Supplementary Materials for

High-resolution spatial and genomic characterization of coral-associated microbial aggregates in the coral *Stylophora pistillata*

Naohisa Wada *et al.*

Corresponding author: Sen-Lin Tang, sltang@gate.sinica.edu.tw

*Sci. Adv.* 8, eabo2431 (2022)
DOI: 10.1126/sciadv.abo2431

The PDF file includes:

Supplementary Methods
Figs. S1 to S11
Tables S2, S3, S5, S7, and S9
Legends for tables S1, S4, S6, and S8
References

Other Supplementary Material for this manuscript includes the following:

Tables S1, S4, S6, and S8
Supplementary Methods

Sample collections

A total of 16 *Stylopora pistillata* coral colonies were collected from two locations (Kenting, Taiwan 21˚56’20.2” N 120˚44’44.6” E, and Okinawa, Japan 26˚37’42.2” N 127˚51’35.6” E; Fig. S1) in September and October, respectively, 2017. Three nubbins (nubbin: 20 mm length) were cut from each colony and immediately subjected to three different fixation processes; 1) nubbins for histological and genomic analysis were fixed in modified methacarn solution (MM) that consisted of 55% methanol (Macron Fine Chemicals), 30% chloroform (Sigma Aldrich), 10% glacial acetic acid (Sigma Aldrich), and 5% DMSO (Sigma Aldrich) for 6 hours at 4 °C; 2) nubbins for fluorescent in-situ hybridization (FISH) and FISH-NanoSIMS were fixed in 4% paraformaldehyde in 100 mM phosphate-buffered saline (PFA; Wako) for 8 hours at 4 °C followed by rinsing in 70% ethanol (Sigma Aldrich) twice, and storage in 70% ethanol at 4 °C; 3) nubbins for bacterial community analysis of coral tissues were placed in 100% ethanol.

Molecular characterization of bacteria associated with coral tissues and individual CAMAs

Coral tissues (< 0.8 cm area) were removed using a sterile scalpel blade from the skeletons of nubbins stored in the 100% ethanol and genomic DNA extracted with the DNeasy PowerBiofilm kit (Qiagen) according to the manufacturer’s protocol. Genomic DNA was also extracted from individual CAMAs dissected from tissues of MM-fixed nubbins using a laser microdissection microscope system (PALM MicroBeam, Zeiss). The MM-fixed nubbins were first rinsed in 10 mM phosphate-buffered saline (PBS) twice before being decalcified using 14% EDTA (w/v; pH 7.6, Sigma Aldrich), with samples being placed in a new solution every two days until the specimen completely decalcified. Following decalcification, the MM-fixed specimens were rinsed with 10 mM PBS twice, embedded in OCT compound (CRYOMATRIX, Thermo Scientific), and stored at -80 °C until sectioning. The MM-fixed samples were sectioned at 10 μm thickness (with a spacing of >100 μm between each section) using a cryostat (Leica...
CM3050S) with the obtained sections placed on individual slides (MembraneSlide 1.0 PEN, Zeiss) and stored at -20 °C. Sections were subsequently cleared of the OCT compound with nucleotide-free water, stained with SYBR gold (0.2% v/v, Invitrogen) for 10 min, rinsed with water and 100% ethanol sequentially and then air-dried. CAMAs were identified by staining of the ribonucleotide material and the arrangement in tissues under 10x magnification in the laser microdissection microscope system (LMD; PALM MicroBeam, Zeiss). Individual CAMAs were dissected from the tissues and placed into a tube (AdhesiveCup 200 clear, Zeiss) by the “dry collection method”. Coral tissue regions without CAMAs were also collected with the same area of the largest micro-dissected CAMA from each colony at each experiment. Micro-dissected CAMAs and tissues were stored at -20 °C. A total of 67 individual micro-dissected CAMAs from 3 colonies were sampled from Kenting and Okinawa along with 29 micro-dissected control coral tissue regions from colonies from the two locations. DNA was extracted from the LMD collected CAMAs and control coral tissues using the OlAamp DNA Micro Kit (Qiagen) according to the manufacturer’s protocol.

**Direct sanger sequencing of genomic DNA from individual micro-dissected CAMAs**

We performed PCR on the genomic DNA derived from individual micro-dissected CAMAs (n=67) and the control coral tissue regions (n=29) sourced by LMD. The V6-V8 region of the 16S rRNA gene was amplified by PCR using the universal primer set for bacteria (968F: 5’- AAC GCG AAG AAC CTT AC -3’, and 1391R: 5’- ACG GGC GGT GWG TRC -3’) (52, 53). PCR amplifications were carried out in a 30 µl reaction mixture containing 3 µl of genomic DNA, 22.02 µl of dH2O, 0.6 µl of each primer (100 µM), 0.18 µl of Taq DNA Polymerase (5 U/µl; TaKaRa Ex Taq HS, TaKaRa), 3 µl of 10x buffer (TaKaRa Ex Taq HS, TaKaRa), and 2.4 µl of dNTP (each 2.5 mM; TaKaRa Ex Taq HS, TaKaRa). The thermocycling program consisted of an initial denaturation at 94 °C for 5 min, 35 cycles of 94 °C for 30 sec, 52 °C for 20 sec and 72 °C
for 45 sec, and a final extension at 72 °C for 10 min. The amplified products were detected by electrophoresis on a 2% agarose gel containing HealthView nucleic acid stain (Genomics GN-NAS-100). All samples from individual micro-dissected CAMAs were confirmed to produce an amplified PCR product of the correct size while no products were obtained from the control micro-dissected coral tissues.

Amplified products from the CAMAs were excised and purified with a universal gel extraction kit before directly sequencing both DNA strands using the BigDye Terminator V3.1 Cycle Sequencing kit (Applied Biosystems) with the primers 968F and 1391R. The cycle sequencing reactions were carried out in a 10 µl mixture containing 100 ng of the amplified template, 50 ng of the primer, 1 µl of BigDye mix, and 1 µl of 5x reaction buffer. The reaction program consisted of an initial denaturation at 96 °C, 35 cycles of 96 °C for 10 sec, 50 °C for 10 sec and 60 °C for 4 min. The sequencing products were analyzed by an ABI 3730x1 DNA analyzer (Applied Biosystem). The sequences from the DNA strands were assembled and edited to a consensus sequence of each individual micro-dissected CAMAs using the ATSQ software (GENETYX). A Blast search of each consensus sequence was performed against the nr/nt GenBank National Center for Biotechnology Information (NCBI) database to determine the closest matching bacterial 16S rRNA sequence for identification. The consensus sequences from the 67 individuals were aligned using SSU-ALIGN (68) and the pair-wise identities calculated in MEGA7.0.26 (57).

**Bacterial community analysis of individual CAMAs and coral tissues**

DNA derived from four individual randomly selected CAMAs of each colony (n=24 total) and DNA derived from coral tissues (n=16 total) were amplified and subjected to 16S rRNA gene metabarcoding sequencing. Primers 968F and 1391R targeting the V6-V8 hypervariable region included unique barcode sequences for each sample and PCR was conducted using the same...
conditions previously described (see direct Sanger sequencing of genomic DNA from individual micro-dissected CAMAs). The amplified products were sequenced using an Illumina MiSeq platform (Illumina).

Sequences were processed using the USEARCH v11 (54, 55) and MOTHUR v.1.39.5 (56) combination pipelines. Sequences were quality filtered with the following criteria: 1) read lengths between 410 and 450 bp; 2) read quality avg. 27; and 3) homopolymer read length < 8 bp, and then clustered to zero-radius operational taxonomic units (zOTUs) with USEARCH at a threshold of 100% identity and re-clustered into operational taxonomic units (OTUs) at a 99.5% similarity level. The OTUs were also assigned to known taxonomic groups by mapping onto the Silva SSU 138 database (https://www.arb-silva.de/) using MOTHUR with a cut-off value of 80. Sequences obtained from unknown, eukaryotes, and chloroplasts were eliminated (38,260, 53, and 2,257 reads respectively).

*Metagenome-assembled genome analysis*

Genomic DNA of individual micro-dissected CAMAs were pooled from three nubbins sampled from each location, which represented between 8 to 48 CAMAs from each colony. The dominant 16S rRNA sequence from the individual CAMAs was confirmed as identical to the consensus sequence by direct sanger sequencing. Whole Genome Amplification (WGA) was performed on respective samples using the REPLI-g FFPE Kit (Qiagen) without the random ligation step. DNA was purified using MinElute columns following the clean-up protocol in the QIAamp DNA Micro Kit (Qiagen). We again confirmed the 16S rRNA gene sequence in the resulting WGA products was identical to the Sanger-derived consensus sequences. Before pooling the WGA products from each location (n = 84 and 57 individual micro-dissected CAMAs from Kenting and Okinawa, respectively). The mixed WGA products were sequenced on the Illumina HiSeq (pair-end 150 bp) platform following DNA quality checks.
Sequenced reads were quality checked with FastQC (69) and trimmed with trimmomatic v.0.36 (70), using a minimum phred score of 30 (SLIDINGWINDOW:4:30) and MINLEN:30. High-quality trimmed reads were de novo assembled using Megahit v.1.2.9 (71), with kmers (21,29,39,59,79,99,119) and a minimum contig length (min-contig-len) of 500 bp. Assembled contigs were then subjected to the metawrap v.1.3.2 (72) pipeline for binning using Maxbin2 v.2.2.7 (73), Concoct v.1.00.00 (74), and Metabat2 v.2.15 (75), minimum contig length was kept at (-l) 500 bp (metabat2 was kept at 1500 bp). Bins obtained from the three binning algorithms were consolidated via the bin-refinement step of metawrap keeping bins only with minimum completeness (-c) of 70% and contamination (-x) to 10% as estimated from CheckM v.1.1.3 (76).

Subsequently, the reassemble_bins module was used to map reads to refined bins and generate fresh assemblies using SPAdes v3.13.0 (77) using default parameters. Since metawrap reassemble_bins use “permissive” and “strict” algorithms, the “strict” algorithm provided better results in our case based on genome completeness and contamination. Further, an additional curation step for both the bins was incorporated where contigs were taxonomically annotated using CAT/BAT v5.2.3 (78) with default parameters and any non-bacterial contigs were removed. This step removed 22 and 27 contigs from HY_Ok and W5_Kt respectively. After removal of contigs, CheckM v.1.1.3 lineage-specific workflow (lineage_wf) was used to calculate completeness and contamination statistics, no changes were reported. Coverage of individual contigs in a bin was calculated by mapping high-quality trimmed reads using bbmap (79) with default settings. Taxonomic classification of bins was performed using classify workflow (classify_wf) from GTDB-Tk v.1.6.0 (80). Taxonomically annotated bins were then subjected to gene prediction and annotation using PROKKA v.1.14.6 (81), functional annotation for genes was obtained using blastkoala (82) utilizing KEGG orthology (83) and subsystem level classification was obtained using Rapid Annotation Search Tool (RAST) (84) with the ClassicRAST annotation scheme. The predicted protein domain sequences also were searched using the web-based Batch
Conserved Domain-Search (CD-search) tool (85) with the CDD-58235 PSSMs and an e-value threshold of 1e-5 (June 2021).

The full-length 16S rRNA gene sequences were retrieved from the reconstructed MAGs. One sequence with the samples from Kenting fragmented into two sequences which were manually assembled after confirming a 52 bp overlap. The two 16S rRNA genes recovered from the MAGs were aligned to 1289 reference 16S rRNA gene sequences (>1400 bp) assigned to the family Endozoicomonadaceae compromising the genus Endozoicomonas, Kistimonas and Zooshikella in the Silva SSU r138 database (31) using the software infernal v.1.1.3 (86). The multiple alignment was trimmed using trimal tool v.1.4. rev22 setting with “gappyout” (87) and a maximum likelihood tree was constructed in IQ-tree v.1.6.12 (88) using the GTR+G4+F model and node support values were calculated by 1000 ultrafast bootstrap (89). The tree was visualized with interactive Tree of Life (iTOL) v4 (90). We additionally manually obtained meta-information of sequence sources including host association, environment and ocean on all 1289 reference sequences from the NCBI (National Center for Biotechnology Information) and literature reviews.

The average nucleotide identity (ANI) values of our two MAGs were calculated among the 24 available genomes of the family Endozoicomonadaceae and the closest related bacteria (obtained from JGI/IMG) (91) using the online ANI calculator (92). The plot and heatmap of ANI values were generated with the statistical software R v.4.0.2 (63) with the package ggplot2 v.3.3.5 (93).

The overall putative metabolic functions of the reconstructed MAGs were inferred from predicted proteins within RAST based on the SEED subsystem level 1. Shared predicted proteins were compared between the two reconstructed MAGs based on assignments in RAST. Additional predictions of metabolic pathways and secretion systems were based on protein sequences annotated with KEGG orthology (calculated with blastkoala), RAST, and the domains from a CD
search. Effectors in the Type III secretion system and T3SS putative secretion proteins from the two MAGs were identified by EffectiveT3 (94) and annotated with the KEGG BRITE assignment. For analysis of T6SS, we first gleaned 18 available genomes of the family Endozoicomonadaceae within NCBI and predicted protein sequences using PROKKA v.1.14.6 (81). All protein sequences among those genomes were annotated with KEGG orthology using EnrichM v.0.6.4 (https://github.com/geronimp/enrichM) with hmmer and orthologs related to predicted T6SS proteins identified based on the KEGG orthology with presence/absence of predicted protein in the MAGS and available genomes determined. Moreover, a T6SS phylogenetic tree was inferred based on a concatenated alignment of the sheath proteins TssB and TssC, two highly conserved proteins in T6SS. Top 40 Blast Hits for the homology search of TssB and TssC in the MAG HY-Ok bin were retrieved from the nr database using blastp (95). These retrieved sequences were analyzed together with other TssB and TssC protein sequences predicted from other Endozoicomonas genomes. The TssB and TssC sequences were aligned with muscle individually, and the concatenated alignment was trimmed using Gblocks (96). A tree was constructed using IQ-TREE v1.6.12 with 1000 bootstrap replicates (88, 97) and visualized with the interactive Tree of Life (iTOL) v4 (90).

Biofilm- and quorum sensing-associated proteins were collected from the UniProt database (98) based on the annotation category in the QuickGO database (99) and clustered using CD-HIT v.4.8.1 (100) with 70% sequence identity. In total, 2904 biofilm-associated and 4620 quorum sensing-associated bacterial proteins were compiled. We tested the predicted biofilm- and quorum sensing-associated proteins sequences within the two MAGs against the compiled databases using BlastP (95) within e-value cutoff 1e-4. To filter the blast hits, a cutoff of 70% coverage for both the query and the subject protein sequences was used along with a bit score cutoff of 50.
Fluorescent in-situ Hybridization (FISH) characterization of CAMAs

Histological preparation of samples for FISH

The PFA-samples from five colonies sampled from both Kenting and Okinawa were decalcified by Morse’s solution (10% sodium citrate and 22.5% formic acid) (101), with the solution exchanged twice within 3 days at 4 °C before rinsing twice with 10 mM PBS and stored in 70% ethanol. For whole polyp imaging, six decalcified polyps from each colony (n=60 total) were dissected under a stereomicroscope Leica EZ4 (Leica). Dissected PFA fixed polyps were then rinsed twice in PBS, dehydrated sequentially through 70%, 90%, and two changes in abs. 100% ethanol series for 1 hour each before being processed through two changes in Xylene for 3 hours each, and embedded in paraffin.

-Whole-mount three-dimensional (3D) imaging of coral polyps using FISH

Polyps preparations were decolorized and permeabilized as described by (58). Briefly, samples were dehydrated in 50%, 75%, and 90% methanol in 10 mM PBS for 10 min each, and 100% methanol for 300 min. After the dehydration step, the polyps were re-hydrated in 90%, 75%, and 50% methanol in 10 mM PBS containing 0.2% Triton X-100 (PBST solution) for 10 min each. The polyps were washed with PBST solution twice for 10 min and twice with PBS for 10 min each with gentle shaking at 55 rpm.

QuickHCR-FISH of polyps was performed as described by (59). Briefly the polyps were pre-hybridized here time in hybridization buffer (20 mM Tris-HCl [pH 8.0], 0.9 NaCl, 0.01% SDS, 30% formamide, 10% dextran sulfate, 1% blocking reagent) for 15 min at 46 °C. After the pre-hybridization, the polyps were immersed in hybridization buffer with the oligonucleotide probe Non338-initiatorC adapter (see Table S9, final concentration: 0.5 µM) at 46 °C for 1 hour, and washed exceed Non338 probes by hybridization buffer at 46 °C for 15 min thrice. The polyps were then hybridized overnight with oligonucleotide probe EUB338mix-initiatorH adapter (see
Table S9, final concentration: 0.5 µM) at 46 °C. After the incubation, the polyps were washed briefly twice with washing buffer (20 mM Tris-HCl [pH 8.0], 0.112 M NaCl, 5 mM EDTA [pH 8.0], 0.01% SDS) before incubated again in the washing buffer at 48 °C for 1 hour. Subsequently, the polyps were rinsed three times with amplifier buffer (50 mM Na₂HPO₄, 0.9 M NaCl, 0.01% SDS, 10% dextran sulfate, 1% blocking reagent) at 35 °C for 15 min. During the rinsing, the corresponding four amplifier probes labeled with Cy3 (H1 and H2 amplifier probes) and Alexea488 (C1 and C2 amplifier probes) for H and C initiator adapters (see Table S9), respectively, were independently prepared in the amplifier buffer, incubated for 1.5 min at 95 °C and then for 30 min at 25 °C and mixed (final concentration: 0.5 µM of each probe). The amplifier probe mixture was applied to each polyp and the polyps were incubated at 35 °C overnight followed by washing three times with 10 mM PBS at 4 °C. Subsequently, the polyps were immersed in gel solution (10 mM PBS, 4% acrylamide, 0.05% bis-acrylamide) at 4 °C overnight and degassed under vacuum at room temperature for 20 min. Ammonium persulfate (final concentration: 0.1%) and TEMED (final concentration: 1%) were added, gently mixed and incubated at 25 °C for 90 min to allow polymerization of the gel. The samples were washed with PBST solution at room temperature for 10 min and twice with PBS at room temperature for 10 min. For refractive index matching, the samples were soaked in RapiClear 1.47 (RI = 1.47, SUNJin Lab) at 4 °C for 7 hours, exchanged with new RapiClear, and then placed at 4 °C overnight.

Image acquisition was conducted with a 5x magnification objective lens and zoom function 0.8 (resolution: 1.14 µm/pixel) in Lightsheet Z.1 (Zeiss) with ZEN2014 (black edition) software (Zeiss). The samples were immersed with the RapiClear in a sample chamber, excited with the laser at wavelengths 405 nm, 488 nm and 561 nm excitation and acquired images in the emission lengths; >660 nm for autofluorescence of coral and Symbiodiniaceae; 505 – 545 nm for Alexea488; 575 – 615 nm for Cy3. In addition, the samples in one experiment were acquired
under identical conditions of laser power (405 nm: 5%, 488 nm: 5%, 561 nm: 2%), exposure time (299.61 ms for autofluorescence of coral and Symbiodiniaceae; 99.87 ms for Alexea488; 49.94 ms for Cy3), and z-stepping (5 µm interval).

The 3D images were reconstructed with the Imaris software 8.0.2 (BitPlane Inc.) and assessed manually for the number and distribution of CAMAs in the whole coral polyp. For measurement of the volume of each CAMA, we extracted only the single-channel Cy3 signals from each z-stepped image of the polyp, focusing on the region of interest that encompassed each CAMA in the coral tissue, as a subset of the remaining original pixel size in the ZEN2011 (blue edition) software (Zeiss). The subset image then was processed to provide an accurate measurement of the volume of each CAMA in the software ImageJ 1.53d (62). The image was first passaged through a smooth function, converted to black and white pixels using the auto-threshold function with Huang2 (102), Li (103), and Mean (104) algorithms. Each binary z-stepped image of the CAMA was subjected to a function fill-hole in the inside of CAMA with the volume of CAMA measured using the 3D object counter function (size filter used from 100 to 1,000,000). Each result from the three different auto-threshold algorithms for each single CAMA was selected based on the minimum volume to avoid over-estimation.

**3D imaging of bacterial cells within the CAMAs**

The paraffin PFA-sample were sectioned to a 10 µm thickness using a Leica RM2125 RTS (Leica) microtome. Tissue sections were dewaxed through two changes in Xylene for 15 min, dehydrated through two changes in abs. 100% ethanol for 5 min and then dried. The dried sections were immersed with a 20 mM Tris-HCl solution (pH 8.0) at room temperature for 10 min, incubated with a lysozyme solution (2 mg/ml lysozyme in 20 mM Tris-HCl [pH 8.0]) at 37 °C for 30 min to allow bacterial cell wall permeabilization and then rinsed in a 20 mM Tris-HCl solution at room temperature for 10 min. After removal of the solution, quickHCR-FISH was
performed as described by (59). Briefly, sections were pre-hybridized in hybridization buffer (20 mM Tris-HCl [pH 8.0], 0.9 NaCl, 0.01% SDS, 30% formamide, 10% dextran sulfate, 1% blocking reagent) at 46 °C for 10 min and then hybridized with oligonucleotide probe EUB338mix-initiatorH adapter (see Table S9, final concentration: 0.5 µM) in a moist chamber at 46 °C for 2 hours. After the incubation, the sections were washed with washing buffer (20 mM Tris-HCl [pH 8.0], 0.112 M NaCl, 5 mM EDTA [pH 8.0], 0.01% SDS) at 48 °C for 30 min. During the washing step, corresponding two H-amplifier probes labeled with Cy3 (see Table S9) were independently prepared in the amplifier buffer (50 mM Na2HPO4, 0.9 M NaCl, 0.01% SDS, 10% dextran sulfate, 1% blocking reagent), incubated for 1.5 min at 95 °C, then for 30 min at 25 °C and mixed (final concentration: 0.5 µM of each probe). After the washing step, the section was washed with amplifier buffer at 35 °C for 10 min before covering with amplifier buffer and the amplifier probes (final concentration: 0.5 µM) in a moist chamber at 35 °C for 30 min. The amplified section was immediately soaked in cool PBS for 10 min and stained with DAPI (2 µg/ml, PureBlu DAPI Nuclear Staining Dye, Bio-Rad) in a moist chamber overnight at room temperature. The section was then rinsed with PBS briefly, dehydrated through distilled water and 100% ethanol for 5 min each and then air-dried. The section was subsequently mounted with ProLong Diamond Antifade Mountant (Invitrogen).

Images were acquired by scanning sections with a 63x objective lens in z-stack mode (0.1 µm interval) using a confocal laser scanning microscope Zeiss LSM 880 (Zeiss) and Zen SP2.3 software (Zeiss). DAPI signal and autofluorescence signals from both coral and Symbiodiniaceae were obtained by excitation with a 405 nm laser and fluorescence detected in the 410 – 508 nm and 627 – 710 nm emission wavelengths respectively. Cy3 signals was exited with a 561 nm laser and fluorescence signals detected in the 561 – 597 nm emission wavelength range. We confirmed the Cy3 fluorescence signals were from the CAMA structures (see Fig. S10A). Quantitative analysis of bacterial cells within CAMAs was achieved through 3D acquisition with an XY-
resolution of 2048 x 2048 (0.009 µm/pixel) at a depth of 2.5 µm (see Fig. S10B) with the region of interest cropped for the image acquisition (see Fig. S10C below).

For image processing, we extracted the single DAPI channel signal from each z-stack image using ZEN2011 (blue edition) software (Zeiss) (see Fig. S10D). The z-stack DAPI image was subdivided into small grids (xy size 5 x 5 µm) by the rectangle tool in ImageJ 1.53d (62) (see Fig. S10E), with three grids randomly selected and the z-stack image in each grid further extracted with a 5 x 5 x 2 µm cube. Each grided z-stack image (see Fig. S10F) was reconstructed into a 3D image in the Imaris software 9.6.0 (BitPlane Inc.) with the surface rendering function using the following settings: (1) gaussian filter: 0.01, (2) smooth: 0.05, (3) background: 0.5, (4) auto-threshold, (5) enable split: 0.5. On the reconstructed 3D image and its surface rendering image (see Fig. S10G and H respectively), we excluded the cell objects on the edges of one side in the xyz orientation and counted manually the number of cells in the 5 x 5 x 2 µm cube.

-Specific probe design and visualization of Endozoicomonas phylotypes

The dominant four 16S rRNA gene sequences (OTU 2, OTU 16, OTU 5 and OTU 18) identified through sequencing approaches were aligned using infernal v1.1.4 (86) and applied to a function Design Probes in DECIPHER (65). Specific probes for each target OTU were not possible to design due to high sequence identity (ranging from 97.8% to 99.1%). Two specific probes could be designed, however, Endo-Group A (targeting OTU 2 and OTU 16) and Endo-Group B (targeting OTU 5 and OTU 18) (see Table S9 and Fig. S11A). To improve probe specificity, we also manually designed competitor probes due to mismatches at 2 sites for each probe (see Table S9 and Fig. S11A). The two probe candidates were tested for specificity using the web-based tool TestProbe3.0 (https://www.arb-silva.de/search/testprobe/) with the Silva SSU r138.1 database with parameter “maximum number of Mismatches: 0”. The results showed that the Endo-Group A probe and the Endo-Group B probe were matched to 16 sequences (including 2
from the *Endozocomonas* genus) and 146 sequences (including 133 from the *Endozocomonas* genus) respectively (see Fig. S11B). The ∆G° overall values associated with the combination of each specific probe and respective competitor probe were also calculated using the mathFISH web server (105). Both combinations exhibited high hybridization efficiencies (probe-target: -14.41 kcal/mol, probe-non-target: -8.12 kcal/mol, competitor-target: -7.55 kcal/mol, and competitor-non-target: -13.03 kcal/mol for Endo-Group A set, and probe-target: -14.24 kcal/mol, probe-non-target: -11.74 kcal/mol, competitor-target: -11.76 kcal/mol, and competitor-non-target: -14.74 kcal/mol for Endo-Group B set).

The Endo-Group B probe was identical to a region in the 16S rRNA gene of the isolate *Endozoicomonas montiporae* CL33^T* (see Fig. S11A below). To determine the hybridization-specificity, we applied both probes to the cultured bacterial cells of *E. montiporae* CL33^T* using quickHCR-FISH (59) with specific probes (Endo-Group A-initiatorH and Endo-Group B-initiatorR, see Table S9), the competitor probes (Comp. Endo-Group A probe and Comp. Endo-Group B, see Table S9), and corresponding amplifier probes across four hybridization condition with variable formamide concentrations (15%, 20%, 25% and 30% formamide [FA]). *E. montiporae* CL33^T* was cultivated on Modified Marine Broth Version 4 medium (21) and fixed with 4% PFA. The fixed cells were dropped on four microscope slides and dried completely. The cells were permeabilized with lysozyme, pre-hybridized in hybridization buffer (20 mM Tris-HCl [pH 8.0], 0.9 NaCl, 0.01% SDS, 10% dextran sulfate, 1% blocking reagent, variable formamide concentrations [15%, 20%, 25%, and 30% FA]) and subjected to hybridization in respective hybridization buffers with the two specific probes and the two competitor probes, as described previously. After the initial hybridization, each slide at the four different formamide concentrations ((1)15%, (2) 20%, (3) 25% and (4) 30% FA) were washed with respective washing buffers (20 mM Tris-HCl [pH 8.0], 0.01% SDS, (1) 0.318 M NaCl for 15% FA, (2) 0.225 M NaCl and 5 mM EDTA [pH 8.0] for 20% FA, (3) 0.159 M NaCl and 5 mM EDTA [pH 8.0] for 25%
FA, and (4) 0.112 M NaCl and 5 mM EDTA [pH 8.0] for 30% FA) at 48 °C for 15 min. Prior to the amplifier step, the corresponding H- and R-amplifier probes labeled with Cy3 and Cy5 respectively, (see Table S9) were independently prepared in the amplifier buffer (50 mM Na₂HPO₄, 0.9 M NaCl, 0.01% SDS, 10% dextran sulfate, 1% blocking reagent) and incubated for 1.5 min at 95 °C and then for 30 min at 25 °C and mixed (final concentration: 0.5 µM of each probe). Subsequently, after the pre-amplifier hybridization without any probe, the slides were subjected to amplifier hybridization with the two corresponding amplifier probes labeled with Cy3 (H1 and H2 amplifier probes) and Cy5 (R1 and R2 amplifier probes) for H and R initiator probes (see Table S9) respectively and processed through dehydration steps and mounted with antifade solution, as described in the Cell density section above. The FISH-stained cells were observed under a 40x magnification lens on a confocal laser scanning microscope LSM880 (Zeiss) with ZEN2.3 (black edition) software (Zeiss). The Cy3 signal emitted by the Endo-Group A probe was excited with the 561 nm laser (0.1%) and emission signals detected in the range 565 – 641 nm (master gain: 650). The Cy5 signal emitted by the Endo-Group B probe was excited with the 633 nm laser (0.05%) and detected in the range 652 – 735 nm (master gain: 600) for all image acquisitions. The results showed that both signals were consistent among four different formamide concentrations (see Fig. S11C).

Based on the FISH results on the bacterial culture, the following approach was applied to the paraffin PFA fixed coral samples. Initial hybridization was conducted at two formamide concentrations (20% and 25%) using quickHCR-FISH (59) with three initial probes (EUB338mix-initiatorC, Endo-Group A-initiatorH, and Endo-Group B-initiatorR, see Table S9), the competitor probes (Comp. Endo-Group A probe and Comp. Endo-Group B, see Table S9 below) and corresponding amplifier probes. The PFA fixed coral samples from Kenting and Okinawa were serially sectioned at 5 µm thickness using a Leica RM2125 RTS (Leica) microtome. Tissues sections were dewaxed through two changes in Xylene for 15 min,
dehydrated through two changes in abs. 100% ethanol for 5 min and air-dried. Each of the two serial sections was rinsed in a 20 mM Tris-HCl (pH 8.0) solution, and incubated with lysozyme solution (2 mg/ml lysozyme in 20 mM Tris-HCl [pH 8.0]) at 37 °C for 30 min to allow bacterial cell wall permeabilization and then rinsed in a 20 mM Tris-HCl solution (pH 8.0) at room temperature for 10 min. The serial sections were pre-hybridized in hybridization buffer (20 mM Tris-HCl [pH 8.0], 0.9 NaCl, 0.01% SDS, 10% dextran sulfate, 1% blocking reagent, variable formamide concentrations [20% FA for one serial section, 25% FA for other serial section]), and subsequently subjected to initial hybridization in respective hybridization buffers with the three initial probes and the two competitor probes, as described in the Cell density section above. The sections at the two different formamide concentrations were washed by respective washing buffers (20 mM Tris-HCl [pH 8.0], 0.01% SDS, 5 mM EDTA [pH 8.0], 0.225 M NaCl for 20% FA and 0.159 M NaCl for 25% FA) at 48 °C for 15 min. Prior to the amplifier step, the corresponding C-, H- and R-amplifier probes labeled with Alexea488 (C1 and C2 amplifier probes), Cy3 (H1 and H2 amplifier probes) and Cy5 (R1 and R2 amplifier probes) respectively (see Table S9), were independently prepared in the amplifier buffer (50 mM Na2HPO4, 0.9 M NaCl, 0.01% SDS, 10% dextran sulfate, 1% blocking reagent) and incubated for 1.5 min at 95 °C and then for 30 min at 25 °C and mixed (final concentration: 0.5 µM of each probe). The sections were subjected to pre-amplifier hybridization with only the buffer and then amplifier hybridization with the probe mixture, as described in the Cell density section above. The hybridized sections were washed in a cool PBS solution, dehydrated sequentially through water and abs. 100% ethanol and then mounted with antifade solution after air-dried, as described in the Cell density section. The hybridized sections were observed under a 63x objective lens on a confocal laser scanning microscope Zeiss LSM 880 (Zeiss) with ZEN2.3 (black edition) software (Zeiss). Detection of coral autofluorescence, the EUB338mix probe labeled with Alexea488, the Endo-Group A probe labeled with Cy3 and the Endo-Group B probe labeled with Cy5 was
achieved with the following settings: (1) laser: 405 nm (0.4%), 488 nm (0.2%), 561 nm (0.06%) and 633 nm (0.07%), (2) emission spectrum ranges: 417 – 500 nm (master gain 650), 501 – 543 nm (700), 565 – 641 nm (600), and 652 – 735 nm (650), respectively. The results were confirmed across the different hybridization optima, with the Endo-Group A probe signal detected from samples from both Kenting and Okinawa while the Endo-Group B probe signal was only detected in Kenting samples (see Fig. S11D). In addition, the signals detected from the EUB338mix probe overlapped completely with the signals from the Endo-Group A probe and/or the Endo-Group B probe (see Fig. S11D). Given that this result showed similar patterns to the molecular-based bacterial community analysis, we decided both specific probes are applicable and to be chosen in the hybridization condition with 25% FA due to higher specificity.

Finally, we performed FISH on samples from three colonies from each of the Kenting and Okinawa sites, using quickHCR-FISH (59) with two specific probes (Endo-Group A-initiatorH, and Endo-Group B-initiatorR, see Table S9), the competitor probes (Comp. Endo-Group A probe and Comp. Endo-Group B, see Table S9), and corresponding amplifier probes, with images acquired using the confocal laser scanning microscope as described above.

**NanoSIMS imaging**

The paraffin PFA-samples (that were decalcified by the Morse’s solution) of KS2 and OS6 colonies from Kenting and Okinawa, respectively, were cut into two serial sections (5 µm thickness, one for FISH and another for NanoSIMS) using a microtome Lecia RM2125 RTS (Leica). The sections were mounted onto a microscope slice for FISH observation and onto ITO (Indium-Tin-Oxide) glass coated with poly-L-lysine solution (Sigma-Aldrich) for NanoSIMS imaging, respectively. Those sections were dewaxed through two changes in Xylene for 15 min, dehydrated through two changes in abs. 100% ethanol for 5 min and then dried in a lamina flow cabinet (the FISH sample for 15 min and the NanoSIMS sample overnight). One serial section for
NanoSIMS was kept in a desiccator until the analysis. To identify *Endozoicomonas* and record the position of the CAMAs within the coral tissues, another serial section was performed with the quickFISH with two specific probes (Endo-Group A and B probes) according to the method above. Overall coral tissue images were acquired with a 10x objective lens in tile scan mode using a confocal laser scanning microscope Zeiss LSM 880 (Zeiss) and Zen SP2.3 software (Zeiss). Each CAMA image was also acquired with a 40x objective lens using the confocal laser scanning microscope. The detection of coral autofluorescence, the Endo-Group A and the Endo-Group B probes was achieved according to the method above.

NanoSIMS imaging was conducted by referring to the map obtained from confocal images to confirm the position of the CAMAs. The elemental distribution of CAMAs within coral tissue was acquired on a NanoSIMS 50L (Cameca-Ametek, Gennevilliers) instrument. Each 50×50 μm square region of interest around the CAMA was pre-sputtered with a 145~175 pA Cs+ ion beam for 2 minutes and sputtered by 1.1~1.5 pA Cs+ ion beam for 1.5 hours to acquire image data at 512×512 pixel resolution. For single-cell level imaging within the CAMAs, each 8×8 μm square area was pre-sputtered by a 15~18 pA Cs+ ion beam for 3 minutes and sputtered by a 0.3~0.4 pA Cs+ ion beam for 2~4 hours to obtain image data at 256×256 pixels resolution. Secondary ions of $^{12}\text{C}_2^+,\,^{12}\text{C}^{14}\text{N}^-,\,^{31}\text{P}^-,\,\text{and\,}^{32}\text{S}^-$. were collected simultaneously by electron multipliers. More than $5\times10^5$ and $2\times10^6$ counts of $^{12}\text{C}_2^-$ and $^{12}\text{C}^{14}\text{N}^-$ were respectively collected for each region of interest (ROI) associated with either a confirmed CAMA, Symbiodiniaceae cell, or coral tissue. L’Image software (developed by Larry Nittler, Carnegie Institution of Washington, Washington D.C.) was used to compare the secondary ion images of CAMA, Symbiodiniaceae and coral tissue ROI. A total of 22 CAMA regions (n = 11 CAMAs each of two location), 61 coral tissue regions (n = 32 in Kenting and n = 29 in Okinawa), and 61 Symbiodiniaceae cells (n = 33 in Kenting and n = 28 in Okinawa) were obtained. Elemental ratios of each region are calculated from total accumulated counts collected of each secondary ion.
Fig. S1. Schematic flowchart showing the workflow of this study. (A) Three *Stylophora pistillata* coral nubbins were sampled from 8 colonies from Kenting, Taiwan and 8 colonies from Okinawa, Japan. (B) Nubbins were fixed in either modified methacarn solution (MM), 4% paraformaldehyde solution (PFA), or 100% ethanol (EtOH). (C) Bacterial cell numbers within CAMAs were determined from PFA–fixed samples subjected to fluorescence *in situ* hybridization (FISH) (both polyps and histological sections) using the EUB338mix probe. Lightsheet determined the number and volume of CAMAs and confocal microscopic images determined cell density in CAMAs. (D) MM samples were cryosectioned and intact CAMA regions were dissected using laser microdissection (LMD) to facilitate DNA extraction from CAMA genomes.
CAMAs. (E) The 16S rRNA gene (V6-V8 region) was amplified from DNA extracted from CAMAs and directly Sanger sequenced. (F) Amplicon 16S rRNA gene (V6-V8 region) sequencing was undertaken from coral tissues fixed in 100% EtOH and intact single CAMAs to profile the bacterial community. (G) 16S rRNA gene probes were designed based on bacterial community analysis, FISH was performed on PFA-fixed samples to visualize bacterial phylotypes within CAMAs. (H) Total genomic DNA from CAMAs were subjected to whole genome amplification, metagenome sequencing and MAG binning. (I) NanoSIMS imaging of PFA samples profiled the natural elemental distribution of CAMAs.
Fig. S2 (A) Three-dimensional (3D) images of CAMAS within tentacles of a single polyp of *Stylophora pistillata*. CAMAs were detected by fluorescence in situ hybridization (FISH) using EUB338mix and Non338 probes labeled with Cy3 and Alexea488. Images are taken in Lightsheet using four channels, including autofluorescence of coral (blue) and Symbiodiniaceae (blue), Cy3 (red), and Alexea488 (white). The scale bars indicate 500 µm. (B) Distribution of CAMAs within coral tissues was assessed from five colonies from Kenting and Okinawa. (C) Dot plots showing the volume of CAMAs with coral tissues were determined using three different auto-threshold algorithms (Huang2, Li, and Mean algorithms).
Fig. S3. Fluorescence *in situ* hybridization images showing probe binding within a CAMA from a Kenting sample. (A) displays a merged image from signals from the Endo-Group A probe (B) and the Endo-Group B probe (C). Arrows indicate distinguished bacterial cells hybridized by the Endo-Group B probe (C), while at the edge of CAMA it is difficult to observe the signaling cells (both signals are manifested by orange-yellow overlap regions in the merged image a). Scale bars represent 10 µm.
Fig. S4. Heatmap (A) and principal component analysis (PCA) ordination (B) based on the pairwise average nucleotide identity (ANI) values across two MAGs (W5_Kt and HY_Ok) from CAMAs and 24 reference genomes from the family **Endozoicomonadaceae**. The 24 reference genomes categorized with *Endozoicomonadaceae* are available in the JGI/IMG database.
Fig. S5. Map illustrating type VI secretion system (T6SS) gene cluster present in three contigs of W5_Kt MAG and one contig of HY_Ok MAG. The gene clusters are represented with the functional annotations using different colors; black (annotated with T6SS proteins in RAST), grey (hypothetical proteins based on the RAST), and blue dot line (annotated with T6SS proteins in a domain search).
Fig. S6. (A) Heatmap showing presence/absence of type VI secretion system (T6SS) among members of the family Endozoicomonadaceae. The heatmap of the presence/absence of proteins encoding T6SS is based on KEGG orthology and shows a color gradient for the absence or copy number of each protein. (B) Maximum likelihood generated tree based on both TssB and TssC orthologs in T6SS. The tree contains sequences from the Endozoicomonadaceae genomes (showing the presence of T6SS) and close relative sequences from the Oceanospirillaceae, Vibrionaceae, and several families from the order Alteromonadales. Bootstrap values greater than 50% are shown at the nodes (based on 1000 replications). On the right side of both panels, the source of the genome is described.
Fig. S7. NanoSIMS ion maps of single elements $^{12}\text{C}_2$, $^{12}\text{C}^{14}\text{N}$, $^{31}\text{P}$ and $^{32}\text{S}$ from three different regions: CAMAs, coral tissues, and Symbiodiniaceae. The upper and lower panels correspond to images from Kenting and Okinawa samples, respectively in Figure 6A.
Fig. S8. Box plots representing mapped NanoSIMS elemental ratios $^{12}\text{C}/^{12}\text{C}^{14}\text{N}$ (C/N), $^{31}\text{P}/^{12}\text{C}$ (P/C), $^{31}\text{P}/^{12}\text{C}^{14}\text{N}$ (P/N), and $^{32}\text{S}/^{12}\text{C}$ (S/C) from (A) coral tissues (n = 32 regions in Kenting and n = 29 regions in Okinawa), and (B) Symbiodiniaceae (n = 33 cells in Kenting and n = 28 cells in Okinawa). Asterisk indicates statistically significant differences (Wilcoxon signed-rank test, $p < 0.05$ [*] and $p < 0.01$ [**]).
Fig. S9. Representative NanoSIMS ion maps of $^{12}$C/$^{14}$N, $^{32}$S, and the ratios with $^{12}$C/$^{14}$N (C/N), $^{31}$P/$^{14}$N (P/N), and $^{32}$S/$^{12}$C (S/C) showing bacterial cells within CAMAs at the single-cell level. Three ROIs (region of interest) correspond to figure 6d. ROI 1-2 and ROI 3 show the images from Kenting and Okinawa, respectively.
**Fig. S10. Image processing for cell density analysis.** (A) Image showing a single CAMA (red) in the tentacle of the coral *Stylophora pistillata* (DAPI and autofluorescence: blue, Symbiodiniaceae cells: green). (B) Coral tissue image with a 63x objective (C) cropping the region in the middle of the bacterial aggregation. (D) The extracted image with DAPI channel (E) creating grids for random selection. (F) Each selected grid proceeding in Imaris software (g) for constructing a 3D image from the DAPI signals and (h) its surface rendering image.
Fig. S11. (A) Position and alignment of sequences for FISH probe designs, (B) the probe specificity (C) and the hybridization-optimizations using bacterial culture with *Endozoicomonas montiporae* CL-33T (D) and coral specimens. (A) The alignment (V6-V8 region of the 16S rRNA gene) showing the targeted sequences of OTU 2, OTU 16, OTU 5 and OTU 18 which were identified through bacterial community analysis and *Endozoicomonas montiporae* CL33 for two specific probes (Endo-Group A and Endo-Group B) and two competitor probes (Comp. Endo-Group A and Comp. Endo-Group B). (B)Venn diagram showing probe specificity using the TestProbe 3.0 with the silva SSU r138,1 database. (C) Microphotographs showing hybridization-optimization on the *E. montiporae* culture and (D) coral specimens across different formamide concentrations (FA) in the initiator probe reaction.
Table S2. Coral-associated bacterial aggregates play a role in phosphate sequestration inside corals.

|                             | W5_Kt    | HY_OK    |
|-----------------------------|----------|----------|
| Size (bp)                   | 3,425,338| 3,930,706|
| Coverage (median value)     | 237.37x  | 224.19x  |
| Completeness (%)            | 88.53    | 90.89    |
| Contaminations (%)          | 1.8      | 1.9      |
| Strain heterogeneity in the contamination (%) | 41.67 | 30.77 |
| G+C content (%)             | 51.36    | 50.84    |
| N50                         | 10,556   | 14,184   |
| L50                         | 99       | 81       |
| Number of contigs (with PEGs) | 792   | 785     |
| Number of coding sequences (CDS) | 3181 | 3561 |
| Number of rRNAs            | 3        | 2        |
| Number of tRNAs            | 49       | 61       |
| Number of tmRNAs           | 1        | 1        |
| CRISPR (spacers)           | 3        | 7        |
Table S3. Identification of 16S rRNA gene from the two bin genomes by comparison of 16S rRNA gene sequences from *Endozoicomonadaceae* OTUs in the bacterial community analysis.

| Bin ID | Length (bp) | Targeted OTU (length: 440 bp) | E value | Identity (%) |
|--------|-------------|--------------------------------|---------|--------------|
| W5_Kt  | 1492        | OTU 2                          | 0.0     | 99.55        |
|        |             | OTU 5                          | 0.0     | 98.18        |
|        |             | OTU 16                         | 0.0     | 99.09        |
|        |             | OTU 18                         | 0.0     | 98.64        |
| HY_Ok  | 1483        | OTU 2                          | 0.0     | 99.55        |
|        |             | OTU 5                          | 0.0     | 98.18        |
|        |             | OTU 16                         | 0.0     | 99.09        |
|        |             | OTU 18                         | 0.0     | 98.64        |
Table S5. Number of proteins that contain eukaryotic-like domains in the metagenomes.

| Domain                          | W5_Kt | HY_Ok |
|---------------------------------|-------|-------|
| Ankyrin repeat proteins         | 26    | 29    |
| WD40 domain proteins            | 1     | 1     |
| Tetratricopeptide repeat proteins| 14    | 13    |
| Leucine-rich repeat proteins    | 0     | 1     |
Table S7. Portions of putative T3SS proteins with KEGG BRITE assignment in two bin-genomes.

| Protein families            | Sub families                                      | Proportion of assigned proteins in KEGG BRITE functional hierarchy (%) |
|-----------------------------|--------------------------------------------------|-----------------------------------------------------------------------|
|                             |                                                  | bin_W5_Kt (n=186 proteins) | bin_HY_Ok (n=209 proteins) |
| Metabolism                  | Enzyme                                           | 40.00                     | 42.99                     |
|                             | Protein kinases                                  | 0.36                      | 0.62                      |
|                             | Protein phosphatases and associated proteins     | 0.36                      | 0.31                      |
|                             | Peptidases and inhibitors                       | 4.29                      | 3.12                      |
|                             | Glycosyltransferases                            | 0.00                      | 0.31                      |
|                             | Lipopolysaccharide biosynthesis                 | 0.36                      | 0.31                      |
|                             | Peptidoglycan biosynthesis and degradation proteins | 0.71                      | 0.62                      |
|                             | Lipid biosynthesis proteins                     | 0.36                      | 0.93                      |
|                             | Prenyltransferases                              | 1.07                      | 0.93                      |
|                             | Amino acid related enzymes                      | 1.43                      | 2.18                      |
|                             | Photosynthesis proteins                         | 0.36                      | 0.00                      |
| Genetic information         | Transcription factors                           | 1.07                      | 1.25                      |
| processing                  | Transcription machinery                         | 0.71                      | 0.62                      |
|                             | Messenger RNA biogenesis                        | 1.43                      | 1.87                      |
|                             | Ribosome                                        | 0.71                      | 0.62                      |
|                             | Ribosome biogenesis                             | 2.86                      | 2.80                      |
|                             | Transfer RNA biogenesis                         | 4.64                      | 5.61                      |
|                             | Translation factors                             | 0.71                      | 0.62                      |
|                             | Chaperones and folding catalysts                | 1.07                      | 1.56                      |
|                             | Membrane trafficking                            | 0.36                      | 0.62                      |
|                             | Ubiquitin system                                | 0.71                      | 1.25                      |
|                             | DNA replication proteins                        | 1.07                      | 0.31                      |
|                             | Chromosome and associated proteins              | 1.43                      | 1.56                      |
|                             | DNA repair and recombination proteins           | 3.93                      | 2.49                      |
|                             | Mitochondrial biogenesis                        | 0.71                      | 0.62                      |
| Signaling and cellular      | Transporters                                     | 7.50                      | 5.30                      |
| processes                   | Secretion system                                | 1.43                      | 1.56                      |
|                             | Bacterial toxins                                | 0.36                      | 0.31                      |
|                             | Bacterial motility proteins                     | 0.71                      | 0.31                      |
|                             | Cytoskeleton proteins                           | 2.14                      | 3.43                      |
|                             | Exosome                                         | 2.14                      | 2.49                      |
|                             | Prokaryotic defense system                      | 1.79                      | 1.56                      |
|                             | Antimicrobial resistance genes                  | 0.36                      | 0.00                      |
| Category                                      | Description                                           | Value 1 | Value 2 |
|----------------------------------------------|-------------------------------------------------------|---------|---------|
| Unclassified: metabolism                     | Domain-containing proteins not elsewhere classified   | 0.36    | 0.00    |
|                                              | Enzymes with EC numbers                               | 3.57    | 3.74    |
|                                              | Energy metabolism                                     | 1.07    | 0.93    |
|                                              | Cofactor metabolism                                   | 0.36    | 0.31    |
|                                              | Secondary metabolism                                  | 0.36    | 0.00    |
| Unclassified: genetic information processing | Protein processing                                    | 0.36    | 0.31    |
|                                              | Replication and repair                                 | 0.00    | 0.31    |
| Unclassified: signaling and cellular processes| Transport                                             | 0.36    | 0.31    |
|                                              | Cell growth                                           | 0.36    | 0.31    |
|                                              | Signaling proteins                                    | 0.36    | 0.31    |
| Poorly characterized                         | General function prediction only                      | 1.79    | 1.25    |
|                                              | Function unknown                                      | 3.93    | 3.12    |
Table S9. Probes used in this study

| Probe name          | Probe sequence (5'-3') | Label | Ref.                        |
|---------------------|------------------------|-------|-----------------------------|
| **Initial probes**  |                        |       |                             |
| EUB338mix-initiatorH| CCGAATACAAAGCATCAACGACTAGA AAA AAGCWGCCWCCC CGTAGG GWT | (56) for targeted probe sequence and (99) for initiator sequence |
| EUB338mix-initiatorC| CCAGTTATCAGTAGTCGTCCTCATTTT GCWGCCWCCC CGTAGG GWT | (56) for targeted probe sequence and (99) for initiator sequence |
| Non338-initiatorC   | CCAGTTATCAGTAGTCGTCCTCATTTT ACATCTAGGGGAGGC | (57) for targeted probe sequence and (99) for initiator sequence |
| Endo-Group A-initiatorH | CCGAATACAAAGCATCAACGACTAGA AAA AAGCTTCGCCGACCGTCTGT | This study for targeted probe sequence and (99) for initiator sequence |
| Endo-Group B-initiatorR | TACGCCCTAAGAATCCGAACCCCTATGAAAT AATGCACTTT TCTCAGATTAGCTCC | This study for targeted probe sequence and (99) for initiator sequence |
| **Competitor probes** |                        |       |                             |
| Comp. Endo-Group A  | GCTTGGCAAACCGTCTGT     |       | This study                  |
| Comp. Endo-Group B  | ACGCACTTTT TCTCAGATTAGCTCC | This study                  |
| **Amplifier probes** |                        |       |                             |
| H1 amplifier probe  | TCTAGTCGTTgatgctttgtattcggCGACAGATAAcc gaatacaaaacate | Cy3 at the 3' end (99) |
| H2 amplifier probe  | ccaatacaaaacateAACGACTAGAgatgctttgtattcggT TATCTGTCG | Cy3 at the 5' end (99) |
| C1 amplifier probe  | ATGAAAGGACGgactactgataactggGACTTCCATAc cagttatcagagtc | Alexea48 8 at the 3' end (99) |
| C2 amplifier probe  | ccagttatcagagtcGTCCCTTCATgactactgataactggT ATGGAAGTC | Alexea48 8 at the 5' end (99) |
| R1 amplifier probe  | CATAGGGTTTgattcctaggcgaGAACGCATCAAt aegccctaaagatc | Cy5 at the 3' end (99) |
| R2 amplifier probe  | tacgccctaaagatcGAACCTATGgattcctaggcgaT TGATGCTGC | Cy5 at the 5' end (99) |

*1 Underlined sequences indicate the initiator sequence. *2 Double underlined sequences are the respective targeted sequence. *3 Lowercase letters represent the stem structure of the amplifier probe.
Other External Supplementary Material for this manuscript includes the following:

Table S1. The bacterial abundance containing all OTUs from the CAMAs in the coral *Stylopora pistillata* colonies from Kenting and Okinawa and those sequences.

Table S4. Gene components of biofilm and quorum sensing.

Table S6. Gene components of the bacterial secretion system.

Table S8. Metabolisms of *Endozoicomonas* sp. W5_Kt and HY_Ok MAGs based on KEEG.
REFERENCES AND NOTES

1. A. Elbeltagy, K. Nishioka, T. Sato, H. Suzuki, B. Ye, T. Hamada, T. Isawa, H. Mitsui, K. Minamisawa, Endophytic colonization and in planta nitrogen fixation by a *Herbaspirillum* sp. isolated from wild rice species. *Appl. Environ. Microbiol.* **67**, 5285–5293 (2001).

2. E. Kondorosi, P. Mergaert, A. Kereszt, A paradigm for endosymbiotic life: Cell differentiation of rhizobium bacteria provoked by host plant factors. *Annu. Rev. Microbiol.* **67**, 611–628 (2013).

3. R. Stouthamer, J. A. J. Breeuwer, G. D. D. Hurst, *Wolbachia* *pipientis*: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* **53**, 71–102 (1999).

4. S. M. Kuechler, P. Renz, K. Dettner, S. Kehl, Diversity of symbiotic organs and bacterial endosymbionts of Lygaeoid bugs of the families Blissidae and Lygaeidae (Hemiptera: Heteroptera: Lygaeoidea). *Appl. Environ. Microbiol.* **78**, 2648–2659 (2012).

5. S. Balmand, C. Lohs, S. Aksoy, A. Heddi, Tissue distribution and transmission routes for the tsetse fly endosymbionts. *J. Invertebr. Pathol.* **112** Suppl, S116–S122 (2013).

6. J. S. Graf, S. Schorn, K. Kitzinger, S. Ahmerkamp, C. Woehle, B. Huettel, C. J. Schubert, M. M. M. Kuypers, J. Milucka, Anaerobic endosymbiont generates energy for ciliate host by denitrification. *Nature* **591**, 445–450 (2021).

7. N. Dubilier, C. Mülders, T. Ferdelman, D. de Beer, A. Pernthaler, M. Klein, M. Wagner, C. Erséus, F. Thiermann, J. Krieger, O. Giere, R. Amann, Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* **411**, 298–302 (2001).

8. H. Felbeck, Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* **213**, 336–338 (1981).

9. M. D. Tianero, J. N. Balaich, M. S. Donia, Localized production of defence chemicals by intracellular symbionts of Haliclona sponges. *Nat. Microbiol.* **4**, 1149–1159 (2019).

10. C. Schuett, H. Doepke, A. Grathoff, M. Gedde, Bacterial aggregates in the tentacles of the sea anemone Metridium senile. *Helgol. Mar. Res.* **61**, 211–216 (2007).
Endozoicomonas are specific, facultative symbionts of sea squirts. *Front. Microbiol.* **7**, 1042 (2016).

R. Ponnudurai, M. Kleiner, L. Sayavedra, J. M. Petersen, M. Moche, A. Otto, D. Becher, T. Takeuchi, N. Satoh, N. Dubilier, T. Schweder, S. Markert, Metabolic and physiological interdependencies in the Bathymodiolus azoricus symbiosis. *ISME J.* **11**, 463–477 (2017).

I. Cano, D. Ryder, S. C. Webb, B. J. Jones, C. L. Brosnahan, N. Carrasco, B. Bodinier, D. Furones, T. Pretto, F. Carella, B. Chollet, I. Arzul, D. Cheslett, E. Collins, K. B. Lohrmann, A. L. Valdivia, G. Ward, M. J. Carballal, A. Villalba, I. Marigómez, S. Mortensen, K. Christison, W. C. Kevin, E. Bustos, L. Christie, M. Green, S. W. Feist, Cosmopolitan distribution of *Endozoicomonas*-like organisms and other intracellular microcolonies of bacteria causing infection in marine mollusks. *Front. Microbiol.* **11**, 577481 (2020).

M. J. Neave, R. Rachmawati, L. Xun, C. T. Michell, D. G. Bourne, A. Apprill, C. R. Voolstra, Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales. *ISME J.* **11**, 186–200 (2017).

J. Engebrecht, K. Nealson, M. Silverman, Bacterial bioluminescence: Isolation and genetic analysis of functions from Vibrio fischeri. *Cell* **32**, 773–781 (1983).

N. Miyamoto, M.-A. Yoshida, H. Koga, Y. Fujiwara, Genetic mechanisms of bone digestion and nutrient absorption in the bone-eating worm *Osedax japonicus* inferred from transcriptome and gene expression analyses. *BMC Evol. Biol.* **17**, 17 (2017).

J.-H. Shiu, J.-Y. Ding, C.-H. Tseng, S.-P. Lou, T. Mezaki, Y.-T. Wu, H.-I. Wang, S.-L. Tang, A newly designed primer revealed high phylogenetic diversity of *Endozoicomonas* in coral reefs. *Microbes Environ.* **33**, 172–185 (2018).

B. Glasl, C. E. Smith, D. G. Bourne, N. S. Webster, Disentangling the effect of host-genotype and environment on the microbiome of the coral *Acropora tenuis*. *PeerJ* **7**, e6377 (2019).
19. N. Miller, P. Maneval, C. Manfrino, T. K. Frazer, J. L. Meyer, Spatial distribution of microbial communities among colonies and genotypes in nursery-reared Acropora cervicornis. *PeerJ* **8**, e9635 (2020).

20. M. Sweet, H. Villela, T. Keller-Costa, R. Costa, S. Romano, D. G. Bourne, A. Cárdenas, M. J. Huggett, A. H. Kerwin, F. Kuek, M. Medina, J. L. Meyer, M. Müller, F. J. Pollock, M. S. Rappé, M. Sere, K. H. Sharp, C. R. Voolstra, N. Zaccardi, M. Ziegler, R. Peixoto, Insights into the cultured bacterial fraction of corals. *mSystems* **6**, e01249-20 (2021).

21. J.-Y. Ding, J.-H. Shiu, W.-M. Chen, Y.-R. Chiang, S.-L. Tang, Genomic insight into the host–endosymbiont relationship of *Endozoicomonas montiporae* CL-33T with its coral host. *Front. Microbiol.* **7**, 251 (2016).

22. I. Cano, R. van Aerle, S. Ross, D. W. Verner-Jeffreys, R. K. Paley, G. S. E. Rimmer, D. Ryder, P. Hooper, D. Stone, S. W. Feist, Molecular characterization of an *Endozoicomonas*-like organism causing infection in the king scallop (*Pecten maximus* L.). *Appl. Environ. Microbiol.* **84**, e00952-17 (2018).

23. J. G. Klinges, S. M. Rosales, R. McMinds, E. C. Shaver, A. A. Shantz, E. C. Peters, M. Eitel, G. Wörheide, K. H. Sharp, D. E. Burkepile, B. R. Silliman, R. L. Vega Thurber, Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, *Candidatus* Aquarickettsia rohweri, gen. nov., sp. nov. *ISME J.* **13**, 2938–2953 (2019).

24. K. Tandon, C.-Y. Lu, P.-W. Chiang, N. Wada, S.-H. Yang, Y.-F. Chan, P.-Y. Chen, H.-Y. Chang, Y.-J. Chiou, M.-S. Chou, W.-M. Chen, S.-L. Tang, Comparative genomics: Dominant coral-bacterium *Endozoicomonas acroporae* metabolizes dimethylsulfoniopropionate (DMSP). *ISME J.* **14**, 1290–1303 (2020).

25. T. Work, G. Aeby, Microbial aggregates within tissues infect a diversity of corals throughout the Indo-Pacific. *Mar. Ecol. Prog. Ser.* **500**, 1–9 (2014).

26. J. C. Harshbarger, S. C. Chang, S. V. Otto, Chlamydiae (with phages), mycoplasmas, and rickettsiae in Chesapeake Bay bivalves. *Science* **196**, 666–668 (1977).
27. R. Cruz-Flores, J. Cáceres-Martínez, Rickettsiales-like organisms in bivalves and marine gastropods: A review. *Rev. Aquac.* **12**, 2010–2026 (2020).

28. P. M. Costa, S. Carreira, J. Lobo, M. H. Costa, Molecular detection of prokaryote and protozoan parasites in the commercial bivalve *Ruditapes decussatus* from southern Portugal. *Aquaculture* **370-371**, 61–67 (2012).

29. N. Wada, M. Ishimochi, T. Matsui, F. J. Pollock, S.-L. Tang, T. D. Ainsworth, B. L. Willis, N. Mano, D. G. Bourne, Characterization of coral-associated microbial aggregates (CAMAs) within tissues of the coral *Acropora hyacinthus*. *Sci. Rep.* **9**, 14662 (2019).

30. L. Karstens, M. Asquith, S. Davin, D. Fair, W. T. Gregory, A. J. Wolfe, J. Braun, S. McWeeney, Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems* **4**, e00290-19 (2019).

31. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

32. R. Cunning, R. A. Bay, P. Gillette, A. C. Baker, N. Traylor-Knowles, Comparative analysis of the *Pocillopora damicornis* genome highlights role of immune system in coral evolution. *Sci. Rep.* **8**, 16134 (2018).

33. H. R. Nelson, A. H. Altieri, Oxygen: The universal currency on coral reefs. *Coral Reefs* **38**, 177–198 (2019).

34. C. Ferrier-Pagès, J.-P. Gattuso, S. Dallot, J. Jaubert, Effect of nutrient enrichment on growth and photosynthesis of the zooxanthellate coral *Stylophora pistillata*. *Coral Reefs* **19**, 103–113 (2000).

35. K. Koop, D. Booth, A. Broadbent, J. Brodie, D. Bucher, D. Capone, J. Coll, W. Dennison, M. Erdmann, P. Harrison, O. Hoegh-Guldberg, P. Hutchings, G. B. Jones, A. W. D. Larkum, J. O’Neil, A. Steven, E. Tentori, S. Ward, J. Williamson, D. Yellowlees, ENCORE: The effect of nutrient enrichment on coral reefs. Synthesis of results and conclusions. *Mar. Pollut. Bull.* **42**, 91–120 (2001).
36. F. Zhang, L. C. Blasiak, J. O. Karolin, R. J. Powell, C. D. Geddes, R. T. Hill, Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4381–4386 (2015).

37. F. R. Cianfanelli, L. Monlezun, S. J. Coulthurst, Aim, load, fire: The type VI secretion system, a bacterial nanoweapon. *Trends Microbiol.* **24**, 51–62 (2016).

38. C. S. Bernard, Y. R. Brunet, E. Gueguen, E. Cascales, Nooks and crannies in type VI secretion regulation. *J. Bacteriol.* **192**, 3850–3860 (2010).

39. M. Gallique, M. Bouteiller, A. Merieau, The type VI secretion system: A dynamic system for bacterial communication? *Front. Microbiol.* **8**, 1454 (2017).

40. F. D. Cassan, A. Coniglio, E. Amavizca, G. Maroniche, E. Cascales, Y. Bashan, L. E. de-Bashan, The *Azospirillum brasilense* type VI secretion system promotes cell aggregation, biocontrol protection against phytopathogens and attachment to the microalgae *Chlorella sorokiniana*. *Environ. Microbiol.* **23**, 6257–6274 (2021).

41. D. Linares, N. Jean, P. Van Overtvelt, T. Ouidir, J. Hardouin, Y. Blache, M. Molmeret, The marine bacteria *Shewanella frigidimarina* NCIMB400 upregulates the type VI secretion system during early biofilm formation. *Environ. Microbiol. Rep.* **8**, 110–121 (2016).

42. C.-M. Ryu, Against friend and foe: Type 6 effectors in plant-associated bacteria. *J. Microbiol.* **53**, 201–208 (2015).

43. L. Speare, A. G. Cecere, K. R. Guckes, S. Smith, M. S. Wollenberg, M. J. Mandel, T. Miyashiro, A. N. Septer, Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8528–E8537 (2018).

44. A. Salinero-Lanzarote, A. Pacheco-Moreno, L. Domingo-Serrano, D. Durán, E. Ormeño-Orrillo, E. Martínez-Romero, M. Albareda, J. M. Palacios, L. Rey, The Type VI secretion system of *Rhizobium etli* Mim1 has a positive effect in symbiosis. *FEMS Microbiol. Ecol.* **95**, fiz054 (2019).

45. C. Elton, Competition and the structure of ecological communities. *J. Anim. Ecol.* **15**, 54–68 (1946).
46. C. Bongrand, E. G. Ruby, Achieving a multi-strain symbiosis: Strain behavior and infection dynamics. *ISME J.* **13**, 698–706 (2019).

47. H. Itoh, S. Jang, K. Takeshita, T. Ohbayashi, N. Ohnishi, X.-Y. Meng, Y. Mitani, Y. Kikuchi, Host–symbiont specificity determined by microbe–microbe competition in an insect gut. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22673–22682 (2019).

48. F. J. Pollock, R. McMinds, S. Smith, D. G. Bourne, B. L. Willis, M. Medina, R. V. Thurber, J. R. Zaneveld, Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat. Commun.* **9**, 4921 (2018).

49. T. Bayer, M. J. Neave, A. Alsheikh-Hussain, M. Aranda, L. K. Yum, T. Mincer, K. Hughes, A. Apprill, C. R. Voolstra, The microbiome of the Red Sea coral *Stylophora pistillata* is dominated by tissue-associated *Endozoicomonas* bacteria. *Appl. Environ. Microbiol.* **79**, 4759–4762 (2013).

50. P. A. O’Brien, S. Tan, C. Yang, P. R. Frade, N. Andreakis, H. A. Smith, D. J. Miller, N. S. Webster, G. Zhang, D. G. Bourne, Diverse coral reef invertebrates exhibit patterns of phylosymbiosis. *ISME J.* **14**, 2211–2222 (2020).

51. P. A. O’Brien, N. Andreakis, S. Tan, D. J. Miller, N. S. Webster, G. Zhang, D. G. Bourne, Testing cophylogeny between coral reef invertebrates and their bacterial and archaeal symbionts. *Mol. Ecol.* **30**, 3768–3782 (2021).

52. C.-P. Chen, C.-H. Tseng, C. A. Chen, S.-L. Tang, The dynamics of microbial partnerships in the coral *Isopora palifera*. *ISME J.* **5**, 728–740 (2011).

53. S. L. Jorgensen, B. Hannisdal, A. Lanzén, T. Baumberger, K. Flesland, R. Fonseca, L. Ovreås, I. H. Steen, I. H. Thorseth, R. B. Pedersen, C. Schleper, Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2846–E2855 (2012).

54. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
55. R. C. Edgar, UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).

56. P. D. Schloss, S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, C. F. Weber, Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).

57. S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

58. C. Liu, S. H. Cheng, S. Lin, Illuminating the dark depths inside coral. *Cell. Microbiol.* **22**, e13122 (2020).

59. T. Yamaguchi, B. M. Fuchs, R. Amann, S. Kawakami, K. Kubota, M. Hatamoto, T. Yamaguchi, Rapid and sensitive identification of marine bacteria by an improved in situ DNA hybridization chain reaction (quickHCR-FISH). *Syst. Appl. Microbiol.* **38**, 400–405 (2015).

60. H. Daims, A. Brühl, R. Amann, K. H. Schleifer, M. Wagner, The domain-specific probe EUB338 is insufficient for the detection of all bacteria: Development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).

61. G. Wallner, R. Amann, W. Beisker, Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14**, 136–143 (1993).

62. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

63. R Core Team, R: A Language and environment for statistical computing (R Foundations for Statistical Computing, 2020); www.R-project.org/.

64. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (Sage Publications, ed. 3, 2019).
65. E. S. Wright, L. S. Yilmaz, A. M. Corcoran, H. E. Ökten, D. R. Noguera, Automated design of probes for rRNA-targeted fluorescence in situ hybridization reveals the advantages of using dual probes for accurate identification. *Appl. Environ. Microbiol.* **80**, 5124–5133 (2014).

66. J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, R. B. O’hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, H. Wagner, Vegan: Community ecology package (2020); https://cran.r-project.org/web/packages/vegan/index.html.

67. P. Martinez Arbizu, pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.0.1 (2017).

68. E. Nawrocki, “Structural RNA Homology Search and Alignment Using Covariance Models,” thesis, Washington University in Saint Louis, School of Medicine (2009).

69. S. Andrews, FastQC: A quality control tool for high throughput sequence data (2010); www.bioinformatics.babraham.ac.uk/projects/fastqc/.

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

71. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

72. G. V. Uritskiy, J. DiRuggiero, J. Taylor, MetaWRAP—A flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).

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

74. J. Alneberg, B. S. Bjarnason, I. de Bruijn, M. Schirmer, J. Quick, U. Z. Ijaz, L. Lahti, N. J. Loman, A. F. Andersson, C. Quince, Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
75. D. D. Kang, F. Li, E. Kirton, A. Thomas, R. Egan, H. An, Z. Wang, MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7, e7359 (2019).

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

77. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477 (2012).

78. F. A. B. von Meijenfeldt, K. Arkhipova, D. D. Cambuy, F. H. Coutinho, B. E. Dutilh, Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* 20, 217 (2019).

79. B. Bushnell, BBMap: A fast, accurate, splice-aware aligner (LBNL-7065E, Lawrence Berkeley National Lab, 2014); www.osti.gov/biblio/1241166.

80. P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927 (2020).

81. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069 (2014).

82. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731 (2016).

83. M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30 (2000).

84. R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A.
Wilke, O. Zagnitko, The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).

85. A. Marchler-Bauer, S. H. Bryant, CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* **32**, W327–W331 (2004).

86. E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

87. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

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

89. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

90. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

91. I.-M. A. Chen, K. Chu, K. Palaniappan, A. Ratner, J. Huang, M. Huntemann, P. Hajek, S. Ritter, N. Varghese, R. Seshadri, S. Roux, T. Woyke, E. A. Elof-Fadrosh, N. N. Ivanova, N. C. Kyprides, The IMG/M data management and analysis system v.6.0: New tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).

92. S.-H. Yoon, S.-M. Ha, J. Lim, S. Kwon, J. Chun, A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**, 1281–1286 (2017).

93. R. A. M. Villanueva, Z. J. Chen, ggplot2: Elegant graphics for data analysis. *Interdiscip. Res. Perspect.* **17**, 160–167 (2019).
94. V. Eichinger, T. Nussbaumer, A. Platzer, M.-A. Jehl, R. Arnold, T. Rattei, EffectiveDB—Updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res.* **44**, D669–D674 (2016).

95. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

96. J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).

97. J. Felsenstein, Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791 (1985).

98. The UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

99. D. Binns, E. Dimmer, R. Huntley, D. Barrell, C. O’Donovan, R. Apweiler, QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).

100. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

101. A. Morse, Formic acid-sodium citrate decalcification and butyl alcohol dehydration of teeth and bones for sectioning in paraffin. *J. Dent. Res.* **24**, 143–153 (1945).

102. L.-K. Huang, M.-J. J. Wang, Image thresholding by minimizing the measures of fuzziness. *Pattern Recogn.* **28**, 41–51 (1995).

103. C. H. Li, P. K. S. Tam, An iterative algorithm for minimum cross entropy thresholding. *Pattern Recogn. Lett.* **19**, 771–776 (1998).

104. C. A. Glasbey, An analysis of histogram-based thresholding algorithms. *Graph. Models Image Process.* **55**, 532–537 (1993).
105. L. S. Yilmaz, S. Parnerkar, D. R. Noguera, mathFISH, a web tool that uses thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **77**, 1118–1122 (2011).