Dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis* revealed by its hologenome

Yi Lan[1], Jin Sun[1], Chong Chen[2], Yanan Sun[1], Yadong Zhou[3], Yi Yang[1], Weipeng Zhang[4], Runsheng Li[5], Kun Zhou[1], Wai Chuen Wong[1], Yick Hang Kwan[1], Aifang Cheng[1], Salim Bougouffa[6,7], Cindy Lee Van Dover[8], Jian-Wen Qiu[5], Pei-Yuan Qian[1*]

[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 Prefecture 237-0061, Japan

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

[4] College of Marine Life Science, Ocean University of China, 5 Yushan Road, Qingdao, 266003, China

[5] Department of Biology, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[6] Computational Bioscience Research Centre, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia

[7] King Abdullah University of Science and Technology (KAUST), Core Labs, Thuwal, 23955-6900, Saudi Arabia

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

* Corresponding author
Prof. Pei-Yuan Qian
E-mail: boqianpy@ust.hk
Tel: +852-2358-7331; Fax: +852-2358-1552

Abstract

Animals endemic to deep-sea hydrothermal vents often form obligatory relationships with bacterial symbionts, maintained by intricate host-symbiont interactions. Endosymbiosis with more than one symbiont is uncommon, and most genomic studies focusing on such 'dual symbiosis' systems have not investigated the host and the symbionts to a similar depth simultaneously. Here, we report a novel dual symbiosis among the peltospirid snail *Gigantopelta aegis* and its two Gammaproteobacteria endosymbionts – one being a sulphur oxidiser and the other a methane oxidiser. We assembled high-quality genomes for all three parties of this holobiont, with a chromosome-level assembly for the snail host (1.15 Gb, N50 = 82 Mb, 15 pseudo-chromosomes). In-depth analyses of these genomes reveal an intimate mutualistic relationship with complementarity in nutrition and metabolic codependency, resulting in a system highly versatile in transportation and utilisation of chemical energy. Moreover, *G. aegis* has an enhanced immune capability that likely facilitates the possession of more than one type of symbiont. Comparisons with *Chrysomallon squamiferum*, another chemosymbiotic snail in the same family but only with one sulphur-oxidising endosymbiont, show that the two snails' sulphur-oxidising endosymbionts are phylogenetically distant, agreeing with previous results that the two snails have evolved endosymbiosis independently and convergently. Notably, the same capabilities of biosynthesis of specific nutrition lacking in the host genome are shared by the two sulphur-oxidising endosymbionts of the two snail genera, which may be a key criterion in the selection of symbionts by the hosts.

Keywords: cryptometamorphosis, methane-oxidising, sulphur-oxidising, Mollusca

Introduction

Animals live in harmony with microorganisms in and around them, forming ecological units referred to as holobionts. Some metazoans, for instance reef-forming corals, have established intricate obligatory symbiotic relationships with single-celled organisms that enable the holobiont to more effectively utilise the available energy source around them[1]. In deep-sea chemosynthetic environments such as hydrothermal vents, many endemic animals harbour endosymbiotic bacteria within their cells in order to directly access chemosynthesis utilising reducing substances such as hydrogen sulphide, hydrogen, thiosulfate, and methane[2]. Most of the species have evolved long-term symbiotic relationships with one dominant phylotype of microbial endosymbiont[3,4,5,6]. Several species of annelids and molluscs, however, are capable of hosting multiple co-occurring endosymbionts[7,8,9,10]. A number of studies have used genomic tools to explore the molecular mechanisms that enable and support symbioses[3,4,11], but most studies focused more closely on either the symbiont or the host. Studying the symbiotic relationship by a total 'hologenome' approach combining high-quality genomes of both the host and the symbiont is key to understanding how the partners cooperate and collaborate to adapt and flourish in extreme environments.

Among deep-sea vent molluscs housing endosymbionts, the two peltospirid genera *Gigantopelta* and *Chrysomallon* are perhaps the most extreme 'holobionts'. They host the symbionts inside a specialised organ within the gut (a modified oesophageal gland) rather than on the gill epithelium like all others[4,8,12,13,14,15]. This means that their endosymbionts are unable to exchange material directly with the vent fluid and are completely reliant on the host for supplies of reducing substances. The two genera are thought to have evolved this symbiosis independently through convergent evolution[15]. A crucial difference between the two is that while *Chrysomallon* relies on endosymbionts for nutrition throughout its post-settlement life[3,14], *Gigantopelta* initially adopts grazing for nutrients and only later (at around 5-7 mm shell length) rapidly shifts to hosting endosymbionts[16]. This process, dubbed 'cryptometamorphosis', occurs with a dramatic reorganisation of its digestive system where the oesophageal gland is greatly enlarged while the other parts of the digestive system effectively stop growing[16].

The fact that representatives of the two genera, *Gigantopelta aegis* and the Scaly-foot Snail *Chrysomallon squamiferum*, co-occur in great abundance in the Longqi vent field on the ultra-slow spreading Southwest Indian Ridge (SWIR) makes them ideal candidates to compare molecular adaptations in two similar-yet-different holobionts. From previous studies, we know that *C. squamiferum* hosts only one phylotype of sulphur-oxidising Gammaproteobacteria whose genome has been sequenced[3]. The whole genome of the host was also recently sequenced at chromosome-level[17], but little has been explored in terms of its adaptations related to symbiosis. In the present study, we reveal that *Gigantopelta aegis* has a dual endosymbiosis, where the snail houses two Gammaproteobacteria endosymbionts, one sulphur-oxidising and one methane-oxidising, in its oesophageal gland. Previously, the 16S rRNA sequencing of the closely related *Gigantopelta chessoia* only revealed a single sulphur-oxidising Gammaproteobacteria symbiont[18], although the transmission electron micrograph of the oesophageal gland hinted a second, rarer, methane-oxidising symbiont[15]. Here, we use high-quality genome assemblies of both the *G. aegis* snail host and its two physiologically distinct

95   endosymbionts to elucidate interactions and co-operations among this 'trinity' of parties in the
96   holobiont. Furthermore, we compare these results to the interactions between *C. squamiferum*
97   and its single sulphur-oxidising endosymbiont in order to understand how may *Gigantopelta*
98   *aegis* benefit from symbiosis from more than one symbiont.
99

100   **Results and Discussion**
101   *A Holobiont with Three Parties and Hologenome Features*
102   The two genera, *Gigantopelta aegis* (dark orange) and the Scaly-foot Snail *Chrysomallon*
103   *squamiferum* (black), occur in great abundance side-by-side at Tiamat chimney in the Longqi
104   vent field (**Fig. 1a**). We obtained high-quality genome assemblies of all three parties in the *G.*
105   *aegis* holobiont using an adult snail 4 cm in shell length, for in-depth analyses of its symbiosis.
106   The host genome had a size of 1.15 Gb, comprising 15 pseudo-chromosomes with an N50 of
107   ~82 Mb (93.8% genome completeness; **Fig. 1b, Table S1**). This is one of only a few
108   chromosome-level genome assemblies currently available in Gastropoda or even Mollusca.
109   The genome size of *G. aegis* is more than double of the *C. squamiferum* genome (~444.4 Mb)[17]
110   mainly due to the larger contribution of repetitive regions (**Fig. S1, Fig. S2, Table S2**). The
111   genomes of both *G. aegis* and *C. squamiferum* snails had 11 *hox* gene clusters in the same order
112   (**Fig. S3**), 13 protein-coding genes of mito-genome in the same order (**Fig. S4**), and one-to-one
113   chromosomal synteny (**Fig. 1c**). These indicate that much of the gene order and synteny are
114   conserved within the family Peltospiridae, and such clear correspondences exemplify the high
115   quality of the genome assembly for both genomes (**Fig. S5**). In addition, the genomic
116   rearrangements that were marked by gene synteny were exclusively restricted to intra-
117   chromosomal rather than inter-chromosomal (**Fig. 1c**). This result is also in line with our
118   former comparative genomic analysis among four congener apple snail genomes, and the intra-
119   chromosomal genomic rearrangements were more likely to be dominant than inter-
120   chromosomal in shaping the genomic differences in Mollusca[19]. The genetic regulatory
121   networks and related functional differences among them are underpinned by intra-
122   chromosomal rearrangements, but how exactly warrants further studies. A total of 21,472 genes
123   (**Table S3**) were predicted from the *G. aegis* genome, of which 1,783 genes were highly
124   expressed in the oesophageal gland. No particular chromosomal distribution bias was observed
125   among these genes (**Fig. 1b**).
126

127   Transmission electron micrographs of the oesophageal gland from *Gigantopelta aegis*
128   confirmed the existence of intracellular endosymbionts densely packed inside bacteriocytes
129   (**Fig. 2a, Fig. S6**). Two types of endosymbionts with distinct morphology features were found
130   within a single bacteriocyte (**Fig. 2a**). Sequences of the 16S ribosomal RNA from oesophageal
131   glands of three *G. aegis* individuals revealed that the two types of symbionts both belong to
132   Gammaproteobacteria, the more abundant one being sulphur-oxidising bacteria (SOB hereafter)
133   and the rarer one being methane-oxidising bacteria (MOB hereafter) in the family
134   Methylococcaceae. Indeed, of the two symbionts identified from TEM, one exhibited
135   intracellular stacked membranes (**Fig. 2a**) which is a characteristic of Type I methanotroph[12],
136   and the other lacked it. Fluorescence *in situ* hybridisation (FISH) carried out on transverse
137   sections of the oesophageal gland from *G. aegis* yielded positive signals of both the SOB and
138   the MOB (**Fig. 2b, Fig. S6**). These results confirm that *G. aegis* is a holobiont with three active

139   parties, including the host and two physiologically distinct types of endosymbionts from
140   Gammaproteobacteria. Genome binning results showed the SOB, the MOB, and the host could
141   be separated with high fidelity by their GC content and sequencing coverage (**Fig. S7**). Both
142   the TEM results and the sequencing coverage indicate that both endosymbionts are common
143   in the oesophageal gland, with the SOB being approximate seven times more abundant than
144   the MOB in this individual (**Table S4**). However, the relative abundance of the two symbionts
145   possibly varies due to the fluctuant concentration of sulphur substances and methane, as was
146   observed in *Bathymodiolus* mussels which have two symbionts[20,21].

147

148   The assembled genomes of the SOB (98.55% genome completeness; 11 scaffolds of 4.91 Mb;
149   5,518 genes) and the MOB (99.25% genome completeness; 10 scaffolds of 2.93 Mb; 3,102
150   genes) (**Table S4, Table S5, Table S6**) showed different genomic features in terms of GC
151   content of the deviation from the average 37.22%, GC skew, and annotated Clusters of
152   Orthologous Groups (COG) (**Fig. 2c**; for details of assembly and functional annotation see
153   Supplementary Materials). Bacterial sequences were barely detectable in the scaffolds of *G.*
154   *aegis*, suggesting a lack of horizontal gene transfer from the prokaryote. The CheckM analysis
155   confirmed the high quality of the two symbionts genomes, as the potential contamination was
156   low (3.25% in SOB, 1.67% in MOB, **Table S4**). In the metaproteomic analyses of the three
157   parties, a total of 704 proteins were identified in the host (**Table S7**), 462 proteins in the SOB
158   (**Table S8**), and 119 proteins (**Table S9**) in the MOB, which provide additional protein
159   evidence to trace the metabolism of *G. aegis* holobiont.

160

161   *Evolution of Symbiosis in Deep-sea Peltospirid Snails*
162   Molecular clock methods based on 1,066 single-copy genes suggests that the two deep-sea
163   peltospirid snails, *G. aegis* and *C. squamiferum*, diverged approximately 117.94 million years
164   ago (Ma; 95% confidence intervals: 68.73~187.03 Ma) (**Fig. 3a**), which is consistent with
165   previous evidence that the ancestor of peltospirids extends to the Late Cretaceous and the order
166   Neomphalida likely had a mid-Jurassic origin[15]. Despite occupying the same microhabitat, the
167   peltospirid snails *G. aegis* and *C. squamiferum* host different lineages of sulphur-oxidising
168   endosymbionts (**Fig. 3b, Table S10**), consistent with the horizontal acquisition of symbionts[8,18]
169   and host-specific selectivity of symbionts relying on the pattern recognition receptors[22]. The
170   host-specific selectivity for different lineages of gammaproteobacterial endosymbionts in
171   peltospirids is in line with a previous study indicating that they have independent origins of
172   endosymbioses[15]. In contrast, provannid *Alviniconcha* snails (*Alviniconcha strummeri, A.*
173   *kojimai*) from vent fields at the Eastern Lau Spreading Center[8] have been found to be capable
174   of accepting a wide range of endosymbionts in each host species, and the endosymbiont has an
175   impact on the holobiont's niche utilisation[23]. This difference between peltospirids and
176   provannids likely indicates that housing endosymbionts in an internal organ is associated with
177   higher host selectivity of endosymbionts.

178

179   Despite the divergent phylogenetic distances between the sulphur-oxidising endosymbionts in
180   the two peltospirid snails, the gene content encoding for nutrition metabolism in the symbionts
181   are similar, and the same goes for the two peltospirid hosts (**Table 1**). Divergence of hosts and

182 their SOBs at the genome level together with the convergence of host and SOB gene content
183 relating to nutrition again indicate that the peltospirid snails established their symbiotic
184 relationships through convergent evolutionary pathways. Therefore, the selection of the
185 symbionts may have been based on their capability of the biosynthesis of specific nutrients,
186 especially key amino acids and vitamins that the hosts are unable to synthesise themselves.
187
188 *Genetic Control of Symbiont Acquisition*
189 The immune response of the host is critical for symbiont infection and colonisation. Among
190 the shared gene families of *G. aegis* and 18 other lophotrochozoan genomes (**Fig. S8**), several
191 immune-related gene families were expanded in *G. aegis* (**Table S11**). For example, the gene
192 families fucolectin (FUCL) and galectin (Gal) were prominently expanded in *G. aegis* (**Fig.**
193 **4a**). These lectins allow host cells to bind with carbohydrate recognition domains on the
194 bacterial surface, resulting in recognition of beneficial bacteria and elimination of detrimental
195 microbes via antimicrobial activity[22]. In comparison to *C. squamiferum*, *G. aegis* harbours
196 much more abundant pattern recognition receptors, including peptidoglycan recognition
197 proteins (PGRP), toll-like receptors (TLR), fibrinogen-related proteins (FBG), and C-type
198 lectins (CLEC) (**Fig. 5**). These receptors are thought to help other chemosymbiotic host
199 animals, such as siboglinid tubeworms and bathymodiolin mussels, to respond to the microbes
200 appropriately as well as assist the acquisition and maintenance of their symbiont populations[22].
201 Pattern recognition receptors that are highly expressed in the oesophageal glands are likely
202 related to the recognition of the endosymbionts. From gene expression patterns, all PGRPs are
203 highly expressed in the oesophageal glands of both peltospirid snails, suggesting PGRPs play
204 an important role in the establishment of the symbiosis in deep-sea peltospirid snails. The
205 abundance and high expression of PGRPs may help the host to regulate the symbiont
206 population[22] via the digestion process (**Table S12**), resulting in efficient utilisation of
207 environmental chemical compounds whose concentrations are expected to fluctuate. Besides
208 PGRPs, many other pattern recognition receptors, particularly those belonging to the gene
209 family TLR, are highly expressed in the oesophageal gland of *G. aegis* compared to *C.*
210 *squamiferum* (**Fig. 5**). Although *G. aegis* occupies the same habitat in the Longqi vent field as
211 *C. squamiferum*, these additional pattern recognition receptors may enable *G. aegis* to host two
212 distinct endosymbionts.
213
214 There is also genomic evidence that the symbionts of the *G. aegis* peltospirids play a role in
215 the invasion of the host. The peptidoglycan-associated lipoprotein Pal that can bind with TLR
216 protein families[8] was expressed in the transcriptomes and proteomes of the *G. aegis* SOB
217 (**Table S5, Table S8**). Furthermore, both the SOB and the MOB of *G. aegis* contained OmpA
218 family proteins and OmpH family proteins (**Table S5**, **Table S6**), which could interact with
219 TLR and scavenger receptors to help the bacteria invade and adapt to the intracellular
220 environment[24]. These symbiont attributes could help the host establish an endosymbiotic
221 lifestyle.
222
223 *Energy Resources*
224 Previous reports showed that the concentrations of reduced components (i.e., CH_4, sulphur
225 substances, and hydrogen) in the endmember fluid from the Longqi field are comparable or

226 higher than those in the vents along the Central Indian Ridge[25,26,27]. Although methane is
227 enriched in Longqi compared to other vents on the Central Indian Ridge[26], further biogenic
228 methane input from microbes living below and around the community may be sufficient to
229 sustain the MOB. The blood vascular system of the host snail transports these essential
230 dissolved compounds from the vent fluid, together with oxygen from the surrounding seawater,
231 to the SOB and MOB **(Table S13)**. The dual symbionts are highly versatile in utilising those
232 chemical substances, respectively (**Fig. 6a**). The central metabolism, associated gene
233 expression levels, and protein abundance levels of the symbionts are summarised in **Figure 6**.

235 The genome data indicate that the sulphur metabolism in the SOB of *G. aegis* (**Fig. 6a**) is
236 similar to that of the *C. squamiferum* snail SOB, *Alviniconcha* snail gamma SOB and siboglinid
237 tubeworm gamma SOB[3,5]. The proteins of *soxA* (35th in proteome), *dsrB* (21th in proteome),
238 *sqr* (34th in proteome), *aprA* (6th in proteome), *aprB* (63th in proteome), and *sat* (57th in
239 proteome) (**Fig. 6**) have a high expression level, indicating their active engagement in oxidising
240 thiosulfate, sulphite, and sulphide in the environment and to detoxify sulphide for the holobiont.

242 Like many sulphur-oxidising endosymbionts of vent molluscs[3,4,28], the SOB of *G. aegis* relies
243 on the complete Calvin-Benson-Bassham (CBB) cycle to fix CO_2 by using a form II ribulose-
244 bisphosphate carboxylase (the protein of *cbbM* gene: 2nd in proteome) that was found the
245 highest protein expression level and lacks a complete rTCA cycle (**Fig. 6**).
246 Phosphoribulokinase (the protein of *prkB* gene: 15th in proteome), type I glyceraldehyde-3-
247 phosphate dehydrogenase (the protein of *gapA* gene: 16th in proteome), and transketolase (the
248 protein of *tkt* gene: 43th in proteome) playing important roles in CBB cycle for carbon fixation
249 also exhibited high protein expression levels in the SOB (**Fig. 6**).

251 The MOB of *G. aegis* has neither a complete CBB cycle nor an rTCA cycle pathway for CO_2
252 fixation. However, the *G. aegis* MOB possesses the methane monooxygenase operon (*pmoA*
253 gene: 5th in transcriptome, 50th in proteome; *pmoB* gene: 6th in transcriptome, 1st in proteome;
254 *pmoC* gene: 1st in transcriptome, 10th in transcriptome) with top 1% gene expression enabling
255 it to oxidise methane (**Fig. 6b, Fig. S9**), confirming that methane oxidation is among the most
256 prominent metabolic processes in the MOB. In the tetrahydromethanopterin (H4MPT) pathway,
257 the high expression level of the formaldehyde activating enzyme (*fae* gene: 65th in
258 transcriptome, 20th in proteome) indicates an active generation of CO_2 via formaldehyde
259 oxidation. Then CO_2 could be redirected to the CBB cycle of the SOB to form an internal
260 carbon cycle among the symbionts. Internal carbon cycling among symbionts has been detected
261 in the internal symbiotic organs (trophosomes) of vent siboglinid tubeworms[29]. This would
262 allow the symbionts to maximise carbon efficiency within the trophosome-like oesophageal
263 glands and trophosomes in the absence of direct interaction with CO_2-rich seawater and vent
264 fluids.

266 Aerobic methanotrophs are classified into two groups based on their formaldehyde assimilation
267 pathways: Type I group (part of Gammaproteobacteria) using a ribulose
268 monophosphate (RuMP) cycle and Type II group (part of Alphaproteobacteria) using a serine
269 cycle. Methane-oxidising symbionts usually belong to Type I and only use the RuMP cycle;

270 examples include the methane-oxidising symbionts of deep-sea *Bathymodiolus* mussels, snails,
271 and sponges (referred-to genomes in **Fig. 3b**). The MOB of *G. aegis*, however, surprisingly
272 possesses a complete serine cycle in addition to a complete RuMP cycle (**Fig. 6a)**. This can
273 mitigate the accumulation and toxicity of formaldehyde when formaldehyde is in excess for
274 the RuMP pathway. In the RuMP cycle, high expression of 3-hexulose-6-phosphate synthase
275 (*hxlA* gene: 20th in transcriptome) and 6-phospho-3-hexuloisomerase (*hxlB* gene: 25th in
276 proteome) in the Entner-Doudoroff (EDD) pathway indicate an important role for
277 formaldehyde oxidation (**Fig. 6**). Furthermore, the MOB genome possesses an additional
278 Embden-Meyerhof-Parnas (EMP) pathway for glycolysis (**Fig. 6a)**, absent in other methane-
279 oxidising endosymbionts (referred in **Fig. 3b**). These results indicate that MOB of *G. aegis* are
280 highly efficient and versatile in assimilating carbon sources from methane.
281
282 *Respiration*
283 Both the SOB and the MOB housed in the oesophageal glands of *G. aegis* are capable of using
284 nitrate as an electron acceptor for anaerobic respiration. Their genomes encode genes for nitrate
285 reduction, such as nitrate reductase (*napAB* and *nasA* in SOB, *narGHI* in MOB), nitrite
286 reductase (*nirBDS* in SOB, *nirBDK* in MOB), and nitric oxide reductase (*norBC* in the SOB
287 and the MOB) (**Fig. 6, Fig. S10**), which are also found in sulphur-oxidising symbionts of
288 tubeworms, mussels, and other gastropods but not in the methane-oxidising symbionts of
289 mussels and sponges[3,5,28,30]. As the *G. aegis* symbionts rely on the host's blood vessels for
290 oxygen supply while input from the hydrothermal vent fluctuates due to the constant interplay
291 between the oxygen-poor vent fluid and the oxygen-rich seawater, there may be periods when
292 the oxygen supply runs low in the oesophageal gland. Indeed, hemocyanin, hemoglobin, and
293 globin-like genes playing a role in oxygen binding and transport have lower gene expression
294 in the oesophageal gland and higher expression in the foot and mantle while five sialin genes
295 used for transporting nitrate were highly expressed in the oesophageal gland (**Fig. S11**).
296 Therefore, these symbionts might have developed abilities to avoid oxygen competition with
297 the host, especially under potential periodic hypoxic conditions.
298
299 *Holobiont Nutritional Interdependence*
300 The SOB do not have *mdh*, *pdhD* and *sucA* genes in the TCA cycle while the homologous
301 genes of those missing in the SOB are highly expressed in the MOB. The SOB does have a C4-
302 dicarboxylate TRAP transport system (DstPQM) that would allow the uptake of four-carbon
303 compounds involved in the TCA cycle[8], indicating that the SOB likely replenishes the missing
304 intermediates of the TCA cycle from the other symbiont partner.
305
306 Both the *G. aegis* and *C. squamiferum* host genomes lack the capability of the biosynthesis of
307 7 amino acids (isoleucine, valine, leucine, lysine, histidine, phenylalanine, and tryptophan), 6
308 vitamins (vitamin B1, vitamin B2, vitamin B5, vitamin B7, vitamin B9, and vitamin B12), and
309 2 cofactors (siroheme and coenzyme A; **Table 1**). However, these nutrients, except vitamin B5
310 and coenzyme A, can be obtained from their SOB through host lysis symbiont cells (**Table**
311 **S12**). The MOB of *G. aegis* has a complete pathway to synthesise vitamin B5 and coenzyme
312 A, whereas the SOBs of two peltospirids lacks a key gene 2-dehydropantoate 2-reductase

313   (*panE*) for generating pantoate, the key substrate for producing pantothenate (**Fig. S12**). We
314   speculate two possibilities: 1) the SOB, particularly of *C. squamiferum*, has a replacement of
315   2-dehydropantoate 2-reductase or replenishment for the biosynthesis of pantothenate and
316   coenzyme A like in the pea amphid holobiont[31], or 2) in the *G. aegis* holobiont, MOB serves
317   as the major source for providing pantothenate and coenzyme A to both the SOB and the host.
318   In return, the host can provide the SOB and the MOB with methionine, tyrosine, vitamin B3,
319   and vitamin B6 through the symbionts' ATP-binding cassette transporters of amino acids.
320   These host-supplied nutrients are critical for fulfilling the nutritional demands of the
321   endosymbionts housed in an internal organ lacking direct contact with the outside world.
322
323   *Transposase Expansion in Three Partners of the* Gigantopelta aegis *Holobiont*
324   The enrichment of the transposable elements (TE) provides beneficial genomic plasticity for
325   adaptation to an intracellular symbiotic life by regulating nearby genes[32]. In the results from
326   the gene family analysis, expansion of TEs were found in all three parties comprising *G. aegis*
327   holobiont (**Fig. 4**). For the host *G. aegis*, these expansions correspond to 342 transposase genes
328   (1.6% of total host genes) and 510 DNA transposons (2.4% of total host genes); for SOB
329   symbionts, these expansions correspond to 704 transposase genes (12.8% of total SOB genes);
330   and for the MOB symbionts, these expansions correspond to 345 transposase genes (11.1% of
331   total MOB genes). Numerous transposase genes were unusually inserted into the flagellar
332   operons and chemotaxis operons playing an important role in host infection by both SOB and
333   MOB of *G. aegis*. This may contribute to the regulation of the symbionts motility for infection
334   and colonisation in the host environment. To our knowledge, the expansion of transposase
335   genes occurring across multiple symbiotic partners within the same holobiont has not been
336   reported in any marine invertebrate holobionts.
337

338   **Conclusions**
339   We revealed that the deep-sea peltospirid snail *Gigantopelta aegis* endemic to hydrothermal
340   vent exhibits a dual symbiotic lifestyle, which includes dominant sulphur-oxidising
341   Gammaproteobacteria and less dominant methane-oxidising Gammaproteobacteria (Type I
342   methanotroph). Furthermore, we provide a high-quality hologenome including a chromosome-
343   level assembly for the snail host and an in-depth analysis of the genomic interdependencies
344   among the trinity of symbiotic partners in this holobiont, finding an intimate mutualistic
345   relationship with complementarity in nutrition and metabolic codependency. The two
346   endosymbionts live in an organ lacking direct interaction with the ambient environment. Both
347   symbionts as well as the host are highly versatile with regards to the transportation and
348   utilisation of chemical energy, increasing the efficiency of carbon fixation by forming an
349   internal carbon cycle among the two symbionts. Although *G. aegis* and another
350   chemosymbiotic peltospirid snail *Chrysomallon squamiferum* occupy the same habitat in the
351   Longqi vent field, *C. squamiferum* has a single sulphur-oxidising symbiont. The fact that the
352   sulphur-oxidising endosymbionts of both snails are not particularly closely related
353   phylogenetically but have essentially the same capabilities of the biosynthesis of specific
354   nutrition suggest that this may be a key criterion for the selection of symbionts by the snail
355   hosts.
356

Materials and Methods

Deep-Sea Sampling

Gigantopelta aegis snails were collected by the manned submersible HOV *Jiaolong* on-board the R/V *Xiangyanghong 9* cruise 35II from Longqi hydrothermal vent field (37.7839°S, 49.6502°E; 2,761 m depth) on the Southwest Indian Ridge in January 2015. Snails were immediately flash-frozen in liquid nitrogen once they were recovered on the ship. They were then transferred to -80 °C until DNA and RNA extraction. Specimens of *G. aegis* for Hi-C sequencing and fluorescence *in situ* hybridisation (FISH) experiments were collected also from Longqi ('Tiamat' chimney) during Leg 3 of the COMRA R/V *Dayang Yihao* expedition 52 in April 2019. Tissue from the foot was immediately dissected from one individual, cut up, washed in phosphate-buffered saline buffer, and stored at -80 °C until Hi-C library preparation. The oesophageal gland tissue was dissected from the same individual and fixed in 4% paraformaldehyde overnight, followed by dehydration through an ethanol series (20%, 40%, 60%, and 80%) for 15 minutes each and stored at -80 °C until fluorescence *in situ* hybridisation experiments. Samples for transmission electron microscopy (TEM) were immediately fixed in 10% buffered formalin after recovery on-board the research vessel. Identity of the specimens was confirmed by both mitochondrial *COI* sequences and morphology.

TEM

The oesophageal gland tissue was dehydrated using a series of graded acetone and then transferred into Epon resin (Sigma-Aldrich) for embedding. An ultramicrotome (Reichert Ultracut S, Leica) was used to slice ultrathin (70 nm) sections that were used for staining in 2% aqueous uranyl acetate with lead stain solution (0.3% lead acetate and 0.3% lead nitrate, Sigma-Aldrich). The presence of intracellular symbionts was confirmed using a Tecnai 20 Transmission Electron Microscopy (FEI) at an acceleration voltage of 120 kV. For more detailed methods for thin-section preparation see ref 15.

FISH

The oesophageal gland tissues were dehydrated in 100% ethanol and embedded in Shandon Cryomatrix Frozen Embedding Medium (6769006; Thermo Fisher Scientific). After full embedding, they were quickly frozen in dry ice. Cryostat sections 10 μm in thickness (CryoStar NX70; Thermo Fisher Scientific) were cut and mounted on glass slides. Two 16S rRNA-based probes were designed to target the endosymbionts specifically. The Cy3-labelled SOB1 probe (5'-AGCATATTAAACTTGTACCC-3') was used to target the sulphur-oxidising endosymbiont, and the Cy5-labelled MOB1 probe (5'-CGTGTGTTTTCCTCCCTTCT-3') was used to bind the methane-oxidising endosymbiont. The sections were rehydrated in a decreasing ethanol series (95%, 80%, and 70%) for 15 minutes each and hybridised at 46 °C with 50 ng/ml of each probe in a hybridisation buffer (0.9 M NaCl, 0.02M Tris-HCl, 0.01% sodium dodecyl sulphate and 20% formamide) for 3 hours. Then, the slides were washed in a wash buffer (0.1 M NaCl, 0.02 M Tris-HCl, 0.01% sodium dodecyl sulphate, 5 mM EDTA) at 48 °C for 15 min. Two drops of 4',6-diamidino-2-phenylindole (DAPI) were added on each slide followed by incubation at room temperature for 3 min. After washing and air drying, the slides were mounted by coverslips with SlowFade™ Diamond Antifade Mountant medium

400   (Invitrogen, Carlsbad, CA, USA). Images were taken using a confocal microscope (Leica
401   Microsystems, Wetzlar, Germany) with a 3D model, and post-processed by LAS X software
402   (Leica Microsystems, Wetzlar, Germany).
403

404   *DNA Extraction and Illumina Sequencing of the Hologenome*
405   The MagAttract High-Molecular-Weight DNA Kit (QIAGEN, Hilden, Netherlands) was used
406   to extract high-molecular-weight genomic DNA from the oesophageal gland and the foot
407   separately for sequencing of the genome of the symbionts and the host, respectively, following
408   manufacturer's protocols. The extracted DNA was further purified by Genomic DNA Clean &
409   Concentrator^{TM}-10 kit (ZYMO Research, Irvine, CA, USA). The DNA was finally eluted in
410   10mM Tris-HCl buffer (pH=8.5). The quality of the genomic DNA was evaluated using a
411   BioDrop μLITE (BioDrop, Cambridge, UK) with the OD 260/280 of 1.8 and the OD 260/230
412   of 2.0–2.2. The DNA concentration was assessed using a Quibit^{TM} 3 Fluorometer (Thermo
413   Fisher Scientific, Singapore). The sizes of DNA fragments were assessed by pulsed-field gel
414   electrophoresis. Approximate 1 μg DNA of each tissue, including foot and oesophageal gland,
415   were used for the short-insert library (350 bp and 500 bp) in an Illumina NovaSeq 6000
416   platform to generate 171 Gb and 49 Gb paired-end reads with a length of 150 bp, respectively.
417

418   *Oxford Nanopore Technologies (ONT) library preparation and MinION sequencing*
419   The DNA of the oesophageal gland was used to prepare the long-read library (Oxford
420   Nanopore MinION, UK) for genome sequencing of endosymbionts. To construct the MinION
421   library, a NEBNext Ultra II End-Repair/dA-tailing Module (NEBE7546) kit was used to
422   perform DNA repair and end-prep. A Ligation Sequencing Kit (SQK-LSK109) was used to
423   perform adaptor ligation and clean-up with the large DNA fragments. A Flow Cell Priming Kit
424   (EXP-FLP001) was used to prepare the priming buffer to prime the MinION flow cells (FLO-
425   MIN106D R9 version). The Library Loading Bead Kit R9 version (EXP-LLB001) was used to
426   assist in loading the libraries into the flow cells. Sequencing was started immediately in the
427   MinION^{TM} portable single-molecule nanopore sequencer (Oxford Nanopore Technologies
428   Limited, UK) using the software MinKNOW version 3.1.8 (Oxford Nanopore Technologies
429   Limited, UK). The read event data was base-called by Oxford Nanopore basecaller Guppy
430   version 2.1.3[33] with default settings.
431

432   *Single-Molecule Real-Time (SMRT) Library Construction and PacBio Sequel Sequencing*
433   For genome sequencing of the host animal, the purified DNA of the foot from one individual
434   was utilised to construct PacBio single-molecule real-time (SMRT) library (Pacific
435   Biosciences, USA). BluePipin (Sage Science, USA) was utilised to select long DNA fragments
436   with sizes between 8 kb to 50 kb for PacBio library construction. Universal hairpin adapters
437   were used to ligate the DNA fragments. The adapter dimers and hairpin dimers were both
438   removed using magnetic bead in PacBio's MagBead kit. Exonucleases were further used to
439   remove the failed ligation DNA fragments. The purified library was sequenced by binding the
440   sequencing polymerase in the SMRT sequencing cells and generated long subreads with a
441   length distribution (**Fig. S13**).
442

443   *Hi-C Sequencing*

444   To construct Hi-C library, the frozen foot tissue of *G. aegis* was thawed slowly on ice and
445   suspended in 45 mL of 37% formaldehyde in serum-free Dulbecco's modified Eagle's medium
446   (DMEM) for chromatin cross-linking. After incubation at room temperature for 5 minutes,
447   glycine was added to quench the formaldehyde, followed by incubation at room temperature
448   for another 5 minutes and subsequently on ice for over 15 minutes. The cells were further lysed
449   in pre-chilled lysis buffer (10 mM NaCl, 0.2% IGEPAL CA-630, 10 mM Tris-HCl, and 1×
450   protease inhibitor solution) using a Dounce homogenizer. The chromatin was digested by Mbo
451   I restriction enzyme, marked with biotin and ligated[34]. The Hi-C library was sequenced on a
452   NovaSeq 6000 platform (Illumina, USA) and generated reads with a length of 150 bp. The
453   genome sequencing strategy of the host was summarised in **Table S14**.
454

455   *RNA Extraction, Metatranscriptome Sequencing, and Eukaryotic Transcriptome Sequencing*
456   The total RNA of oesophageal glands dissected from three individuals of *Gigantopelta aegis*
457   were separately extracted using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA). The
458   quality and quantity of RNA were checked by agarose gel electrophoresis and BioDrop μLITE
459   (BioDrop, Cambridge, UK), respectively. To enrich the mRNA of endosymbionts, the rRNA
460   of both eukaryote and bacteria were removed by both the Ribo-Zero[TM] Magnetic Kit
461   (Human/Mouse/Rat) (Epicenter, USA) and the Ribo-Zero[TM] Magnetic Kit (Bacteria)
462   (Epicenter, USA) following the manufacturers' protocols. The remaining RNA was used to
463   construct cDNA libraries for metatranscriptome sequencing. The sequencing data are shown
464   in **Table S15**.
465

466   Four *G. aegis* individuals were dissected to obtain different organs (foot, mantle, ctenidium,
467   oesophageal gland, etc) for eukaryotic transcriptome sequencing (**Table S15**). Total RNA of
468   these samples were separately extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA)
469   following the manufacturers' protocol. mRNA from each host tissue was enriched by Oligo-
470   dT probes and further transferred to cDNA for eukaryotic library construction. All cDNA
471   libraries were sequenced in a NovaSeq 6000 Illumina platform to generate paired-end reads
472   with a length of 150 bp.
473

474   *Hologenome Assembly and Scaffolding*
475   Adaptors and low-quality bases of raw Illumina sequencing reads were trimmed using
476   Trimmomatic version 0.36[35] with the settings "ILLUMINACLIP: Truseq3-PE-2.fa". To obtain
477   the genome of symbionts, clean Illumina reads from the oesophageal gland were used for initial
478   assembly in metaSPAdes[36] with a series of kmers of 51, 61, 71, 81, and 91 bp. Illumina short
479   reads belonging to the symbionts were grouped by genome binning according to both the
480   sequencing coverage and the GC content of the initial contigs[5]. The grouped Illumina short
481   reads and all ONT sequencing reads were reassembled together by Unicycler version 0.4.7[37].
482   The assembled contigs were grouped again via genome binning. Each group of contigs were
483   further used in genome scaffolding by SSPACE-LongRead version 1.1[38] with the ONT reads
484   and mapped PacBio reads that are over 10 kb. CheckM version 1.0.13[39] was used to assess the
485   completeness and the contamination of the assembled bacterial genomes.
486

487   Several assembly pipelines were used for assembly, and the best contig assembly result was

488   generated by "Canu correction + wtdbg2" (**Table S16**). PacBio subreads over 4 kb were

489   corrected by Canu version 1.7.1[40] with settings of "genomeSize = 1.27 Gb, corMhapSensitivity

490   = normal, corMinCoverage = 0, corMaxEvidenceErate = 0.15, correctedErrorRate = 0.065,

491   minReadLength = 8000" and then used in genome assembly using wtdbg2 version 2.1[41] with

492   settings of "-e 2 --tidy-reads 5000 -S 1 -k 15 -p 0 --rescue-low-cov-edges --aln-noskip". To

493   improve the accuracy of the initially assembled contigs, two polishing rounds using Racon

494   version 1.3.1[42] were performed and followed by an additional round using Pilon version 1.22[43]

495   with Illumina short reads that were mapped by Bowtie2 version 2.3.5[44] in the "--sensitive"

496   mode. Bacterial contamination was further filtered using MaxBin version 2.2.5[45].

497

498   Raw Hi-C sequencing reads were first trimmed by Trimmomatic version 0.36[35]. To remove the

499   invalid pairs of Hi-C reads without effective ligation, the reads were mapped to the corrected

500   assembled contigs using Bowtie2 version 2.3.5[44], and Hi-C contact maps were generated based

501   on the mapped reads using HiC-Pro[46]. Juicer version 1.5[47] was used to filter and deduplicate.

502   The remaining valid reads were used for contig scaffolding using the 3D *de novo* assembly

503   (3D-DNA) pipeline[48] in diploid mode. The scaffolds were manually corrected based on the Hi-

504   C contact maps (**Fig. S14**). The continuity and completeness of the assembled genome were

505   evaluated with assemblathon_stats.pl (https://github.com/ucdavis-

506   bioinformatics/assemblathon2-analysis/blob/master/assemblathon_stats.pl) and BUSCO

507   version 3.0.2[49] using the "metazoa_odb9" database, respectively.

508

509   *Genes Prediction and Functional Annotation*

510   The coding sequences and proteins of symbionts genomes were predicted by Prodigal version

511   2.6.3[50] with default settings. The host genome was first hard-masked in the repetitive regions

512   and then trained with the boundary of the introns and exons in order to predict coding genes of

513   *G. aegis* (for details see Supplementary Materials).

514

515   Repeats in the *Gigantopelta aegis* genome were *de novo* identified and classified using

516   RepeatModeler version 1.0.11 (http://www.repeatmasker.org/RepeatModeler/) pipeline and

517   soft-masked using RepeatMasker version 4.0.8

518   (http://www.repeatmasker.org/RMDownload.html) with the parameter "-xsmall" (for details

519   see Supplementary Materials).

520

521   To predict the gene models, we run two rounds of the MAKER pipeline version 2.31.10[51] on

522   the soft-masked genome. To guide the prediction, we used *de novo* assembled transcripts,

523   Metazoan protein sequences downloaded from the Swiss-Prot database, as well as protein

524   sequences of *Chrysomallon squamiferum*[17] (for details see the Supplementary Materials).

525

526   The predicted protein sequences were searched against NCBI Non-Redundant (NR) database

527   using BLASTp with an *E*-value cutoff of 1e-5. The sequences were also used to search the

528   Kyoto Encyclopedia of Genes and Genomes (KEGG) database via KEGG Automatic

529   Annotation Server (KAAS) using bi-directional best hit methods of BLASTp[52]. The protein

530   sequences of the host were further searched against the EuKaryotic Orthologous Groups (KOG)

531 database. The protein sequences of symbionts were annotated against the functional Clusters
532 of Orthologous Groups (COG) using eggNOG-mapper v2[53]. Pfam database[54] with hmmscan
533 (http://hmmer.org/) was used to identify the functional domains of proteins.
534
535 *Gene Family and Phylogenomic analysis*
536 To perform gene family analysis of *G. aegis*, Orthofinder version 2.3.3[55] was used to identify
537 the orthologous groups shared between proteins of *G. aegis* and those of other molluscan
538 genomes (see Supplementary Materials). Only single-copy orthologs were used to construct
539 the phylogenetic tree. The protein sequences of each gene were aligned separately using
540 MAFFT v7.407[56] and were concatenated for phylogenetic analysis using FastTree version
541 2.1.10[57] with partitions. The software MCMCtree[58] was used to yield the time-calibrated tree
542 by calibrating the phylogenetic tree with seven fossil records and geographic events (see
543 Supplementary Materials).
544
545 Both CAFE 3[59] and Fisher's Exact Test were used for gene family analysis. The time-calibrated
546 tree and gene family numbers shared by these species were used for gene family analysis in
547 CAFE 3[59]. The Fisher's Exact Test was also used for gene family analysis based on the gene
548 family numbers of *G. aegis* and the average gene family numbers of other species included.
549 Only the gene families with a FDR corrected *P*-value smaller than 0.05 were considered as
550 expanded or contracted.
551
552 To explore the phylogenetic relationship of the symbionts, 30 genomes of bacterial symbionts
553 belonging to Gammaproteobacteria from deep-sea invertebrate taxa were included (details of
554 referred symbionts see **Table S17**). Same pipelines of orthologs cluster and sequence
555 alignments of *G. aegis* were applied in the symbionts analyses. The IQ-TREE multicore
556 version 1.6.10[60] with LG+I+G4+F were used to perform the phylogeny analysis of the
557 symbionts with 1000 ultrafast bootstraps.
558
559 *Gene Expression Analysis of the Holobiont*
560 The coding DNA sequences of the predicted genes were used as references to assess gene
561 expression levels. Kallisto version 0.45.1[61] was used to calculate transcripts per million (TPM)
562 value according to the mapped reads count for each gene. In symbionts, the TPM values were
563 directly used for gene expression level comparison. For the host, the highly expressed genes in
564 the oesophageal gland were identified by differential expression analysis versus foot, ctenidium,
565 and mantle ($n = 4$) in edgeR[62]. If genes had a fold change larger than two and a significant FDR
566 *P*-value (< 0.05), they were designated as highly expressed genes. To compare the genes
567 expression of *G. aegis* and *C. squamiferum*, the same analysis pipeline was applied to the
568 transcriptome sequencing data of foot tissue, ctenidium, mantle, and oesophageal gland from
569 the *C. squamiferum*[17].
570
571 *Synteny Analysis*
572 Proteins of *G. aegis* and *C. squamiferum* were searched against each other via BLASTp with
573 an *E*-value cutoff of 1e-5. The hits were used to detect collinear blocks and synteny shared

574 between these two deep-sea gastropods by MCScanX[63]. JCVI
575 (https://github.com/tanghaibao/jcvi) was utilised to generate the chromosome-scale synteny
576 plot.
577
578 *Metaproteomics*
579 The oesophageal glands from three *Gigantopelta aegis* individuals were used for protein
580 extraction using the methanol-chloroform method[64]. SDS-PAGE gel was used to separate
581 different size (ranging from 10 kDa to 150 kDa) of ~30 μg extracted protein from each sample
582 and stained by colloidal coomassie blue. The peptide for LC-MS/MS was obtained through
583 protein reduction, alkylation and digestion, peptide extraction and further dry. Dionex UltiMate
584 3000 RSLCnano coupled with an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher)
585 was utilised to analyse each protein fraction. The search database contains the protein
586 sequences predicted from the genome and the corresponding reversed sequences (decoy) of
587 both *Gigantopelta aegis* and its two endosymbionts. Mascot version 2.3.0 was used to identify
588 and quantify the protein via the raw mass spectrometry data. Proteins were identified with the
589 assigned peptides' identification confidence level over 0.95 and a false discovery rate of less
590 than 2.5% (see Supplementary Materials).
591

604

605 **Data Availability**
606 All sequencing data, assembly data, predicted genes, and proteins of *Gigantopelta aegis* and
607 its two symbionts were submitted to the database of National Centre for Biotechnology
608 Information under BioProject PRJNA612619.
609

610 **Author Contributions**
611 P-YQ conceived the project. YL, JS, CC, and P-YQ designed the experiments. YZ, JS, and YS
612 collected the samples of *Gigantopelta aegis* gastropods. CC, JS, YL, and YY dissected the
613 *Gigantopelta aegis* samples. CC performed TEM experiments. YL, JS, AC, and WZ
614 contributed to the FISH experiments. YL, SB, RL, and JS performed host genome assembly.
615 YL and YS performed the gene prediction of *G. aegis*. RL drew Circos plots of the hologenome.
616 YL and JS contributed to phylogenomic analysis and calibration of the molecular clock of the
617 host. YL, KZ, YY, and WZ contributed to the symbiont genome assembly. WCW, YHK, JS,

618 and YL performed the proteome analysis. YL performed the DNA extraction, RNA extraction,
619 ONT sequencing, gene expression analysis, gene family analysis, synteny analysis, immune
620 response analysis, and drafted the manuscript. J-WQ and CVD contributed to manuscript
621 editing. All authors contributed to the manuscript and approved it for submission and
622 publication.
623

624 **Competing interests**
625 The authors declare no competing interests.
626

627 **References**

628 1. Robbins, S. J. et al. A genomic view of the reef-building coral *Porites lutea* and its
629 microbial symbionts. *Nat. Microbiol.* **4**, 2090–2100 (2019).
630 2. Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of
631 harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6**, 725–740 (2008).
632 3. Nakagawa, S. et al. Allying with armored snails: the complete genome of
633 gammaproteobacterial endosymbiont. *ISME J.* **8**, 40–51 (2014).
634 4. Newton, I. L. G. et al. The *Calyptogena magnifica* chemoautotrophic symbiont
635 genome. *Science* **315**, 998–1000 (2007).
636 5. Yang, Y. et al. Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-
637 sea tubeworm holobionts. *ISME J.* **14**, 135–150 (2019).
638 6. Lan, Y. et al. Host-symbiont interactions in deep-sea chemosymbiotic vesicomyid clams:
639 insights from transcriptome sequencing. *Front. Mar. Sci.* **6**, 680 (2019).
640 7. Fujiwara, Y., Kato, C., Masui, N., Fujikura, K. & Kojima, S. Dual symbiosis in the cold-
641 seep thyasirid clam *Maorithyas hadalis* from the hadal zone in the Japan Trench, western
642 Pacific. *Mar. Ecol. Prog. Ser.* **214**, 151–159 (2001).
643 8. Beinart, R. A., Luo, C., Konstantinidis, K., Stewart, F. & Girguis, P. R. The bacterial
644 symbionts of closely related hydrothermal vent snails with distinct geochemical habitats
645 show broad similarity in chemoautotrophic gene content. *Front. Microbiol.* **10**, 1818
646 (2019).
647 9. Ansorge, R. et al. Functional diversity enables multiple symbiont strains to coexist in deep-
648 sea mussels. *Nat. Microbiol.* **4**, 2487–2497 (2019).
649 10. Geier, B. et al. Spatial metabolomics of in situ host–microbe interactions at the micrometre
650 scale. *Nat. Microbiol.* **5**, 498–510 (2020).
651 11. Sun, J. et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel
652 genomes. *Nat. Ecol. Evol.* **1**, 121 (2017).
653 12. Won, Y. J. et al. Environmental acquisition of thiotrophic endosymbionts by deep-sea
654 mussels of the genus *Bathymodiolus*. *Appl. Environ. Microbiol.* **69**, 6785–6792 (2003).
655 13. Goffredi, S. K., Warén, A., Orphan, V. J., Van Dover, C. L. & Vrijenhoek, R. C. Novel
656 forms of structural integration between microbes and a hydrothermal vent gastropod from
657 the Indian Ocean. *Appl. Environ. Microbiol.* **70**, 3082–3090 (2004).
658 14. Chen, C., Linse, K., Copley, J. T. & Rogers, A. D. The 'scaly-foot gastropod': a new genus
659 and species of hydrothermal vent-endemic gastropod (Neomphalina: Peltospiridae) from
660 the Indian Ocean. *J. Molluscan Stud.* **81**, 322–334 (2015).
661 15. Chen, C., Uematsu, K., Linse, K. & Sigwart, J. D. By more ways than one: rapid

662 convergence at hydrothermal vents shown by 3D anatomical reconstruction of
663 *Gigantopelta* (Mollusca: Neomphalina). *BMC Evol. Biol.* **17**, 62 (2017).

664 16. Chen, C., Linse, K., Uematsu, K. & Sigwart, J. D. Cryptic niche switching in a
665 chemosymbiotic gastropod. *Proc. Biol. Sci.* **285**, 20181099 (2018).

666 17. Sun, J. et al. The scaly-foot snail genome and the ancient origins of biomineralised armour.
667 *Nat. Commun.* **11**, 1657 (2020).

668 18. Heywood, J. L., Chen, C., Pearce, D. A. & Linse, K. Bacterial communities associated with
669 the Southern Ocean vent gastropod, *Gigantopelta chessoia*: indication of horizontal
670 symbiont transfer. *Polar Biol.* **40**, 2335–2342 (2017).

671 19. Sun, J. et al. Signatures of divergence, invasiveness, and terrestrialization revealed by four
672 apple snail genomes. *Mol. Biol. Evol.* **36**, 1507–1520 (2019).

673 20. Salerno, J. L. et al. Characterization of symbiont populations in life-history stages of
674 mussels from chemosynthetic environments. *Biol. Bull.* **208**, 145–155 (2005).

675 21. Duperron, S. et al. Diversity, relative abundance and metabolic potential of bacterial
676 endosymbionts in three *Bathymodiolus* mussel species from cold seeps in the Gulf of
677 *Mexico. Environ. Microbiol.* **9**, 1423–1438 (2007).

678 22. Wippler, J. et al. Transcriptomic and proteomic insights into innate immunity and
679 adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis. BMC*
680 *Genomics* **17**, 942 (2016).

681 23. Beinart, R. A. et al. Evidence for the role of endosymbionts in regional-scale habitat
682 partitioning by hydrothermal vent symbioses. *Proc. Natl. Acad. Sci. USA* **109**, E3241–
683 E3250 (2012).

684 24. Jeannin, P. et al. Complexity and complementarity of outer membrane protein A recognition
685 by cellular and humoral innate immunity receptors. *Immunity* **22**, 551–560 (2005).

686 25. Kawagucci, S. et al. Fluid chemistry in the Solitaire and Dodo hydrothermal fields of the
687 Central Indian Ridge. *Geofluids* **16**, 988–1005 (2016).

688 26. Ji, F. et al. Geochemistry of hydrothermal vent fluids and its implications for subsurface
689 processes at the active Longqi hydrothermal field, Southwest Indian Ridge. *Deep-Sea Res.*
690 *PT. I* **122**, 41–47 (2017).

691 27. Tao, C. et al. Deep high-temperature hydrothermal circulation in a detachment faulting
692 system on the ultra-slow spreading ridge. *Nat. Commun.* **11**, 1300 (2020).

693 28. Ponnudurai, R. et al. Metabolic and physiological interdependencies in the *Bathymodiolus*
694 *azoricus* symbiosis. *ISME J.* **11**, 463–477 (2017).

695 29. Fisher, C. R., Kennicutt, M. C. & Brooks, J. M. Stable carbon isotopic evidence for carbon
696 limitation in hydrothermal vent vestimentiferans. *Science* **247**, 1094–1096 (1990).

697 30. Rubin-Blum, M. et al. Fueled by methane: deep-sea sponges from asphalt seeps gain their
698 nutrition from methane-oxidising symbionts. *ISME J.* **13**, 1209 (2019).

699 31. Price, D. R. & Wilson, A. C. A substrate ambiguous enzyme facilitates genome reduction
700 in an intracellular symbiont. *BMC Biol.* **12**, 110 (2014)

701 32. Kleiner, M., Young, J. C., Shah, M., VerBerkmoes, N. C. & Dubilier, N. Metaproteomics
702 reveals abundant transposase expression in mutualistic endosymbionts. *mBio* **4**, e00223-13
703 (2013).

704 33. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for
705 Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).

34. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

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

36. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

37. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595 (2017).

38. Boetzer, M. & Pirovano, W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).

39. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

40. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

41. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2019).

42. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).

43. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS ONE* **9**, e112963 (2014).

44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

45. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2015).

46. Servant, N. et al. HiC-Pro: An optimized and flexible pipeline for Hi-C processing. *Genome Biol.* **16**, 259 (2015).

47. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

48. Dudchenko, O. et al. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

49. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

50. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

51. Cantarel, B. L. et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).

52. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).

53. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology

750   assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).

751 54. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**,
752   D427–D432 (2019).

753 55. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome
754   comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157
755   (2015).

756 56. Katoh, K., Kuma, K. I., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy
757   of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).

758 57. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum-evolution
759   trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009)

760 58. dos Reis, M. & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian
761   estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).

762 59. Han, M. V., Thomas, G. W., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and
763   loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol.*
764   *Biol. Evol.* **30**, 1987–1997 (2013).

765 60. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and
766   effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol.*
767   *Evol.* **32**, 268–274 (2015).

768 61. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
769   quantification. *Nat. Biotechnol.* **34**, 525 (2016).

770 62. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
771   differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–
772   140 (2010).

773 63. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny
774   and collinearity. *Nucleic Acids Res.* **40**, e49–e49 (2012).

775 64. Wessel, D. M. & Flügge, U. I. A method for the quantitative recovery of protein in dilute
776   solution in the presence of detergents and lipids. *Anal Biochem.* **138**, 141–143 (1984).

777

Tables

Table 1 Genome capabilities of *Gigantopelta aegis* and *Chrysomallon squamiferum* holobionts[3,17] in the biosynthesis of different nutrients, including amino acids, cofactors and vitamins (SOB: sulphur-oxidising endosymbiont, MOB: methane-oxidising endosymbiont).

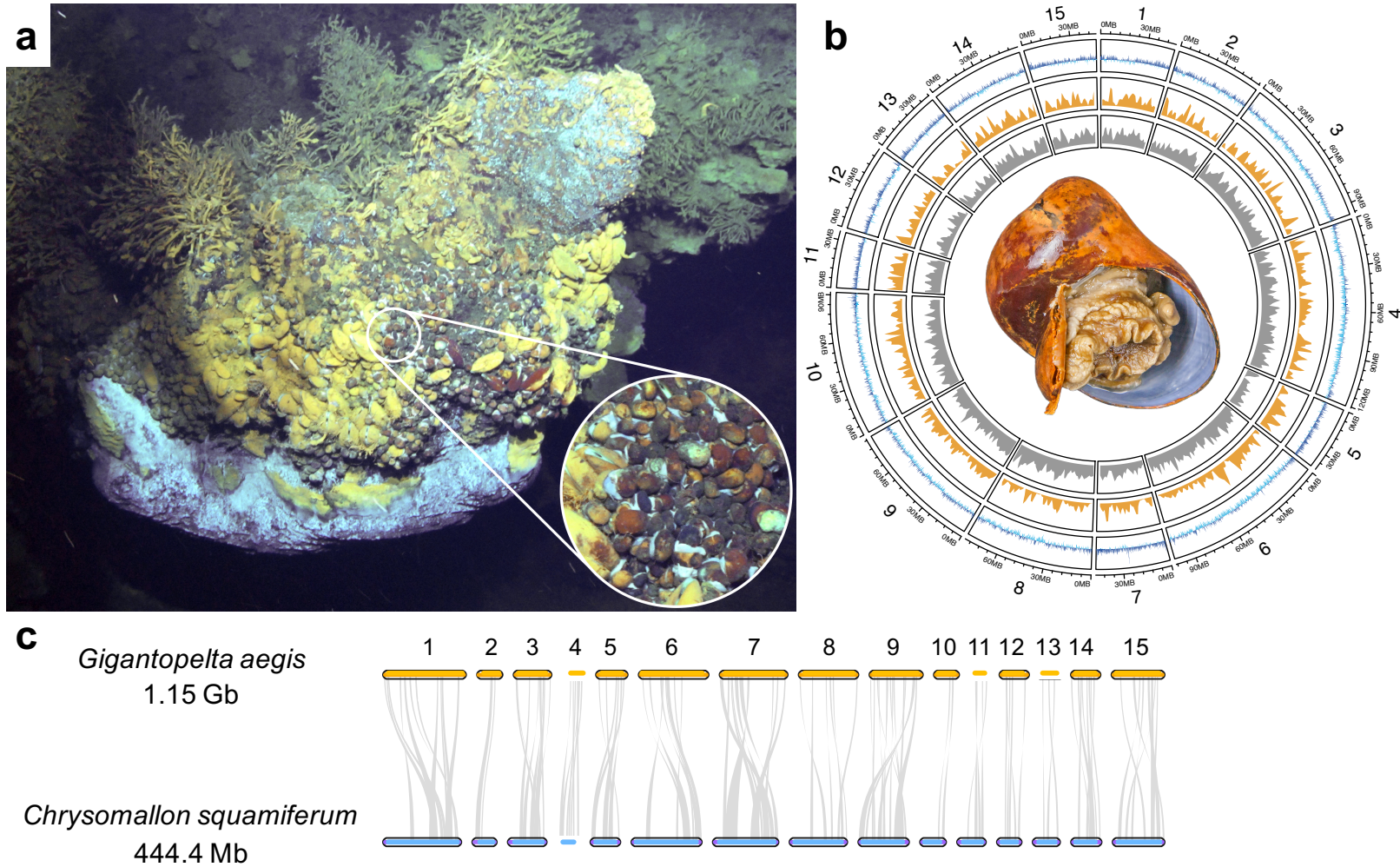| Nutrients	*Gigantopelta aegis*			*Chrysomallon squamiferum*	
	SOB	MOB	Host	SOB[3]	Host[17]
Amino acid					
Alanine	+	+	+	+	+
Glutamine	+	+	+	+	+
Glutamate	+	+	+	+	+
Serine	+	+	+	+	+
Glycine	+	+	+	+	+
Threonine	+	+	+	+	+
Cysteine	-	+	+	-	+
Methionine	-	-	+	-	+
Isoleucine	+	+	-	+	-
Valine	+	+	-	+	-
Leucine	+	+	-	+	-
Lysine	+	+	-	+	-
Arginine	+	+	+	+	+
Proline	+	+	+	+	+
Asparagine	+	+	+	+	+
Aspartate	+	+	+	+	+
Histidine	+	+	-	+	-
Phenylalanine	+	+	-	+	-
Tryptophan	+	+	-	+	-
Tyrosine	-	-	+	-	+
Cofactor / Vitamin					
Vitamin B1	+	+	-	+	-
Vitamin B2	+	+	-	+	-
Vitamin B3	-	-	+	-	+
Vitamin B5	-	+	-	-	-
Vitamin B6	-	-	+	-	+
Vitamin B7	+	+	-	+	-
Vitamin B9	+	+	-	+	-
Vitamin B12	+	+	-	+	-
Coenzyme A	-	+	-	-	-
Ubiquinone	+	+	+	+	+
Siroheme	+	+	-	+	-

782

783
Figure 1 a. Two chemosymbiotic peltospirid snails, *Gigantopelta aegis* and the Scaly-foot Snail *Chrysomallon squamiferum*, occuring in great abundance side-by-side at Tiamat chimney in the Longqi vent field on the ultra-slow spreading Southwest Indian Ridge (*G. aegis* in dark orange; *C. squamiferum* in black). **b**. A circos plot shows key features of 15 pseudo-chromosomal linkage groups of the *Gigantopelta aegis* genome showing an adult snail at the centre, the inner circle (gray) indicating the gene density of each pseudo-chromosome, the second circle (orange) indicating the gene density of highly expressed genes in the oesophageal gland ($n = 4$), and the outer circle (blue) showing the GC content of the deviation from the average 37.22%. A sliding window of 100 kb in a step of 50kb was applied in the calculation of the GC content. **c**. Synteny blocks shared between *G. aegis* (orange) and *C. squamiferum* (blue). The pseudo-chromosome numbers of both *G. aegis* and *C. squamiferum* are 15. The synteny blocks are distributed in one-to-one correspondence among the pseudo-chromosomes without cross-chromosomal synteny.
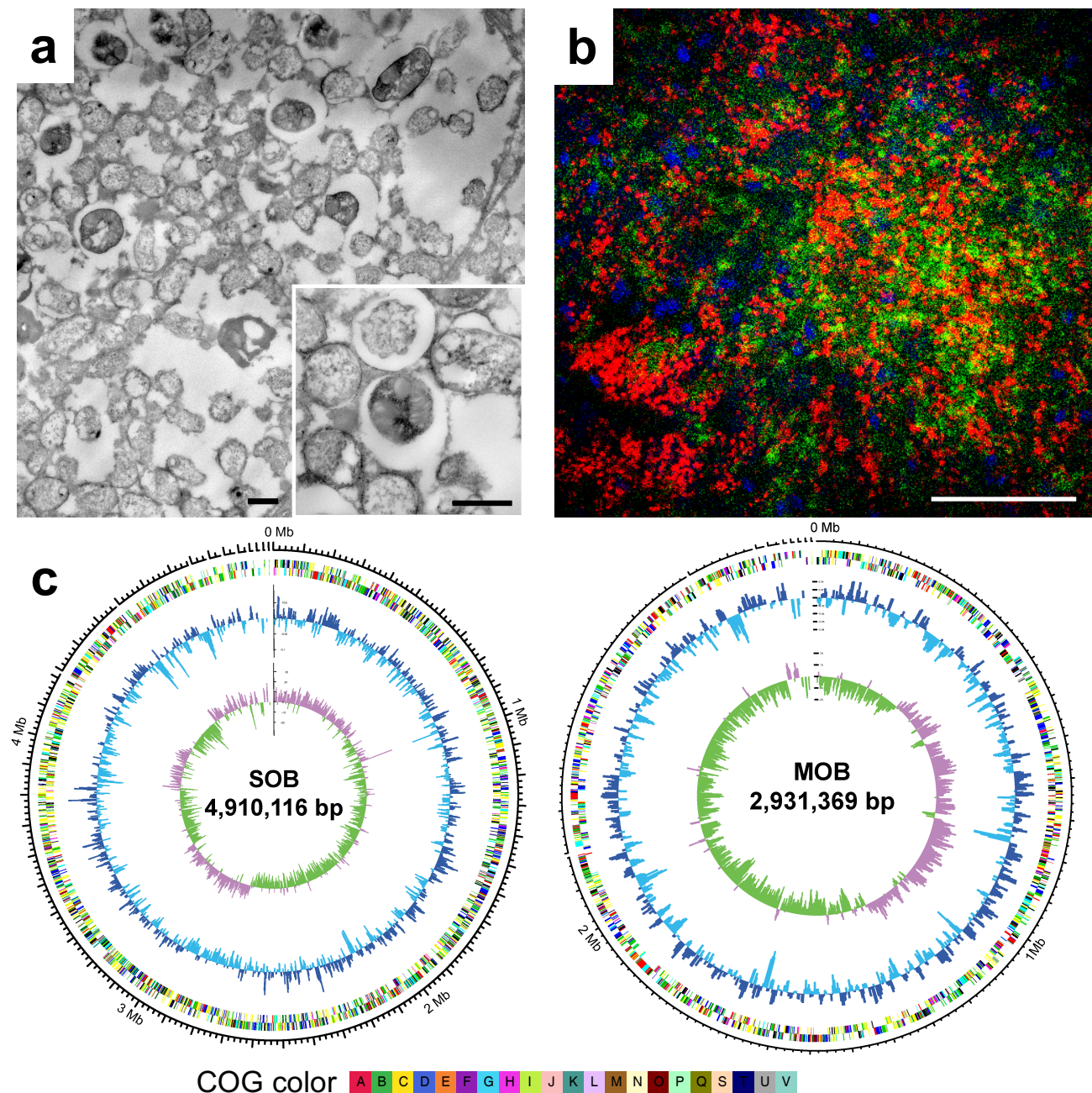
Figure 2 a. Transmission electron microscopy (TEM) images of intracellular endosymbionts showing two distinct morphological types, including one endosymbiont with intracellular stacked membranes and another without (scale bar: 1 μm). **b.** Fluorescence *in situ* hybridisation (FISH) image yielding signals of sulphur-oxidising endosymbionts (SOB; Cy3: green) and the methane-oxidising endosymbionts (MOB; Cy5: red) on transverse sections of the oesophageal gland from *Gigantopelta aegis* (scale bar: 50 μm). Host nuclear DNA with DAPI staining is blue. More detailed TEM and FISH images are shown in Figure S6. **c**. Key genome features of both the SOB and the MOB of *G. aegis* with the inner circle showing the GC skew, the second circle showing the GC content of the deviation from the average (SOB: 39.93%; MOB: 45.26%), and the outer circle showing the COG annotation categories with different colours. A sliding window of 5 kb in a step of 2.5 kb was applied in the calculation of both the GC skew and the GC content.
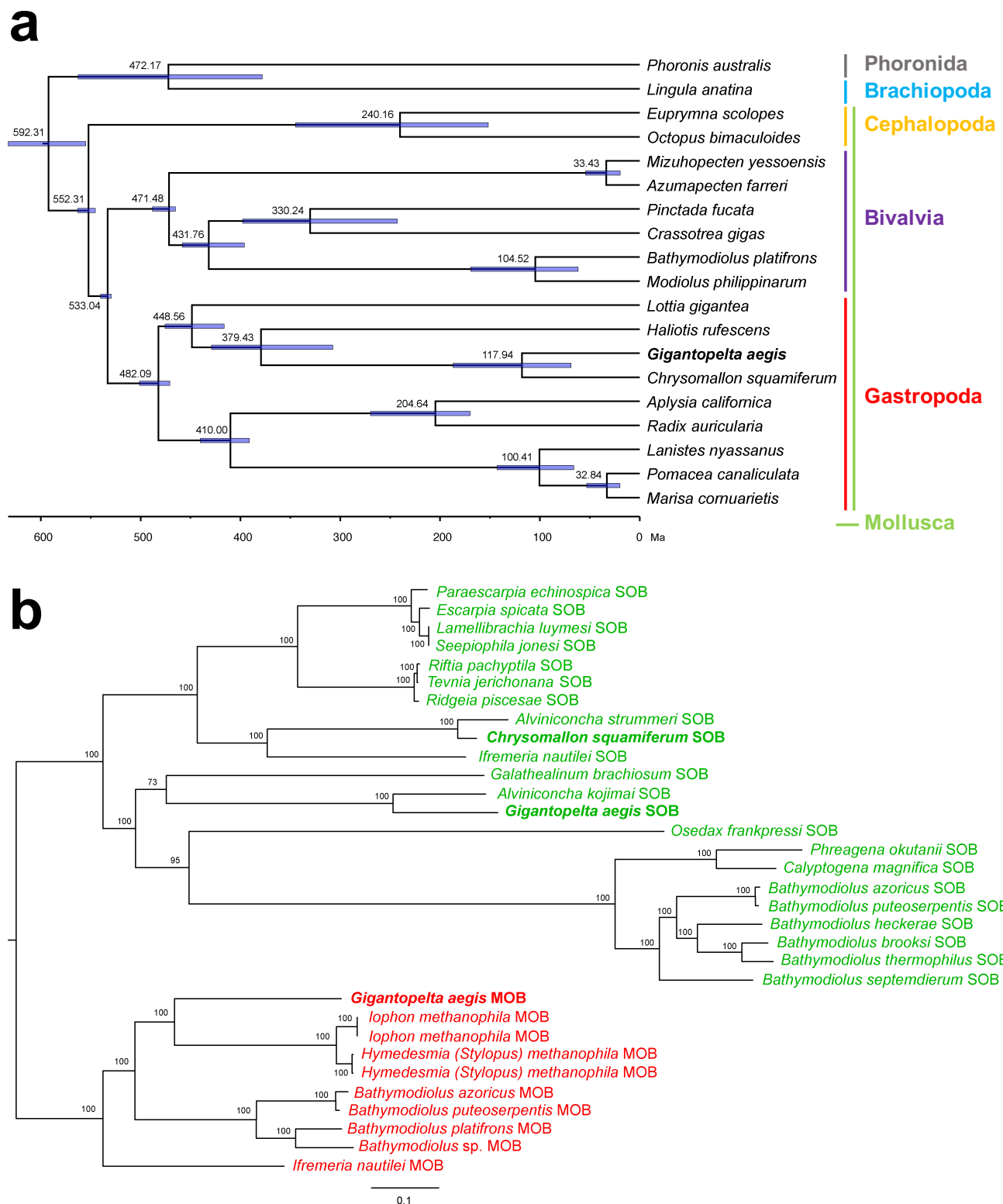
Figure 3 a. Time-calibrated phylogeny showing the estimated divergence time of 19 lophotrochozoan taxa, with a focus on 17 molluscs. *Phoronis australis* (Phoronida) and *Lingula anatina* (Brachiopoda) served as outgroups for the molluscs. A total of 1,066 single-copy orthologs were used. The purple horizontal lines indicate the 95% confidence intervals of the divergence times. For genome resources, fossil records and geographic events used in the time calibration see Supplementary Materials. **b**. Phylogenetic reconstruction of 32 symbionts in Gammaproteobacteria (See Supplementary Materials) from invertebrate taxa living in deep-sea chemosynthetic environments. A total of 424 single-copy orthologs were used to construct the tree. The phylogenetic tree includes two major lineages of the sulphur-oxidising bacteria (green) and the methane-oxidising bacteria (red).
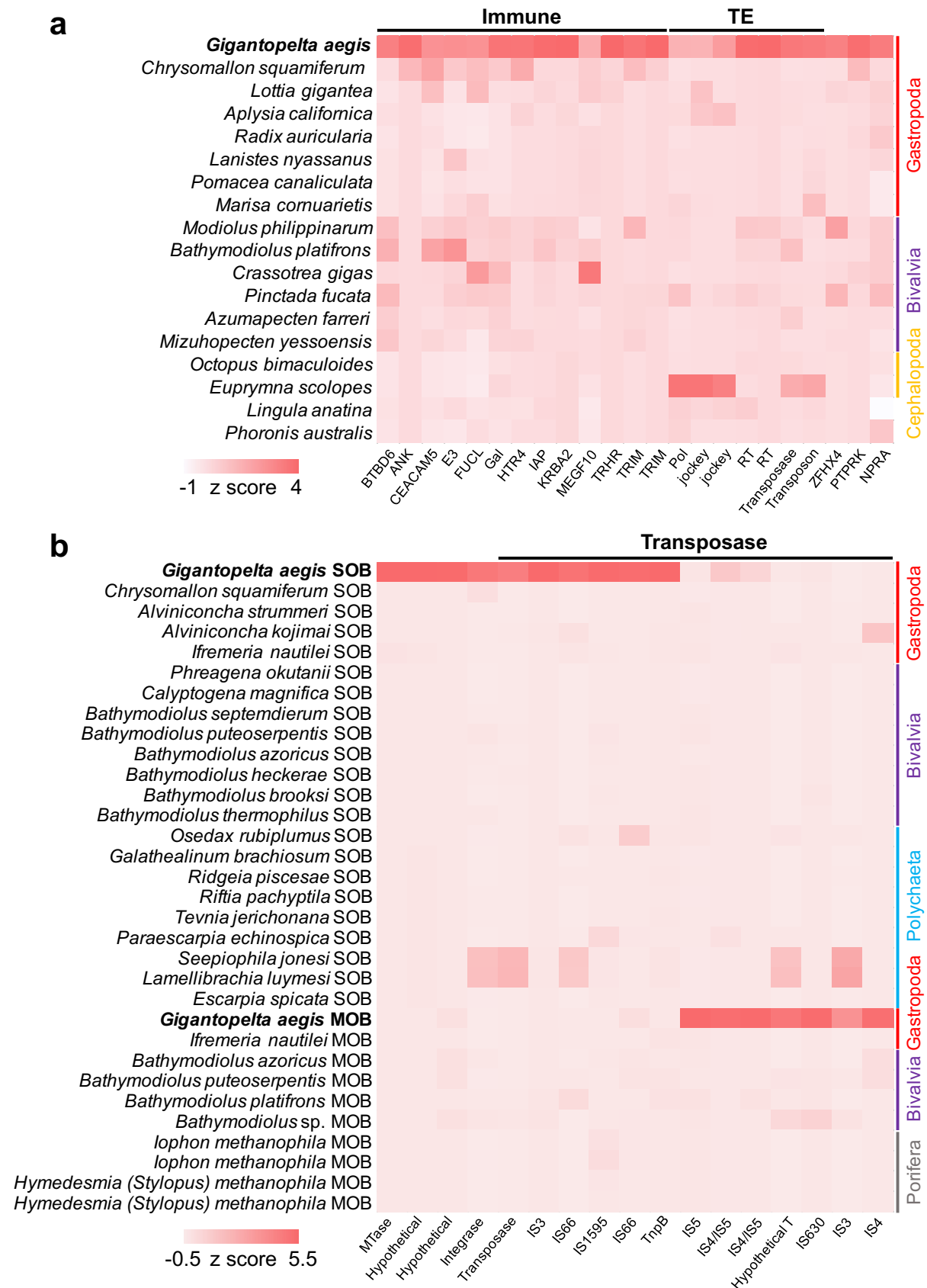
Figure 4 Heat maps showing the expanded gene families in *Gigantopelta aegis* and its two endosymbionts. **a**. Compared to other 15 molluscan taxa (systematic affinity at the class level shown on the right side), *Phoronis australis* (Phoronida), and *Lingula anatina* (Brachiopoda) with available genomes (used in the molecular clock analysis), the gene families expanded in *G. aegis* mainly include immune-related gene families (Immune) and transposable elements (TEs). **b**. Both the sulphur-oxidising endosymbionts and the methane-oxidising endosymbionts exhibit the transposases expansion. References are 30 symbionts in Gammaproteobacteria (See Supplementary Materials) from invertebrate taxa living in deep-sea chemosynthetic environments. The systematic affinity of the host taxa is shown on the right side.
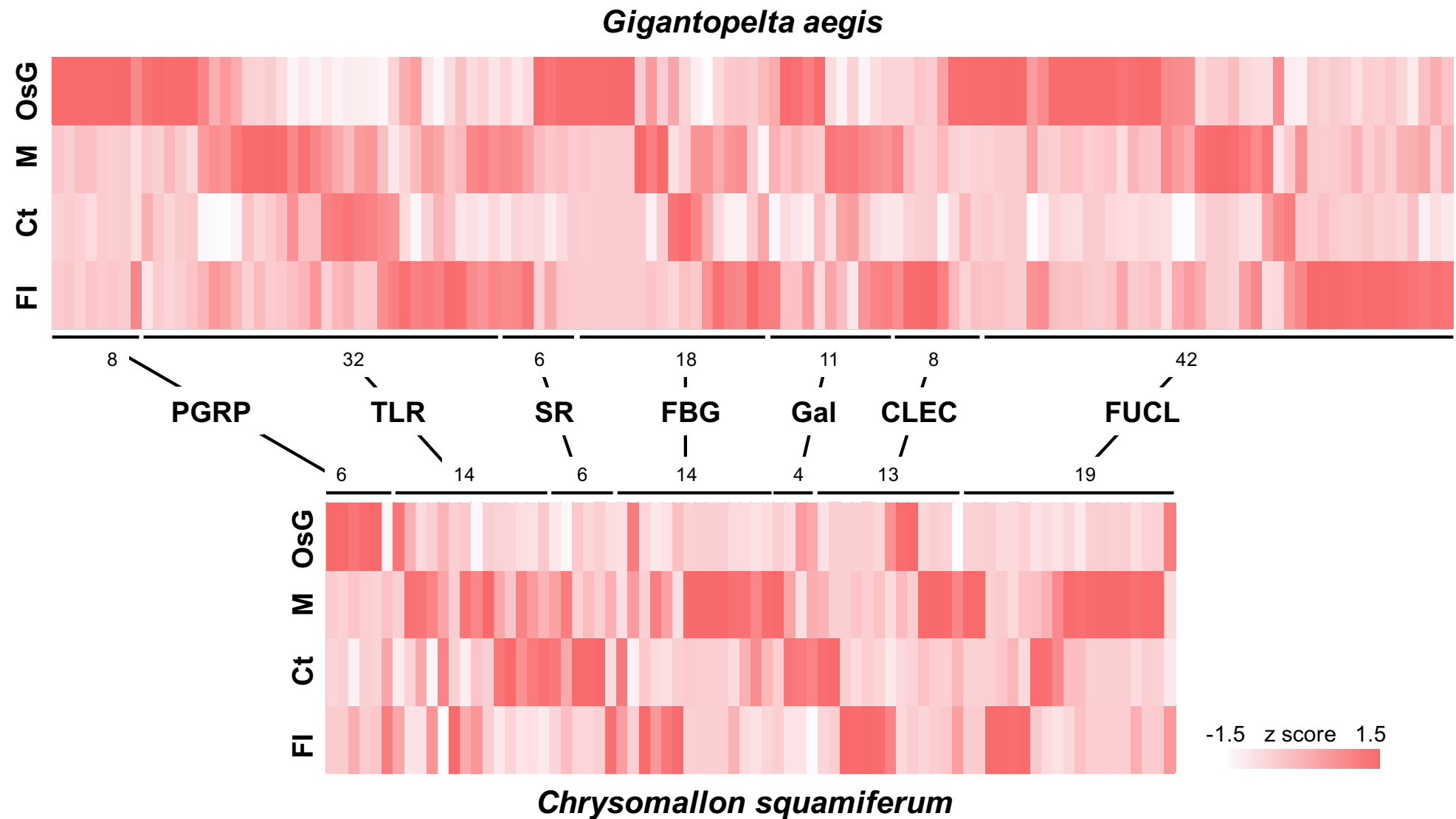
Figure 5 Two heat maps showing gene abundances and expression levels of pattern recognition receptors, including peptidoglycan recognition proteins (PGRP), toll-like receptors (TLR), scavenger receptor (SR), fibrinogen-related proteins (FBG), galectins (Gal), C-type lectins (CLEC), and fucolectin (FUCL), in *Gigantopelta aegis* ($n = 4$) and *Chrysomallon squamiferum* ($n = 3$)[17] (OsG: oesophageal gland; M: mantle; Ct: ctenidium; FI: internal tissue of foot). The numbers near the lines indicate the gene numbers of the gene families.
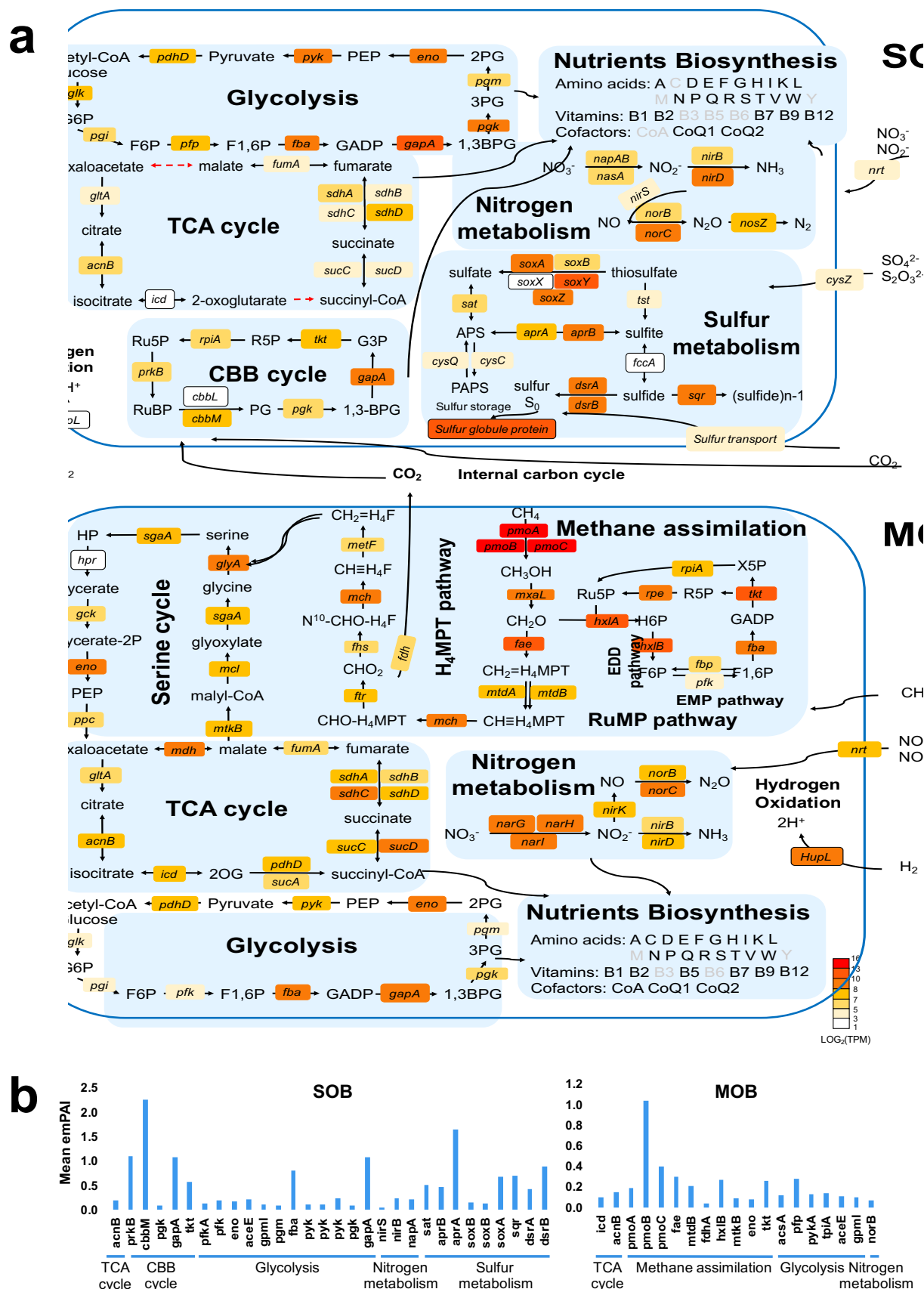
Figure 6 Central metabolism of the dual symbionts of *Gigantopelta aegis*. **a**. Central metabolism pathways in the sulphur-oxidising endosymbiont (SOB) and the methane-oxidising endosymbiont (MOB). The boxes represent genes involved in the respective process, and they are colour coded according to the Log-transformed normalized transcripts per kilobase million (TPM) value ($n = 3$) showing the gene expression level in the metatranscriptome analysis (colour from red to light white represents high expression level to low expression level; dash arrow represents that the gene is missing in the genome). **b**. Protein abundances involved in the central metabolism of the SOB (left) and the MOB (right). Protein abundances are reported as mean emPAI values ($n = 3$) assessed in the metaproteome analysis.