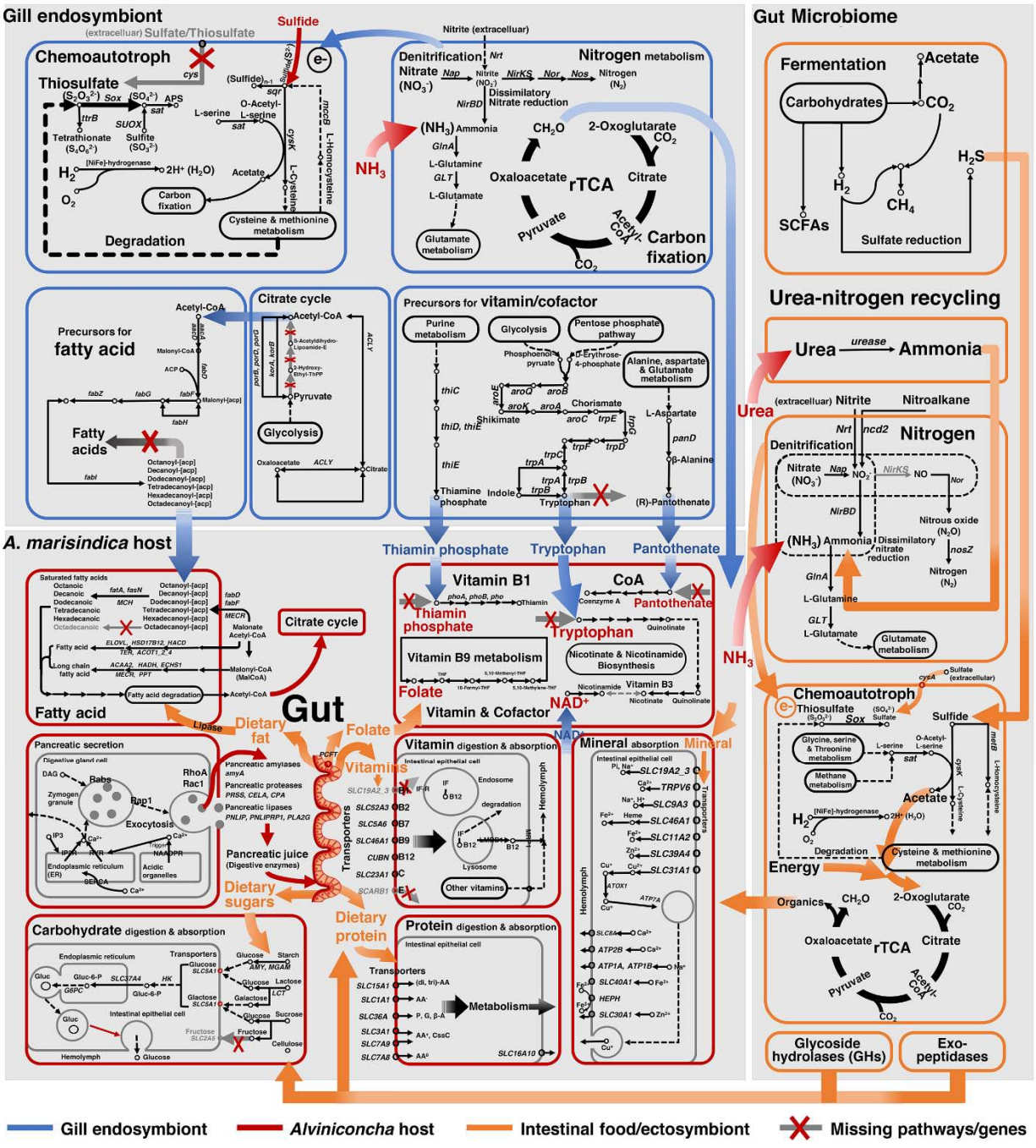
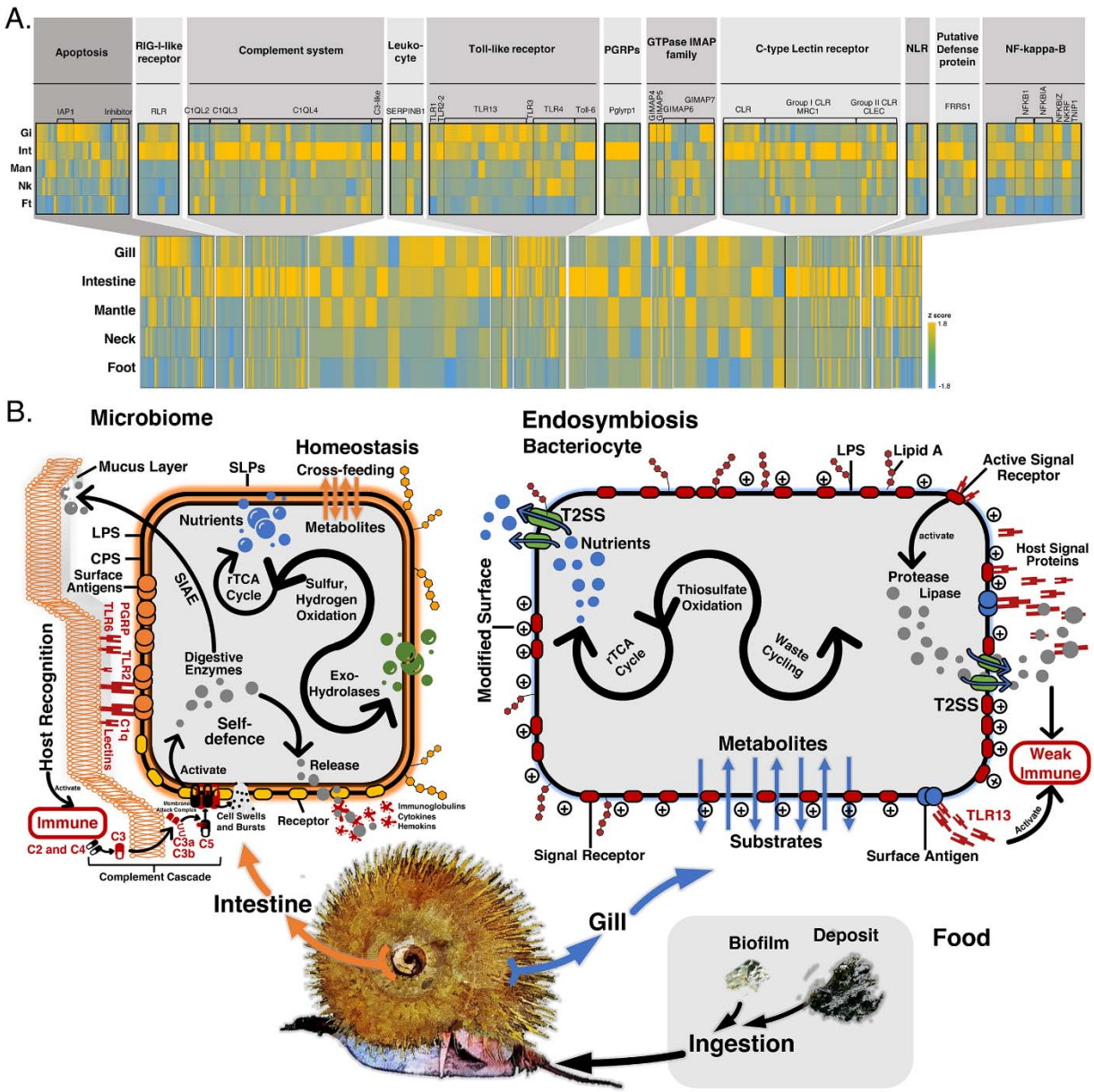


Figure 4.



1256

1257 Figure 5.



1258

1259 Figure 6.