TITLE:

A universal subcuticular bacterial symbiont of a coral predator, the crown-of-thorns starfish, in the Indo-Pacific

RUNNING TITLE:

Subcuticular bacterium of a coral predator

AUTHORS:

Naohisa WADA[1,2†], Hideaki YUASA[3†], Rei KAJITANI[3], Yasuhiro GOTOH[4], Yoshitoshi OGURA[4], Dai YOSHIMURA[3], Atsushi TOYODA[5], Sen-Lin TANG[2], Yukio HIGASHIMURA[1], Hugh SWEATMAN[6], Zac FORSMAN[7], Omri BRONSTEIN[8], Gal EYAL[9,10], Nalinee THONGTHAM[11], Takehiko ITOH[3*], Tetsuya HAYASHI[4*], Nina YASUDA[1*]

AFFILIATIONS:

1: Faculty of Agriculture, University of Miyazaki, 1-1 Gakuenkibanadai-Nishi, Miyazaki, Miyazaki 889-2192, Japan

2: Biodiversity Research Center, Academia Sinica, No.128, Sec 2, Academia Rd, Nangang, Taipei 11529 Taiwan

3: School of Life Science and Technology, Department of Life Science and Technology, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan

4: Department of Bacteriology, Faculty of Medical Sciences, Kyushu University, 3-1-1 Maidashi Higashi-ku Fukuoka 812-8582, Japan

5: Center for Information Biology, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan

6: Australian Institute of Marine Science, PMB No.3, Townsville, QLD 4810, Australia

7: Hawai'i Institute of Marine Biology, School of Ocean & Earth Sciences & Technology, University of Hawai'i at Mānoa, Coconut Island, Kāne'ohe, HI, United States

8: School of Zoology, Faculty of Life Sciences & Steinhardt Museum of Natural History, Tel Aviv University, Tel Aviv 6997801, Israel

9: ARC Centre of Excellence for Coral Reef Studies, School of Biological Sciences, University of Queensland, St. Lucia, QLD 4072, Australia

10: The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel

11: Phuket Marine Biological Center, Box 60, Phuket 83000 Thailand

†CO-FIRST AUTHORS:

Naohisa Wada & Hideki Yuasa

***CORRESPONDING AUTHORS:**

Takehiko ITOH (takehiko@bio.titech.ac.jp)

Tetsuya HAYASHI (thayash@bact.med.kyushu-u.ac.jp)

Nina YASUDA (nina27@cc.miyazaki-u.ac.jp)

Abstract

Background:

Population outbreaks of the crown-of-thorns starfish (*Acanthaster planci* sensu lato; COTS), a primary predator of reef-building corals in the Indo-Pacific Ocean, are major concerns in coral reef management. While biological and ecological knowledge of COTS has been accumulating since the 1960s, little is known about its associated bacteria. The aim of this study was to provide fundamental information on dominant COTS-associated bacteria through a multifaceted molecular approach.

Methods:

A total of 205 COTS individuals from 17 locations throughout the Indo-Pacific Ocean were examined for the presence of COTS-associated bacteria. We conducted 16S rRNA metabarcoding of COTS to determine the bacterial profiles of different parts of the body, and generated a full-length 16S rRNA gene sequence from a single dominant bacterium, which we designated COTS27. We performed phylogenetic analysis to determine the taxonomy, screening of COTS27 across the Indo-Pacific, FISH to visualize it within the COTS tissues, and reconstruction of the chromosome from the hologenome sequence data.

Results:

We discovered that a single bacterium exists at high densities in the subcuticular space in COTS forming a biofilm-like structure between the cuticle and the epidermis. COTS27 belongs to a clade that presumably represents a distinct order (so-called marine spirochetes) in the phylum *Spirochaetes* and is universally present in COTS throughout the Indo-Pacific Ocean. The reconstructed genome of COTS27 includes some genetic traits that are probably linked to adaptation to marine environments and evolution as an extracellular endosymbiont in subcuticular spaces.

Conclusions:

COTS27 can be found in three allopatrically speciated COTS species, ranging from northern Red Sea to the Pacific, implying that symbiotic relationship arose before the speciation (approximately 2 million years ago). The universal association of COTS27 with COTS and nearly mono-specific association at least with the Indo-Pacific COTS potentially provides a useful model system for studying symbiont-host interactions in marine invertebrates.

Keywords:

Crown-of-thorns starfish, subcuticular bacteria, marine spirochetes

Introduction

Coral reefs support almost one-third of the world's marine coastal species [1,2]. However, population outbreaks of a coral predator, the crown-of-thorns starfish (*Acanthaster planci* sensu lato; COTS), are a great threat to Indo-Pacific coral reef ecosystem integrity and biodiversity [3–5]. A 27-year study of the Great Barrier Reef concluded that COTS outbreaks and tropical cyclones were the main causes of the loss of reef-building corals (1985-2012) [6]. While some aspects of the biology of COTS, such as its reproduction, larval ecology, phylogeography, and behaviour, have been studied intensively [5], little is known about its associated microbiota.

The bacterial symbionts of marine invertebrates have been shown to be important to their host organisms [7]. In echinoderms, bacterial communities may play a role in larval settlement [8], amino acid uptake on the integument [9], and digestive strategies in the gut [10,11], and these communities may even drive morphological variations in their host [12]. Bacterial symbionts are prevalent on the body surfaces of echinoderms [13], showing high host specificity [14,15]. Notably, extracellular endosymbionts known as subcuticular bacteria (SCB [16]) have been shown to reside under the cuticular layer of echinoderm fauna from all five extant classes, and it has been postulated that these bacteria provide dissolved free amino acids to their echinoderm hosts [9,17]. To date, molecular genetic approaches targeting the 16S rRNA gene have revealed that several proteobacteria (*Alphaproteobacteria* and *Gammaproteobacteria*) are SCB that are distributed in the subcuticular space in two brittle star species [13,18], one holothurian species [19], and one asteroid species (19).

Despite their potential biological importance, the studies of the bacteria associated with COTS have been mostly culture-based, and only two culture-independent studies have been published to date. Carrier et al. reported shifts in the COTS larval microbiomes associated with diet [20]. Høj et al. found that adult COTS exhibit tissue-specific bacterial communities, largely comprising four major bacterial groups: *Mollicutes* in male gonads, *Spirochaetales* in the body wall, *Hyphomonadaxeae* in the tube feet, and *Oceanospirillales* in all tissues [21]. Although these studies significantly increased our understanding of the COTS microbiome, there is still a great lack of knowledge regarding COTS-associated bacteria, particularly SCB, despite being common in many echinoderm taxa, where they may play an important role for their host organisms.

In the current study, we aimed to obtain primary information on the indigenous bacteria of the body surface of COTS. We carried out a comprehensive analysis of bacterial symbionts associated with COTS in a total of 205 individuals collected from the northern Red Sea to the Pacific over a 13-years period. We highlighted the existence of dominant SCB in COTS, its novel phylogenetic status, universal distribution in the Indo-Pacific COTS, and its genomic characteristics, all of which provide insights into interactions between the COTS host and the SCB.

Results

Identification of a single OTU (COTS27) that dominates the body surface microbiota of COTS using 16S rRNA metabarcoding analysis

We used 16S rRNA metabarcoding to analyse the bacterial composition of the microbiota in the body parts (7-8 body parts; disc spines [top and base], arm spines [top and base], ambulacral spines [top and base for Okinawa, or the whole spine for Miyazaki], tube feet, and pyloric stomachs; **Fig.1b**) of six COTS individuals that were collected in Miyazaki and Okinawa, Japan (three individuals from each location). Seawater samples from the same locations were similarly analysed for their bacterial compositions (three samples from each location). After quality filtering, 1,427,570 sequences of bacterial origins were obtained from the COTS samples (n=130 for all body parts in replicates or duplicates; **Suppl. table S1**) and 108,334 bacterial sequences from seawater samples (n=6) with an averages of 10,981 and 18,056 sequences per sample, respectively (**Suppl. table S2**). From the abovementioned sequences, 671 bacterial OTUs were identified, 503 and 401 of which were found in the COTS and seawater samples, respectively. There were 233 OTUs that were common to both. The OTUs that were identified in the COTS and seawater samples represented the bacterial taxa with 144 and 96 families, 29 and 22 OTUs were unclassified, 7 and 12 OTUs were as unknown (classified as bacteria by Silva SINA [22]), respectively (see more detail in **Suppl. table S3**). The rarefaction curves based on the OTUs indicated that all samples reached saturation points (**Suppl. fig. S1**).

In the six COTS individuals that were examined, the relative abundance showed that a single unclassified OTU (OTU 1) occupied 61.8% of the total sequences on average, predominantly in most body parts of both the Okinawan and Miyazaki COTS populations (60.3% and 63.8% of the total sequences on average were assigned to OTU 1 in the Okinawa and Miyazaki COTS collections, respectively; **Fig. 2**), despite the fact that these populations were separated by more than 720 km separating these populations. The high abundance of OTU 1 in all individuals was attributed to the surface body parts (68.8% and 79.1% of the sequences from all spine and tube foot samples, respectively), with 8.0% of these sequences originating from the pyloric stomach samples (**Fig. 2** and **Suppl. table S4**). OTU 1 was abundant at both the aboral (discs and arm spines) and oral (ambulacral spines and tube feet) sides (**Suppl. fig. S2** and **Suppl. table S4**) of the COTS. The tips and bases of the spines showed roughly the same levels of OTU 1 abundance (**Suppl. fig. S2** and **Suppl. table S4**). Five of the 88 spine samples that were examined (containing both tip and base) exhibited no or only a low abundance of OTU 1 (**Suppl. fig. S2**); however, OTU 1 was abundant in the other two DNA preparations of the triplicates from the same sample in all cases, suggesting that the exceptional data from the five preparations were due to some technical problems. OTU 1 was only detected in the Okinawan seawater samples, in which it showed a low abundance (0.026%;

143 **Suppl. fig. S2**). The relatively abundant bacteria other than OTU 1 are described in **Appendix 1**. In
144 total, we identified 41 different OTUs, including OTU 1, in all COTS individuals from the two
145 locations, and these OTUs may represent the core members of the bacterial community of COTS
146 (**Suppl. fig. S3**). The core bacterial OTUs other than OTU 1 accounted for up to 18.4% (the
147 abundance of each OTU was less than 3.5%) of the total reads from all COTS samples (**Suppl. fig.**
148 **S3d**). These results indicate that a single bacterium (OTU 1) predominantly colonizes the body
149 surface of COTS.
150

151 ***Phylogeny of the dominant OTU 1 (COTS27) based on 16S rRNA gene sequences***

152 To elucidate the phylogenetic status of the dominant OTU 1, we determined the full-length 16S
153 rRNA gene sequences of OTU 1 in five tube foot samples obtained from Miyazaki (n=3) and
154 Okinawa (n=2). The five sequences were largely identical (99.9-100% similarity), and there was a
155 partial sequence overlap with the 16S rRNA gene sequence of a spirochete-like bacterium (GenBank
156 accession No. PRJNA420398) that was a dominant bacterium on the body wall of COTS from the
157 Great Barrier Reef [21]. The maximum likelihood (ML) phylogenetic tree (**Fig. 3a**) based on full-
158 length 16S rRNA gene sequences showed that the five sequences related the OTU 1 formed a
159 distinct subclade within one of the three clades of the unclassified spirochete cluster (named clade I;
160 **Fig. 3a**). All sequences in this unclassified spirochete cluster originated from marine environments
161 and marine invertebrates (see **Appendix 2** for more details of clade I) with the exception of a single
162 sequence obtained from a wetland soil sample (GenBank accession No. FQ660021.1). Hereafter, we
163 refer to these spirochetes as "marine spirochetes", as referred to by Høj et al. (2018) [21]. These
164 marine spirochetes formed a distinct cluster within the phylum *Spirochaetes*, with the order
165 *Brachyspirales* being their closest relative (**Fig. 3a**). Notably, the 16S rRNA gene sequences of the
166 marine spirochetes, including the OTU 1 group, showed only a 76.3–78.1% identity to those of the
167 order *Brachyspirales*, which is well below the proposed threshold for defining a novel order (82.0%)
168 [23]. Thus, the marine spirochetes most likely represent a distinct order in the phylum *Spirochaetes*.
169 Hereafter, we refer to the bacterium corresponding to the OTU 1 as COTS27.
170

171 ***Universal association of COTS27 with COTS throughout the Indo-Pacific Ocean***

172 The presence of COTS27 or COTS27-like bacteria in COTS individuals inhabiting various
173 geographic regions was determined in a PCR assay designed to amplify a specific 261 bp fragment
174 of the COTS27 16S rRNA gene. PCR products were obtained from all 195 COTS individuals that
175 were collected at 15 locations throughout the Indo-Pacific Ocean comprising three known species of
176 COTS (**Figs. 1a** and **c**). The sequencing of the PCR products from 53 randomly selected individuals
177 confirmed the presence of COTS27 or very close relatives. The ML tree based on these 261 bp

178 sequences (**Suppl. fig. S4**) revealed that all sequences formed a tight cluster with the six COTS27
179 sequences from the abovementioned phylogenetic analysis and with those obtained from the genome
180 reconstruction described below. However, the sequences from the Israeli COTS population (Red Sea
181 species) formed a clade separate from those of the Indo-Pacific populations from the northern Indian
182 Ocean or Pacific Oceans. Among the northern Indo-Pacific species, only one single-nucleotide
183 polymorphism (SNP) was detected in one sequence obtained from Japan (Wakayama C29 adult JPN;
184 **Suppl. fig. S4**). These results indicate the universal association of COTS27 with the Indo-Pacific
185 COTS species.
186

187 ***Localization and biofilm-like structure formation of COTS27 in subcuticular spaces***
188 ***across the body surface of COTS***

189 We observed the localization of COTS27 in COTS tissues using fluorescence *in situ* hybridization
190 (FISH), as demonstrated by the binding of the general bacterial probe and a COTS27-specific probe
191 that we designed (the binding signals on the COTS central disc spines are shown in **Fig. 4**). COTS27
192 was consistently present in the subcuticular spaces on both the aboral side (**Figs. 5a-d**; spines of the
193 discs and arms, dermal papulae, and pedicellariae, see **Fig. 1b** and **Suppl. fig. S5** for their anatomical
194 locations and structures) and oral side (**Figs. 5e**; the stems of the tube feet) and the pattern was
195 similar for all three COTS individuals (**Figs. 4** and **5**). No COTS27 signal or any other bacterial
196 signals were detected in the pyloric caeca and gonads (**Fig. 5g-h**). Likewise, no COTS27 was found
197 in the pyloric stomachs, although numerous cyanobacteria-like bacteria were detected (**Fig. 5f**).
198 In the cross-sections, COTS27 displayed continuous layer-like signals (**Figs. 4b, 5a-c** and **5e**),
199 although a patchy distribution was also occasionally observed. Furthermore, three-dimensional (3D)
200 images showed that COTS27 formed a biofilm-like structure on the epidermis of the pedicellariae
201 (**Fig. 5d**). These observations indicate that COTS27 is an SCB that covers nearly all the surface area
202 (the epidermis) of COTS by forming a biofilm-like structure. COTS27 cells appear to have
203 filamentous or long rod-like shapes (**Figs. 5c** and **e**), but different approaches such as electron
204 microscopy are required to accurately determine their cell morphology.
205

206 ***Reconstruction of the COTS27 chromosome***

207 We have not yet succeeded in isolation of COTS27, but were able to reconstruct the chromosome
208 sequence of COTS27 from the hologenome sequences, which contained sequences derived from the
209 host genome and the associated microbes (**Suppl. table S5**), of a COTS sample collected in
210 Miyazaki (**Fig. 6** and **Suppl. table S6**), with 90.66% completeness and 0.26% contamination, as
211 evaluated by CheckM [24]. The structural accuracy of the chromosome was validated based on the
212 physical coverage of the 15 kbp-mate-pairs (**Suppl. Materials and Methods Fig. 1**), and the circular

213 structure was also confirmed using PCR and Sanger sequencing. We also tested other assembly
214 pipelines consisting of removal of reads from the host genome, metagenome assemblers, and binning
215 tools; however, none generated a higher-quality genome of COTS27 nor an obvious chromosome of
216 a different bacterium (**Suppl. Materials and Methods table 1**). Although 23 gaps remained in the
217 final assembly, all were derived from tandem repeats in the genic regions, and the estimated gap
218 sizes were less than 28 bp. The COTS27 chromosome was 2,684,921 bp in length, with a 39.6%
219 average GC-content, and contained 1,650 protein-coding genes, three rRNA genes, and 35 tRNA
220 genes. No transposable elements or prophages were detected. The 1,650 protein-coding genes
221 included one giant gene (53,043 bp in length; COTS27_01023), but its function is currently
222 unpredicted (see **Suppl. materials and methods** for the details). Among the three rRNA genes, the
223 16S rRNA gene was located separately from the 23S and 5S rRNA genes. The 35 tRNA genes
224 covered all 20 basic amino acids. Phylogenetic analysis using the sequences of 43 conserved marker
225 genes with 5,656 reference bacterial and archaeal genomes placed COTS27 in the phylum
226 *Spirochaetes* (**Fig. 3b**), supporting the results of the 16S rRNA sequence-based analysis (**Fig. 3a**).
227

228 ***Biological features of COTS27 inferred from the gene repertoire***

229 In the Clusters of Orthologous Groups (COG) functional category-based principal component
230 analysis of COTS27 performed using 716 high-quality *Spirochaetes* genomes obtained from the
231 IMG database [25], COTS27 was placed in a distinct position with regard to all *Spirochaetes* (**Suppl.**
232 **fig. S6**). This indicates potential biological features unique to COTS27. Subsequently, we performed
233 a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis to obtain basic information
234 on the biology of COTS27 (**Suppl. table S7**). Complete or near-complete biosynthesis pathways for
235 18 of the 20 basic amino acids were identified, excluding those for asparagine and aspartic acid.
236 Although the guanine ribonucleotide biosynthesis pathway was not complete (one block missing), all
237 other nucleotide biosynthesis pathways were detected. For vitamin and cofactor biosynthesis, the
238 complete biosynthetic pathways of nicotinamide adenine dinucleotide (NAD), coenzyme A, and
239 riboflavin and the C1-unit interconversion were identified. Pathways for fatty acid biosynthesis,
240 beta-oxidation, and phosphatidylethanolamine biosynthesis were also detected. The conservation of
241 these metabolic pathways suggests that COTS27 is not strongly metabolically dependent on the host
242 COTS.
243 Regarding energy production, COTS27 exhibited the complete glycolysis pathway and TCA cycle.
244 Genes for succinate dehydrogenase, cytochrome c oxidase, and F-type ATPase were also identified;
245 however, no genes for NADH dehydrogenase were detected. Instead, COTS27 presented an operon
246 encoding a sodium-pumping NADH:ubiquinone oxidoreductase (Na^+-NQR) (**Suppl. fig. S7**).

247 Consistent with the general characteristics of *Spirochaetes*, which are generally Gram-
248 negative, helical or spiral-shaped, and motile, with periplasmic flagella [26], COTS27 contained sets
249 of genes for the biosynthesis of DAP-type peptidoglycan, lipopolysaccharide and
250 phosphatidylethanolamine. While a set of genes for flagellar biosynthesis was identified, no gene for
251 chemotaxis, such as genes encoding methyl-accepting chemotaxis proteins and chemotaxis-related
252 signal transduction conponents, was detected.
253

254 **Discussion**

255 We identified a single bacterium that forms a subcuticular biofilm-like structure in Indo-Pacific
256 COTS although we can not completely exclude a possibility that the signals detected in FISH
257 analyses included some false-positive signals (because of only *in silico* verification). This bacterium
258 is universally present and numerically dominant and likely represents a previously undefined order
259 within the phylum *Spirochaetes*. The universal association of COTS27 with the Indo-Pacific COTS
260 species suggests a long history of the COTS-COTS27 association. COTS are thought to have
261 allopatrically diverged into four species during the Pliocene-Early Pleistocene (1.95–3.65 Myr ago)
262 in the Indo-Pacific Ocean [27]. Therefore, the association of COTS27 with at least three of the four
263 extant COTS species (data for the fourth species are currently not available) implies that the mutual
264 relationship between COTS and COTS27 emerged prior to the Pliocene or Early Pleistocene eras.
265 This hypothesis is supported by the finding that COTS27 from the northern Red Sea (forming a
266 different cluster from the Indo-Pacific regions; **Suppl. fig. S4**) was notably different from COTS27
267 from other regions. Additional comparative genomic analyses of the Indo-Pacific COTS and
268 COTS27 from different geographic regions would provide more detailed insights into the possible
269 co-evolutionary history. In addition, further studies linking environmental conditions with COTS27
270 abundance and microbial composition, will help to understand the ecological roles of COTS27.
271 Regarding the evolutionary and functional aspects of COTS27 and its association with COTS,
272 we obtained two key findings from the genome sequence analysis: 1) the presence of Na^+-NQR and
273 2) the selective loss of chemotaxis genes. 1) Some bacteria living in Na^+-rich environments (e.g.,
274 marine or intercellular environments) exhibit an Na^+-NQR that oxidizes NADH to NAD^+ and pumps
275 Na^+ out of cells, thus functioning in the respiratory chain and in the maintenance of intercellular
276 homeostasis in Na^+-rich environments [28]. In line with these features of Na^+-NQR, only one
277 genome out of the 716 high-quality *Spirochaetes* genomes (above), which was also reconstituted
278 from the metagenome sequences of a seawater sample (*Spirochaetaceae* bacterium NP120, IMG
279 Genome ID: 2509276057), contained the Na^+-NQR operon. The acquisition of Na^+-NQR may
280 represent one of the mechanisms responsible for the adaptation of COT27 to marine environments.
281 2) Besides, the selective loss of chemotaxis genes is very unusual in *Spirochaetes*; most of the high-

282 quality *Spirochaetes* genomes mentioned above (>98%) contained gene sets for both flagellar
283 biosynthesis and chemotaxis. The remaining genomes, such as those from the genus *Sphaerochaeta*,
284 lack genes for both flagellar biosynthesis and chemotaxis, suggesting that chemotaxis genes have
285 been selectively lost from the COTS27 genome. It has been proposed that the active migration and
286 colonization by symbionts through motility and chemotaxis are often required for the acquisition of
287 microbial partners by host organisms from environments [29]. However, the selected lost chemotaxis
288 genes appear to represent a specific adaptation strategy of COT27 as an SCB. COTS27 may require
289 flagella to spread and stably and widely colonize subcuticular spaces, but chemotaxis is no longer
290 required after specially adapting to the subcuticular spaces of COTS. These findings are informing
291 because these features are likely linked to the adaptation to marine environments and evolution as an
292 extracellular endosymbiont in the subcuticular space, respectively. Additional genome sequences of
293 marine spirochetes are required to verify this hypothesis and elucidate the evolutionary and
294 functional aspects of the COTS27-COTS association.
295 Our study implied that COTS27 as SCB forms a nearly mono-specific symbiotic relationship
296 with COTS, at least with the Indo-Pacific COTS. SCB, however, are commonly found in echinoderm
297 fauna [13,19] and have been classified into three morphotypes [13,15,16]. Among these
298 morphotypes, COTS27 most likely belongs to the SCB Type 2, which exhibits a long spiral shape
299 and is commonly found among all five echinoderm classes [15,30]. Jackson et al. suggested the
300 presence of a highly dominant *Spirochaetae* in the hard tissues (including the body wall) of some
301 starfish species in the United States and Australia [31]. Such a wide distribution of spirochetes or
302 spirochete-like bacteria in echinoderms suggests that many echinoderms may have established
303 symbiotic relationships with marine spirochetes that are similar to that between COTS27 and COTS.
304 Further explorations of SCB in a wider range of echinoderms would provide more detailed insights
305 into the association between echinoderm hosts and marine spirochetes.
306 The outer body surfaces of marine organisms often represent a highly active interface between
307 an organism (host) and the surrounding marine environment regarding aspects such as light
308 exposure, gas exchange, nutrient uptake and interactions with other fouling organisms, consumers,
309 and pathogens [32]. The presence of SCB among different echinoderms has been reviewed in
310 different bacterial taxa [13,18,19]. Although it is also plausible that SCB play hypothetical role in
311 their interactions with the host such as nutrition transfer [9,33], or antibiotics production [14,34], or
312 even that their presence may be vestigial, remaining as leftovers from a previously mutualistic
313 partnership [15], the physiological and potential ecological roles of SCB are largely unclear and
314 remain unexplored. The universal and nearly mono-specific association of COTS27 with COTS
315 would provide an ideal model system for further exploring the roles of SCB as well as symbiont-host

316 interactions in marine invertebrates. Moreover, COTS27 could be used as an environmental marker
317 to monitor and/or predict population outbreaks of COTS.
318

319 Conclusions

320 Despite the fact that the 205 COTS individuals utilized in our current analyses were collected over a
321 13-year period (2004–2017) and from 17 different locations across the Indo-Pacific, the COTS27
322 association remained exceptionally ubiquitous both spatially and temporally. Additionally, it is likely
323 that COTS hosted COTS27 as an extracellular endosymbiont for more than 2 million years before
324 allopatric speciation occurred during the Pliocene-Early Pleistocene suggesting a strong association.
325 COTS27 is likely an extracellular endosymbiotic bacteria strongly associated with COTS as an SCB.
326 COTS27 also would acquire the Na^+-NQR system for adapting to marine environment since the
327 speciation within phylum *Spirochaetes.* The lack of chemotaxis genes in COTS27 would
328 physiologically associate with optimization in subcuticular space of COTS. Although the functional
329 role of COTS27 as an SCB is still unclear, this close relationship and chromosome genome
330 information of COTS27 described here will significantly contribute to testing the hypotheses of
331 symbiotic function in SCB, and may also provide as a model system for studying endosymbionts in
332 marine invertebrates more broadly.
333

334 Materials and Methods

335 *Sample collection and preparation for DNA analyses and histology*

336 We collected 205 individual COTS from 17 locations throughout the Indo-Pacific collected over a 13
337 year period (2004-2017; **Fig. 1a**, and **Suppl. table S8**). For 16S rRNA metabarcoding, six
338 individuals were collected in Okinawa and Miyazaki (three from each location) in Japan (**Fig.1a**).
339 The specimens were dissected to facilitate sampling from different body parts (7-8 body parts; **Fig.**
340 **1b,c**) which was done in triplicate or duplicate (**Suppl. table S1**). Seawater samples were also
341 collected in triplicate from each of the last two locations. Consequently, 130 DNA samples were
342 prepared from the six COTS individuals (**Suppl. table S1**) and six seawater DNA samples and used
343 for the metabarcoding analysis. The tube foot DNA samples from five of the six individuals were
344 used to determine the full-length 16S rRNA gene sequence of the dominant OTU 1. DNA samples
345 prepared from the tube feet of 195 individuals collected from 15 geographic locations were used to
346 examine the presence of COTS27 (the dominant OTU 1) in three species of COTS [35] (**Fig. 1a, c,**
347 and **Suppl. table S8**). DNA from the tube feet of one individual collected in Miyazaki (Japan) was
348 used for the COTS hologenome sequencing (**Suppl. table S8**). [36]. Note that the experiment was
349 designed to capture both the host and an associated microbiome genomes as a hologenome. Samples

350 of six different body parts (**Fig. 1b, c**) were prepared from the remaining other three individuals
351 collected in Miyazaki for the FISH analyses.
352 We provide a more detailed description of the above collection and preparation methods in the
353 **Suppl. materials and methods**.
354

355 ### *16S rRNA metabarcoding*

356 16S rRNA amplicon libraries (V4 region) were prepared as previously described [37,38] using the
357 primers listed in **Suppl. table S9** and **Suppl. fig. S8** and subjected to paired-end (PE) sequencing (2
358 x 300 bp) using the Illumina MiSeq platform. In total, 130 DNA samples including the samples
359 collected from 7-8 body parts of six COTS individuals (**Suppl. table S1**), and six seawater DNA
360 samples were analysed. The obtained PE sequences were processed using software USEARCH
361 v8.1.1861 [39] and MOTHUR v.1.36.1 [40] software for the merging, the filtering, the OTU
362 clustering, and the taxonomic assignment of the sequence (see more detail in **Suppl. materials and**
363 **methods)**. Finally, a total of 1,535,904 sequences were assigned to bacteria. The others were
364 assigned to eukaryotes (55,377 reads containing 74.7% of COTS genes), Archaea (12,686 reads),
365 chloroplasts (19,715 reads), or unknown origins (476,795 reads containing 99.4% of COTS genes),
366 and were excluded from our study (**Suppl. table S2**).
367

368 ### *Phylogenetic analysis of OTU 1 using the full-length 16S rRNA gene sequence*

369 Full-length 16S rRNA gene sequences of OTU 1 were obtained from each tube foot of five COTS
370 individuals using a specific primer set for OTU 1 that was designed in this study (**Suppl. materials**
371 **and methods**, **Suppl. table S9** and **Suppl. fig. S8**). The sequences were used to reconstruct the
372 phylogenetic tree using the maximum likelihood (ML) method (**Suppl. materials and methods**). As
373 OTU 1 was revealed to represent a unique clade of bacteria present in COTS, we hereafter refer to
374 this bacterium as COTS27.
375

376 ### *PCR screening and sequencing of COTS27*

377 In total 195 COTS individuals were screened for the presence of COTS27 on their tube feet by PCR
378 using primers that were designed to specifically amplify a 261 bp fragment of the 16S rRNA gene
379 (**Suppl. table S9** and **Suppl. fig. S8**). The PCR products obtained from 53 randomly selected
380 samples from all COTS27-positive samples (n=195) were sequenced and used for phylogenetic
381 reconstruction (**Suppl. materials and methods**).
382

383 ### *Fluorescence in situ hybridization (FISH)*

384 The FISH experiments were performed on three serial sections (thickness of 5 μm) from the six body
385 parts of the three individuals (**Fig. 1b,c**) as previously described [41]. FISH was performed
386 separately with three different probes: COTS27-specific oligonucleotide probe (COTSsymb; for
387 more detail of the probe design, see in **Suppl. materials and methods, Suppl. table S9** and **Suppl.**
388 **fig. S8**), a Eubacterial probe (EUB338mix [42]), and a nonsense probe (Non338 [43]). Bacterial
389 localization was observed using a confocal laser scanning microscope (LSM 550; Zeiss, Germany)
390 (see more detail in **Suppl. materials and methods).** In addition, we reconstructed three-dimensional
391 (3D) structures from thick sections (thickness: 50 μm) of the disc spines using a confocal laser
392 scanning microscope (LSM770; Zeiss, Germany),
393

394 ***Reconstitution of the COTS27 chromosome from the hologenome sequences of a***
395 ***COTS sample***

396 Two PE libraries and six mate-pair libraries from the tube foot of one individual were prepared and
397 sequenced using Illumina HiSeq 2500 sequencers. *De novo* assembly was performed using Platanus
398 v. 1.2.3 [44]. To identify the COTS27-derived sequences in the hologenome assembly, scaffolds
399 with a high coverage depth (≥×200), which probably reflected the high abundance of the species,
400 were selected. The average coverage depth for all scaffolds was ×130, and it was assumed that most
401 of the scaffolds from the host COTS genome exhibted coverage depths <×200. The longest scaffold,
402 which was identified as the COTS27 chromosome, was closed by Sanger sequencing, and an
403 alternative assembly was obtained using Platanus-allee v. 2.0.0 [45]. A circular view of the COTS27
404 chromosome was generated using the CGView Server [46] with manual processing. The
405 completeness of the final assembly was evaluated using CheckM v. 1.0.11 [24], and the structural
406 accuracy of the assembly was validated based on the physical coverage of the 15 kbp-mate-pairs
407 (See Suppl. materials and methods for the details). We also tested other assembly pipelines
408 consisting of removal of reads from the host genome, metagenome assemblers, and binning tools
409 (See Suppl. materials and methods for details).
410

411 ***Gene prediction and functional annotation***

412 Protein-coding sequences (CDSs) were predicted by using PROKKA v. 1.12 [47], followed by
413 manual curation. For functional annotation, Clusters of Orthologous Groups (COG) were assigned
414 by querying the CDSs against the Conserved Domain Database (CDD) with COG position-specific
415 scoring matrices (PSSMs) using RPS-BLAST. Additionally, K numbers of Kyoto Encyclopedia of
416 Genes and Genomes (KEGG) were assigned to each CDS; BlastKOALA [48] and KofamKOALA
417 [49] were used to perform searches in the KEGG GENES and KOfam databases, respectively.
418

Principal component analysis and phylogenetic analysis based on the genome sequences

Principal component analysis was performed based on the compositions of the COG functional categories. The genome sequences of the *Spirochaetes* bacteria were retrieved from the DOE-JGI IMG database, and 716 high-quality genomes (completeness >90% and contamination <5% as evaluated by CheckM v. 1.0.11 [24]) were retained (see **Suppl. materials and methods** for the details). Whole-genome sequence-based phylogenetic analysis was performed using CheckM to obtain the ML tree of COTS27 and 5,656 bacterial and archaeal genomes based on the sequences of 43 conserved marker genes. The tree was visualized using FigTree v. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

List of abbreviations

CDD: Conserved Domain Database
CDSs: Protein-coding sequences
COG: Clusters of Orthologous Groups
COTS: Crown-of-thorns starfish
FISH: Fluorescence *in situ* hybridization
KEGG: Kyoto Encyclopedia of Genes and Genomes
ML: Maximum likelihood
NAD: Nicotinamide adenine dinucleotide
Na+-NQ: Sodium-pumping NADH:ubiquinone oxidoreductase
OTU: Operational taxonomic unit
SCB: Subcuticular bacteria

Declarations

• ***Ethics approval and consent to participate***

Not Applicable

• ***Consent for publication***

Not applicable

• ***Availability of data and material***

All sequences produced for this study have been deposited in the DDBJ under BioProject accession number PRJDB4009 for the 16S metabarcoding and COTS27 chromosome data and accession

numbers LC490103 - LC490107 and LC495323 – LC495375 for the 16S rRNA gene sequence-based phylogeny.

• ***Competing interests***

The authors declare that they have no competing interests.

• ***Funding***

This study was supported by JSPS KAKENHI 221S0002, 16H06279 (PAGS), Grant-in-aid for Young Scientists (A) (17H04996) and (B) (25870563), the Program to Disseminate Tenure Tracking System in University of Miyazaki, the Environment Research and Technology Development Fund (4RF-1501), MOE, Japan, the postdoctoral fellows programme of Academia Sinica (2nd session, 2017), Grant-in-Aid for JSPS Fellows (18J23317), and the European Union's Horizon 2020 research and innovation programme under MSCA No. 796025.

• ***Authors' contribution***

N.W., H.Y., T.I., T. H, and N.Y. conceived the research idea and designed the experiments. N.W., H.Y., Y. H, H. S, Z.F., O. B, G.E., N.T., and N.Y. conducted the field sampling. N.W., Y.G., and N.Y. performed the 16S rRNA metabarcoding. N.W. and N.Y. performed the phylogenetic analysis using the full-length 16S rRNA gene. N.W. H. S, Z.F., OB, N.T., and N.Y. performed the PCR screening and sequencing of the PCR products. N.W. conducted the FISH experiments. H.Y., R.K., D.Y., and A.T. conducted the sequencing and analysis of the COTS27 genome. N.W., H.Y., T.I., T. H, and N.Y. made major contributions to the manuscript writing and figure making. R.K., Y.G., Y.O., S.T., H.S., Z.F., O.B., and G.E. contributed to writing and editing the manuscript. All authors critically reviewed, revised and ultimately approved this final version.

• ***Acknowledgements***

We thank members of Yasuda Laboratory and Atsushi Toyoda Laboratory, J. Suzuki, H. Kagawa, S. Nagai, Y. R. Kubota, Y. Uchida, K. Sakai, T. Miki, C. Shinzato, L. Høj, and D. Bourne for their logistical and technical support. We also thank M. Timmers, O. Ben-Zvi, N. Phongsuwan, S. Ayalon, the Seaver Institute, and the Interuniversity Institute for Marine Sciences of Eilat for local sampling and analysis.

References

1. Spalding MD, Ravilious C, Green EP. World atlas of coral reefs. Prepared at the UNEP World Conservation Monitoring Centre. Univ Calif Berkeley EEUULinks. 2001;

489 2. Wilkinson C. Status of coral reefs of the world: 2004. Australian Institute of Marine Science;
490 2004.

491 3. Moran P, Bradbury R. The crown-of-thorns starfish controversy. Search. 1989;20:3–6.

492 4. Birkeland C, Lucas J. *Acanthaster planci*: Major Management Problem of Coral Reefs. CRC
493 Press; 1990.

494 5. Pratchett MS, Caballes CF, Sweatman JAR-P& HPA. Limits to Understanding and Managing
495 Outbreaks of Crown- of- Thorns Starfish (*Acanthaster* spp.). Oceanogr Mar Biol. 2014;52:133–200.

496 6. De'ath G, Fabricius KE, Sweatman H, Puotinen M. The 27–year decline of coral cover on the
497 Great Barrier Reef and its causes. Proc Natl Acad Sci. 2012;109:17995–9.

498 7. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al.
499 Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci.
500 2013;110:3229–36.

501 8. Bryan PJ, Rittschof D, McClintock JB. Bioactivity of echinoderm ethanolic body-wall extracts: an
502 assessment of marine bacterial attachment and macroinvertebrate larval settlement. J Exp Mar Biol
503 Ecol. 1996;196:79–96.

504 9. Lesser MP, Walker CW. Comparative study of the uptake of dissolved amino acid in sympatric
505 brittle stars with and without endosymbiotic bacteria. Comp Biochem Physiol Part B Comp
506 Biochem. 1992;101:217–23.

507 10. Thorsen MS. Abundance and biomass of the gut-living microorganisms (bacteria, protozoa and
508 fungi) in the irregular sea urchin *Echinocardium cordatum* (Spatangoida: Echinodermata). Mar Biol.
509 1999;133:353–60.

510 11. Thorsen MS, Wieland A, Ploug H, Kragelund C, Nielsen PH. Distribution, identity and activity
511 of symbiotic bacteria in anoxic aggregates from the hindgut of the sea urchin *echinocardium*
512 *cordatum*. Ophelia. 2003;57:1–12.

513 12. Balakirev ES, Pavlyuchkov VA, Ayala FJ. DNA variation and symbiotic associations in
514 phenotypically diverse sea urchin *Strongylocentrotus intermedius*. Proc Natl Acad Sci.
515 2008;105:16218–23.

516 13. Burnett WJ, McKenzie JD. Subcuticular bacteria from the brittle star *Ophiactis balli*
517 (Echinodermata: Ophiuroidea) represent a new lineage of extracellular marine symbionts in the alpha
518 subdivision of the class *Proteobacteria*. Appl Environ Microbiol. 1997;63:1721–4.

519 14. McKenzie JD, Kelly MS. Comparative study of sub-cuticular bacteria in brittlestars
520 (Echinodermata: Ophiuroidea). Mar Biol. 1994;120:65–80.

521 15. Kelly MS, McKenzie JD. Survey of the occurrence and morphology of sub-cuticular bacteria in
522 shelf echinoderms from the north-east Atlantic Ocean. Mar Biol. 1995;123:741–56.

523 16. Holland ND, Nealson KH. The Fine Structure of the Echinoderm Cuticle and the Subcuticular
524 Bacteria of Echinoderms. Acta Zool. 1978;59:169–85.

525 17. McKenzie JD, Black KD, Kelly MS, Newton LC, Handley LL, Scrimgeour CM, et al.
526 Comparisons of fatty acid and stable isotope ratios in symbiotic and non-symbiotic brittlestars from
527 Oban Bay, Scotland. J Mar Biol Assoc U K. 2000;80:311–20.

528 18. Morrow KM, Tedford AR, Pankey MS, Lesser MP. A member of the Roseobacter clade,
529 *Octadecabacter* sp., is the dominant symbiont in the brittle star *Amphipholis squamata*. FEMS
530 Microbiol Ecol. 2018;94:fly030.

531 19. Lawrence SA, O'Toole R, Taylor MW, Davy SK. Subcuticular Bacteria Associated With Two
532 Common New Zealand Echinoderms: Characterization Using 16S rRNA Sequence Analysis and
533 Fluorescence *in situ* Hybridization. Biol Bull. 2010;218:95–104.

534 20. Carrier TJ, Wolfe K, Lopez K, Gall M, Janies DA, Byrne M, et al. Diet-induced shifts in the
535 crown-of-thorns (*Acanthaster* sp.) larval microbiome. Mar Biol. 2018;165:157.

536 21. Høj L, Levy N, Baillie BK, Clode PL, Strohmaier RC, Siboni N, et al. Crown-of-Thorns Sea Star
537 *Acanthaster* cf. *solaris* Has Tissue-Characteristic Microbiomes with Potential Roles in Health and
538 Reproduction. McBain AJ, editor. Appl Environ Microbiol. 2018;84:e00181-18.

539 22. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence
540 alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.

541 23. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the
542 classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat
543 Rev Microbiol. 2014;12:635–45.

544 24. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality
545 of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res.
546 2015;25:1043–55.

547 25. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0: an
548 integrated data management and comparative analysis system for microbial genomes and
549 microbiomes. Nucleic Acids Res. 2019;47:D666–77.

550 26. Paster BJ. Phylum XV. Spirochaetes Garrity and Holt 2001. In: Krieg NR, Staley JT, Brown DR,
551 Hedlund BP, Paster BJ, Ward NL, et al., editors. Bergey's Manual® Syst Bacteriol Vol Four
552 Bacteroidetes Spirochaetes Tenericutes Mollicutes Acidobacteria Fibrobacteres Fusobacteria
553 Dictyoglomi Gemmatimonadetes Lentisphaerae Verrucomicrobia Chlamydiae Planctomycetes. New
554 York, NY: Springer New York; 2010. p. 471–566.

555 27. Vogler C, Benzie J, Lessios H, Barber P, Wörheide G. A threat to coral reefs multiplied? Four
556 species of crown-of-thorns starfish. Biol Lett. 2008;4:696–9.

557 28. Reyes-Prieto A, Barquera B, Juárez O. Origin and Evolution of the Sodium -Pumping NADH:
558 Ubiquinone Oxidoreductase. PLOS ONE. 2014;9:e96696.

559 29. Raina J-B, Fernandez V, Lambert B, Stocker R, Seymour JR. The role of microbial motility and
560 chemotaxis in symbiosis. Nat Rev Microbiol. 2019;17:284.

561 30. Kelly MS, Barker MF, McKenzie JD, Powell J. The Incidence and Morphology of Subcuticular
562 Bacteria in the Echinoderm Fauna of New Zealand. Biol Bull. 1995;189:91–105.

563 31. Jackson EW, Pepe-Ranney C, Debenport SJ, Buckley DH, Hewson I. The Microbial Landscape
564 of Sea Stars and the Anatomical and Interspecies Variability of Their Microbiome. Front Microbiol.
565 2018;9:1829.

566 32. Wahl M, Goecke F, Labes A, Dobretsov S, Weinberger F. The Second Skin: Ecological Role of
567 Epibiotic Biofilms on Marine Organisms. Front Microbiol. 2012;3.

568 33. Walker CW, Lesser MP. Nutrition and development of brooded embryos in the brittlestar
569 Amphipholis squamata: do endosymbiotic bacteria play a role? Mar Biol. 1989;103:519–30.

570 34. Strahl ED, Dobson WE, Lundie Jr LL. Isolation and Screening of Brittlestar-Associated Bacteria
571 for Antibacterial Activity. Curr Microbiol. 2002;44:450–9.

572 35. Haszprunar G, Spies M. An integrative approach to the taxonomy of the crown-of-thorns starfish
573 species group (*Asteroidea: Acanthaster*): A review of names and comparison to recent molecular
574 data. Zootaxa. 2014;3841:271–84.

575 36. Yasuda N, Taquet C, Nagai S, Yoshida T, Adjeroud M. Genetic connectivity of the coral-eating
576 sea star *Acanthaster planci* during the severe outbreak of 2006–2009 in the Society Islands, French
577 Polynesia. Mar Ecol. 2015;36:668–78.

578 37. Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene
579 primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol. 2015;75:129–37.

580 38. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved
581 Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene
582 Primers for Microbial Community Surveys. mSystems. 2016;1:e00009-15.

583 39. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics.
584 2010;26:2460–1.

585 40. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing
586 mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and
587 Comparing Microbial Communities. Appl Environ Microbiol. 2009;75:7537–41.

588 41. Wada N, Pollock FJ, Willis BL, Ainsworth T, Mano N, Bourne DG. *In situ* visualization of
589 bacterial populations in coral tissues: pitfalls and solutions. PeerJ. 2016;4:e2424.

590 42. Daims H, Brühl A, Amann R, Schleifer KH, Wagner M. The domain-specific probe EUB338 is
591 insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive
592 probe set. Syst Appl Microbiol. 1999;22:434–44.

593 43. Wallner G, Amann R, Beisker W. Optimizing fluorescent in situ hybridization with rRNA-
594 targeted oligonucleotide probes for flow cytometric identification of microorganisms. Cytometry.
595 1993;14:136–43.

596 44. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo
597 assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res.
598 2014;24:1384–95.

599 45. Kajitani R, Yoshimura D, Okuno M, Minakuchi Y, Kagoshima H, Fujiyama A, et al. Platanus-
600 allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous
601 regions. Nat Commun. 2019;10:1702.

602 46. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView.
603 Bioinformatics. 2005;21:537–9.

604 47. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.

605 48. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for
606 Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428:726–31.

607 49. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA:
608 KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics.
609 2019;btz859.

610

611 **Figure Legends**

612

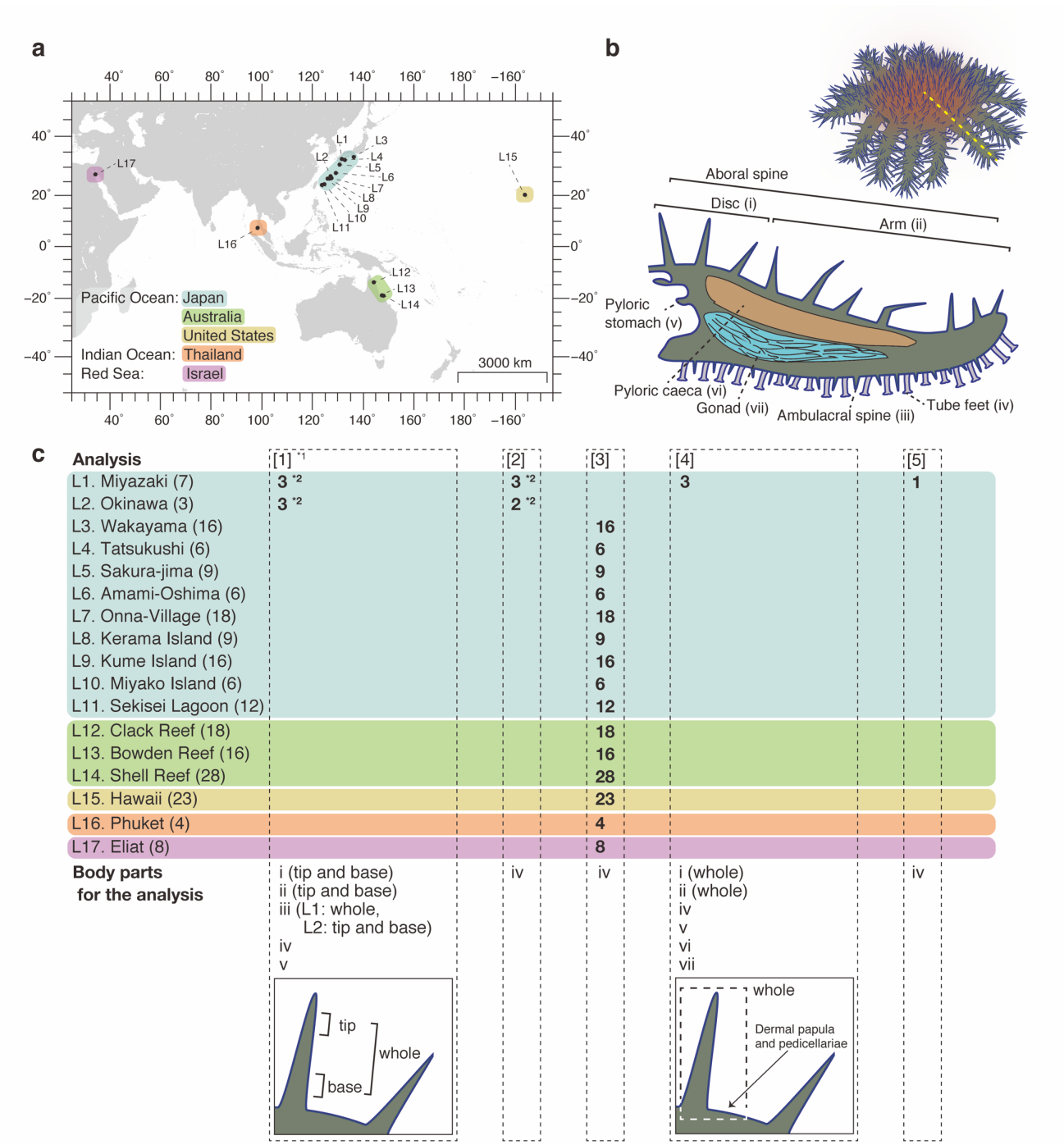

Fig. 1 Geographic and anatomical distributions of COTS individuals and the COTS body parts analysed in this study.

The seventeen locations where the COTS individuals were collected (**a**) and the dissected body parts of COTS for the analyses (**b**) are shown. The dashed yellow line (panel **b**) indicates the dissection line for the cross-sectional view. In panel (**c**), details of the samples used in each analysis are shown: [1] 16S rRNA metabarcoding, [2] phylogenetic analysis using the full-length 16S rRNA gene sequences, [3] PCR screening and sequencing of the 16S rRNA gene sequences of COTS27, [4]

FISH analysis, and [5] hologenome sequencing analysis. *1: This analysis was performed in triplicate for each sample. *2: The same individuals were used in analyses [1] and [2].

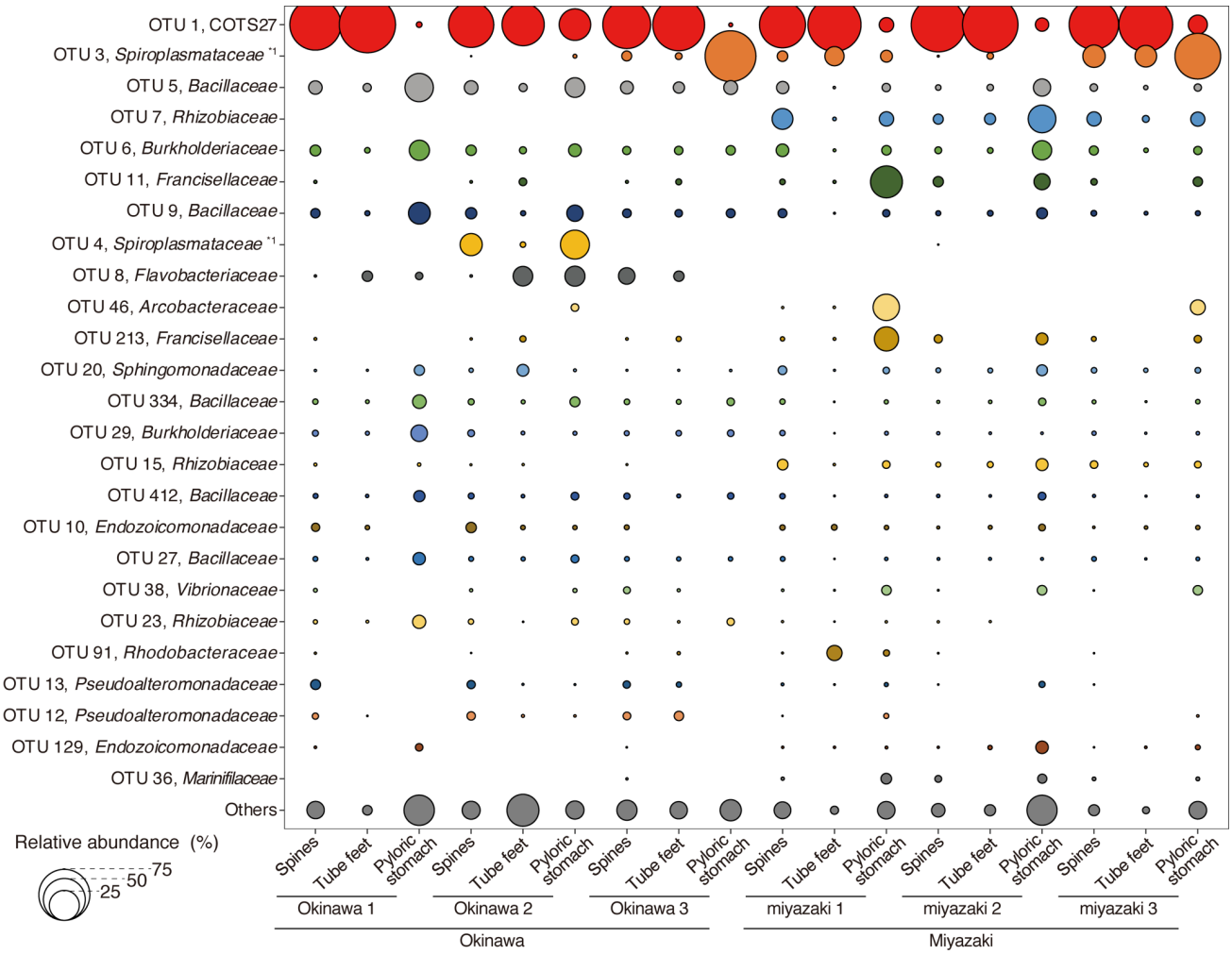

Fig. 2 The relative abundances of the 25 most abundant OTUs, including COTS27 (OTU 1; red), in the total samples analysed in this study.

The bubble chart of the relative abundances was calculated from the merged replicates of each body part (spines, tube feet, and pyloric stomachs) in each COTS individual. The phylogenies of each OTU were determined based on the results (best hit) of BLAST searches against the NCBI nr/nt database. *1: The phylogenies of OTU3 and OTU4 were determined in the All-species Living Tree Project and RDP databases, respectively.

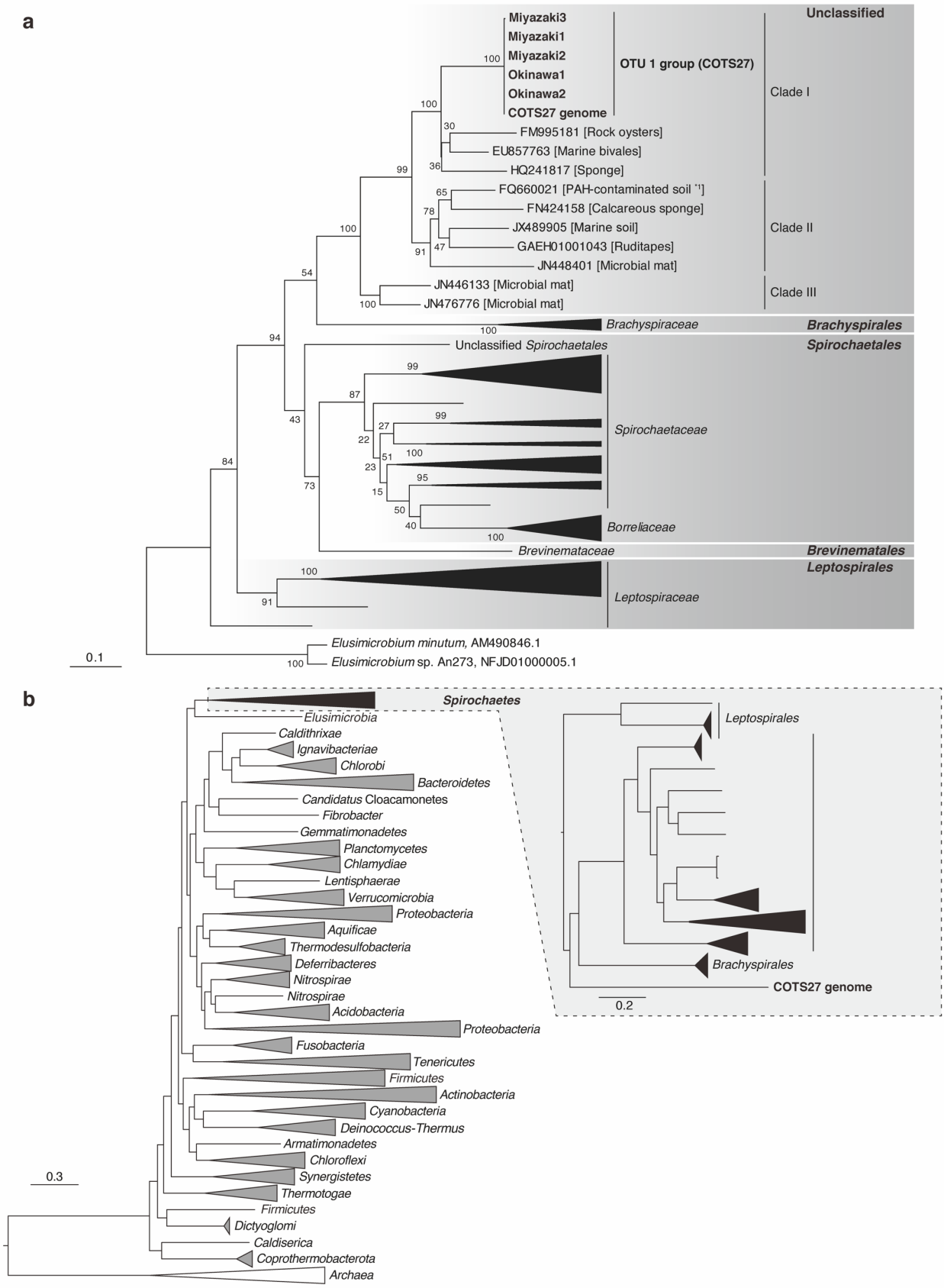

Fig. 3 The phylogenetic position of the dominant OTU 1 (COTS27) in the phylum *Spirochaetes*.

Maximum likelihood (ML) trees were constructed based on the full-length 16S rRNA gene sequences (**a**) and the sequences of 43 conserved marker genes identified by CheckM (**b**). The bootstrap values in (**a**) were calculated by resampling 1,000 times. The scale bars indicate substitutions per site. *1: The gene with accession No. FQ660021.1 in panel (a) was obtained from a polycycle aromatic hydrocarbon (PAH)-contaminated soil sample in a mitigated wetland.

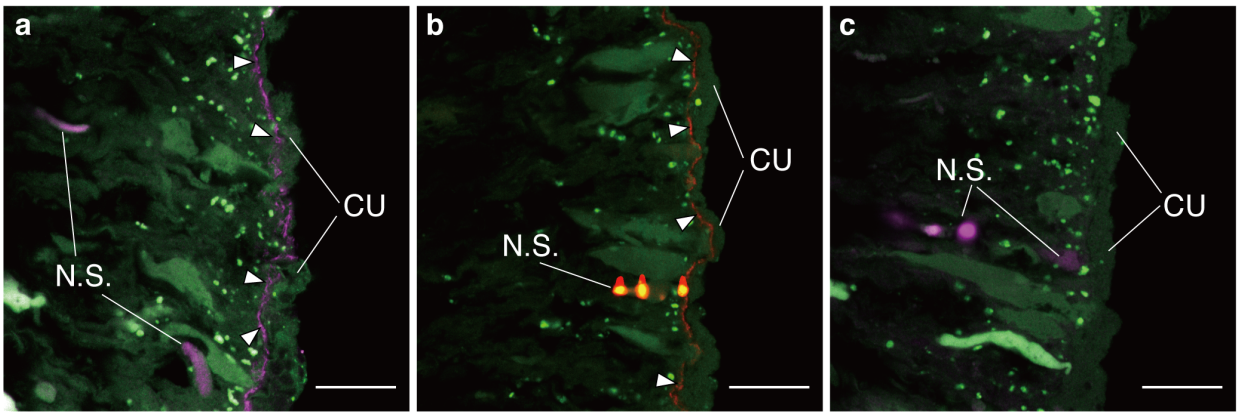

Fig. 4 FISH analysis of three serial sections of a COTS disc spine.
Each section was hybridized with the EUB338mix (**a,** purple; a general probe for bacteria), COTSsymb (**b**, red; a COTS27-specific probe), or Non338 (**c**, purple; a negative control to detect non-specific binding) probes. The probes were labelled with Cy3 in all panels and coloured with purple in panels **a** and **c** and with red in panel **b**. The green signals are tissue-derived autofluorescence. The arrowheads in panels **a** and **b** indicate layer-like signals from the general probe for bacteria (**a**) and the COTS27- specific probe (**b**). N.S. and CU indicate regions with non-specific binding and the outer cuticle complex, respectively. The scale bars represent 20 μm (**a**-**c**).

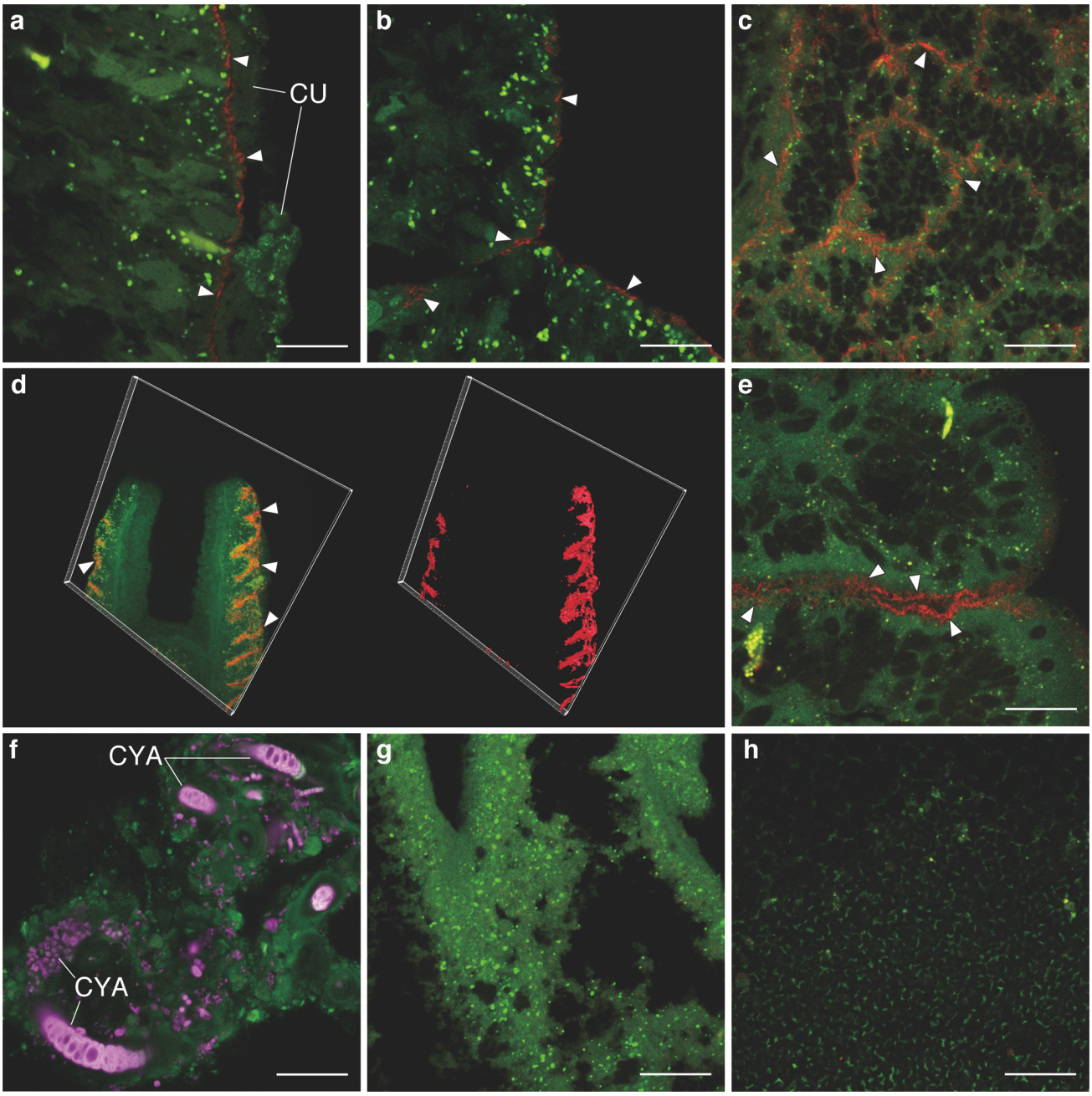

Fig. 5 Visualization of the COTS27 cells in different body parts of COTS using FISH. COTS27 cells (red) residing in the subcuticular spaces of the body walls were detected with COTSsymb, a COTS27-specific probe, in the tips (**a**) and bases (**b**) of aboral spines on the discs and arms, respectively, dermal papula (**c**), pedicellariae on the aboral side (**d**; 3D image [left] and 3D rendering image [right]), and tube feet (**e**). Many non-COTS27 bacteria (purple) were detected in the pyloric stomachs (**f**) using the EUB338mix probe. No visible bacteria were detected in the pyloric caeca (**g**) and gonads (**h**) in this study (the images were obtained applying the COTS27-specific probe). The arrowheads indicate signals from COTS27. The green signals are tissue-derived autofluorescence. CU: outer cuticle complex; CYA: cyanobacteria-like cells. Scale bars (**a–c** and **e–h**) indicate 20 µm. The 3D image in panel (**d**) was taken with an original objective of x40.

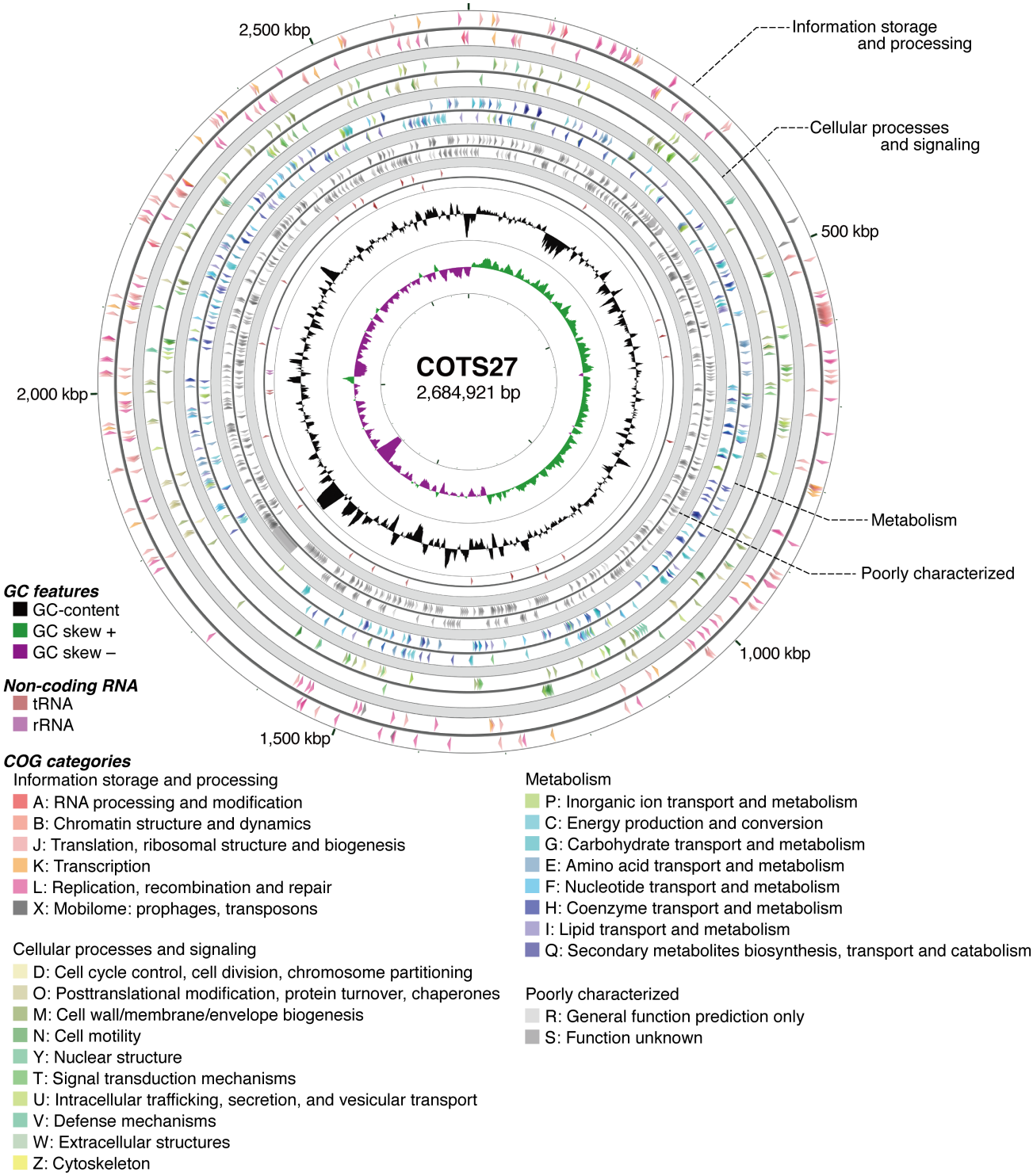

Fig. 6 Circular view of the COTS27 chromosome.

From the outside to the centre, each circle indicates forward strand CDSs; reverse strand CDSs; forward strand tRNA and rRNA genes; reverse strand tRNA and rRNA genes; GC-content; and GC skew. The CDSs were coloured according to the COG functional category of each CDS. The circular maps were created using CGView Server and the designations were then superposed manually.