



**HAL**  
open science

## Ecology of Endozoicomonadaceae in three coral genera across the Pacific Ocean

Corentin Hochart, Lucas Paoli, Hans-Joachim Ruscheweyh, Guillem Salazar, Emilie Boissin, Sarah Romac, Julie Poulain, Guillaume Bourdin, Guillaume Iwankow, Clémentine Moulin, et al.

► **To cite this version:**

Corentin Hochart, Lucas Paoli, Hans-Joachim Ruscheweyh, Guillem Salazar, Emilie Boissin, et al.. Ecology of Endozoicomonadaceae in three coral genera across the Pacific Ocean. *Nature Communications*, 2023, 14 (1), pp.3037. 10.1038/s41467-023-38502-9 . hal-04176293

**HAL Id: hal-04176293**

**<https://hal.science/hal-04176293v1>**

Submitted on 12 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Ecology of *Endozoicomnadaceae* in three coral genera across the Pacific Ocean

Received: 18 October 2022

Accepted: 26 April 2023

Published online: 01 June 2023

 Check for updates

Corentin Hochart<sup>1</sup>, Lucas Paoli<sup>2</sup>, Hans-Joachim Ruscheweyh<sup>2</sup>, Guillem Salazar<sup>2</sup>, Emilie Boissin<sup>3</sup>, Sarah Romac<sup>4</sup>, Julie Poulain<sup>5,6</sup>, Guillaume Bourdin<sup>7</sup>, Guillaume Iwankow<sup>3</sup>, Clémentine Moulin<sup>8</sup>, Maren Ziegler<sup>9</sup>, Barbara Porro<sup>10,11</sup>, Eric J. Armstrong<sup>3</sup>, Benjamin C. C. Hume<sup>12</sup>, Jean-Marc Aury<sup>5,6</sup>, Claudia Pogoreutz<sup>3,12</sup>, David A. Paz-García<sup>13</sup>, Maggy M. Nugues<sup>3</sup>, Sylvain Agostini<sup>14</sup>, Bernard Banaigs<sup>3</sup>, Emmanuel Boss<sup>15</sup>, Chris Bowler<sup>6,16</sup>, Colomban de Vargas<sup>4,6</sup>, Eric Douville<sup>17</sup>, Michel Flores<sup>18</sup>, Didier Forcioli<sup>10,11</sup>, Paola Furla<sup>10,11</sup>, Eric Gilson<sup>10,11,19</sup>, Fabien Lombard<sup>6,20,21</sup>, Stéphane Pesant<sup>22</sup>, Stéphanie Reynaud<sup>11,23</sup>, Olivier P. Thomas<sup>24</sup>, Romain Troublé<sup>8</sup>, Patrick Wincker<sup>5,6</sup>, Didier Zoccola<sup>11,23</sup>, Denis Allemand<sup>11,23</sup>, Serge Planes<sup>3,6</sup>, Rebecca Vega Thurber<sup>12,25</sup>, Christian R. Voolstra<sup>12</sup>, Shinichi Sunagawa<sup>2</sup> & Pierre E. Galand<sup>1,6</sup> ✉

Health and resilience of the coral holobiont depend on diverse bacterial communities often dominated by key marine symbionts of the *Endozoicomnadaceae* family. The factors controlling their distribution and their functional diversity remain, however, poorly known. Here, we study the ecology of *Endozoicomnadaceae* at an ocean basin-scale by sampling specimens from three coral genera (*Pocillopora*, *Porites*, *Millepora*) on 99 reefs from 32 islands across the Pacific Ocean. The analysis of 2447 metabarcoding and 270 metagenomic samples reveals that each coral genus harbored a distinct new species of *Endozoicomnadaceae*. These species are composed of nine lineages that have distinct biogeographic patterns. The most common one, found in *Pocillopora*, appears to be a globally distributed symbiont with distinct metabolic capabilities, including the synthesis of amino acids and vitamins not produced by the host. The other lineages are structured partly by the host genetic lineage in *Pocillopora* and mainly by the geographic location in *Porites*. *Millepora* is more rarely associated to *Endozoicomnadaceae*. Our results show that different coral genera exhibit distinct strategies of host-*Endozoicomnadaceae* associations that are defined at the bacteria lineage level.

Tropical coral reefs are hotspots of biodiversity in the world's oceans. They provide food and shelter for many marine animals, and are important to humans who benefit from the reefs' goods and services<sup>1</sup>. In recent decades, global warming and human activities have, however, increasingly threatened coral reefs<sup>2–5</sup>. Their resilience

and adaptability rely heavily on their ecosystem engineers, the corals, and their ability to resist damaging perturbations. In turn, the coral's adaptive capabilities are anchored in its relationship with associated microorganisms that together with the host form the holobiont<sup>6–9</sup>.

A full list of affiliations appears at the end of the paper. ✉ e-mail: [pierre.galand@obs-banyuls.fr](mailto:pierre.galand@obs-banyuls.fr)

Coral holobionts include complex microbial communities composed of viruses, fungi, archaea, bacteria and micro-algae<sup>10</sup>. The symbiosis between the Symbiodiniaceae algae and corals is the best characterized association within the coral holobiont<sup>11</sup>, but bacteria are also important partners. Coral-associated bacteria contribute significantly to the health of the host by participating in nutrient acquisition, metabolic (re)cycling, and protection against pathogens<sup>6,7,12,13</sup>. Understanding the composition, the diversity and the functions of coral-associated microorganisms can thus provide clues about the health status, but also the resilience and adaptive capabilities of corals<sup>14</sup>. Among the large diversity of coral-associated bacteria<sup>15</sup> some are noteworthy because of their abundance and commonness. Bacteria from the *Endozoicomonadaceae* family (*Gammaproteobacteria*, *Oceanospirillales*) are frequent in corals<sup>7,16</sup>, but also in a number of other marine invertebrates such as sponges<sup>17,18</sup>, bivalves<sup>19</sup>, ascidians<sup>20</sup> or gorgonians<sup>21</sup>. Their global distribution in a large number of hosts make them potential key bacterial symbionts in the marine ecosystem<sup>16</sup>. A strict host-bacteria interdependency remains, however, to be demonstrated since *Endozoicomonadaceae* have some characteristics of non-obligate symbionts including possible free-living stages<sup>22</sup>, and relatively large genome sizes<sup>23,24</sup>.

The family *Endozoicomonadaceae* was described only recently<sup>25</sup> and comprises the genera *Parendoicomonas*<sup>25</sup>, *Kistimonas*<sup>26</sup>, *Sansalvadorimonas*<sup>27</sup> and *Endozoicomonas*<sup>28</sup>. Although the definition of the genus *Endozoicomonas* is recent, retrospective analysis show that it was commonly detected in early coral microbiome studies<sup>29–32</sup>, and it continues to be frequently reported in a number of different coral species<sup>23,33–42</sup>. Different corals may harbor different *Endozoicomonadaceae* 16S rRNA types that could show patterns of co-phylogeny with the coral host<sup>41,43</sup>. The patterns of host-symbiont associations have also been hypothesized to be linked to the reproductive mode of the host<sup>43</sup>, or the environmental condition<sup>44</sup>, while other studies suggest a correlation between *Endozoicomonas* and Symbiodiniaceae<sup>45</sup>. The question of host-specificity remains open, however, since some *Endozoicomonadaceae* types could be shared between different hosts<sup>41</sup>, different *Endozoicomonadaceae* types can dominate within a same host<sup>43</sup>, and changes in their relative abundance seems to be independent of the symbiotic algae<sup>46</sup>. If the host is not the driver of microbiome composition, external factors could play a role<sup>47</sup>. *Endozoicomonadaceae* communities could be shaped by environmental factors as they were shown to be less abundant in *Acropora millepora* at lower seawater pH<sup>48,49</sup>, or during increased temperatures and subsequent bleaching<sup>31,50</sup>, as well as anthropogenic impact and habitat suitability<sup>44,51</sup>. In *Porites astreoides*, lesioned colonies also contained fewer *Endozoicomonas* sequences, compared to non-lesioned colonies<sup>52</sup>. These findings suggest that *Endozoicomonas* are under-represented in stressed corals<sup>14</sup>, however, it's not always the case. In *Pocillopora verrucosa* and *Acropora hemprichii*, although *Endozoicomonadaceae* decreased at sites impacted by sedimentation, they increased at sites impacted by municipal wastewater<sup>39</sup>. *Endozoicomonadaceae* also increased in abundance in *Porites* spp. under lower pH<sup>53</sup>, and under natural stressful conditions of shallow hydrothermal vent<sup>54</sup>, and were not impacted by coral bleaching or severe tissue sloughing in *P. verrucosa* in response to eutrophication<sup>55</sup>. *Endozoicomonadaceae* also proliferated after warm summer months in French Polynesia<sup>42</sup>. These contrasted observations between individual studies conducted locally illustrate the need for a large-scale approach on multiple coral species to unveil *Endozoicomonadaceae* diversity and biogeography, and better understand the factors controlling their presence and community composition. The widespread prevalence of *Endozoicomonadaceae* in corals indicates that they are important for the host, and the analyses of their genomes have given some clues regarding their potential interactions and the establishment of symbiosis with corals<sup>56</sup>. *Endozoicomonadaceae* have large genome sizes that suggest a possibility for a non-symbiotic living stage before host invasion<sup>23,24,57</sup>.

Host infection could be promoted by type III secretion system, like the one detected in a strain of *Endozoicomonas montiporae*<sup>57</sup>, and by potential effector proteins that that could help interactions<sup>23,56,58</sup>. A comparative analysis of several *Endozoicomonadaceae* genomes showed an enrichment of genes associated with carbon sugar transport and utilization, protein secretion, and synthesis of amino acids that could potentially be transferred to the host<sup>43</sup>. In addition, *Endozoicomonas acroporae* has the potential to degrade dimethylsulfoniopropionate (DMSP), which bacteria could use as a carbon source<sup>24</sup>. Various studies also suggested that *Endozoicomonadaceae* play a role in regulating the overall microbiome structure either by direct competition with other bacteria, or by producing antimicrobial compounds<sup>48,59</sup>. The extent of metabolic capabilities of *Endozoicomonadaceae* across species remains, however, to be mapped, and the variability of these pathways at ocean scale is not known.

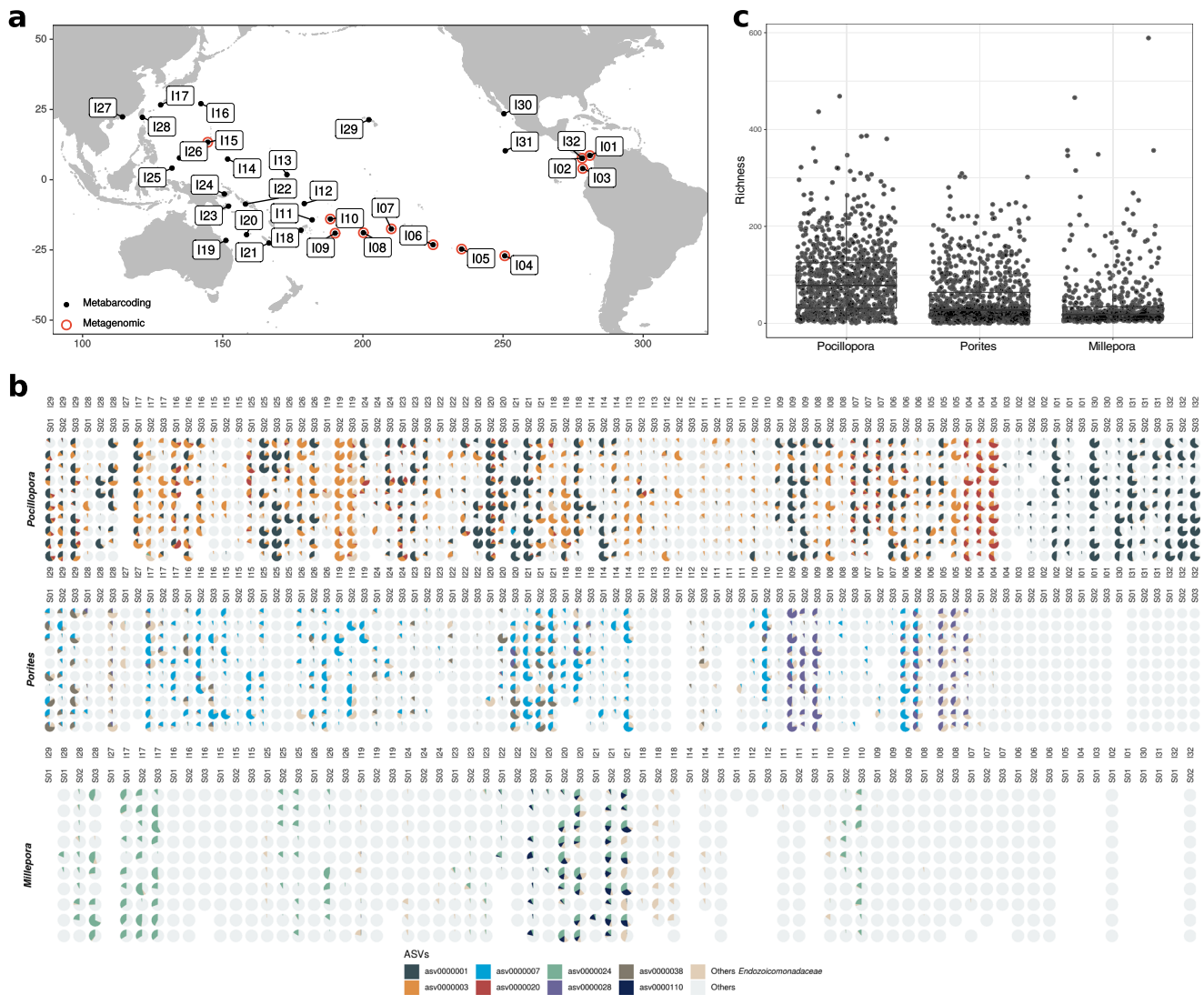
In this work, we study the ecology of *Endozoicomonadaceae* at an ocean-basin scale and in different coral genera. Our hypothesis was that different corals may have evolved different strategies of host-bacteria relationships. To assess this, we conducted an unprecedented campaign to methodically sample three globally distributed coral species. Based on morphology, we targeted the complex *Pocillopora meandrina*, the robust *Porites lobata*, and the fire coral *Millepora platyphylla* on 99 reefs from 32 islands across the Pacific Ocean. We precisely described the diversity of *Endozoicomonadaceae* in more than 2400 coral colonies by metabarcoding the 16S rRNA gene. In addition, we sequenced 270 metagenomes to assemble high-quality draft genomes of novel *Endozoicomonadaceae* species, and putative lineages. Here we show that the most common *Endozoicomonadaceae* lineage, found in *Pocillopora*, is globally distributed and has distinct metabolic capabilities, including the synthesis of amino acids and vitamins not produced by the host. The other lineages are structured partly by the host genetic lineage in *Pocillopora* and mainly by the geographic location in *Porites*. Our results demonstrate that different coral genera exhibit distinct strategies of host-*Endozoicomonadaceae* associations that are defined at the bacteria lineage level.

## Results

### *Endozoicomonadaceae* distribution and abundance in corals and seawater across the Pacific Ocean

Overall, *Endozoicomonadaceae* ASVs were detected in 99% of the coral samples (n = 2447). Despite their prevalence, their relative abundance varied greatly between samples (Fig. 1b). *Pocillopora* had the highest proportion of ASVs affiliated to *Endozoicomonadaceae* (53% of all *Pocillopora* sequences) followed by *Porites* (30%) and then *Millepora* (11%). *Pocillopora* also had the highest *Endozoicomonadaceae* richness, followed by *Porites* and then *Millepora* (Fig. 1c). In *Pocillopora*, *Endozoicomonadaceae* were present in all islands, but with low relative abundances in Coiba (I02), Malpelo (I03), and Guam (I15). In *Porites*, *Endozoicomonadaceae* were not detected east of Rapa Nui (I04), and in Kiribati (I13) and Southwest Palau Islands (I25). In *Millepora*, *Endozoicomonadaceae* were almost absent in corals from islands located east of Fiji (I18), with the exception of Samoa (I10), and almost absent from Guam (I15) and Ogasawara Islands (I16). We observed high variations in relative abundance between sites in some islands, like in Las Perlas (I01), Cook (I08) and Kiribati (I13) for *Pocillopora*, and Gambier (I06), Upolu (I10) and Chuuk Island (I14) for *Porites*, and New Caledonia (I21), Southwest Palau Islands (I25) for *Millepora*. In other cases, there was a high relative abundance and low variability between sites (Fig. 1b). At any given site, the presence of *Endozoicomonadaceae* in one coral genus was not predictive of its presence in the other genera (e.g., Rapa Nui I04 or Moorea I07), although at other sites, *Endozoicomonadaceae* were present and abundant in all three coral genera (e.g., Chesterfield I20 and New Caledonia I21).

Overall, asv0000001 was the most abundant *Endozoicomonadaceae* (12% of the sequences in the dataset) and the most common



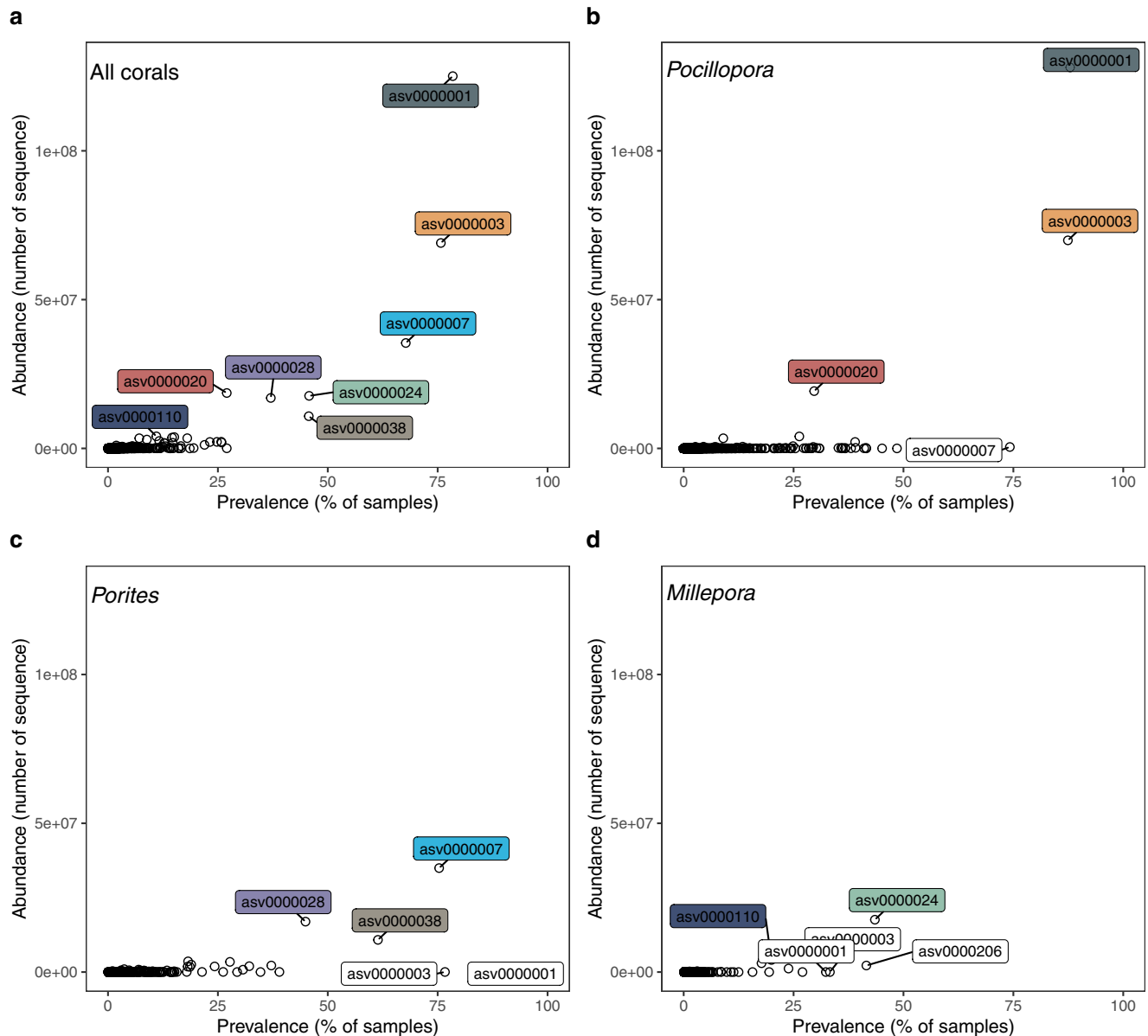
**Fig. 1 | Endozoicomnadaceae distribution and abundance in corals across the Pacific. a** Map showing the islands sampled during the *Tara* Pacific expedition. Black and red symbols indicate whether samples were analyzed by metabarcoding (16S rRNA gene) and/or metagenomics. **b** Proportion of the most abundant *Endozoicomnadaceae* ASVs in *Pocillopora*, *Porites*, and *Millepora* sampled across 99 reefs from 32 islands across the Pacific Ocean. The islands are ordered

longitudinally from west to east. **c** *Endozoicomnadaceae* community richness in the corals *Pocillopora* ( $n = 959$  samples), *Porites* ( $n = 944$ ), and *Millepora* ( $n = 544$ ). The box plot horizontal bars show the median value, the box indicates the first and third QRs, and the whiskers indicate  $1.5 \times \text{IQR}$ . Source data are provided as a Source Data file.

(found in 78% of the samples) (Fig. 2). *Pocillopora* had 3 dominant ASVs (abundance >1%, asv0000001, asv0000003, asv0000020) of which asv0000020 was more abundant at Rapa Nui (I04) and asv0000003 at the Great Barrier Reef (I19) (Fig. 1). *Porites* also had 3 dominant ASVs (asv0000007, asv0000028, asv0000038); asv0000028 dominated at Ducie Island (I05) and Niue (I09). *Millepora* had two dominant ASVs (asv0000024, asv0000110), of which asv0000110 was found mostly in Chesterfield (I20), New Caledonia (I21) and Solomon Islands (I22) (Fig. 1). All dominant ASVs had a broad geographic distribution since the same ASVs were found across the entire Pacific Ocean. Each of the ASVs defined as dominant in one coral host were also detected at very low proportions (>0.1%) in the other coral species (Fig. 2).

When the relative abundance of *Endozoicomnadaceae* in corals was low, the diversity of the overall microbial community reached higher values (Supplementary Fig. 1a). In *Pocillopora*, corals with low abundance of *Endozoicomnadaceae* were often dominated by *Flavobacteriaceae* (Supplementary Fig. 1b). In *Porites*, there were more *Kiloniellaceae* and *Rhodobacteraceae*, while in *Millepora* there were more *Spirochaetaceae*.

We also looked for *Endozoicomnadaceae* ASVs in the water surrounding the *Pocillopora* colonies, in surface water over the reef, and in the surface water off the island (Fig. 3). *Endozoicomnadaceae* sequences were detected in the water, but their relative abundance decreased rapidly with increasing distance from the colonies (Fig. 3a). While *Endozoicomnadaceae* represented on average 30% of the bacteria in *Pocillopora*, their relative abundances decreased to -0.5% in coral surrounding water, 0.05% in surface water over the reef, and 0.001% in surface water off the island (Fig. 3a). Additionally, we observed a significant positive correlation ( $r = 0.52$ ) between the relative abundance of *Endozoicomnadaceae* in *Pocillopora* and in the colony surrounding water (Fig. 3b). The *Endozoicomnadaceae* ASVs that were abundant in *Pocillopora* were also detected in the water, but in varying proportions (Fig. 3c). When moving away from the *Pocillopora* colonies, the proportion of typical *Pocillopora Endozoicomnadaceae* (asv0000001, asv0000003 and asv0000020) decreased, while the proportion of other *Endozoicomnadaceae* increased. In some cases, *Endozoicomnadaceae* were not detected at all in the surface water off the island.



**Fig. 2 | Prevalence and relative abundance of Endozoicomonadaceae.** Endozoicomonadaceae ASVs found in **a** all corals, **b** Pocillopora, **c** Porites and **d** Millepora.

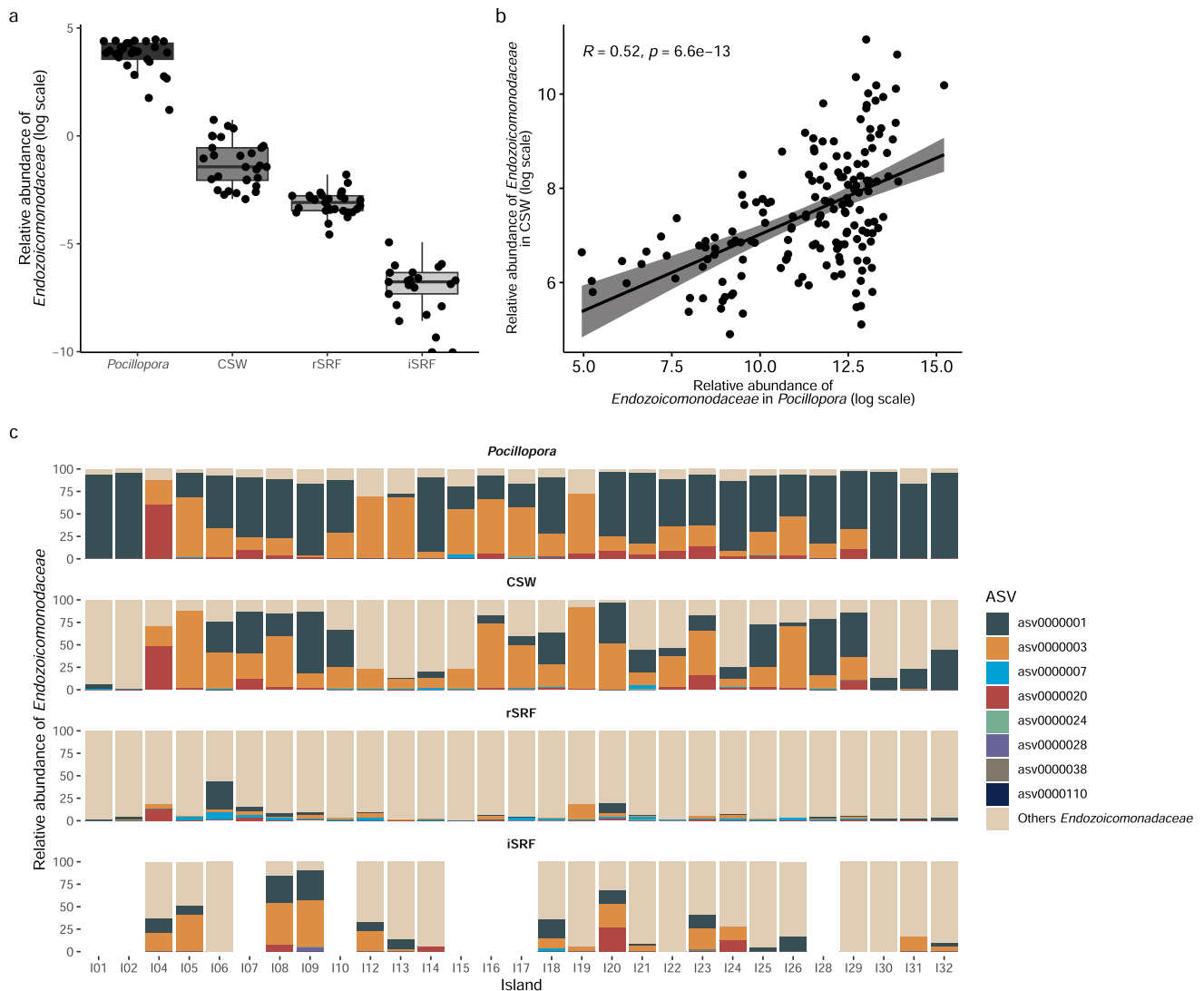
### Correlation between *Endozoicomonadaceae*, environment, islands, host genetic lineages, Symbiodiniaceae symbionts, and phylotypic biomarkers

There was no strong correlation between *Endozoicomonadaceae* relative abundance at the family level and chlorophyll *a*, pH, phosphate, silanol, sea surface salinity and temperature in *Pocillopora* and *Millepora* (Fig. 4a). In *Porites*, *Endozoicomonadaceae* had the highest correlation to salinity. At the ASV level, however, all correlation values were low, but there were differences within coral genera. In *Pocillopora*, asv0000001 had the strongest positive correlation to sea surface temperature and was negatively associated with pH, while asv0000003 and asv0000020 had the highest correlations to salinity and were negatively correlated to SiOH (Fig. 4a). In *Porites*, asv0000007 had the highest positive correlation to salinity, while asv0000028 and asv0000038 only had very low correlation values to environmental conditions. In *Millepora*, asv0000024 had the highest positive correlation to SiOH and asv0000110 only had low correlation values (Fig. 4a).

Information about assemblages of Symbiodiniaceae, phylotypic biomarkers indicative of coral health, and host genetic lineages was

obtained from a subset of 11 islands (Fig. 1a—red circle). The host genetic lineages identified through genome-wide SNP calling (material and methods) separated *Pocillopora* in five genetic lineages (Poc-SVD1 to Poc-SVD5), *Porites* in nine lineages (Por-K1a to Por-K3d) and *Millepora* in six lineages (Mil-SVD1 to Mil-SVD6). Overall, the composition of *Endozoicomonadaceae* communities was best explained by the genetic lineage of the host for both *Pocillopora* and *Porites* (Fig. 4b). Symbiodiniaceae had low significance with regard to *Endozoicomonadaceae* assemblage in *Pocillopora*, and the environment had low significance in *Porites*. Phylotypic biomarkers were always the lowest explanatory factor.

At the ASV level, variance partitioning showed that individual ASVs had a higher proportion of variance explained than overall *Endozoicomonadaceae* (Fig. 4c). For *Pocillopora*, asv0000001 and asv0000003 had the largest proportion of their variation explained by the factor 'island', while asv0000020 was explained mostly by the genetic lineage of the host (Fig. 4c). In *Porites*, the overall *Endozoicomonadaceae* relative abundance was best explained by the host genetic lineage, but all individual ASVs (asv0000007, asv0000028 and asv0000038) had most of their variance explained by island.



**Fig. 3 | Relative abundance of *Endozoicomnadaceae* in *Pocillopora* corals and the surrounding environment.** **a** *Endozoicomnadaceae* relative abundance in the coral *Pocillopora* host, the colony surrounding water (CSW), the surface water over the reef (rSRF), and the surface water off the island (iSRF) in the 0.2–3  $\mu$ m planktonic size fraction. The box plot horizontal bars show the median value, the box indicates the first and third QRs, and the whiskers indicate 1.5\*IQR. **b** Spearman

correlation between relative abundance of *Endozoicomnadaceae* in *Pocillopora* and in colony surrounding water (CSW). The lines represent the result of linear regression with the gray shadow as the 95% confidence interval. A two-sided *p*-value is given. **c** Relative abundance of the dominant *Endozoicomnadaceae* ASVs in the different sampling environments at each island. Source data are provided as a Source Data file.

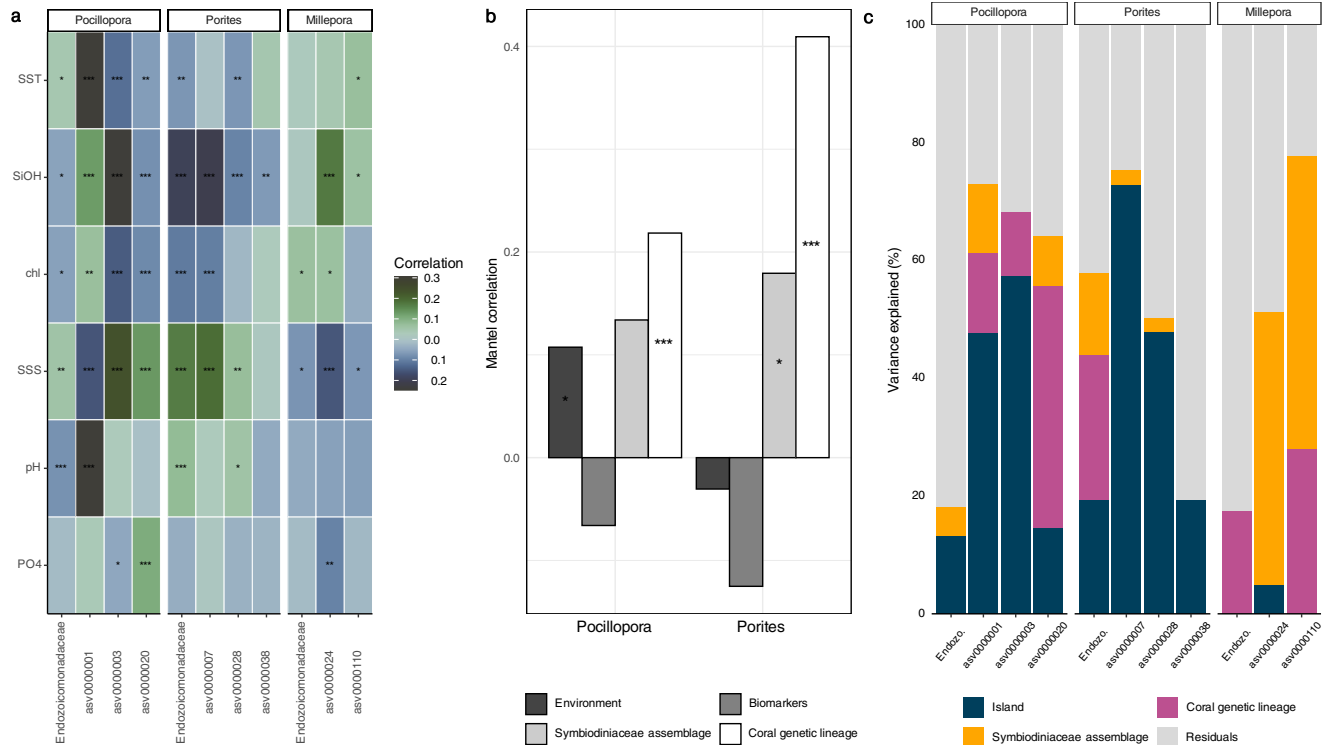
Asv0000038 had only a small portion of the variance explained by the factors that we tested. In *Millepora*, the Symbiodiniaceae assemblage was the strongest explanatory factor for both asv0000024 and asv0000110.

On the 11 islands we could also correlate individual ASV relative abundance to individual coral biomarkers. Different ASVs showed distinct correlation patterns (Supplementary Fig. 2). In *Pocillopora*, asv0000001 had strongest positive correlations to Symbiodiniaceae cell content and negatively to protein carbonylation (Supplementary Fig. 2). Asv0000003 was positively correlated to total antioxidant content and asv0000020 to protein ubiquitination. In *Porites*, asv0000007 was most positively correlated to Symbiodiniaceae content, and asv0000038 to protein carbonylation (Supplementary Fig. 2).

### Genomic characteristics and phylogeny of *Endozoicomnadaceae*

A total of 270 coral metagenomes were sequenced from a subset of 11 islands covering 33 reefs (Fig. 1a). We reconstructed a total of

24 MAGs belonging to the *Endozoicomnadaceae* family. They had a completeness ranging from 51.2% to 93.3%, and contamination levels from 0% to 4.9% (Supplementary Data 1). To conduct comparative genomics analyses, all publicly available *Endozoicomnadaceae* genomes were downloaded from NCBI or recovered from the literature ( $n = 30$ ). The taxonomic annotation of the MAGs and the phylogenetic tree based on 71 marker genes separated the *Endozoicomnadaceae* into five distinct genera: *Endozoicomonas*, *Parendozoicomonas*, *Kistimonas*, *Sansalvadorimonas*, and one unnamed genus (Fig. 5, Supplementary Data 1). Our *Tara* Pacific MAGs fell within two of these genera: the *Endozoicomonas* genus for *Millepora* and *Pocillopora* MAGs, and the *Parendozoicomonas* genus for *Porites* MAGs. All *Endozoicomnadaceae* genomes had similar coding density (80.8–91.6%), but distinct guanine-cytosine (GC) contents (33.2–57.3%) between genera (Supplementary Data 1). *Endozoicomonas* and *Parendozoicomonas* had similar GC contents with, respectively, 48.5% and 48.2%, while the only *Kistimonas* and *Sansalvadorimonas* genomes available had 57.3% and 50.3%, respectively (Supplementary Data 1).



**Fig. 4 | Correlation between *Endozoicomnadaceae*, environment, islands, host genetic lineages, Symbiodiniaceae symbionts, and phylotypic biomarkers.**

**a** Kendall correlations across 32 islands between abundant *Endozoicomnadaceae* ASVs and environmental factors. *Pocillopora* ( $n = 889$  samples), *Porites* ( $n = 903$ ), and *Millepora* ( $n = 515$ ). SST: Sea Surface Temperature; SiOH: Silanol; chl: Chlorophyll a; SSS: Sea Surface Salinity; pH; PO4: phosphate. **b** Mantel correlation across 11 islands between *Endozoicomnadaceae* community composition and

environmental conditions, phylotypic biomarkers, Symbiodiniaceae assemblages and coral host genetic lineages for the two coral genera *Pocillopora* ( $n = 77$ ) and *Porites* ( $n = 80$ ) phylotypic biomarkers data for *Millepora* are missing). **c** Variance partition analysis explaining the variance of the relative abundance of *Endozoicomnadaceae* and individual ASVs across 11 islands in relation to island, coral genetic lineages and Symbiodiniaceae assemblages in *Pocillopora* ( $n = 96$ ), *Porites* ( $n = 99$ ), and *Millepora* ( $n = 55$ ). Source data are provided as a Source Data file.

The *Parendoziomonas* genomes formed a monophyletic group with average ANI and AAI values of 73.80% (73.41–74.12%) and 71% (70.33–71.41%), respectively (Supplementary Fig. 3, Supplementary Data 2). The *Endozoicomonas* genomes also formed a well defined group, and they had average ANI and AAI values of 76% (71.08–99.99%) and 73% (60.96–99.99%), respectively. The genus *Sansalvadorimonas* has only one representative, *Sansalvadorimonas verongulae*. Because its genome is not present in the latest version of GTDB database, it was first annotated as *Parendoziomonas*, but our genomic analysis confirmed its position as a distinct genus.

Our *Tara* Pacific MAGs from *Pocillopora*, *Porites*, and *Millepora* were divided into three well-supported monophyletic clusters that corresponded to the three coral hosts (Fig. 5a, Supplementary Fig. 3). *Endozoicomonas* MAGs obtained from *Pocillopora* all grouped with sequences from the *Pocilloporidae* family earlier obtained from one *P. verrucosa* and two *S. pistillata*<sup>23</sup>. The two *Endozoicomonas* genomes from *S. pistillata* represent two distinct species (93% ANI similarity). *Parendoziomonas* MAGs from *Porites* clustered with a MAG obtained earlier from *Porites lutea*<sup>12</sup>. It was previously annotated as *Endozoicomonas*, but our analysis shows that it is a *Parendoziomonas* with closer similarity to *Parendoziomonas* from *P. haliclonae* (73.8% ANI and 70.69% AAI) than with other *Endozoicomonas* genomes (71.83% ANI and 60.50% AAI). *Millepora* MAGs were distantly related to all genomes published to date (Fig. 5a).

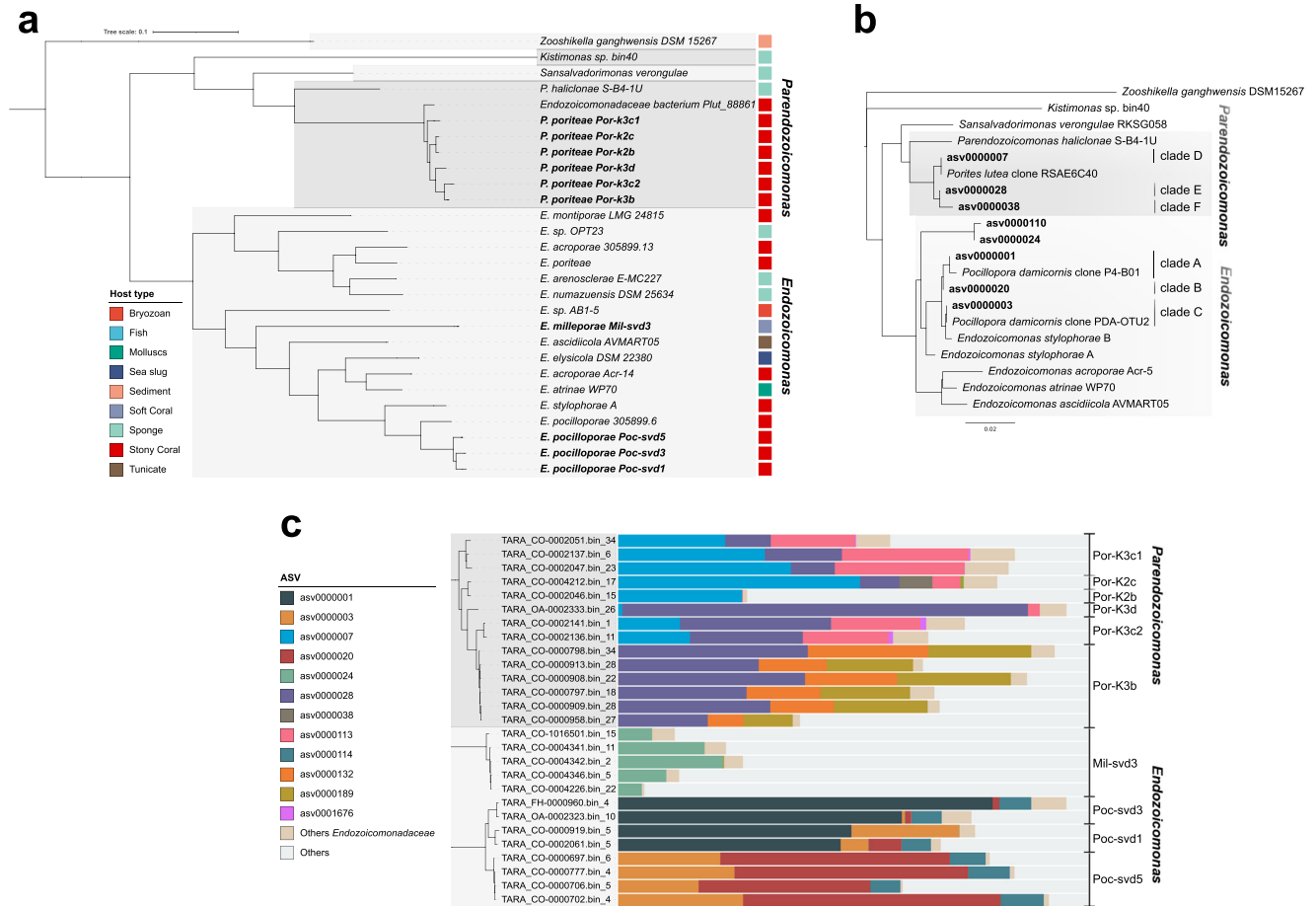
Although *Endozoicomonas* genomes from the same host species clustered together, there was no clear pattern of co-phylogeny between the bacteria and their host. *Endozoicomonas* genomes from different coral genera were often more related to genomes from other marine invertebrates (Supplementary Fig. 3). For instance, the *Endozoicomonas* symbionts of two *Acropora* species, *A. humilis* and *A.*

*muricata*, were separated in two distinct clades and were more closely related to a sponge and an ascidian symbiont than to each other. There were, however, exceptions and *Endozoicomonas* symbionts from *S. pistillata*, *P. verrucosa*, and *P. meandrina* clustered according to host phylogeny.

In the MAGs from this study, ANI was >95% in coral host-specific clusters, which corresponds to the threshold used for species delineation<sup>60</sup>. The specl method confirmed that each coral species harbored one specific *Endozoicomnadaceae* species (Supplementary Data 1). We could also delineate subclades within host-specific *Endozoicomnadaceae* species. Within *Pocillopora*, the MAGs were separated into three subclades that corresponded to three distinct *Pocillopora* genetic lineages (Poc-SVD5, Poc-SVD1 and Poc-SVD3) (Fig. 5c). For *Porites*, the MAGs separated into 6 subclades (Fig. 5c). The subclades that contained most MAGs corresponded exclusively to the *Porites* genetic lineages K3b. For *Millepora*, all MAGs shared a high identity and thus belonged to the same group.

Based on the phylogenomic tree, Specl results and ANI/AAI values we thus propose three new species of *Endozoicomnadaceae*: *Candidatus Endozoicomonas pocilloporae* sp. nov., *Candidatus Parendoziomonas poriteae* sp. nov. and *Candidatus Endozoicomonas milleporae* sp. nov. Additionally, we propose three lineages within *E. pocilloporae*, Poc-SVD1, Poc-SVD3 and Poc-SVD5, named according to the host genetic lineages (Fig. 5c, Supplementary Fig. 4). We also propose 6 lineages within *P. poriteae*, also named according to lineages, of which Por-K3c2 was not host lineage specific (Fig. 5c).

We constructed a phylogenetic tree based on 16S rRNA gene sequences (ASVs) obtained from *Tara* Pacific amplicon sequencing data and references from the literature. The 16S rRNA gene tree had a similar topology than the marker gene tree (Fig. 5b) separating



**Fig. 5 | Distance tree of Endozoicomonadaceae. a** Based on 71 concatenated single-copy marker proteins from MAGs and **b** based on 16S rRNA genes from ASVs. **c** Proportion of ASVs in the samples from which the MAGs were reconstructed.

Clades were defined in this study. ASVs and MAGs from this study are in bold. **c** Proportion of ASVs in the samples from which the MAGs were reconstructed.

the 4 named *Endozoicomonadaceae* genera. ASVs from the 3 coral species were separated in host-specific clades under the *Endozoicomonas* and *Parendoziomonas* genera. Within host-specific clades, ASVs grouped in distinct subclades. For *Pocillopora*, we delineated 3 clades. Clade A included the most common ASV (asv0000001), as well as sequences from *Pocillopora* of the Red Sea (clone P4-B01, KC668593<sup>36</sup>). Clade A was found in most *Pocillopora* genetic lineages (Supplementary Fig. 5). Clade B contained asv0000020 found mostly in the genetic lineage SVD5. Clade C contained asv0000003, as well as sequences from the Great Barrier reef (clone PDA-OTU2, AY700600<sup>30</sup>) (Fig. 5b), and was found mostly in host lineages SVD1, SVD2 and SVD5 (Supplementary Fig. 5). For *Porites*, we delineated 3 groups. The most common ASV was in clade D (asv0000007) together with a sequence from the Indian Ocean (clone RSAE6C40, KF180129<sup>41</sup>); it was more common in genetic lineage K2. Other common ASVs were in clade E (asv0000028), clade F (asv0000038) that were more common in the *Porites* lineage K3 (Supplementary Fig. 5). *Millepora* ASVs grouped in two clades, clade G and H, corresponding to the two most common ASVs (asv0000024 and asv0000110) (Fig. 4b).

Since the *Endozoicomonadaceae* MAGs did not contain 16S rRNA genes, we could not directly link the genomes to the ASVs. However, by looking at the ASV relative abundance in the samples from which the MAGs were reconstructed, we could associate groups of ASVs to the MAGs (Fig. 5c). In *Pocillopora*, the *Endozoicomonas* strains Poc-SVD1 and Poc-SVD3 corresponded to the dominance of asv0000001 (clade A), and Poc-SVD5 contained a majority of asv0000020 (clade B), but also asv0000003 (clade C). In *Porites*, Por-K3b, Por-K3c2 and Por-K3d had a majority of asv0000028 (clade E), while Por-K2b, Por-K2c and Por-K3c had more asv0000007 (D) (Fig. 5c).

**Endozoicomonadaceae mOTUs and mTAGs**

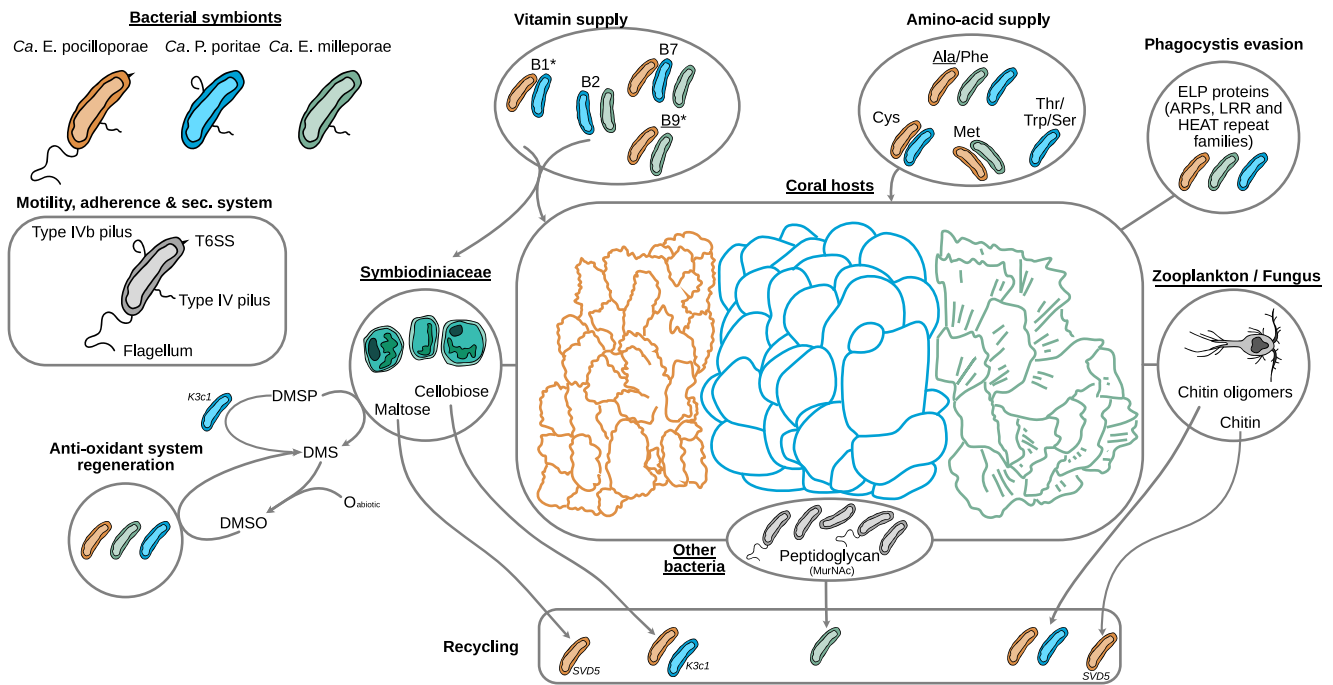
*Endozoicomonadaceae* MAGs were also grouped into marker gene-based operational taxonomic units (mOTUs). All *Parendoziomonas* MAGs from *Porites* grouped within one mOTU (mOTU\_3), *Endozoicomonas* MAGs from *Pocillopora* were separated in two mOTUs (mOTU\_2 and mOTU\_67) and *Millepora* MAGs in one mOTU (mOTU\_23) (Supplementary Fig. 6). Only one *Endozoicomonas* MAGs belonged to the mOTU\_67 and was recovered from a SVD5 lineage coral while five of them grouped into the mOTU\_2 and were recovered from SVD1 and SVD3 lineage corals (Supplementary Fig. 5). Finally, two MAGs did not have enough marker genes (<6) to be grouped into mOTUs. The mOTUs were specific to the coral host.

The relative abundance of *Endozoicomonadaceae* were determined based on mOTUs and mTAGs, which are 16S rRNA gene sequences extracted from metagenomes, in our three different coral species from 11 islands across the Pacific Ocean (Supplementary Fig. 7). We observed broadly similar patterns between mOTUs, mTAGs and ASVs profiles. We detected, however, some small differences: mTAGs and ASVs allowed the detection of *Endozoicomonadaceae* at very low abundance that were not seen with mOTU in Coiba (I02) and Guam (I15) islands for *Pocillopora*, in Coiba (I02), Cook (I08) and Rapa Nui (I04) for *Porites* and in Moorea (I07), Cook (I08), Niue (I09) and Guam (I15) islands for *Millepora* (Fig. 1, Supplementary Fig. 7).

**Metabolic capabilities of Endozoicomonadaceae**

Among the pathways common to all *Endozoicomonadaceae* we noted the Embden-Meyerhof pathway, the pyruvate dehydrogenase, the citrate cycle, the non-oxidative phase of the pentose phosphate pathway, and the Type I, II secretion systems. Sequences coding for





**Fig. 6 | Schematic summary of the inferred metabolisms and appendages characterizing the three new species and lineages (SVD5, K3c1) of *Endozoicomonadaceae* found in *Pocillopora* (orange), *Porites* (blue), and *Millepora* (green).** Underlined names: vitamin or amino acid also produced by the coral host; asterisk:

genes for export not known. T6SS: type 6 secretion system, ELP: eukaryotic like protein, ARPs: actin-related proteins, LRR: leucine-rich repeats, DMSP: dimethylsulfoniopropionate, DMSO: dimethylsulfoxide, DMS: Dimethyl sulfide.

subunits of the DmsABC enzyme, which catalyzes the reduction of dimethylsulfoxide (DMSO) to dimethyl sulfide (DMS), were also detected in all *Endozoicomonadaceae* (Fig. 6, Supplementary Fig. 8). Most *Endozoicomonadaceae* encoded different eukaryotic-like proteins (ankyrin, leucine-rich, tetratricopeptide, HEAT and WD40 repeats). We did not find any genes involved in the assimilation of nitrate in the *E. pocilloporae*, *P. poritae*, and *E. milleporae*, species, but they were present in *E. acroporae* and *P. haliclona* (Supplementary Fig. 8).

The three new *Endozoicomonadaceae* genomes had genes for biotin (vitamin B7) biosynthesis, from pimeloyl-Coa and the pimeloyl-[acp], and export (Supplementary Fig. 9a). The biosynthesis pathway of the pimeloyl precursor through the BioI, BioW and BioC-BioH pathways were, however, incomplete in all the *Endozoicomonadaceae* genomes. The genes to synthesize this precursor through BioI (long chain acyl pathway) were present in all the genomes of the coral hosts that we analyzed (Supplementary Fig. 9b).

Most of the *Endozoicomonadaceae* genomes had genes for the biosynthesis of riboflavin (vitamin B2). *E. milleporae* and *P. poritae* had the potential to export riboflavin, but not *E. pocilloporae* (Supplementary Fig. 9, Fig. 6). Genes for the synthesis of the two flavoproteins, derived from riboflavin, the flavin adenine dinucleotide (FAD) and the flavin mononucleotide (FMN) were found in all three *Endozoicomonadaceae* (Supplementary Fig. 9). The genes for riboflavin synthesis were not present in our coral host genomes (Supplementary Fig. 9b).

*E. pocilloporae* and *P. poritae* also had the potential to synthesize thiamine (vitamin B1), while the coral host could not. Finally, *E. pocilloporae* and *E. milleporae* and the coral hosts had the potential of folate (vitamin B9) synthesis (Supplementary Fig. 9b).

Alanine and phenylalanine biosynthesis, and their potential export by genes with an AlaE and EamA domain, were found in the three *Endozoicomonadaceae* species (Fig. 6, Supplementary Fig. 10). Only *E. pocilloporae* and *P. poritae* harbored genes for the cysteine synthesis, and had the potential to export it through the cydDC complex or bcr gene. Only *E. pocilloporae* and *E. milleporae* had genes involved in the

biosynthesis of methionine as well as genes with AzlCD domains, which could allow the export of the amino acid. Genes for serine (Ser) biosynthesis were present in *P. poritae* and *E. pocilloporae* Poc-SVD5 and Poc-SVD3. Threonine (Thr) biosynthesis were present in some *P. poritae* and *E. pocilloporae*. However, genes for export with ThrE domains (Thr and Ser export) were only found in *P. poritae*. Tryptophan (Trp) biosynthesis genes were found only in *P. poritae* that also had the gene carrying the EamA domain for putative Trp export.

The coral hosts only had the potential for alanine synthesis (Supplementary Fig. 10b).

Regarding motility, the presence of genes coding for pili and flagella varied between *Endozoicomonadaceae*. *E. milleporae* and *P. poritae* did not have sequences for the flagella genes, while *E. pocilloporae* did. *E. milleporae* and *P. poritae* had genes for the type IVb pilus, but *E. pocilloporae* did not. All had genes for the type IV pilus (Fig. 6). Some *Endozoicomonadaceae*, like *E. montiporae* and *P. haliclona*, had all the genes for these three appendages (Supplementary Fig. 8). Type VI secretion system (T6SS) was only found in *E. pocilloporae* (Fig. 6).

Among genes coding for transporters, phosphate, zinc and methionine transporters were found in the *Endozoicomonadaceae* genomes from the 3 coral species (Supplementary Fig. 8). Inversely, sequence coding for the transport of Arginine were found only in *P. poritae*, and L-Amino acid transport genes only in *E. pocilloporae* (Supplementary Fig. 8).

Genes for different subunits of the phosphotransferase system (PTS) were present in *Endozoicomonadaceae*. Most of the *Endozoicomonadaceae* species had the potential to assimilate fructose, glucose and beta-glucoside. Alpha-glucoside, trehalose and maltose PTS genes were found in *E. pocilloporae* and *P. poritae* (Supplementary Fig. 8). *E. pocilloporae* and the Por-K3c1 clade of *P. poritae* could assimilate cellobiose, while maltose assimilation potential was only present in the Poc-SVD5 clade within *E. pocilloporae*.

Protein coding sequences for the degradation of chitin residue were detected in *E. milleporae*, *E. pocilloporae* and *P. poritae* species.

However, among chitin residues, *E. milleporae* has only the potential to assimilate a N-Acetylglucosamine (GlcNAc) monomer, while *E. pocilloporae* and *P. poriteae* have the capabilities to assimilate GlcNAc dimers (chitobiose) and trimers (chitotriose) (Supplementary Fig. 8). Although *E. pocilloporae* and *P. poriteae* have the capability to actively import chitin oligomers (trimmer to hexamer) into their periplasm, only *E. pocilloporae* Poc-SVD5 has the chitinase gene to degrade GlcNAc tetramer (Chitodextrin) and larger polymers. Among amino sugar metabolisms, N-Acetylmuramic acid (MurNAc) phosphotransferase system was only detected in *E. milleporae* MAGs. MurNAc and GlcNAc can be both used for the biosynthesis of peptidoglycan, however, GlcNAc contrary to MurNAc can be redirected to the glycolysis pathway.

Genes involved in the uptake of the L-Ascorbate (vitamin C) and its anaerobic degradation in D-Xylulose-5P were found in *E. milleporae* and *P. poriteae* only (Supplementary Fig. 8). The type II L-ascorbate degradation (aerobic) was partially complete. In fact only the genes common to anaerobic degradation were present. Key genes required for degrading testosterone were not detected in the *E. milleporae*, *E. pocilloporae*, and *P. poriteae* species. The Endozoicomonadaceae *E. montiporae* and *E. elysicola* had all genes required for the degradation of androstenedione, but *E. pocilloporae*, *E. milleporae*, and *E. milleporae* did not (Supplementary Fig. 8).

*E. milleporae*, *P. poriteae*, and *E. pocilloporae* like *E. elysicola* genomes showed the potential for arsenate detoxification through the Type III arsenate detoxification and the arsenite methylation pathways (Supplementary Fig. 8).

Almost all Endozoicomonadaceae species had the capabilities of heme B biosynthesis from the L-glutamate (Supplementary Fig. 8). However, for *E. milleporae*, the first protein involved in the pathway, which catalyzes the reaction of transforming L-Glutamate in L-Glutamate-tRNA was not found. Additionally, none of our recovered MAGs contains the genes allowing siroheme biosynthesis. The genes implied in heme A biosynthesis from the heme were found in all Endozoicomonadaceae species except for *E. milleporae*. The biosynthesis of biliverdin, another product of the heme, was found in most of the Endozoicomonadaceae species apart from *E. milleporae*, *Kistimonas* and genomes from the unnamed Endozoicomonadaceae genus (Supplementary Fig. 8).

We also annotated the MAGs against the CAZy database specifically targeting carbohydrate active enzymes. A PCA based on CAZy composition separated the *Porites* MAGs from *Millepora* and *Pocillopora*, which could not be separated from each other (Supplementary Fig. 11). The CAZy that best explained the separation between groups are listed in Supplementary Data 2.

## Discussion

The unprecedented systematic sampling effort of the *Tara* Pacific expedition revealed that each studied coral genus was closely associated with one specific Endozoicomonadaceae species with broad cross-ocean geographic distribution despite varying environmental conditions. These new host-specific bacteria were composed of different lineages that varied in proportion between reefs, and the lineages' biogeography showed that the host-bacteria association patterns differed between host species. Endozoicomonadaceae strains had specific metabolic potentials, and they were associated with the host's genetic lineage, or the islands, rather than to environmental conditions.

### *Pocillopora*

*Pocillopora* was the coral genus that contained the highest proportion of Endozoicomonadaceae, and more precisely, it was specifically associated with the newly described species *E. pocilloporae* within the genus *Endozoicomonas* (Supplementary discussion). Among the three main *E. pocilloporae* strains, the most common, clade A (asv0000001, 90% of

the *Pocillopora* samples), appears like an extremely successful coral-associated bacterium. It was abundant across the entire Pacific Ocean, and also found, although at very low relative abundance, in *Porites* and *Millepora* (in 82% of all coral samples), suggesting an important function in the coral holobiont (explored further below). The clade A sequence was similar to sequences found earlier in *P. verrucosa* and *P. damicornis* in the Red Sea and American Samoa<sup>36,43</sup> indicating the presence of a global *E. pocilloporae* strain shared between different *Pocillopora* species. We also identified sequences 100% identical to our clade A in earlier studies on the effect of stress on coral microbiomes. Our clade A corresponded to an abundant *Pocillopora* OTU that did not decrease in abundance under elevated nitrate and urea concentrations in the Pacific Ocean (Moorea) or the Red Sea<sup>42,55</sup>, although it decreased under excess dissolved organic carbon (OTU2<sup>55</sup>) and elevated temperature (OTU-Endo<sup>62</sup>). The success of *E. pocilloporae* clade A may thus, in part, rely on its tolerance to specific environmental perturbations, but it could also benefit from its ability to disperse and effectively colonize corals. We found it abundant in the plankton surrounding the coral colonies, or 'coral ecosphere'<sup>22</sup>, and as far away as in water off the islands, yet in lower abundances. In addition, the genomes of *E. pocilloporae* harbour flagellar assembly genes, which are not found in *P. poriteae* and *E. milleporae*. The potential ability to produce a flagellum could allow bacteria to swim towards hosts<sup>63,64</sup>, and facilitate effective colonization of corals able to spawn like *Pocillopora*, which can acquire prokaryotes from the environment<sup>56,65-68</sup>. Finally, *E. pocilloporae* was also characterized by the presence of a type VI secretion system, an antibacterial apparatus delivering toxins<sup>69</sup>, that would give a competitive benefit by preventing other bacteria from multiplying in the coral host<sup>70</sup>.

Clade A was, however, not always dominant. It was thought that *Pocillopora* was always associated with a single *Endozoicomonas*<sup>43</sup>, but our study reveals that other clades (16S rRNA gene clade B and C) replaced the dominant clade A in some islands. These clades were associated with specific genetic lineages of *Pocillopora*. Clade B (asv0000020) was most abundant in the genetic lineage SVD5 found in Easter Island (I04), where Clade A was completely absent. Clade B could represent a bacterium that has co-evolved with its SVD5 host. SVD5 was also present in Moorea (I07), but there, clade B was present in association with Clade A. The host-bacteria association could thus reflect an interplay between co-evolution that is strongly marked in isolated islands, such as Easter Island, and other selection factors leading to mixed communities, potentially via 'shuffling' of abundant *Endozoicomonas* phylotypes<sup>71</sup> in more interconnected islands. Clade C (asv0000003) was most abundant at the Great Barrier Reef (I19), and was closely related to a bacterium earlier detected in *P. damicornis* in eastern Australia<sup>30</sup>. Our field observations indicate that the corals sampled at the Great Barrier Reef during the *Tara* Pacific expedition may have been *P. damicornis*. Clade C could thus be specific to that coral species. Clade B and C had, together with clade A, the potential for type IV pili synthesis, which could help host sensing and surface motility<sup>72</sup>.

Our genome analysis showed that *E. pocilloporae* have genes coding for the production of the essential vitamins B1, B2 and B7, and specific amino acids (Phe, Cys, Met), while the coral hosts did not have these metabolic capabilities. The coral holobiont must thus acquire these essential metabolites from heterotrophic feeding and/or its associated prokaryotes, including *E. pocilloporae*. Many algae like Symbiodiniaceae are auxotrophs for different B vitamins<sup>12,56,73,74</sup> and could also need vitamins produced by bacteria.

Our data also provide clues about *E. pocilloporae* lifestyle. They had the genomic potential for cellobiose assimilation, suggesting that they could use cellulose originating from the degradation of Symbiodiniaceae cells<sup>23</sup>. However, our study also revealed differences in potential metabolisms between *E. pocilloporae* lineages. Poc-SVD1, contrary to Poc-SVD5 and Poc-SVD3, had the potential to degrade chitin, the most abundant polysaccharide in the ocean<sup>75</sup>, and which for instance could originate from zooplankton taken by the coral via

heterotrophic feeding. The Poc-SVD5 genomes, corresponding to 16S rRNA gene Clade B and C defined in this study, had the potential to assimilate maltose, a sugar resulting from the hydrolysis of starch. Starch could originate from Symbiodiniaceae's carbon and energy storage. Differences in maltose metabolisms between different *E. pocilloporae* lineages may thus reflect the presence of different Symbiodiniaceae genera that have different energy storage strategies, and thus different starch content<sup>76</sup>. Poc-SVD5 was associated with Symbiodiniaceae Cladocopium C42, while Poc-SVD1 was reconstructed from samples with higher proportion of Cladocopium C1.

### Porites

In *Porites*, although *Endozoicomonas* was the most common and abundant bacterial symbiont, it was rare or absent from a large number of islands. *Endozoicomonas* has earlier been shown to be abundant in *Porites* from the Caribbean<sup>34,38</sup> and the Pacific Ocean<sup>53</sup>, but our data reveal that their overall abundance can vary greatly geographically. The pattern of host-bacteria association thus appears different from the one seen in *Pocillopora*, with different species of associated *Endozoicomnadaceae*. The new species *P. poriteae*, genus *Par-endozicomonas*, was dominated by Clade D (asv0000007) found from the Gambier Island (I06) to Sesoko Island in the west Pacific (I19). Clade D sequences were similar to the ones found earlier in the Indian Ocean off South Africa<sup>61</sup> indicating that its geographic spread is probably large. This wide geographic distribution, although not as extensive as *E. pocilloporae* Clade A, may also be due to resistance to some environmental perturbations. Sequence comparison showed that *Endozoicomonas* similar to clade D resisted elevated environmental nitrate and urea concentrations, and elevated temperatures<sup>42</sup>. In *P. lobata*, *Endozoicomnadaceae*, although we could not verify lineage identity, did not decline under nutrient stress<sup>77</sup> or elevated pH at CO<sub>2</sub> seeps<sup>48,53</sup>. Other *P. poriteae* lineages were more limited in their distribution, and contrary to *E. pocilloporae*, the distribution of these less abundant lineages was associated with islands rather than their host's genetic lineage. Clade F was for instance mostly restricted to Ducie Island (I05) and Cook Island (I09).

The *P. poriteae* genomes also showed differences to *E. pocilloporae*; they had genes coding for a type IVb pilus for host colonization and adhesion<sup>78</sup>, and may be able to provide their host with B1, B2 and B7 vitamins, and Thr/Trp amino acids. Interestingly, there were also differences among the different *P. poriteae* lineages. Por-K3c1, which corresponded to asv0000007, was the only *P. poriteae* that had the potential for cellobiose assimilation. The Por-K3c1 lineage was also the only one to have a gene coding for metabolizing dimethylsulfoniopropionate (DMSP) into dimethyl sulfide (DMS). That gene was not found in other *P. poriteae*, *E. pocilloporae* or *E. milleporae* lineages, which indicates that the involvement of *Endozoicomnadaceae* in the sulfur cycle<sup>24</sup> is very much lineage specific.

### Millepora

In *Millepora*, host-bacteria association differed from that of scleractinian corals, with only few sampled fire coral colonies having abundant communities of *Endozoicomonas*. The association seems looser and thus probably less important for the host. Among the 2 main *E. milleporae* lineages, clade G had the widest distribution, while clade H was more common in areas around New Caledonia (I21). Our analyses suggest that *E. milleporae* clades were associated with specific Symbiodiniaceae genera and coral host genetic lineages, but these results remain hypothetical since only few *Endozoicomonas* were present in the 11 islands for which host genetic information was available. Our results contrast with the ones from a recent study in Moorea<sup>47</sup> where differences in bacterial abundance were associated with reef habitats rather than host genotype. Similarly, in the soft coral *Scleronephthya gracillium*, *Endozoicomonas* abundance varied greatly along the Kur-oshio Current Region, and was seemingly related to differences in reef

habitats<sup>79</sup>. In our study, the environmental parameters did not strongly explain the variability of *Endozoicomonas* in *Millepora*. Regarding potential metabolisms, *E. milleporae*, contrary to *E. pocilloporae* and *P. poriteae*, did not have the potential to use GlcNac polymer, cellobiose or starch, but could use peptidoglycans from other bacteria.

Our results draw a fine picture of the patterns of *Endozoicomnadaceae* abundance in corals across the Pacific Ocean with implications for our understanding of the ecology and evolution of host-symbiont relationships. We show that the ecology of this widespread symbiont should be considered at the lineage level to understand the factors structuring communities and infer associated metabolic contributions. Our data suggest that different coral species exhibit distinct host-*Endozoicomnadaceae* relationships ranging from a strong association illustrated by the global, and abundant presence of *Endozoicomonas* in *Pocillopora*, to a rather weak association with a rare and scattered presence in *Millepora*. In all corals, the environment had generally only a small structuring effect on *Endozoicomnadaceae* community composition, while the genetic lineage of the host was important in some corals, arguing for a high level of host specificity putatively shaped by long co-evolutionary histories. Thus, coral-bacterial association at large may range from stable co-dependent relationships that arose through evolutionary time to opportunistic associations that are flexible and determined by the prevailing environment. Even for bacteria within the family *Endozoicomnadaceae*, we find evidence for one and the other. *Pocillopora* was associated with a very abundant *Endozoicomonas* lineage throughout its entire geographic range. The success of this lineage could be explained by its functional potential inferred from genome characteristics including the ability to swim, destroy other bacteria, and thrive on carbon sources produced by Symbiodiniaceae. Other *E. pocilloporae* lineages were specific to the host's genetic lineages, while in *P. poriteae*, the lineages were linked to the location (island). This unique dataset from the Tara Pacific expedition reveals the taxonomic and functional diversity of coral-associated *Endozoicomnadaceae* and emphasizes the need for studies at the lineage level to uncover the bacterial compartment of the coral holobiont.

## Methods

### Sample collection, environmental parameters, and biomarkers

Coral and plankton samples were taken across the Pacific Ocean from three to four different sites at each of the 32 islands visited during the Tara Pacific expedition<sup>80</sup>. The detailed sampling protocols are presented in Lombard et al.<sup>81</sup>. Based on morphology, we targeted the complex *Pocillopora meandrina*, the robust *Porites lobata* and the fire coral *Millepora platyphylla*. Samples from 10 different colonies from each species were collected at each site using a hammer and a chisel. They were stored individually underwater in Ziploc bags, and conditioned in tubes with DNA/RNA shield (Zymo Research, Irvine, CA, USA) at -20 °C once on board. Additionally, plankton was sampled on 0.2–3 µm and 3–20 µm filters by filtering 100 L of water obtained with a pump near the islands and over the reefs, and with a diver held hose at the surface of the *Pocillopora* colonies. All filters were preserved in cryovials in liquid nitrogen. Because some corals within the same genus can be difficult to differentiate by eye, results are aggregated at genus level: *Pocillopora*, *Porites* and *Millepora*. Several environmental parameters were measured at the time of collection following the protocols detailed in Lombard et al.<sup>81</sup>. These include temperature and salinity measured with a CTD (Castaway CTD), nutrients quantified back in the laboratory and chlorophyll a (chl a) concentrations. Sample provenance and environmental context are available on Zenodo<sup>82</sup>.

### DNA extraction and sequencing

DNA was extracted with commercial kits after mechanical cell disruption. The detailed protocols for corals and plankton are presented in Belser et al.<sup>83</sup>. Metabarcoding sequencing targeted all 10 colonies taken at

each site for each coral genus at the 32 islands. The prokaryote 16S rRNA gene was first amplified with a nested PCR approach with a first full-length amplification using the 27F/1492R 16S rRNA prokaryotic primer set<sup>29,84</sup> in order to increase the target prokaryotic DNA, and a second amplification using the 515F-Y/926R primers<sup>85</sup>. The Symbiodiniaceae internal transcribed spacer (ITS2) was amplified with primers SYM-VAR-5.8S2 / SYM-VAR-REV. The sequencing was performed on an Illumina sequencer to obtain 0.5 to 1 M of ~250 pb 16S rRNA paired-end reads<sup>83</sup>, and ~20,000 of ~250 bp paired-end reads for ITS2. Details are provided in Belser et al.<sup>83</sup>.

Metagenomic sequencing was performed on three coral replicates from each site of a subset of 11 islands (I01-Islas de las Perlas, I02-Coiba, I03-Malpelo, I04-Rapa Nui, I05-Ducie, I06-Gambier, I07-Moorea, I08-Aitutaki, I09-Niue, I10-Upolu, and I15-Guam). The sequencing was performed on a NovaSeq6000 or HiSeq4000 Illumina sequencer (Illumina, San Diego, CA, USA) in order to produce 100 M of ~150 pb paired-end reads per sample. Details are provided in Belser et al.<sup>83</sup>.

All sequencing files were submitted to the European Nucleotide Archive (ENA) at the EMBL European Bioinformatics Institute (EMBL-EBI) under the *Tara* Pacific Umbrella project PRJEB47249. Samples and their metadata were registered in the ENA biosample database. Sample provenance and environmental context are available on *Zenodo*<sup>82</sup>.

### 16S rRNA gene and ITS2 sequence analysis

For 16S rRNA gene analysis<sup>15</sup>, adapters were removed with cutadapt (version 2.8)<sup>86</sup> and an ASV abundance table was built with DADA2 (version 1.14)<sup>87</sup> as detailed in the scripts published in *Zenodo*<sup>88</sup>. Samples were grouped by sequencing lane to learn errors and infer ASVs. ASVs representing less than six inserts were tagged as being spurious and removed. Taxonomic annotation was performed with IDTAXA<sup>89</sup> with a confidence threshold of 40 against the SILVA v.138 database. Eukaryotic ASVs (chloroplast and mitochondria) were identified based on taxonomic annotation following the criteria published in *Zenodo*<sup>88</sup> and removed from the dataset prior to analysis. In addition, bacterial sequences annotated at the family level as Oxalobacteraceae, Comamonadaceae, Cutibacterium and Yersiniaceae were identified as reagent contaminants and removed from the dataset.

Symbiodiniaceae profiles were obtained using SymPortal<sup>90</sup>, which uses the intragenomic diversity of Symbiodiniaceae ITS2 to define ITS2 type profiles based on consistent co-occurrence of intragenomic ITS2 variants across all samples. A hierarchical clustering, using *hclust* function from the R “stats” package with Complete Linkage method, was performed on the unifrac matrix distance of the ITS2 abundance table to obtain Symbiodiniaceae community clusters.

### Host lineage assignment and phenotypic signature

A set of genome-wide unlinked single nucleotide polymorphisms (SNPs), represented in >100 colonies of *Pocillopora* and *Porites*, were identified in the metagenomics dataset from the subset of 11 islands. For *Millepora* only target genes SNP of around 60 colonies were produced. From the analysis of these SNPs<sup>91</sup> we identified five independent genetic lineages in *Pocillopora*, three in *Porites* (nine geographic sub-lineages) and five in *Millepora*. The detailed method for host lineage assignment protocols are presented in ref. 91.

The phenotypic signatures of *Pocillopora* and *Porites* were defined through six biomarkers: the coral protein content, the Symbiodiniaceae content per coral mg of protein biomass<sup>92</sup>, Symbiodiniaceae content per coral surface<sup>92</sup>, protein ubiquitination<sup>93</sup> and carbonylation per coral mg of protein<sup>94</sup> to assess the cellular damages, and total antioxidant content per coral mg of protein to assess the cellular defenses<sup>95</sup>. The detailed protocols are presented in ref. 96.

### 16S rRNA gene data analysis

A total of 2447 metabarcoding samples of corals were analyzed. Prokaryotic and *Endozoicomonadaceae* community diversity were

estimated by computing the Shannon index with the function *diverse* from “diverse” R package<sup>97</sup>. In order to account for the compositional characteristic of the prokaryote abundance table, data were weighted center log ratio transformed (wCLR) before computing statistics. To infer the effect of environmental factors and coral phenotypic biomarkers on *Endozoicomonadaceae* community composition, Kendall’s rank correlation was computed for the 32 islands using *bioenv* function from “vegan” R package<sup>98</sup> and *p*-value were adjusted using Benjamini et Hochberg<sup>99</sup> method.

The correlations between *Endozoicomonadaceae* community composition and environmental factors, phenotypic biomarkers, coral genetic lineage, or Symbiodiniaceae assemblage were tested individually with a mantel test<sup>100</sup> using *mantel* function from “vegan” R package. The matrices were based on Euclidean distance for all datasets except for the genetic lineages that were based on RaxML<sup>101</sup> distances. They were computed for *Pocillopora* and *Porites* on the subset of 11 islands for which data on genetic lineages and phenotypic biomarkers were available.

A variance partitioning analysis was done, using the function *fitExtractVarPartModel* from the “variancePartition” R package<sup>102</sup>, to quantify the contribution of the location (island), the coral genetic lineage, and the Symbiodiniaceae assemblage (community clusters) on *Endozoicomonadaceae* community composition across the subset of 11 islands.

An ASVs distance tree was constructed using 16S rRNA gene sequences. Sequences were aligned with MUSCLE<sup>103</sup>, and the alignment was trimmed manually to remove gaps and get all sequences to start and end at the same position. A distance matrix was computed with the KIMURA model of transition and transversion rates, and the tree constructed based on the Fitch-Margoliash method, based on least squares principle, in PHYLIP<sup>104</sup>.

### Metagenomic assemblies and binning

A total of 270 metagenomes (101 *Pocillopora*, 108 *Porites* and 61 *Millepora*) were analyzed. Sequencing reads were quality filtered using BMap (version 38.71)<sup>105</sup> by removing sequencing adapters from the reads, removing reads that mapped to quality control sequences (PhiX genome) and discarding low-quality reads using the parameters trimq=14, maq=20, maxns=0 and minlength=45. Downstream analyses were performed on quality-controlled reads, or, if specified, merged quality-controlled reads (bbmerge.sh minoverlap=16). Quality-controlled reads were normalized (bbnorm.sh target=40, mindepth=0) before they were assembled with metaSPAdes (version 3.12 or version 3.13 if required)<sup>106</sup>. The resulting scaffolded contigs (hereafter scaffolds) were finally filtered by length ( $\geq 1$  kbp). The quality-controlled metagenomic reads from all samples were individually mapped against the scaffolds of each sample (270 \* 270 mapping) using the Burrows-Wheeler-Algorithm (BWA)(version 0.7.17-r1188)<sup>107</sup>. Alignments were filtered to be at least 45 bases in length, with an identity of  $\geq 97\%$  and covering  $\geq 80\%$  of the read sequence. The resulting BAM files were processed using the *jgi\_summarize\_bam\_contig\_depths* script of MetaBAT2 (version 2.152.1)<sup>108</sup> to provide within- and between-sample coverages for each scaffold.v5. The scaffolds were finally clustered into bins for each metagenomic sample with the automatic algorithm MetaBAT2 (version 2.15) using parameters *-minContig -2000* and *maxEdges 500* to increase sensitivity. The quality of each metagenomic bin was evaluated using both the ‘lineage workflow’ of CheckM (version 1.1.1)<sup>109</sup> and Anvi’o (version 6.1)<sup>110</sup>. Metagenomic bins for which either CheckM/Anvi’o reported a completeness/completion  $\geq 30\%$  and a contamination/redundancy  $\leq 10\%$  were taxonomically annotated using GTDB-Tk (version 1.0.2)<sup>111</sup> with default parameters against the GTDB r89 release<sup>112</sup> in order to select *Endozoicomonadaceae* bacteria. The 27 identified bins are available on *Zenodo* (<https://zenodo.org/record/7840163>).

### Species-level profiling with mOTUs and mTAGs

Metagenomic bins that passed the quality filtering were additionally added to the database (version 2.5.1) of the metagenomic profiling tool mOTUs<sup>113</sup> to generate an extended mOTUs reference database, following the approach of Paoli et al.<sup>114</sup>. Only genomes with at least six out of the ten universal single-copy marker genes as identified by fetchMGs (version 1.1)<sup>115</sup> were used. The species-level profiling was then computed by mapping quality-controlled metagenomic reads against this extended database using default parameters of mOTUs (version 2).

Taxonomic profiling using degenerate consensus reference sequences of ribosomal RNA genes was performed with mTAGs (version 1.0)<sup>116</sup> on metagenomic reads samples.

### Selection of *Endozoicomonadaceae* from databases

All available *Endozoicomonadaceae* genomes (24) and closely related (two *Zooshikella*) were downloaded from the NCBI genomes database (last accessed November 2021). Additionally, five *Endozoicomonas* genome assemblies from Neave et al. 2017<sup>23</sup> were recovered from the authors and two putative *Endozoicomonadaceae* genomes (CAR1\_bin20 and CAR3\_bin23) assembled in Robbins et al.<sup>117</sup> reobtained from [https://data.ace.uq.edu.au/public/sponge\\_mags](https://data.ace.uq.edu.au/public/sponge_mags). External genomes were taxonomically checked using GTDB-Tk as described for the metagenomic bins.

### Quality evaluations of *Endozoicomonadaceae* bins and genomes

The quality of each *Endozoicomonadaceae* bins and added genomes from the literature was assessed using the 'lineage workflow' of CheckM (version 1.1.1)<sup>109</sup> and Anvi'o (version 6.1)<sup>110</sup>. *Endozoicomonadaceae* bins with a completeness  $\geq 50\%$  and a contamination  $\leq 10\%$  were kept for downstream analysis and will be referred to as Metagenome Assembled Genomes (MAGs) or genomes.

The three new *Endozoicomonadaceae* species described in the present study (*Endozoicomonas pocilloporae*, *Parendozoicomonas poriteae* and *Endozoicomonas milleporae*) have been deposited in the SeqCode registry<sup>118</sup>.

### Genes calling and annotation

Gene calling was performed using Prodigal (version 2.6.3)<sup>119</sup> and ribosomal RNAs were extracted using HMMER (version 3.3.2)<sup>120</sup>. These steps were performed with Anvi'o (version 6.1) by following the procedure described here: <https://merenlab.org/2016/06/22/anvio-tutorial-v2>. The predicted proteins recovered from *Endozoicomonadaceae* genomes were annotated by identifying protein domain family with HMMER searches (<http://hmmer.org/>) against the PfamA database (version 33.1)<sup>121</sup> and by assigning orthologous groups (OGs) using eggNOG-mapper (version 2.0.1)<sup>122,123</sup> based on eggNOG database (version 5.0)<sup>124</sup>. KEGG<sup>125</sup> and InterProScan<sup>126</sup> annotation was inferred from the OGs annotation. Based on this annotation, completion of metabolic modules/pathways was accessed using the information available from KEGG Metabolic pathways<sup>125</sup> and MetaCyc Metabolic Pathway Database<sup>127</sup>. Annotation of the predicted proteins was also performed using the HMMs from dbCAN (Version 11) to identify Carbohydrate-Active Enzymes (CAZy)<sup>128</sup>.

### Comparative genomic analysis

The phylogenomic tree of *Endozoicomonadaceae* was inferred using a concatenated alignment of 71 bacterial single-copy marker proteins as detailed by the Anvi'o (<https://anvio.org/>). The tree included MAGs from this study as well as reference genomes from earlier studies that had  $\geq 61$  single-copy markers. The tree was visualized and annotated using the iTOL (version 6.4.1)<sup>129</sup>. Average nucleotide identity (ANI) between *Endozoicomonadaceae* genomes were calculated using PyANI implemented in Anvi'o (version 6.1). Pairwise amino acid identity (AAI) between genomes was computed using *comparem* tool (*similarity* and

*aa1* command) (<https://github.com/dparks1134/CompareM>). In order to provide a species-level clustering of genomes, dRep (version 3.1.1)<sup>130</sup> with a 95% ANI threshold<sup>131</sup> and Spect<sup>132</sup> reused on all genomes. *Endozoicomonadaceae* MAGs within the same genome cluster were grouped into subclades based on mean ANI value  $\geq 98\%$  and host species lineage isolation source. Newly identified *Endozoicomonadaceae* species were named according to the host genus and subclade according to the genetic lineage of the host.

### Coral genomes

Coral genomes assembly and annotation protocols are detailed in Noel et al.<sup>133</sup>. Briefly, *Pocillopora cf. effusa*, *Porites lobata* colonies were collected at Moorea, French Polynesia. An amount of 15 g of each coral colony was stored at  $-80\text{ }^{\circ}\text{C}$ , after flash freezing using liquid nitrogen, until DNA extraction. The genomes were sequenced using a combination of long (Oxford Nanopore Technologies, ONT) and short reads (Illumina). To generate long-reads based genome assemblies four different assemblers were used: Smartdenovo<sup>134</sup>, Redbean<sup>135</sup>, Flye<sup>136</sup>, and Ra<sup>137</sup>. From the assemblies genes calling was performed and were functionally assigned by aligning them against the nr database<sup>138</sup> (NCBI) using diamond<sup>139</sup> (version 0.9.24).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Sample provenance and environmental context are available on Zenodo<sup>82</sup>. *Endozoicomonadaceae* MAGs are available on Zenodo (<https://zenodo.org/record/7840163>). Samples and their metadata were registered in the ENA biosample database. All sequencing files were submitted to the European Nucleotide Archive (ENA) at the EMBL European Bioinformatics Institute (EMBL-EBI) under the Tara Pacific Umbrella BioProject PRJEB47249. All other data supporting the findings of this study are provided in the Supplementary Information or the Source Data. Source data are provided with this paper.

### References

1. Costanza, R. et al. Twenty years of ecosystem services: How far have we come and how far do we still need to go? *Ecosyst. Serv.* **28**, 1–16 (2017).
2. Descombes, P. et al. Forecasted coral reef decline in marine biodiversity hotspots under climate change. *Glob. Change Biol.* **21**, 2479–2487 (2015).
3. Hughes, T. P. et al. Global warming and recurrent mass bleaching of corals. *Nature* **543**, 373–377 (2017).
4. Eddy, T. D. et al. Global decline in capacity of coral reefs to provide ecosystem services. *One Earth* **4**, 1278–1285 (2021).
5. Knowlton, N. et al. *Rebuilding Coral Reefs: A Decadal Grand Challenge*. <https://repository.kaust.edu.sa/handle/10754/670295> <https://doi.org/10.53642/nrky9386> (2021).
6. Thompson, J. R., Rivera, H. E., Closek, C. J. & Medina, M. Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. *Front. Cell. Infect. Microbiol.* **4**, 176 (2015).
7. Bourne, D. G., Morrow, K. M. & Webster, N. S. Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annu. Rev. Microbiol.* **70**, 317–340 (2016).
8. Webster, N. S. & Reusch, T. B. H. Microbial contributions to the persistence of coral reefs. *ISME J.* **11**, 2167 (2017).
9. Voolstra, C. R. et al. Extending the natural adaptive capacity of coral holobionts. *Nat. Rev. Earth Environ.* **2**, 747–762 (2021).
10. Oppen, M. J. Hvan & Blackall, L. L. Coral microbiome dynamics, functions and design in a changing world. *Nat. Rev. Microbiol.* **17**, 557–567 (2019).

11. LaJeunesse, T. C. et al. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Curr. Biol.* **28**, 2570–2580.e6 (2018).
12. Robbins, S. J. et al. A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. *Nat. Microbiol.* **4**, 2090–2100 (2019).
13. Vanwonderghem, I. & Webster, N. S. Coral Reef Microorganisms in a Changing Climate. *iScience* **23**, 100972 (2020).
14. McDevitt-Irwin, J. M., Baum, J. K., Garren, M. & Vega Thurber, R. L. Responses of coral-associated bacterial communities to local and global stressors. *Front. Mar. Sci.* **4**, <https://doi.org/10.3389/fmars.2017.00262> (2017).
15. Galand, P. E. et al. Diversity of the Pacific Ocean coral reef microbiome. *Nat. Commun.* <https://doi.org/10.1038/s41467-023-38500-x> (2023) (in press).
16. Neave, M. J., Apprill, A., Ferrier-Pagès, C. & Voolstra, C. R. Diversity and function of prevalent symbiotic marine bacteria in the genus *Endozoicomonas*. *Appl. Microbiol. Biotechnol.* **100**, 8315–8324 (2016).
17. Appolinario, L. R. et al. Description of *Endozoicomonas arenosclerae* sp. nov. using a genomic taxonomy approach. *Antonie Van Leeuwenhoek* **109**, 431–438 (2016).
18. Nishijima, M., Adachi, K., Katsuta, A., Shizuri, Y. & Yamasato, K. 2013. *Endozoicomonas numazuensis* sp. nov., a gammaproteobacterium isolated from marine sponges, and emended description of the genus *Endozoicomonas* Kurahashi and Yokota 2007. *Int. J. Syst. Evol. Microbiol.* **63**, 709–714 (2016).
19. Hyun, D. W. et al. *Endozoicomonas atrinae* sp. nov., isolated from the intestine of a comb pen shell *Atrina pectinata*. *Int. J. Syst. Evol. Microbiol.* **64**, 2312–2318 (2014).
20. Schreiber, L., Kjeldsen, K. U., Obst, M., Funch, P. & Schramm, A. Description of *Endozoicomonas ascidiicola* sp. nov., isolated from Scandinavian ascidians. *Syst. Appl. Microbiol.* **39**, 313–318 (2016).
21. Bayer, T. et al. Bacteria of the genus *Endozoicomonas* dominate the microbiome of the Mediterranean gorgonian coral *Eunicella cavolini*. *Mar. Ecol. Prog. Ser.* **479**, 75–84 (2013).
22. Weber, L., Gonzalez-Díaz, P., Armenteros, M. & Apprill, A. The coral ecosphere: a unique coral reef habitat that fosters coral–microbial interactions. *Limnol. Oceanogr.* <https://doi.org/10.1002/lno.11190> (2019).
23. Neave, M. J., Michell, C. T., Apprill, A. & Voolstra, C. R. *Endozoicomonas* genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Sci. Rep.* **7**, 4057 (2017).
24. Tandon, K. et al. Comparative genomics: dominant coral-bacterium *Endozoicomonas acroporae* metabolizes dimethylsulfoniopropionate (DMSP). *ISME J.* **14**, 1290–1303 (2020).
25. Bartz, J. O. et al. *Parendozoicomonas haliclona* gen. nov. sp. nov. isolated from a marine sponge of the genus *Haliclona* and description of the family *Endozoicomonadaceae* fam. nov. comprising the genera *Endozoicomonas*, *Parendozoicomonas*, and *Kistimonas*. *Syst. Appl. Microbiol.* **41**, 73–84 (2018).
26. Choi, E. J., Kwon, H. C., Sohn, Y. C. & Yang, H. O. *Kistimonas asteriae* gen. nov., sp. nov., a gammaproteobacterium isolated from *Asterias amurensis*. *Int. J. Syst. Evol. Microbiol.* **60**, 938–943 (2010).
27. Goldberg, S. R., Haltli, B. A., Correa, H. & Kerr, R. G. Description of *sansalvadorimonas verongulae* gen. nov., sp. nov., a gammaproteobacterium isolated from the marine sponge *Verongula gigantea*. *Int. J. Syst. Evol. Microbiol.* **68**, 2006–2014 (2018).
28. Kurahashi, M. & Yokota, A. *Endozoicomonas elysicola* gen. nov., sp. nov., a  $\gamma$ -proteobacterium isolated from the sea slug *Elysia ornata*. *Syst. Appl. Microbiol.* **30**, 202–206 (2007).
29. Rohwer, F., Seguritan, V., Azam, F. & Knowlton, N. Diversity and distribution of coral-associated bacteria. *Mar. Ecol. Prog. Ser.* **243**, 1–10 (2002).
30. Bourne, D. G. & Munn, C. B. Diversity of bacteria associated with the coral *Pocillopora damicornis* from the Great Barrier Reef. *Environ. Microbiol.* **7**, 1162–1174 (2005).
31. Bourne, D., Iida, Y., Uthicke, S. & Smith-Keune, C. Changes in coral-associated microbial communities during a bleaching event. *ISME J.* **2**, 350–363 (2008).
32. Sunagawa, S. et al. Bacterial diversity and white Plague disease-associated community changes in the Caribbean coral *Montastraea faveolata*. *ISME J.* **3**, 512–521 (2009).
33. Kvennefors, E. C. E. et al. Regulation of bacterial communities through antimicrobial activity by the coral holobiont. *Microb. Ecol.* **63**, 605–618 (2012).
34. Morrow, K. M., Moss, A. G., Chadwick, N. E. & Liles, M. R. Bacterial associates of two Caribbean coral species reveal species-specific distribution and geographic variability. *Appl. Environ. Microbiol.* **78**, 6438–6449 (2012).
35. Speck, M. D. & Donachie, S. P. Widespread oceanospirillaceae bacteria in *Porites* spp. *J. Mar. Biol.* **2012**, e746720 (2012).
36. Bayer, T. et al. The microbiome of the Red Sea coral *Stylophora pistillata* is dominated by tissue-associated *Endozoicomonas* bacteria. *Appl. Environ. Microbiol.* **79**, 4759–4762 (2013).
37. Jessen, C. et al. In-situ effects of eutrophication and overfishing on physiology and bacterial diversity of the Red Sea coral *Acropora hemprichii*. *PLoS ONE* **8**, e62091 (2013).
38. Rodriguez-Lanetty, M., Granados-Cifuentes, C., Barberan, A., Bellantuono, A. J. & Bastidas, C. Ecological Inferences from a deep screening of the Complex Bacterial Consortia associated with the coral, *Porites astreoides*. *Mol. Ecol.* **22**, 4349–4362 (2013).
39. Ziegler, M. et al. Coral microbial community dynamics in response to anthropogenic impacts near a major city in the central Red Sea. *Mar. Pollut. Bull.* **105**, 629–640 (2016).
40. Brener-Raffalli, K. et al. Thermal regime and host clade, rather than geography, drive Symbiodinium and bacterial assemblages in the scleractinian coral *Pocillopora damicornis sensu lato*. *Microbiome* **6**, 39 (2018).
41. Pollock, F. J. et al. Coral-associated bacteria demonstrate phyllosymbiosis and cophylogeny. *Nat. Commun.* **9**, 4921 (2018).
42. Maher, R. L. et al. Coral microbiomes demonstrate flexibility and resilience through a reduction in community diversity following a thermal stress event. *Front. Ecol. Evol.* **8**, 356–356 (2020).
43. Neave, M. J. et al. Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales. *ISME J.* **11**, 186–200 (2017).
44. Roder, C., Bayer, T., Aranda, M., Kruse, M. & Voolstra, C. R. Microbiome structure of the fungid coral *Ctenactis echinata* aligns with environmental differences. *Mol. Ecol.* **24**, 3501–3511 (2015).
45. Bourne, D. G. et al. Coral reef invertebrate microbiomes correlate with the presence of photosymbionts. *ISME J.* **7**, 1452–1458 (2013).
46. Shiu, J.-H. et al. Shifting in the dominant bacterial group *Endozoicomonas* is independent of the dissociation with coral symbiotic algae. *Front. Microbiol.* **11**, 1791–1791 (2020).
47. Dubé, C. E. et al. Naturally occurring fire coral clones demonstrate a genetic and environmental basis of microbiome composition. *Nat. Commun.* **12**, 1–12 (2021).
48. Morrow, K. M. et al. Natural volcanic CO<sub>2</sub> seeps reveal future trajectories for host-microbial associations in corals and sponges. *ISME J.* **9**, 894–908 (2015).
49. Webster, N. S. et al. Host-associated coral reef microbes respond to the cumulative pressures of ocean warming and ocean acidification. *Sci. Rep.* **6**, 19324 (2016).
50. Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R. & Voolstra, C. R. Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nat. Commun.* **8**, 14213–14213 (2017).

51. Ziegler, M. et al. Coral bacterial community structure responds to environmental change in a host-specific manner. *Nat. Commun.* **10**, 1–11 (2019).
52. Meyer, J. L., Paul, V. J. & Teplitski, M. Community shifts in the surface microbiomes of the coral *Porites astreoides* with unusual lesions. *PLoS ONE* **9**, e100316 (2014).
53. Shore, A., Day, R. D., Stewart, J. A. & Burge, C. A. Dichotomy between regulation of coral bacterial communities and calcification physiology under ocean acidification conditions. *Appl. Environ. Microbiol.* **87**, e02189–20 (2021).
54. Reigel, A. M., Paz-García, D. A. & Hellberg, M. E. Microbiome of a reef-building coral displays signs of acclimation to a stressful shallow hydrothermal vent habitat. *Front. Mar. Sci.* **8**, <https://doi.org/10.3389/fmars.2021.652633> (2021).
55. Pogoreutz, C. et al. Dominance of Endozoicomonas bacteria throughout coral bleaching and mortality suggests structural inflexibility of the Pocillopora verrucosa microbiome. *Ecol. Evol.* **8**, 2240–2252 (2018).
56. Pogoreutz, C. et al. Coral holobiont cues prime Endozoicomonas for a symbiotic lifestyle. *ISME J.* <https://doi.org/10.1038/s41396-022-01226-7> (2022).
57. Ding, J.-Y., Shiu, J.-H., Chen, W.-M., Chiang, Y.-R. & Tang, S.-L. Genomic Insight into the host–endosymbiont relationship of endozoicomonas montiporae CL-33T with its Coral Host. *Front. Microbiol.* **7**, 251 (2016).
58. Sweet, M. et al. Insights into the cultured bacterial fraction of corals. *mSystems* **6**, e01249–20 (2021).
59. Rua, C. P. J. et al. Diversity and antimicrobial potential of culturable heterotrophic bacteria associated with the endemic marine sponge *Arenosclera brasiliensis*. *PeerJ* **2**, e419 (2014).
60. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
61. Séré, M. G. et al. Bacterial communities associated with *Porites* White Patch Syndrome (PWPS) on Three Western Indian Ocean (WIO) Coral Reefs. *PLoS ONE* **8**, e83746–e83746 (2013).
62. Maher, R. L., Rice, M. M., McMinds, R., Burkepile, D. E. & Vega Thurber, R. Multiple stressors interact primarily through antagonism to drive changes in the coral microbiome. *Sci. Rep.* **9**, 6834 (2019).
63. Gao, C. et al. Coral mucus rapidly induces chemokinesis and genome-wide transcriptional shifts toward early pathogenesis in a bacterial coral pathogen. *ISME J.* **15**, 3668–3682 (2021).
64. Raina, J.-B. et al. Chemotaxis shapes the microscale organization of the ocean’s microbiome. *Nature* **605**, 132–138 (2022).
65. Apprill, A., Marlow, H. Q., Martindale, M. Q. & Rappé, M. S. The onset of microbial associations in the coral *Pocillopora meandrina*. *ISME J.* **3**, 685–699 (2009).
66. Apprill, A., Marlow, H. Q., Martindale, M. Q. & Rappé, M. S. Specificity of Associations between Bacteria and the Coral *Pocillopora meandrina* during Early Development. *Appl. Environ. Microbiol.* **78**, 7467–7475 (2012).
67. Smith, H. A., Moya, A., Cantin, N. E., van Oppen, M. J. H. & Torda, G. Observations of simultaneous sperm release and larval planulation suggest reproductive assurance in the coral *pocillopora acuta*. *Front. Mar. Sci.* **6**, <https://doi.org/10.3389/fmars.2019.00362> (2019).
68. Damjanovic, K., Menéndez, P., Blackall, L. L. & van Oppen, M. J. H. Early life stages of a common broadcast spawning coral associate with specific bacterial communities despite lack of internalized bacteria. *Microb. Ecol.* **79**, 706–719 (2020).
69. Coulthurst, S. 2019. The Type VI secretion system: a versatile bacterial weapon. *Microbiology* **165**, 503–515 (2019).
70. Speare, L. et al. Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proc. Natl Acad. Sci. USA* **115**, E8528–E8537 (2018).
71. Tandon, K. et al. Microbiome restructuring: dominant coral bacterium *Endozoicomonas* species display differential adaptive capabilities to environmental changes. *bioRxiv* <https://doi.org/10.1101/2021.10.31.466697> (2021).
72. Craig, L., Forest, K. T. & Maier, B. Type IV pili: dynamics, biophysics and functional consequences. *Nat. Rev. Microbiol.* **17**, 429–440 (2019).
73. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90–93 (2005).
74. Tang, Y. Z., Koch, F. & Gobler, C. J. Most harmful algal bloom species are vitamin B1 and B12 auxotrophs. *Proc. Natl Acad. Sci. USA* **107**, 20756–20761 (2010).
75. Keller-Costa, T. et al. Metagenomics-resolved genomics provides novel insights into chitin turnover, metabolic specialization, and niche partitioning in the octocoral microbiome. *Microbiome* **10**, 151 (2022).
76. Wang, L. H. et al. Different strategies of energy storage in cultured and freshly isolated *Symbiodinium* sp. *J. Phycol.* **51**, 1127–1136 (2015).
77. Messyasz, A., Maher, R. L., Meiling, S. S. & Thurber, R. V. Nutrient enrichment predominantly affects low diversity microbiomes in a marine trophic symbiosis between algal farming fish and corals. *Microorganisms* **9**, 1873 (2021).
78. Lighthart, K., Belzer, C., de Vos, W. M. & Tytgat, H. L. P. Bridging bacteria and the gut: functional aspects of type IV Pili. *Trends Microbiol.* **28**, 340–348 (2020).
79. Woo, S. et al. Geographical variations in bacterial communities associated with soft coral *Scleronephthya gracillimum*. *PLoS ONE* **12**, e0183663–e0183663 (2017).
80. Planes, S. et al. The Tara Pacific expedition—A pan-ecosystemic approach of the “-omics” complexity of coral reef holobionts across the Pacific Ocean. *PLoS Biol.* **17**, e3000483 (2019).
81. Lombard, F. et al. Open science resources from the Tara Pacific expedition across coral reef and surface ocean ecosystems. *Sci. Data.* <https://doi.org/10.1038/s41597-022-01757-w> (2023).
82. Pesant, S. et al. *Tara Pacific samples provenance and environmental context-version 1.* <https://doi.org/10.5281/zenodo.4068293> (2020).
83. Belser, C. et al. Integrative omics framework for characterization of coral reef ecosystems from the Tara Pacific expedition. *Sci. Data.* <https://doi.org/10.1038/s41597-023-02204-0> (2023).
84. Amann, R. L., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
85. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
86. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
87. Callahan, B. J. et al. DADA2: High resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
88. Ruscheweyh, H.-J. et al. Tara Pacific 16S rRNA data analysis release. <https://doi.org/10.5281/zenodo.6598864> (2022).
89. Murali, A., Bhargava, A. & Wright, E. S. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. *Microbiome* **6**, 140 (2018).
90. Hume, B. C. C. et al. SymPortal: a novel analytical framework and platform for coral algal symbiont next-generation sequencing ITS2 profiling. *Mol. Ecol. Resour.* **19**, 1063–1080 (2019).

91. Hume, B. C. C. et al. Disparate patterns of genetic divergence in three widespread corals across a pan-Pacific environmental gradient highlights species-specific adaptation trajectories. In *bioRxiv* (p. 2022.10.13.512013). <https://doi.org/10.1101/2022.10.13.512013> (2022).
92. Zamoum, T. & Furla, P. Symbiodinium isolation by NaOH treatment. *J. Exp. Biol.* **215**, 3875–3880 (2012).
93. Haas, A. L. & Bright, P. M. The immunochemical detection and quantitation of intracellular ubiquitin-protein conjugates. *J. Biol. Chem.* **260**, 12464–12473 (1985).
94. Buss, D. M. & Shackelford, T. K. From vigilance to violence: Mate retention tactics in married couples. *J. Pers. Soc. Psychol.* **72**, 346–361 (1997).
95. Naguib, Y. M. A. Antioxidant activities of astaxanthin and related carotenoids. *J. Agric. Food Chem.* **48**, 1150–1154 (2000).
96. Porro, B. et al. The Oak and the Reed: Two Environmental Response Strategies in Corals from Pacific Islands. *Commun. Earth & Environ.* (Under Review).
97. Guevara, M. R., Hartmann, D. & Mendoza, M. diverse: an R package to measure diversity in complex systems. *R. J.* **8**, 60–78 (2016).
98. Oksanen, J. et al. Vegan: community ecology package. *R. Package Version 22-1* **2**, 1–2 (2015).
99. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
100. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
101. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
102. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinform.* **17**, 483 (2016).
103. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792–1797 (2004).
104. Felsenstein, J. PHYLIP-Phylogeny inference package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
105. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. (U.S. Department of Energy Office of Scientific and Technical Information, 2014).
106. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
107. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
108. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
109. 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).
110. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
111. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
112. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
113. Milanese, A. et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1–11 (2019).
114. Paoli, L. et al. Biosynthetic potential of the global ocean microbiome. *Nature* **607**, 111–118 (2022).
115. Sunagawa, S. et al. Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* **10**, 1196–1199 (2013).
116. Salazar, G., Ruscheweyh, H.-J., Hildebrand, F., Acinas, S. G. & Sunagawa, S. mTAGs: taxonomic profiling using degenerate consensus reference sequences of ribosomal RNA genes. *Bioinformatics* **38**, 270–272 (2022).
117. Robbins, S. J. et al. A genomic view of the microbiome of coral reef demosponges. *ISME J.* **15**, 1641–1654 (2021).
118. Hedlund, B. P. et al. SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* **7**, 1702–1708 (2022).
119. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
120. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
121. Mistry, J. et al. Pfam: The protein families database in 2021. *Nucl. Acids Res.* **49**, D412–D419 (2021).
122. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
123. Cantalapiedra, J. L. et al. The rise and fall of proboscidean ecological diversity. *Nat. Ecol. Evol.* **5**, 1266–1272 (2021).
124. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucl. Acids Res.* **47**, D309–D314 (2019).
125. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30 (2000).
126. Apweiler, R. et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucl. Acids Res.* **29**, 37–40 (2001).
127. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucl. Acids Res.* **42**, D459–D471 (2014).
128. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucl. Acids Res.* **42**, D490–D495 (2014).
129. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucl. Acids Res.* **49**, W293–W296 (2021).
130. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
131. Olm, M. R. et al. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. *mSystems* **5**, e00731–19 (2020).
132. Mende, D. R., Sunagawa, S., Zeller, G. & Bork, P. Accurate and universal delineation of prokaryotic species. *Nat. Methods* **10**, 881–884 (2013).
133. Noel, B. et al. Pervasive tandem duplications and convergent evolution shape coral genomes. *Genome Biology* <https://doi.org/10.1186/s13059-023-02960-7> (2023).
134. Liu, H. et al. SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte* **2021**, 1–9 (2021).
135. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
136. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
137. Vaser, R. & Šikić, M. Yet another de novo genome assembler. 656306 Preprint at <https://doi.org/10.1101/656306> (2019).
138. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res.* **44**, D7–D19 (2016).



139. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

## Acknowledgements

Special thanks to the Tara Ocean Foundation, the R/V Tara crew, and the Tara Pacific Expedition Participants (<https://doi.org/10.5281/zenodo.3777760>). We are keen to thank the commitment of the following institutions for their financial and scientific support that made this unique 3-year Tara Pacific Expedition possible: CNRS, PSL, CSM, EPHE, Genoscope, CEA, Inserm, Université Côte d'Azur, ANR, agnès b., the Veolia Foundation, the Prince Albert II de Monaco Foundation, Région Bretagne, Lorient Agglomération, L'Oréal, Biotherm, Billerudkorsnas, AmerisourceBergen Company, Oceans by Disney, France Collectivités, Fonds Français pour l'Environnement Mondial (FFEM), UNESCO-IOC, Etienne Bourgois, and the Tara Ocean Foundation teams. P.E.G. acknowledges the support of Agence Nationale de la Recherche (ANR) through the CORALMATES project (ANR-18-CE02-0009), and S.S. acknowledges the support of the Swiss National Science Foundation grant 205321\_184955. C.H. is grateful to the genotoul bioinformatics platform, Toulouse, Occitanie (Bioinfo Genotoul) for providing computing resources (<https://doi.org/10.15454/1.5572369328961167E12>). Tara Pacific would not exist without the continuous support of the participating institutes. The authors also particularly thank Serge Planes, Denis Allemand, and the Tara Pacific consortium. This is publication number #24 of the Tara Pacific Consortium. Sampling permits are provided as supplementary material. CdV, FL and SP, are funded by the European Union's Horizon 2020 research and innovation programme "Atlantic Ecosystems Assessment, Forecasting and Sustainability" (AtlantECO) under grant agreement No 862923. This study was supported in part by FRANCE GENOMIQUE (ANR-10-INBS-09).

## Author contributions

C.H. and P.E.G. designed the experiments, analyzed the data, and wrote the manuscript; L.P., H.J.R., and G.S. conducted the bioinformatic analyses of the raw sequence data and constructed MAGs; M.Z., C.P., C.R.V., D.A.P.G., E. Boissin, L.P. provided useful comments and manuscript edition. C.H., L.P., H.J.R., G.S., E. Boissin, S. Romac, J.P., G.B., G.I., C.M., M.Z., B.P., E.J.A., B.C.C.H., J.M.A., C.P., D.A.P.G., M.M.N., S.A., B.B., E. Boss, C.B., C.d.V., E.D., M.F., D.F., P.F., E.G., F.L., S. Pesant, S. Reynaud,

O.P.T., R.T., P.W., D.Z., D.A., S. Planes, R.V.T., C.R.V., S.S., and P.E.G. reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-38502-9>.

**Correspondence** and requests for materials should be addressed to Pierre E. Galand.

**Peer review information** *Nature Communications* thanks Linda Blackall and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Sorbonne Université, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques (LECOB), Observatoire Océanologique de Banyuls, 66650 Banyuls sur Mer, France. <sup>2</sup>Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zürich, 8093 Zürich, Switzerland. <sup>3</sup>PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Laboratoire d'Excellence CORAIL, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan, Cedex, France. <sup>4</sup>Sorbonne Université, CNRS, Station Biologique de Roscoff, AD2M, UMR 7144, ECOMAP, Roscoff, France. <sup>5</sup>Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France. <sup>6</sup>Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara GOSEE, 75000 Paris, France. <sup>7</sup>School of Marine Sciences, University of Maine, Orono, ME, USA. <sup>8</sup>Fondation Tara Océan, 8 rue de Prague, 75012 Paris, France. <sup>9</sup>Department of Animal Ecology & Systematics, Justus Liebig University Giessen, Heinrich-Buff-Ring 26–32 (IFZ), 35392 Giessen, Germany. <sup>10</sup>CNRS, INSERM, Institute for Research on Cancer and Aging (IRCAN), Université Côte d'Azur, Nice, France. <sup>11</sup>Laboratoire International Associé Université Côte d'Azur-Centre Scientifique de Monaco (LIA ROPSE), Monaco, Principality of Monaco. <sup>12</sup>Department of Biology, University of Konstanz, 78457 Konstanz, Germany. <sup>13</sup>Centro de Investigaciones Biológicas del Noroeste (CIBNOR), La Paz, Baja California Sur 23096, México. <sup>14</sup>Shimoda Marine Research Center, University of Tsukuba, 5-10-1, Shimoda, Shizuoka, Japan. <sup>15</sup>School of Marine Sciences, University of Maine, Orono, ME 04469, USA. <sup>16</sup>Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France. <sup>17</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France. <sup>18</sup>Weizmann Institute of Science, Department of Earth and Planetary Sciences, 76100 Rehovot, Israel. <sup>19</sup>Department of Medical Genetics, CHU Nice, Nice, France. <sup>20</sup>Sorbonne Université, Institut de la Mer de Villefranche sur mer, Laboratoire d'Océanographie de Villefranche, 06230 Villefranche-sur-Mer, France. <sup>21</sup>Institut Universitaire de France, 75231 Paris, France. <sup>22</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK. <sup>23</sup>Centre Scientifique de Monaco, Monaco, Principality of Monaco. <sup>24</sup>School of Biological and Chemical Sciences, Ryan Institute, University of Galway, Galway, Ireland. <sup>25</sup>Microbiology Department, Oregon State University, Corvallis, OR, USA. ✉ e-mail: [pierre.galand@obs-banyuls.fr](mailto:pierre.galand@obs-banyuls.fr)