Fidelity varies in the symbiosis between a gutless marine worm and its microbial consortium

Yui Sato*[1], Juliane Wippler[1], Cecilia Wentrup[2], Rebecca Ansorge[1], Miriam Sadowski[1], Harald Gruber-Vodicka[1], Nicole Dubilier*[1], Manuel Kleiner*[3]

[1]Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

[2]University of Vienna, Department of Microbiology and Ecosystem Science, Althanstr. 14, A-1090 Vienna, Austria

[3]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, North Carolina, USA

***Corresponding authors:**

Yui Sato: ysato@mpi-bremen.de, phone +49 (421) 2028-8240

Nicole Dubilier: ndubilie@mpi-bremen.de, phone +49 (421) 2028 9320

Manuel Kleiner: manuel_kleiner@ncsu.edu, phone +1 919 515 3792

Keywords: microbiome, microbiota, animal-bacterial symbiosis, host-symbiont specificity, partner fidelity, symbiont transmission, phylogenetic congruence, single nucleotide polymorphisms (SNPs), intraspecific genetic variation

Abstract

In obligate symbioses, partner fidelity plays a central role in maintaining the stability of the association across multiple host generations. Fidelity has been well studied in hosts with a very restricted diversity of symbionts, but little is known about how fidelity is maintained in hosts with multiple co-occurring symbionts. The marine annelid *Olavius algarvensis* lives in an obligate association with at least five co-occurring bacterial symbionts that are inherited vertically. The symbionts so efficiently supply their hosts with nutrition that these worms have completely reduced their mouth and digestive tract. Here, we investigated partner fidelity in the *O. algarvensis* symbiosis by sequencing the metagenomes of 80 host individuals from two mitochondrial lineages and two locations in the Mediterranean. Comparative phylogenetic analyses of mitochondrial and symbiont genotypes based on single nucleotide polymorphisms revealed high fidelity for the primary symbiont that dominated the microbial consortium of all 80 *O. algarvensis* individuals. In contrast, the secondary symbionts of *O. algarvensis*, which occurred in lower abundance and were not always present in all host individuals, showed only intermediate to low fidelity. We hypothesize that harbouring symbionts with variable levels of fidelity ensures faithful transmission of the most abundant and nutritionally important symbiont, while flexibility in the acquisition of secondary symbionts enhances genetic exchange and retains ecological and evolutionary adaptability.

Introduction

Faithful, evolutionarily-stable associations between host and symbiont are central to mutualistic symbioses [1]. Many symbioses display macroevolutionary stability as partner specificity, defined here as the restriction at the taxonomic range of symbiont species with which a host species associates and vice versa [2,3]. Examples of such partner specificity at the species-specific level include Aphidoidea aphids and their primary intracellular endosymbiont *Buchnera aphidocola* [4], Stilbonematine nematodes and their ectosymbionts *Candidatus* Thiosymbion (*Ca.* Thiosymbion) [5], vestimentiferan tubeworms and chemoautotrophic endosymbionts [6], deep-sea vesicomyid and bathymodiolin bivalves

with sulphur- and methane-oxidizing bacterial endosymbionts [7,8], and *Euprymna scolopes* squid with the bioluminescent bacterium *Vibrio fischeri* [9,10] to name a few.

To ensure robust symbiotic associations with specific partners over microevolutionary scales (i.e. within a host species), diverse mechanisms have evolved to mediate partner 'fidelity'. Here partner fidelity is defined as the degree that a host associates with particular symbiont genotypes across generations [11]. Mechanisms that mediate fidelity are, for example, reproductive and behavioural mechanisms to transmit symbionts directly from parents to offspring (vertical transmission; e.g. [12-15]), and anatomical, chemical and immunological mechanisms of the host to take up a suitable symbiont lineage selectively from an external source in each generation anew (horizontal transmission; e.g. [9,15-19]). In some cases, these mechanisms function in combination and reinforce partner fidelity; e.g. beewolf wasps that show stringent partner selection during vertical transmission of *Streptomyces* symbionts [20].

Strong partner fidelity through stringent vertical co-transmission of mitochondria and symbionts can be detected as phylogenetic co-divergence between the genomes of host mitochondria and a given symbiont genotype at a microevolutionary scale [21-24]. If such a high-fidelity symbiotic association extends over macroevolutionary timeframes, the congruent phylogenetic pattern can even result in a co-speciation pattern [14,25-27], however this macroevolutionary pattern can be disrupted by rare horizontal transmission events (i.e. mixed-mode transmission [28,29]). In theory, development of a congruent phylogenetic pattern is also possible without stringent vertical transmission of symbionts if horizontal transmission has extremely high selectivity for a symbiont lineage at the genotype level. However, evidence for host-symbiont phylogenetic congruence is extremely limited in symbioses with horizontal symbiont transmission [30], and many horizontally acquired symbioses show weak or no partner fidelity despite their partner specificity [11,15,31]. High partner fidelity can give selective advantages to a heritable symbiosis to exploit a new resource as the host and symbiont co-adapt to the new niche, but it can also pose a long-term fitness cost via ecological niche limitation through irreversible interdependency (i.e. the "evolutionary rabbit hole" [32]). In this process, symbiont genomes undergo dramatic changes that result in loss of genes due to restricted gene flux, population-size bottlenecks and reduced purifying selection associated with host-restricted lifestyle [33]. With less

stringent fidelity, occasional introduction of symbionts from external sources can offer a rescue mechanism to these long-term evolutionary pressures by enabling genome recombination and lateral gene transfer among symbiont populations [29,34].

Symbiotic partner fidelity is an important driver of symbiosis evolution not only in 'single symbiont vs. single host' contexts but also when multiple microbial symbionts establish a stable association with a host. In many of such cases, the host requires interplay among multiple symbionts as they collectively supply essential services [35-39]. For example, several essential amino acids in the mealybug *Planococcus citri* are produced in a patchwork pathway encoded in the genomes of beta- and gamma-proteobacterial symbionts [40]. The primary and oldest symbiont in multibacterial symbioses often shows a high partner fidelity through strict or near-strict vertical transmission, whereas secondary symbionts appear to be acquired horizontally and became co-transmitted with the primary symbiont subsequently. This phenomenon has been observed in many insect-bacterial symbioses, such as adelgids with the primary beta- and secondary gamma-proteobacterial symbionts [41], meadow spittlebugs with *Sulcia* and *Sadalis*-related symbionts [42], diverse mealybug species with *Tremblaya* symbiont and a gammaproteobacterium that nests within *Tremblaya* [43], and *Cinara* aphid species with *Buchnera* and *Erwinia* symbionts [37]. Here, the dynamic nature of multi-member symbioses can provide a mechanism to ameliorate the evolutionary costs of high-fidelity symbiosis, as co-existing symbionts with ancient and recent origins can exchange new genetic material by lateral gene transfer [37]. These aspects of multi-member symbioses have been studied at macroevolutionary scales by comparing multiple host species, but almost nothing is known about the microevolutionary scale. It is currently unknown whether multiple symbiont species display uniform or variable levels of partner fidelity when they cooperatively form an obligate symbiosis with a host animal.

The monophyletic group of marine annelids in the subfamily Phallodrilinae (Clitellata, Naididae, Genera *Olavius* and *Inanidrilus; sensu* Erséus *et al.* [44]) completely lack a digestive tract and excretion organs, and instead live in an obligate symbiosis with several bacterial partners [35,45,46] (Figure 1a and 1b). Bacterial partners of these gutless marine annelids belong to a limited set of taxa in the Alpha-, Delta- and Gamma-proteobacteria, and *Spirochaeta*, and appear to form host species-specific bacterial consortia [45,47-51]. Among six species of gutless marine annelids whose symbionts have been studied, all

species associate with the gammaproteobacterium *Candidatus* (*Ca.*) Thiosymbion, while they harbour species specific combinations of other symbionts [45,47,49-51], suggesting that the *Ca*. Thiosymbion is the primary symbiont for these host species. However, phylogenies of *Ca*. Thiosymbion and host annelids do not show a co-speciation pattern [5], and *Ca*. Thiosymbion has been replaced by the Gamma4 symbiont in *Inanidrilus exumae* [48]. In the best researched species *Olavius algarvensis*, seven endosymbiont genera have been identified thus far - two sulphide-oxidizing Gammaproteobacteria; *Ca.* Thiosymbion [46,47] and Gamma3, four sulphate-reducing Deltaproteobacteria; Delta1a, Delta1b, Delta3 and Delta4, and a spirochete [35,45-47,52,53]. Five or more symbiont genera regularly co-occur within a single host individual, and sulphide-oxidizers and sulphate-reducers always coexist despite variations in symbiont compositions among host individuals [35,47]. These symbionts in concert provide the host with nutrition via chemosynthetic carbon-fixation and also process host waste products [35,45,46]. *O. algarvensis* therefore represents an excellent system to test fidelity of multiple symbionts in an obligate bacteria-animal symbiosis. Morphological studies of other gutless annelids have postulated that symbionts are vertically transmitted during ovipositioning, in which symbionts are smeared onto the egg when it passes through animal tissue layers containing high densities of symbionts termed genital pads [54,55]. This potential vertical transmission could result in a congruence between the phylogenies of a given symbiont and the host mitochondria. However, microscopic observations also suggested that some small-sized symbionts are horizontally acquired from the environment [55,56]. Therefore, transmission modes and levels of partner fidelity of different symbiont species in gutless annelids remain unclear and require further investigation. Considering these observations of flexible symbiotic associations across all symbionts on both macro- and micro-evolutionary scales, we hypothesized that low levels of partner fidelity among co-existing symbionts should be present even at the microevolutionary scale.

To characterize the fidelity of multiple co-residing symbionts in *O. algarvensis*, we sequenced metagenomes of 80 individuals representing two mitochondrial lineages that co-occur at two sampling locations. We generated metagenome-assembled genomes of symbionts, profiled their community structures using single-copy genes, and investigated phylogenetic congruency between the host

mitochondria and each of the symbionts based on their genome-wide single nucleotide polymorphisms (SNPs).

Results

Two co-occurring dominant host mitochondrial lineages form the basis for testing partner fidelity of multiple symbiont species

We initially sequenced the mitochondrial cytochrome c oxidase subunit I (COI) gene of hundreds of *O. algarvensis* individuals to understand the natural diversity of host mitochondrial lineages at the two study locations approximately 16 km apart on the coast of the Island of Elba, Italy (Sant' Andrea and Cavoli; Figure 1c and 1d). We obtained 579 COI sequences from individual specimens and generated a core alignment of 525 bp. A haplotype network based on this alignment revealed that two dominant mitochondrial COI-haplotypes, here termed A and B, occur at both locations (Figure 1e). These dominant COI-haplotypes together encompassed 99% and 71% of local *O. algarvensis* individuals of Sant' Andrea and Cavoli, respectively. We focused on these two co-occurring dominant COI-haplotypes to study symbiotic partner fidelity by comparing phylogenies of symbiotic bacteria and the corresponding host mitochondrial phylogeny. To this end, we obtained 80 metagenomes (2548 ± 715 Mbp (mean ± SD) sequenced per sample) derived from 20 individuals of each COI-haplotype from each of the two locations (i.e. 4 combinations of COI-haplotype and location; hereafter termed 'host groups'). Complete host mitochondrial genomes (mtDNA) were assembled from metagenomes of individuals belonging to COI-haplotype A and B, which were respectively 15,715 bp and 15,730 bp in size and both encoded 13 protein coding genes, 2 ribosomal RNAs, and 21 transfer RNAs with shared synteny. These mtDNAs shared 99.3% nucleotide identity on average, highlighting a close phylogenetic relatedness between these host mitochondrial lineages that enables testing of symbiont fidelity at a fine evolutionary scale.

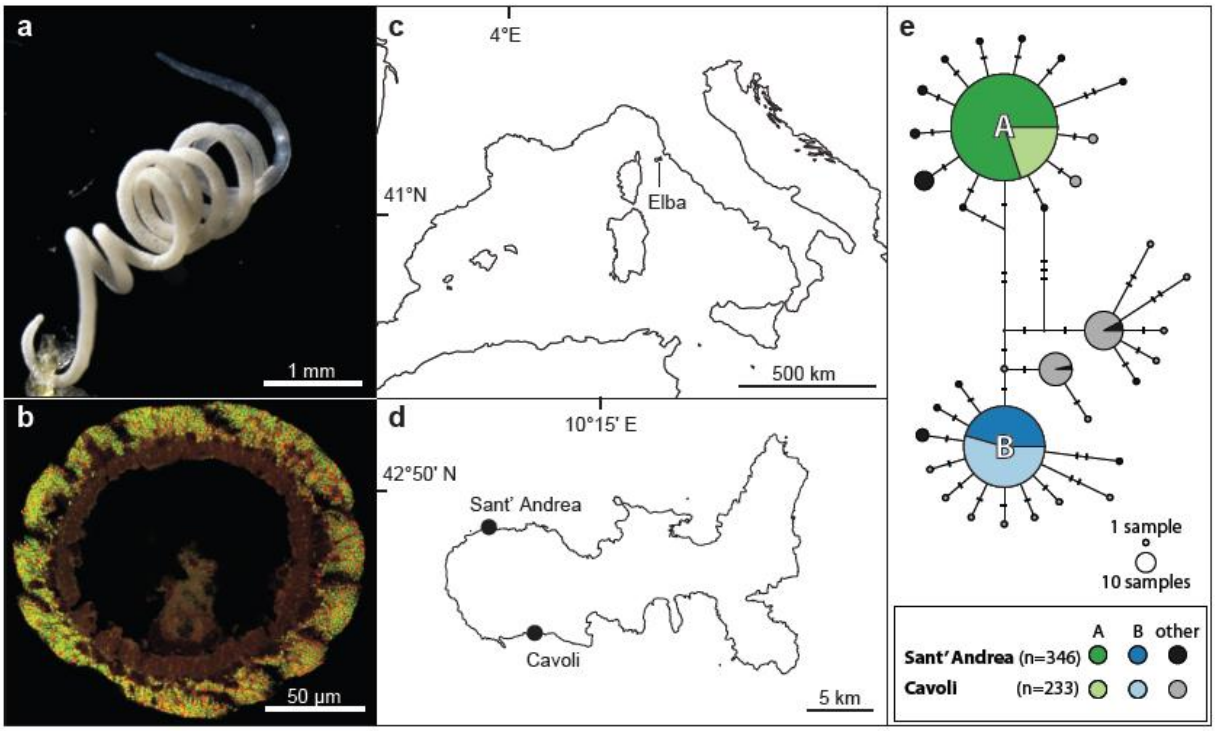


Figure 1. *Olavius algarvensis* populations harboured two co-occurring host mitochondrial lineages. (**a**) Light microscopy image of *Olavius algarvensis*, (**b**) Fluorescence *in situ* hybridization image of an *O. algarvensis* cross section, showing the primary and secondary symbionts (Gammaproteobacteria in green and Deltaproteobacteria in red) just below the cuticle on the outer body surface of the worm. (**c, d**) The location of the two study sites on the Island of Elba in the Mediterranean Sea. (**e**) Haplotype network of mitochondrial cytochrome c oxidase subunit I gene sequences of *O. algarvensis* individuals from Sant' Andrea (darker shades) and Cavoli (brighter shades). The size of pie-charts indicates approximate haplotype frequencies. Hatch marks correspond to the number of point mutations between haplotypes. Nodes depicted by small black dots indicate unobserved intermediates predicted by the haplotype network builder (TCS statistical parsimony algorithm). Note that two dominant haplotypes (A and B, in green and blue shades, respectively) were found at both study sites.

Very low levels of single-nucleotide polymorphisms in symbiont populations within *O. algarvensis* individuals

We generated metagenome-assembled genomes (MAGs) for all seven *O. algarvensis* symbionts belonging to the Gammaproteobacteria, Deltaproteobacteria and *Spirochaeta* (Suppl. Table S1; see Suppl. Figure S1 and Suppl. Figure S2 for taxonomic overview). MAGs for the Delta1a, Delta1b and Delta3 symbionts were assembled in our recent studies [52,53]. Symbiont genome bins were overall high in completeness and low in contamination and had sufficient quality to perform SNP-based phylogenetic analyses (Suppl. Table S1).

Prior to phylogenetic comparisons of symbionts between host individuals, we tested for symbionts' infraspecific genetic diversity within an individual host by examining SNP-site densities across the symbiont genomes using deeply-sequenced metagenomes (Suppl. Table S2a). Average SNP

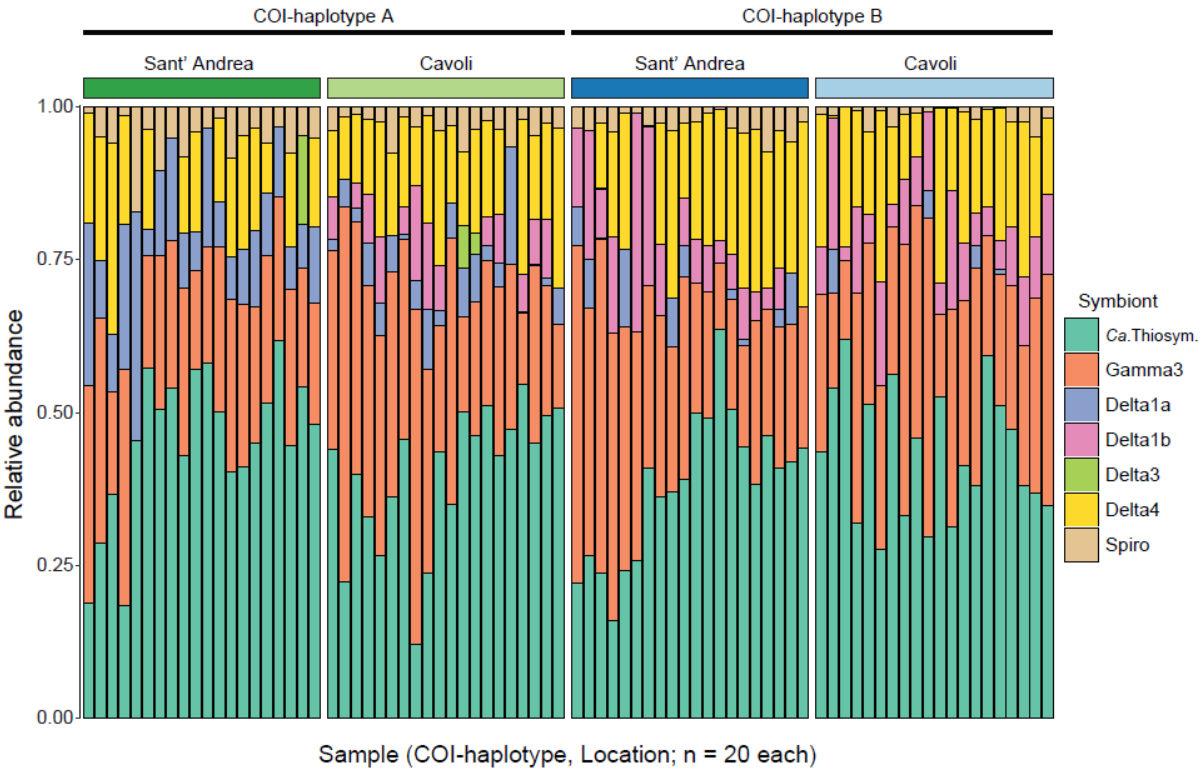
densities showed extremely low values for all the symbiont species (ranging from 0.02 to 0.49 SNPs/Kbp; Suppl. Table S2b). These values are lower than or comparable to SNP densities reported for vertically-transmitted endosymbionts in the shallow water bivalve *Solemya velum* (0.1 – 1 SNPs/Kbp) [29], which regularly experience a strong population bottleneck effect during maternal transmission [57]. SNP densities were also much lower than those reported for horizontally-transmitted symbionts in the giant tubeworm *Riftia pachyptila* (2.9 SNPs/Kbp) [58] and in the deep-sea mussels in the genus *Bathymodiolus* (5 – 11 SNPs/Kbp) [59]. Based on these SNP density values, we treated each of the symbiont species in an individual of *O. algarvensis* as a bacterial population dominated by a single most abundant genotype in the following SNP-based analyses.

Species composition and prevalence of symbionts varied between host groups

In the 80 metagenomes of the *O. algarvensis* individuals, we assessed (**i**) symbiont species composition and (**ii**) symbiont prevalence (the number of host individuals with the respective symbiont), by identifying sequences of single-copy genes (SCGs) that are specific to symbiont species (Figure 2; for details see Supplementary text 1.1)**.** We detected the *Ca.* Thiosymbion and Gamma3 symbionts in all 80 individuals. *Ca.* Thiosymbion was on average the most abundant symbiont among all species across the samples (41.6 ± 11.6%; mean ± SD). The Gamma3 symbiont varied greatly in relative abundance among the individuals, ranging from 0.1 to 61.2% (28.3 ± 11.3%). The spirochete symbiont was also present in all samples, with consistently low relative abundances except for a rare exception with 17.3% (3.4 ± 2.7%; Supplementary text 1.1). In contrast, the deltaproteobacterial symbionts showed high variability in relative abundance within and among the host groups. Delta1a was detected in all host individuals of COI-haplotype A (hereafter ‘A-hosts’) but in less than half of individuals of COI-haplotype B (‘B-hosts’) at both locations. Delta1b was not detected in A-hosts from Sant’ Andrea, but found in 65 to 100% of individuals in the other three host groups. Delta4 was found in all but a few individuals in all host groups, while Delta3 was detected only scarcely in 3 host groups and thus excluded from subsequent phylogenetic analyses. Summed relative abundances of deltaproteobacterial and gammaproteobacterial symbionts accounted for consistent proportions, regardless of the host COI-haplotypes, locations or their combinations (Delta sum; 26.7 ± 7.2%,

Gamma sum; 69.9 ± 7.4%, Delta : Gamma ratios = 1 : 2.92 ± 1.18, Kruskal-Wallis tests: COI-haplotypes; χ^2 = 1.311; p = 0.252, locations; χ^2 = 2.430; p = 0.119, host groups; χ^2 = 5.188; p = 0.159). Symbiont species compositions based on 16S ribosomal RNA gene (SSU) sequences derived from the metagenomes showed the same patterns as the ones based on SCG sequences (see Supplementary text 1.2).

a



b

COI-haplotype, location	*Ca.*Thiosym.	Gamma3	Delta1a	Delta1b	Delta3	Delta4	Spirochete
A, Sant' Andrea	20	20	20	0	1	15	20
A, Cavoli	20	20	20	13	2	19	20
B, Sant' Andrea	20	20	9	16	2	16	20
B, Cavoli	20	20	6	20	0	19	20*
Total	80	80	55	49	5	69	80*

Figure 2. Symbiont community composition varied significantly in 80 *O. algarvensis* individuals. The primary symbiont, *Ca.* Thiosymbion, was dominant in most hosts, followed by the secondary sulfur-oxidizing symbiont, Gamma 3. The relative abundances of the deltaproteobacterial and spirochaetal symbionts were considerably lower. **(a)** Community structures of symbionts in *O. algarvensis* individuals of two mitochondrial lineages (COI-haplotypes A and B) from two locations (Sant' Andrea and Cavoli). **(b)** The number of samples in which the respective symbiont species was detected (n = 20 replicate samples per host group). Relative abundance of the symbiont was estimated based on metagenomic sequences matching to a collection of single-copy gene sequences extracted from symbiont genomes (* See Supplementary text 1.1, '*Ca.* Thiosym'; Ca. Thiosymbion).

Symbiont phylogenies based on genome-wide SNPs showed varying degrees of congruency with host mitochondrial lineages

We identified SNPs across the genomes of host mitochondria and symbionts in metagenomes of 80 *O. algarvensis* individuals. Phylogenies of host mitochondria and symbionts were reconstructed based on SNPs identified with a probabilistic approach to genotyping (Figure 3). This approach takes sequencing errors into account through calculation of posterior genotype probabilities and builds pairwise genetic distance matrices directly from posterior genotype probabilities, and thus has advantages in SNP identification from low-coverage sequence data [60,61]. The resulting phylogeny of mitochondrial genomes (mtDNA) clearly differentiated mtDNA between the two COI-haplotypes A and B (Figures 3a), despite their high similarity (99.3% average nucleotide identity). The pairwise genetic distances of mtDNA among three comparative categories, (**i**) within host group, (**ii**) between COI-haplotypes, and (**iii**) between locations, also supported a clear mtDNA divergence based on COI-haplotypes (Figure 4a; "Within" vs. "A vs. B", Kruskal-Wallis test (KWT); $\chi^2 = 85.77$; df = 2; $p < 2.2e-16$, post-hoc Dunn test (DT); z = 8.98; adjusted p = 8.0e-19). A location-based genetic divergence of mtDNA was also shown (Figure 4a; "Within" vs. "SANT vs. CAVL", DT; z = 2.53; p = 1.1e-02), but the SNP phylogeny of mtDNA showed a location-based monophyletic clustering only for B-hosts from Sant' Andrea (Figure 3a). Phylogenetic relationships of individuals within the same host group were in most cases not resolved with clade support, further highlighting little genetic divergence of mtDNA in our study samples (Suppl. Figure S3).

We expect that symbionts with high partner fidelity show a similar phylogeny as host mtDNA phylogeny and pairwise genetic distances that are positively correlated with host mtDNA divergence (i.e. the symbiont and the host genetically co-diverge), whereas symbionts with low partner fidelity do not show these patterns. The phylogeny based on genome-wide SNPs for *Ca.* Thiosymbion formed highly divergent clades mirroring the divergence in host mtDNA between COI-haplotypes A and B (Figure 3b). Phylogenetic relationships of *Ca*. Thiosymbion in host individuals were similarly unresolved within the A and B lineages as observed for host mtDNA phylogenies (Suppl. Figure S3b), and thus we were not able to examine phylogenetic congruence between mtDNA and *Ca*. Thiosymbion between host individuals. Nonetheless a Mantel test for the correlation between pairwise

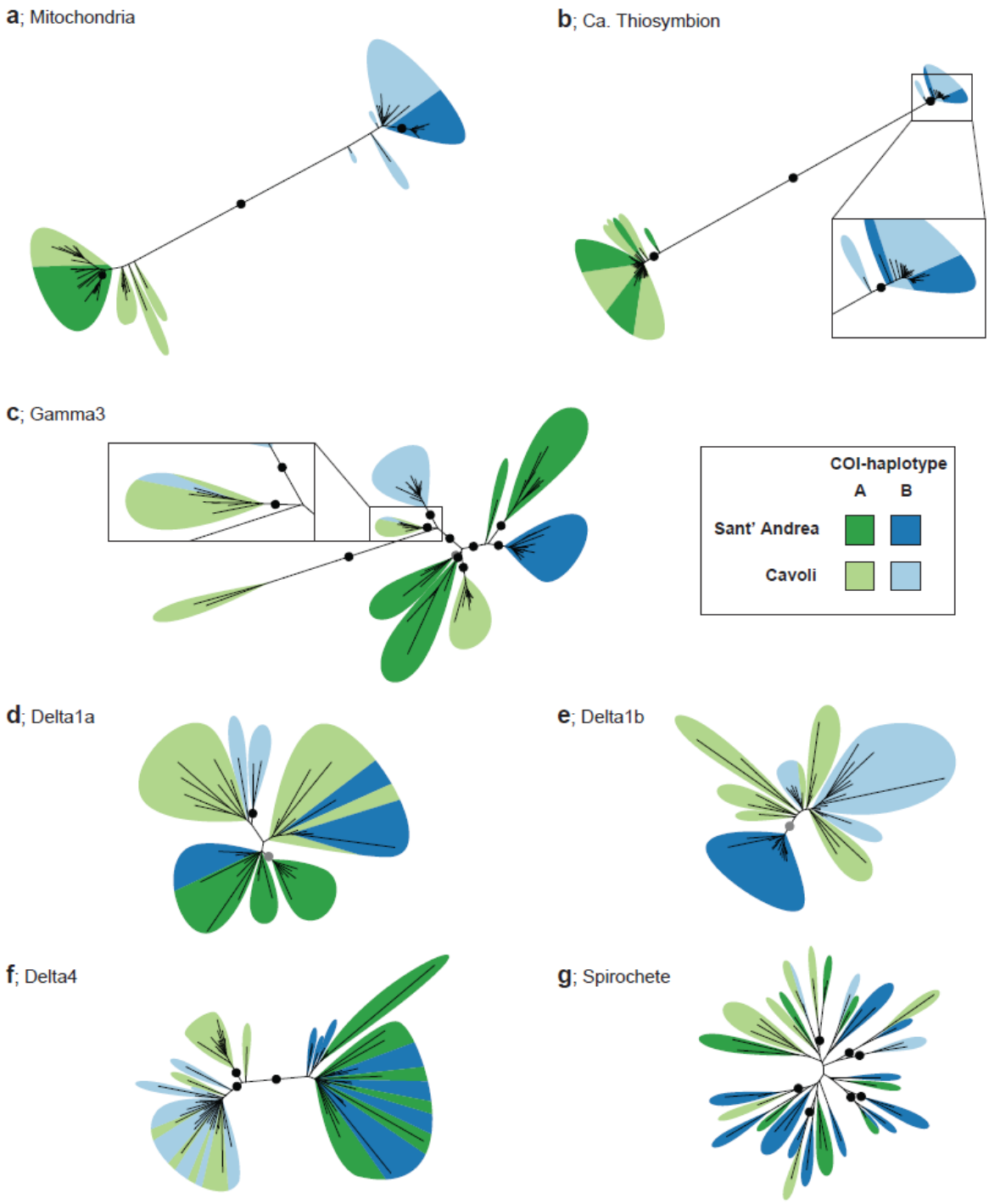


Figure 3. Symbiont clades showed variable levels of fidelity in the *O. algarvensis* association. Comparative phylogenetic analyses of host mitochondria and symbionts revealed a range of co-phylogenetic patterns from strong congruence in the primary symbiont *Ca.* Thiosymbion to no congruence in the spirochaetal symbiont. Phylogenetic trees based on SNPs across genomes of **a**; mitochondria (166 SNPs, n = 80), and **b**; Primary symbiont *Ca.* Thiosymbion (2872 SNPs, n = 80), **c**; Gamma3 (618 SNPs, n = 80), **d**; Delta1a (375 SNPs, >65% cut-off*, n =37), **e**; Delta1b (624 SNPs, >65% cut-off*,n = 46), **f**; Delta4 (675 SNPs, >50% cut-off*, n=67), and **g**; Spirochete symbionts (88 SNPs, >63% cut-off*, n=41) of *O. algarvensis*. The phylogenies were inferred from genetic distances calculated from posterior genotype probabilities without deterministic genotype-calling. Bootstrap support values >95 ~ 100 are shown in black, and <95 ~ >85 in grey. Internal branch support values were omitted for visibility. *Cut-off was applied based on percentage of mapped reference bases to filter out samples with no or extremely low abundance symbionts and thus insufficient data for SNP-detection. (Phylogenies based on core SNPs using a deterministic genotyping method are supplied in Supplementary Figure S4).

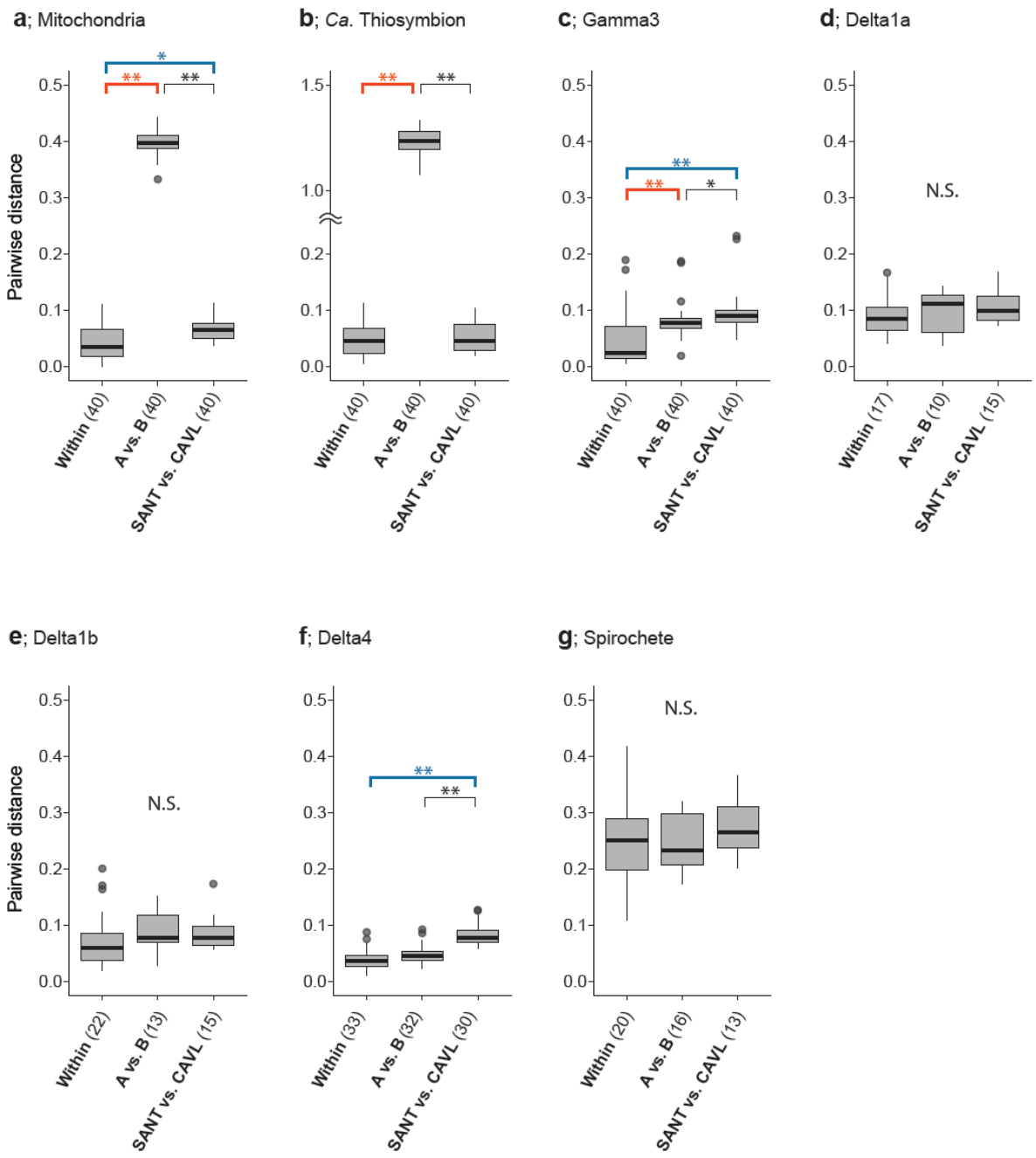


Figure 4. Host mitochondrial genetics and geographic location had a significant effect on some but not all symbionts. Host genetic divergence significantly affected genetic divergence in the primary symbiont *Ca.* Thiosymbion and Gamma3, while geographic location affected the Gamma3 and Delta4 symbionts. Pairwise genetic distances of *O. algarvensis* mitochondria and its symbionts were calculated from pairs of *O. algarvensis* individuals within the same combination of COI-haplotype and location (**Within**), between individuals of the COI-haplotypes A and B from the same location (**A vs. B**), and between individuals from Sant' Andrea and Cavoli but within the same COI-haplotype (**SANT vs. CAVL**). Among these three categories, pairwise genetic distances were compared for **a**; mitochondria, and **b**; *Ca.* Thiosymbion, **c**; Gamma3, **d**; Delta1a, **e**; Delta1b, **f**; Delta4, and **g**; Spirochete symbionts. Pairwise comparisons were made on random pairs without replacement to ensure the data independence. Genetic distances were normalized per SNP site and log-scaled. Thick horizontal lines and grey boxes respectively indicate the median and interquartile range (IQR) of observations. Vertical lines show the IQR ± 1.5 IQR range, and outliers out of this range are shown as circles. Numbers in brackets indicate numbers of pairwise comparisons per category tested. Asterisks respectively denote statistical significance (*; $p<0.05$, **; $p<0.01$, orange and blue highlight host 'matriline effect' (genetic distances between mitochondrial lineages are larger than expected from the variability of genetic distances within the same host group) and 'location effect' (genetic distances between locations are larger than expected from the variability of genetic distances within the same host group). N.S. indicates no significant differences among categories ($p>0.05$).

genetic distances of *Ca.* Thiosymbion and mtDNA genetic distances of the corresponding host pairs confirmed a strong genetic co-divergence with a well-fitting linear regression (r = 0.988; p = 0.001; Suppl. Figure S5a). Correspondingly, pairwise symbiont genetic distances for *Ca*. Thiosymbion showed a large divergence driven by the difference in their host COI-haplotypes, hereafter 'matriline effect', but not by the difference in their sample locations, hereafter 'location effect' (Figure 4b; KWT; $\chi 2 = 79.72$; $p < 2.2e\text{-}16$, matriline effect; DT "Within" vs. "A vs. B"; z = 8.02; p = 3.1e-16, location effect; DT "Within" vs. "SANT vs. CAVL"; z = 0.62; p = 0.54).

The phylogeny of the Gamma3 symbiont supported a number of monophyletic clades, each of which consisted of individuals of the same host group (Figure 3c). Gamma3 in A-hosts clustered into multiple clades per location, whereas ones in B-hosts from Sant' Andrea and Cavoli respectively formed a single clade each. The only exception was a Gamma3 of a B-host from Cavoli that clustered within a clade of A-host symbionts from Cavoli (magnified panel in Figure 3c). Genetic co-divergence between Gamma3 and host mtDNA was supported by a Mantel test (p = 0.001), but its correlation co-efficient (r = 0.230) was much lower than detected between *Ca*. Thiosymbion and mtDNA (r = 0.998). A correlation plot of pairwise genetic distances between Gamma3 and host mtDNA showed three centres of density, indicating the presence of factor(s) driving the genetic divergence of Gamma3 other than the mtDNA divergence (Suppl. Figure S5b). In fact, this was supported by statistical tests of the pairwise genetic divergence of Gamma3 that showed both a matriline effect and a location effect, with the location effect being greater than the matriline effect (Figure 4c; KWT; $\chi^2 = 30.73$; $p < 2.1e\text{-}7$, DTs; matriline effect; z = 3.24; p = 1.2e-3, location effect; z = 5.51; p = 1.0e-7, "A vs. B" vs. "SANT vs. CAVL"; z = 2.28; p = 2.3e-2).

The Delta1a and Delta1b symbionts showed neither clear phylogenetic clustering patterns based on host groups (Figure 3d and 3e), nor significant genetic divergence driven by matriline- and location-effects (Figures 4d and 4e, KWTs; Delta1a; $\chi^2 = 1.97$; p = 0.377, Delta1b; $\chi^2 = 4.29$; p = 0.117). Although Delta1b symbionts in B-hosts from Sant' Andrea formed a monophyletic clade with a weak clade support (88%; Figure 3e), Delta1b was not detected in A-hosts from Sant' Andrea and thus a link between this clade formation and the matriline or location effect remained unclear. Mantel tests nonetheless detected co-diversification patterns between the symbiont and mtDNA for both

Delta1a and Delta1b, albeit with low co-relation coefficients (Delta1a; r = 0.145; p = 0.048, Delta1b; r = 0.283; p = 0.002, Suppl. Figures S5c and S5d).

In contrast to *Ca*. Thiosymbion and Gamma3, the phylogeny of the Delta4 symbiont formed monophyletic clades based on sampling locations (Figure 3f), with a strong location effect on genetic divergence (Figure 4f; KWT; χ^2 = 50.60; p = 1.0e-11, DT; location effect; z = 6.83; p = 2.6e-11). A correlation plot of pairwise genetic distances between Delta4 and mtDNA also displayed four centres of density indicating the location effect (Suppl. Figure S5e). Although a statistical test of pairwise genetic distances did not show a matriline effect (DT; z = 1.56; p = 0.120), a Mantel test detected a weak co-divergence pattern between Delta4 and mtDNA (r = 0.060, p = 0.017). The average pairwise genetic distance within the same host group for Delta4 was at a small level comparable to that of *Ca.* Thiosymbion ("Within" terms in Figure 4b and 4f).

The phylogeny of the spirochete symbiont showed highly intermixed patterns among host mitochondrial lineages and locations (Figure 3g). We found no specific effects explaining genetic divergence of the spirochete symbiont, or a genetic co-divergence pattern between mtDNA and spirochete (Figure 4g; KWT; χ^2 = 2.03; p = 0.362, Suppl. Figure S5f; Mantel test; r = 0.001; p = 0.349). The average pairwise genetic distance of the spirochete within the same host group was more than 2-fold greater than those observed for other symbionts (Figure 4g "Within").

Estimation of the effective population size of symbionts within *O. algarvensis* individuals

We inspected genetic diversity of each symbiont within sample host individuals, as opposed to the genetic diversity of symbionts among individuals in the same host group ("Within" terms in Figure 4). To this end, we inferred effective population size (N_e) within a host individual for each symbiont based on genome-wide SNP abundance (see details in Supplementary Text 1.6). N_e estimates of *O. algarvensis* symbionts were on the order of 10^4 - 10^5, with *Candidatus* Thiosymbion and spirochete symbionts having N_e on the order of 10^5, Delta1b approximately 5×10^4, and Gamma3, Delta1a and Delta4 at the lower 10^4 (Suppl. Figure S6).

Discussion

Increasing sequencing efforts have uncovered variability of symbiont composition in multi-member symbioses among closely related host species [6,8,43,47,48,62] and between conspecific populations [63,64]. While the discovery of variable symbiont communities has raised intriguing questions about the stability of symbiotic associations and thus the partner fidelity in symbioses, it is often challenging to measure degrees of fidelity by studying distant host lineages because an accumulation of infrequent partner-switching can eventually break a signal of otherwise high fidelity. By leveraging highly replicated metagenomes derived from closely related host lineages (matrilines) from two nearby locations, we investigated the variability in symbiont composition and fidelity of symbiotic associations within a single-animal species that harbours multiple bacterial symbionts.

Symbiont composition in closely related host lineages indicates dynamic associations between multiple bacterial symbionts and *Olavius algarvensis*

The close relationship of host lineages, based on distinguishable yet highly similar mtDNA sequences, formed a unique basis for assessing symbiont variability at a fine evolutionary scale. We found surprisingly large variations in symbiont community composition, not only between individuals from the two major mitochondrial lineages A and B, but also within a mitochondrial lineage between two locations (Figure 2). The detection of the *Ca*. Thiosymbion, Gamma3 and spirochete symbionts in all *O. algarvensis* individuals suggests that they are the obligate symbionts necessary for survival of host [36]. In contrast, any given deltaproteobacterial symbionts was only detected in a subset of host individuals, which suggests a facultative symbiotic association. However, consistent proportions of total deltaproteobacterial versus gammaproteobacterial symbionts were found in all samples. This co-occurrence pattern is in line with previous studies showing that the gammaproteobacterial symbionts function as sulphur-oxidizing carbon-fixers and the deltaproteobacterial symbionts as sulphate reducers, thereby leading to internal sulphur cycling in the *O. algarvensis* symbiosis [35,45,46,65,66]. Despite the critical role of the deltaproteobacteria, our results suggest that their taxonomic affiliations appear less important than their function, as some deltaproteobacterial symbionts can be completely replaced by other deltaproteobacterial symbionts even within the same host group (i.e. individuals from the same combination of mitochondrial lineage and location). To our knowledge, such a high

intraspecific variability in the symbiont community composition has not been demonstrated within closely-related host populations at the small spatial scales of our study (i.e. two locations within 16 km). Previous studies on species-level fidelity in insect symbioses have conducted over large numbers of host lineages and at much larger geographical scales (e.g. hundreds to thousands of km [63,64]). Our findings at the microevolutionary scale raise the question if such highly variable symbiont communities are common among hosts that form an obligate symbiosis with multiple bacteria, or if they are a feature of chemosynthetic symbioses in which symbionts have multiple distinct functions such as sulphate reducers and sulphur oxidisers.

Population-level phylogenetic comparisons show various levels of partner fidelity among different symbionts species

Phylogenies reconstructed from genome-wide SNPs for different symbiont species showed a stark contrast in terms of congruence with host mtDNA lineages (Figure 3), from being fully congruent with the major mitochondrial lineages (*Ca*. Thiosymbion), to a mixed pattern explained by both host matrilines and locations (Gamma3), one explained by locations (Delta4), those that could not be explained by matrilines or locations (Delta1a and Delta1b), and a highly intermixed pattern among host groups (spirochete). Instead of the expected general flexibility and low fidelity of symbiotic associations inferred in previous studies of gutless annelid symbioses [5,47-49], these results together indicate various levels of partner fidelity among the symbionts at the microevolutionary scale. The high degree of congruence between host mitochondria and *Ca*. Thiosymbion in *O. algarvensis* shown here is in strong contrast to a previous macroevolutionary-scale study demonstrating that multiple gutless annelid species do not have a co-speciation pattern with *Ca*. Thiosymbion [5]. In this study, *Ca.* Thiosymbion showed an unambiguous matriline effect but no location effect on the genetic divergence, highlighting that the high partner fidelity of *Ca*. Thiosymbion is strongly preserved across the two study locations. However, signatures of phylogenetic congruence between mitochondria and *Ca*. Thiosymbion were not resolved below the scale of mitochondrial lineages A and B. This could be due to incomplete lineage sorting [24] that arises from differences in coalescent times and mutation rates between mtDNA and *Ca*. Thiosymbion genomes, or the lack of their genetic differentiations between

O. algarvensis individuals. The latter is a more likely explanation, given the close relatedness of host mtDNAs for which the phylogeny was largely unresolved within the major clade.

The Gamma3 symbiont in host individuals from the same host lineage and location in many cases clustered together in the SNP-based phylogeny, and patterns of phylogenetic clustering and genetic divergence of the Gamma3 symbiont were explained by both host matrilines and locations. This mixed signal may have resulted from occasional inter-lineage host switching of Gamma3, which was indicated as a rare case in specimens from Cavoli. These results suggest that Gamma3 has a certain level of host fidelity but the fidelity is less stringent than observed in *Ca.* Thiosymbion. A similar mixed-mode symbiont transmission has been demonstrated in other marine symbioses, such as *Solemya* bivalves where the symbiont is regularly transmitted via ovaries but phylogenetic signals indicate occasional host switching [29,67]. Partner fidelity for all the deltaproteobacterial symbionts and spirochete was not supported by phylogenetic reconstruction and examination of the matriline- and location-effects on pairwise genetic distances. However, for Delta1a and Delta1b, (**i**) the symbiont prevalence showed a strong contrast between host groups, and (**ii**) genetic co-divergence patterns between mitochondria and symbiont were still supported by a Mantel test. These two lines of evidence suggest that Delta1a and Delta1b have a certain degree of host-lineage fidelity, despite the lack of support from SNP-based phylogenetic reconstruction. Lineage sorting could again be incomplete for these symbionts, as the formation of distinguishable symbiont lineages linked to mitochondrial lineages can take some time after a heritable symbiont is introduced to the host [24]. Considering the time required for genetic drift to fix symbionts' genetic divergence, it is possible that the introduction of heritable Delta1a and Delta1b symbionts to B-hosts and A-hosts, respectively, were recent events, and that the introduction of Delta1b has not occurred in A-hosts in Sant' Andrea.

Vertical transmission as a likely mechanism for the high fidelity of *Candidatus* Thiosymbion

The strong host-symbiont fidelity observed for *Ca*. Thiosymbion can in theory arise through two scenarios: (**i**) coupling of particular genetic lineages of *O. algarvensis* and symbionts is strongly selected during horizontal symbiont acquisition, or (**ii**) the host employs robust vertical symbiont transmission. For horizontally transmitted symbionts to show a phylogenetic congruence with host

mitochondrial lineages (scenario (**i**)), host lineages should have genetic differences that recognize and select a specific symbiont lineage from the environment, and/or different symbiont lineages recognize genetic signatures of a specific host lineage upon colonization. Such mechanisms would likely need to be maintained in host nuclear genomes of the population. Therefore, this scenario implies no genetic recombination between host individuals with different symbiont specificity, and thus requires hosts of each mitochondrial lineage to be a reproductively isolated population. This scenario is unlikely, given that reproductive isolation at a mtDNA divergence of 0.7% as we observed for *O. algarvensis* is uncommon in clitallate annelids; e.g. Sympatric reproductive isolation in *Lumbricus rubellus* has only been observed with mtDNA sequence divergence that were at least one order of magnitude higher [68,69].

The more likely scenario is (**ii**), predicting that the high fidelity of *Ca.* Thiosymbion is owing to vertical transmission of the symbiont along with maternal mitochondria, a mechanism that has been shown for many animal-bacterial symbioses [14,25,27]. The phylogeny of *Ca.* Thiosymbion based on genome-wide SNPs with two clearly diverged clades linked to two mitochondrial lineages in both of the study locations provides strong evidence for vertical transmission of this symbiont. This was also supported by the linear regression between pairwise genetic distances of *Ca*. Thiosymbion and mitochondria. Genomic features previously reported for *Ca*. Thiosymbion also corroborate its vertical transmission. Characteristic features of symbiont genomes are expected to develop upon the establishment of stringent symbiont vertical transmission linked to host sexual reproduction, due to their host-restricted lifestyle, population-size bottlenecks, reduced selective pressure against mildly deleterious mutations, and limited horizontal gene transfer [70]. Key features of symbiont genome evolution at the early stage of host restriction include proliferation of mobile elements (i.e. transposases and other insertion sequence elements), pseudogenization and gene deletions [33]. Previous genetic and proteomic studies observed abundant transposases and their expression in *Ca.* Thiosymbion in *O. algarvensis* [45,71]. While vertical transmission of the primary symbiont of *Ca.* Thiosymbion spp. has been speculated in gutless annelids based on morphological and microscopic observations of ovipositioning [55], the present study provides the first population-level phylogenomic evidence for vertical transmission of *Ca*. Thiosymbion symbiont in *O. algarvensis*.

If the *Ca*. Thiosymbion symbiont is vertically transmitted in *O. algarvensis*, it is possible that other symbionts are similarly co-transmitted in the same process. We speculate that Gamma3 symbiont is also usually transmitted maternally, as its genome-wide SNPs indicated genetic differentiation partly explained by host mitochondrial divergence while a rare horizontal transmission event between sympatric hosts was also indicated. Vertical transmission may also be at work for Delta1a and Delta1b symbionts, for which notable patterns in prevalence were observed between host mitochondrial lineages. N_e estimates for the Gamma3, Delta1a and Delta1b symbionts were lower than N_e for *Ca*. Thiosymbion (Suppl. Figure S6), which are in agreement with potential population-size bottlenecks associated with vertical transmission (see Supplementary text 1.6). Previous studies have reported abundant transposases in genomes and proteomes of Gamma3 and Delta1a symbionts, suggesting some degree of host restriction in their lifestyle [45,71]. The relative importance of vertical transmission compared to horizontal transmission however appears to vary among symbionts, as we discuss below.

Horizontal symbiont transmission and its source explain the phylogenetic clustering patterns of the Delta4 and spirochete symbionts

Our data indicated that partner fidelity for the Delta4 and spirochete symbionts is low or absent. The data instead showed a location-driven genetic divergence for the Delta4 symbiont that suggests horizontal transmission of this symbiont. There are two possible hypotheses explaining the location-based patterns of the Delta4 symbiont: (**i**) Host individuals take up Delta4 from free-living symbiont populations that are genetically differentiated between locations, or (**ii**) co-occurring *O. algarvensis* individuals often exchange symbionts locally between host maternal lineages (i.e. lineage switching). Interestingly, genetic variation of the Delta4 symbiont among host individuals was as low as that of *Ca.* Thiosymbion within the same host group (Figure 4a and 4f "Within" terms). The *Ca*. Thiosymbion symbiont likely experiences population bottlenecks between host generations, given the strong evidence presented above for vertical transmission. If Delta4 symbionts were horizontally acquired from a large environmental genetic pool of free-living bacterial populations as the explanation (**i**), we would expect a much larger genetic variation for the Delta4 symbiont than that of

Ca. Thiosymbion. Genetic diversity of free-living bacterial species in the coastal environment can be high; e.g., 16S ribosomal RNA gene (SSU) sequences of species of the sulphate-reducer *Desulfobacterium* show high conspecific diversity within a coastal saltmarsh sediment sample [72]. As no comparable population-wide SNP data was available for the free-living sulphate-reducers at our study sites, we tested for signals of nucleotide diversity in the Delta4 SSU sequences, but could not detect any (Suppl. Figure S7). It is unlikely that large and diverse free-living Delta4 populations can maintain the low genetic diversity of Delta4 symbionts among sympatric host individuals. The latter explanation (**ii**) i.e. frequent local lineage-switching of Delta4 symbionts without a free-living state of the bacteria, is thus more likely than horizontal acquisition from free-living populations. The low N_e estimates for the Delta4 symbiont per host individual indicate the low genetic diversity of Delta4 symbionts not only among co-occurring host individuals but also within host individuals (Suppl. Figure S6). Frequent host-lineage switching of Delta4 symbionts, while maintaining their low genetic diversity within and between host individuals, may be mediated by biparental or paternal symbiont-transmission modes as they can facilitate regular symbiont exchanges between co-occurring host individuals where limited numbers of symbionts are exchanged. For example, biparental transmission of nephridial symbionts was demonstrated in the earthworm *Aporrectodea tuberculata* [73], and male-borne transmission of symbionts was observed in some insect symbioses, such as aphids and tsetse fly [74,75].

The spirochete symbiont displayed no phylogenetic patterns linked to location or host mtDNA lineage. Its genetic variation within a host group was more than 2-folder greater than that of all other symbionts. This suggests that the spirochete symbiont samples are derived from genetic pools that are undifferentiated between the two locations and apparently relatively large. Such a lack of phylogenetic evidence for partner fidelity points towards a horizontally transmitted symbiont, as they often present little phylogenetic congruence with host mitochondria [11]. It is thus likely that *O. algarvensis* regularly acquires the spirochete symbiont from free-living populations in the environment. This is also in accordance with relatively large N_e estimated for the spirochete symbiont (Suppl. Figure S6). However, we cannot rule out that inter-host exchange and vertical transmission of spirochete between adjacent host generations may also occur.

Implications of the variable fidelity levels among symbionts in the evolutionary stability of multi-member symbiosis

We showed that fidelity levels for the symbionts of *O. algarvensis* are not uniformly low, as was expected from the macroevolutionary patterns observed for the larger group of gutless annelids and their symbionts. Instead, the fidelity of the symbionts can roughly be ranked as; *Ca.* Thiosymbion > Gamma3 > Delta1a ≈ Delta1b > Delta4 ≈ spirochete. Our results indicated that a difference in transmission mode could, at least partly, explain this variation in partner fidelity among the symbionts (Table 1). Variable partner fidelity levels may be extrapolated to the evolutionary duration of the symbiotic association, with a high partner fidelity linked to a long-established symbiotic association. In fact, the monophyletic group of *Ca*. Thiosymbion symbionts, for which our study showed the

Table 1. Summary of findings of the study and inference of host fidelity and potential transmission mode for each symbiont species.

	Observations			Inference	
Symbiont	**Significant effect on genetic distance**		**Other**	**Host fidelity**	**Potential transmission mode**
Ca. Thiosymbion	Matriline		Detected in all hosts. High genomic divergence between matrilines in two locations.	Very high	Vertical transmission
Gamma3	Matriline	Location	Detected in all hosts. A rare switch between host lineages in Cavoli.	High	Mixed-mode transmission (mostly vertical transmission with rare horizontal transmission events)
Delta1a			Detected in all A-hosts in both locations, but in less than half of B-hosts.	Moderate	Unclear (potentially vertical transmission)
Delta1b			Not detected in A-hosts in Sant' Andrea, but found in 65~100% of individuals in other host groups.	Moderate	Unclear (potentially vertical transmission)
Delta4		Location	Detected in most hosts. Small genetic variability within host group (comparable to that of *Ca.* Thiosymbion).	Low	Mixed-mode transmission (frequent horizontal transmission from co-occurring hosts)
Spirochete			Detected in all hosts. Large genetic variability within host group (>2-fold greater than that of *Ca.* Thiosymbion).	Low	Horizontal transmission, or mixed-mode transmission (frequent horizontal transmission from free-living symbionts)

highest fidelity, form a stable symbiosis in diverse species of gutless annelids (genera *Olavius* and *Inanidrilus*), as well as with nematodes (subfamily Stilbonematinae and genus *Astomonema*) [5], suggesting the symbiotic associations of Ca. Thiosymbion lasting throughout the divergence of respective host animal lineages. Taxonomic compositions of symbionts associated with *O. algarvensis* and *Olavius ilvae* showed that these co-occurring gutless annelid species share several symbiont clades such as *Ca*. Thiosymbion, Gamma3 and Delta1 (Delta1a or/and Delta1b) symbionts, but that the Delta4 and spirochete symbionts were only found in *O. algarvensis* [47]. This host-dependent association of the Delta4 and spirochete symbionts, for which we detected low fidelity levels, could result if they established the symbiotic relationships more recently than the divergence of *O. algarvensis* and *O. ilvae*.

A faithfully-transmitted symbiosis might provide short-term fitness advantages by ensuring offspring can harbour beneficial symbionts but can develop evolutionary costs arising from irreversible host-symbiont co-dependence [1], e.g. via ecological niche limitation that in turn leads to increased extinction rates of the whole system (the 'evolutionary rabbit hole' [32]). Acquisition of a novel facultative symbiont on the other hand can serve as a source of ecological innovations for the symbiosis to exploit new resources and expand to or cope with new habitats [37,76,77], thereby ameliorating the evolutionary costs. Regularly recruiting a novel symbiont strain also re-enables gene flux via horizontal gene transfer and genome recombination among the conspecific symbionts, which can slow down reductive genome evolution of symbiont and reduce the risk of symbiosis extinction [29,32,34,67]. Harbouring multiple symbionts with various degrees of fidelity could therefore provide a certain evolutionary advantage, as low-fidelity symbionts provide new genetic materials while the primary symbiont ensures the long-term survival of the symbiosis. Studies on the mitigation of the evolutionary cost through compositional shifts in a multi-member symbiosis have thus far centred at macroevolutionary scales based on comparisons of multiple host species [41-43]. In contrast to these studies, our population-scale metagenomes from highly replicated, closely related samples revealed hitherto-unreported intraspecific variability in symbiotic composition and partner fidelity, suggesting that such patterns in multi-member symbioses can be observable at finer evolutionary timeframes.

Conclusions and outstanding questions

Based on the whole-genome SNP phylogenies of multiple symbionts and their genetic co-divergence patterns with host mitochondria, we demonstrated that partner fidelity levels can vary in the association between a host and multiple endosymbionts within an obligate animal-bacterial symbiosis. Variable levels of partner fidelity likely confer evolutionary advantages to the symbiotic system through a regular influx of genetic materials that promote the survival of the symbiosis over evolutionally timeframes. With the advancement of sequencing technologies, molecular studies have increasingly revealed the complex and dynamic nature of multi-member symbioses [76,78]. We predict that variability in the host-symbiont fidelity likely has a role in shaping other symbiotic consortia. Several questions emerge from this study in the context of multi-member animal-bacterial symbiosis: (1) Is the level of partner fidelity linked to the level of host dependency in each symbiont, and what functional traits of the symbiont contribute to the dependency? (2) What constrains the diversification of symbiont compositions across geographic distances or host genetic divergence? (3) Does the presence of symbionts with different levels of fidelity facilitate horizontal gene transfer between symbiont and host? If so, does this have a role in evolutionary radiation of a host species complex? Given that the current understanding of the evolution of animal-bacterial symbioses has largely stemmed from well-defined model organisms with a single symbiotic microorganism, answering these questions by scaling our phylogenomic approaches in multi-member symbioses over a range of host taxonomic scales will contribute to better evolutionary understanding of animal-microbial symbiosis and wider animal-microbial interactions in nature.

Methods

Specimen collection and host mitochondrial lineage screening

A total of 579 *O. algarvensis* individuals from 2 locations on Elba, Italy (Sant' Andrea; 42°48'31"N / 10°08'33"E; n=346, and Cavoli, 42°44'05"N / 10°11'12"E; n=233; Figure 1c, 1d) were screened for their mitochondrial lineages based on sequences of the mitochondrial COI gene (COI-haplotypes). *O. algarvensis* individuals were collected between 2010 and 2016 from sandy sediments in the vicinity of

Posidonia oceanica seagrass beds at water depths between 7 and 14 m, as previously described [46]. Live specimens were (**i**) flash-frozen in liquid nitrogen and stored at -80° C, or (**ii**) immersed in RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) and stored at 4° C. DNA of the specimens was individually extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden) according to the manufactures instructions. Approximately 670 bp of COI gene was amplified with PCR using the primer set of COI-1490F (5'-GGT-CAA-CAA-ATC-ATA-AAG-ATA-TTG-G-3') and COI-2189R (5'-TAA-ACT-TCA-GGG-TGA-CCA-AAA-AAT-CA-3') as previously described [79], and sequenced using the BigDye Sanger sequencing kit (Life Technologies, Darmstadt, Germany) with the COI-2189R primer on the Applied Biosystems Hitachi capillary sequencer (Applied Biosystems, Waltham, USA). COI sequences were quality-filtered with a maximum error rate of 0.5% and subsequently aligned using MAFFT v7.45 [80] in Geneious software v11.0.3 (Biomatters, Auckland, New Zealand). A COI-haplotype network was built based on a 525-bp core alignment using the TCS statistical parsimony algorithm [81] implemented in PopART v1.7 [82]. Twenty *O. algarvensis* individuals from each host group (i.e. the combination of sample location, Sant' Andrea or Cavoli, and COI-haplotype A or B; 4 groups in total) were selected for metagenome sequencing (n=80 individuals in total).

Metagenome sequencing

Sequencing libraries were constructed from individual total DNA using a Tn5 transposase purification and tagmentation protocol [83] as previously adopted in a scalable population genetic study [84]. The Tn5 transposase was provided by the Protein Science Facility at Karolinska Institutet Sci Life Lab (Solna, Sweden). Quantity and quality of DNA samples were checked with the Quantus Fluorometer with the QuantiFluor dsDNA System (Promega Corporation, Madison, WI, USA), the Agilent TapeStation System with the DNA ScreenTape (Agilent Technologies, Santa Clara, CA, USA), and the FEMTO Pulse genomic DNA analysis kit (Advanced Analytical Technologies Inc., Heidelberg, Germany) prior to library construction. Insert template DNA was size-selected targeting 400 - 500 bp using the AMPure XP (Beckman Coulter, Indianapolis, IN, USA), and paired-end 150 bp sequences were generated using the Illumina HiSeq3000 System (Illumina, San Diego, CA, USA), with an average

yield of 2 Gbp per sample. Construction and quality control of libraries and sequencing were performed at the Max Planck-Genomic-centre (Cologne, Germany).

Assembly of reference genomes of endosymbionts and host mitochondrion

Metagenome-assembled genomes (MAGs) of the *Ca*. Thiosymbion, Gamma3, Delta4 and spirochete symbionts were *de-novo* assembled from a deeply-sequenced metagenome of an *O. algarvensis* individual available at the European Nucleotide Archive (ENA) project PRJEB28157 (Specimen ID: OalgB6SA; approx.7 Gbp; Suppl. Table S1). Raw metagenomic reads were first adapter-trimmed and quality-filtered with length ≥ 36 bp and Phred quality score ≥ 2 using *bbduk* of BBTools v36.86 (http://jgi.doe.gov/data-and-tools/bbtools), and corrected for sequencing errors using BayesHammer [85] implemented in SPAdes v3.9.1 [86]. Clean reads were assembled using MEGAHIT v1.0.6 [87], and symbiont genome bins were identified with MetaBAT v0.26.3 [88]. Bins were respectively assigned to the *Ca.* Thiosymbion, Gamma3, Delta4 and spirochete symbionts based on 99 ~ 100% sequence matches with their reference SSU sequences (NCBI accession numbers AF328856, AJ620496, AJ620497, AJ620502 [35,47], respectively). Bins were further refined using Bandage v0.08.1 [89] by identifying and inspecting connected contigs on assembly graphs. In addition, MAGs of Delta1a, Delta1b and Delta3 were obtained from the public database deposited under the ENA project accession number PRJEB28157 [52,53]. Completeness and contamination rate of the MAGs were estimated with CheckM version 1.0.7 [90], and assembly statistics of symbiont genomes were calculated with QUAST v5.0.2 [91] (Suppl. Table S1).

Reference complete mitochondrial genomes (mtDNA) were assembled from two metagenomes of *O. algarvensis* from Sant' Andrea representing different COI-haplotypes (Specimen IDs: OalgSANT_A04 and OalgSANT_B04 for COI-haplotypes A and B, respectively). For this process, duplicated sequences due to PCR-amplification during the library preparation were removed using FastUniq v1.1 [92] prior to adapter-clipping and quality-trimming using Trimmomatic v0.36 [93]. A temporary mtDNA scaffold was first generated by iterative mapping of the clean reads to a reference COI sequence of *O. algarvensis* (NCBI accession number KP943854) as a bait sequence with MITObim v1.9 [94]. Mitochondrial reads were then identified by mapping to this mtDNA scaffold using

bbmap of BBTools, and mtDNA was assembled from the identified reads with SPAdes. A circular path of mtDNA was identified on assembly graphs in Bandage, and annotated using MITOS2 webserver (http://mitos2.bioinf.uni-leipzig.de) to confirm the completeness [95].

Characterization of symbiont community compositions

Overall taxonomic compositions of *O. algarvensis* metagenomes were initially screened by identifying SSU sequences using phyloFlash v3.3-beta1 [96] using the SILVA SSU database release 132 [97] as reference. SSU sequences of *O. algarvensis* symbionts were assembled with SPAdes implemented in phyloFlash, and aligned per symbiont species using MAFFT in Geneious to identify SNP sites. Chimeric and incomplete (<1,100 bp) SSU sequences were excluded from the alignments.

Relative abundances of *O. algarvensis* symbionts were estimated by mapping metagenome reads to a collection of symbiont-specific single-copy gene (SCG) sequences that were extracted from genome bins in this study. Orthologous SCG sequences were first identified within each of the reference symbiont MAGs with CheckM. To ensure unambiguous taxon-differentiation, duplicated genes detected in each symbiont MAG (i.e. those labelled as 'contamination' in CheckM) as well as sequences sharing >90% nucleotide identity between multiple symbiont species (checked with CD-HIT v4.5.4 [98]; 8 cases identified between the Delta1a and Delta1b symbionts) were removed from the final SCG reference sequences. Metagenomic reads (quality-filtered with the same processes as for the mtDNA assembly above) matching to the SCG sequences were quantified using Kallisto v0.44.0 [99]. Read coverage was exceptionally high for a small number of the SCGs (especially in *Ca.* Thiosymbion; Suppl. Figure S1); a closer look at sequence alignment revealed that this is likely due to short sequence repeats present in the corresponding symbiont genome that are also contained in small portions of SCG sequences. Therefore, we calculated the mean read coverage from SCGs in the interquartile range (i.e. genes between 25 and 75 percentiles based on coverage ranking) to avoid overestimating abundances of symbionts and obtain a robust estimate for symbiont relative abundance. Symbiont composition estimates were plotted in R using ggplot2 package v3.2.1 [100].

To examine whether our phylogenetic analysis of *O. algarvensis* symbionts reflects a single dominant genotype per symbiont species per host individual, levels of symbiont genotype diversity

within an individual host were assessed by calculating SNP-densities (the number of SNP sites per Kbp of reference genome) with previously established procedures [59]. This analysis was performed using a set of publicly available deeply-sequenced metagenomes of *O. algarvensis* listed in Suppl. Table S2a to ensure that a SNP-density is estimated from sufficient symbiont read-coverages and that the estimates can be compared to those in other studies performing similar assessments [29,58,59].

Identification of single nucleotide polymorphisms and phylogenetic reconstruction

For identification of SNPs in the genomes of symbionts and mitochondria, the quality-controlled reads were first unambiguously split into different symbiont species using *bbsplit* of BBTools, using the reference symbiont MAGs from above. Mitochondrial reads were identified with the reference mtDNA sequence derived from the specimen "OalgSANT_A04" (Sant' Andrea, COI-haplotype A) using *bbmap* with a minimum nucleotide identity of 95%; This step was performed to remove potential contaminations from sequences of nuclear mitochondrial pseudogenes divergent from the mtDNA [101], while ensuring successful mapping of mtDNA reads in the analysis for B-hosts given the high nucleotide identity of mtDNAs between two lineages A and B in this study (>99%; see results).

To reconstruct phylogenies of symbionts and mitochondria, SNPs in genomes of symbionts and host mitochondria were identified using two approaches; based on (**i**) posterior genotype probabilities without genotype calling, and (**ii**) deterministic genotyping only at genetic positions that were deeply sequenced in all samples. For the SNP-identification without genotyping (**i**), the symbiont- and mitochondrial-reads identified above were mapped onto individual reference genomes of symbionts and mitochondria using *bbmap*. Mapping files were filtered based on mapping quality using samtools v1.3.1 [102] and BamUtil v1.0.14 [103], deduplicated with *MarkDuplicates* of Picard Toolkit v2.9.2 (https://github.com/broadinstitute/picard), and realigned around indels with Genome Analysis Tool Kit v3.7 [104]. Posterior genotype probabilities were calculated with ANGSD v0.929 [105]. To deal with haploid genotypes, an inbreeding coefficient F of '1' (i.e. assuming that all genotypes are 'homozygous') and an uniform prior were specified in ANGSD. SNP sites were identified as reference nucleotide positions that were covered ≥1× by all samples and showed a statistically significant support (SNP p-value < 0.01). When no SNP site was found due to the lack of reads from specific

symbionts or their extremely low abundance in some samples, these samples were excluded based on a cut-off of lateral coverages (% reference genetic sites covered by reads). For each symbiont and mitochondria, their pairwise genetic distances among samples were obtained directly from the matrix of genotype probabilities using NGSdist v1.0.2 [61]. Phylogenetic trees with bootstrap-support were computed from the resulting distant matrix using FastME v2.1.5.1 [106] and RAxML v8.2.11 [107].

For the SNP-identification by deterministic genotyping (**ii**), the same symbiont- and mitochondrial-reads were analyzed with the SNIPPY pipeline v3.2 (https://github.com/tseemann/snippy), with the same reference genomes of mtDNA and symbionts as above. In this process, genotypes were first called at all reference nucleotide positions with a minimum coverage of 5 for each sample. Core SNP sites were subsequently identified among these genotype-called sites that were covered ≥5× in all samples. Similar to the approach (**i**) above, when no core SNP site was found, samples with insufficient genotype data were excluded using a cut-off of lateral coverage (% reference sites with coverage ≥5×). Phylogenetic trees with bootstrap-support were computed from resulting core SNP nucleotide alignment with IQ-TREE v1.5.5 [108] using a best-model finder implemented within IQ-TREE. Phylogenetic trees of the symbionts and mitochondria were visualized in the iTOL web tool [109].

Analyses of phylogenetic divergence of symbionts based on host mitochondrial lineages and locations

To examine genetic co-divergence patterns between a symbiont and host mtDNA, regressions of pairwise genetic distances estimated with NGSdist were examined using Mantel tests implemented in the R-package *vegan* v2.4.4 [110]. To inspect which specific factors drive the patterns of genetic divergence in mitochondria and symbionts, the pairwise genetic distances were statistically compared among 3 categories of host pairs; (**a**) within the same host group (samples from the same combination of COI-haplotype and location), (**b**) between COI-haplotypes (A vs. B in Cavoli, and A vs. B in Sant' Andrea), and (**c**) between locations (Cavoli vs. Sant' Andrea in A-hosts, and Cavoli vs. Sant' Andrea in B-hosts). Sample pairs in each category were randomly selected without replacement to ensure the data independence. The statistical analyses were performed for each of the symbionts and

mitochondria using Kruskal-Wallis rank sum test and Dunn post-hoc tests with p-value adjustment by controlling the false-discovery rate, using R core package v3.4.2 [111] and FSA package v0.8.25 [112].

Inference of symbiont effective population sizes within O. algarvensis individuals

Effective population size (N_e) within a host individual for each symbiont was estimated based on genome-wide SNPs. We used the Watterson's estimator θ to calculate N_e using the equation; $\theta = 2\ Ne\ \mu$, where μ is the mutation rate [113], which was estimated as $\mu = 2.73\times10^{-10}$ per site per generation (see Supplementary Text 1.6). Watterson's θ was computed using the R package pegas v0.14 [114] based on the number of SNP-sites within a sample individual animal and the number of sequences (as the mean sequence coverage), both of which were computed as previously described [59]. Samples with mean sequence coverage < 5× were excluded from the calculation of N_e to ensure meaningful identification of within-sample SNPs.

Data and script availability

Raw metagenome sequences, reference genomes of mitochondria and symbionts generated in this study are deposited in the European Nucleotide Archive (ENA) under accession number PRJEB42310. Annotated bioinformatic scripts with all specific parameters, as well as reference SCG sequences, are available in a GitHub depository (https://github.com/yuisato/Oalg_linkage).

Acknowledgements

This work was supported by the Max Planck Society, a Moore Foundation Marine Microbial Initiative Investigator Award to ND (Grant GBMF3811), a U.S. National Science Foundation award to MK (grant IOS 2003107), the USDA National Institute of Food and Agriculture Hatch project 1014212 (MK), and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 660280 (CW).

Author Contributions

ND, MK, CW and JW conceived the study and MK, ND, YS, JW and HGV designed the work. YS, JW, CW, MS and MK acquired materials and data, YS, JW and RA analyzed data. YS, JW, RA, ND and MK interpreted results. YS drafted the work, and MK, HGV and ND provided revisions.

Competing Interests

The authors declare no competing interest.

References

1 Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T. & West, S. A. The evolution of host-symbiont dependence. *Nature Communications* **8**, 15973, doi:10.1038/ncomms15973 (2017).

2 Distel, D. L. *et al.* Sulfur-oxidizing bacterial endosymbionts: analysis of phylogeny and specificity by 16S rRNA sequences. *Journal of Bacteriology* **170**, 2506-2510, doi:10.1128/jb.170.6.2506-2510.1988 (1988).

3 Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Reviews Microbiology* **6**, 725-740 (2008).

4 Munson, M. A. *et al.* Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *Journal of Bacteriology* **173**, 6321-6324, doi:10.1128/jb.173.20.6321-6324.1991 (1991).

5 Zimmermann, J. *et al.* Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. *Molecular Ecology* **25**, 3203-3223, doi:doi:10.1111/mec.13554 (2016).

6 Di Meo, C. A. *et al.* Genetic Variation among Endosymbionts of Widely Distributed Vestimentiferan Tubeworms. *Applied and Environmental Microbiology* **66**, 651-658, doi:10.1128/aem.66.2.651-658.2000 (2000).

7 Distel, D. L., Felbeck, H. & Cavanaugh, C. M. Evidence for phylogenetic congruence among sulfur-oxidizing chemoautotrophic bacterial endosymbionts and their bivalve hosts. *Journal of Molecular Evolution* **38**, 533-542, doi:10.1007/bf00178852 (1994).

8 Raggi, L., Schubotz, F., Hinrichs, K.-U., Dubilier, N. & Petersen, J. M. Bacterial symbionts of *Bathymodiolus* mussels and *Escarpia* tubeworms from Chapopote, an asphalt seep in the southern Gulf of Mexico. *Environmental Microbiology* **15**, 1969-1987, doi:10.1111/1462-2920.12051 (2013).

9 Visick, K. L. & McFall-Ngai, M. J. An Exclusive Contract: Specificity in the *Vibrio fischeri-Euprymna scolopes* Partnership. *Journal of Bacteriology* **182**, 1779-1787, doi:10.1128/jb.182.7.1779-1787.2000 (2000).

10 McFall-Ngai, M. Divining the Essence of Symbiosis: Insights from the Squid-Vibrio Model. *PLOS Biology* **12**, e1001783, doi:10.1371/journal.pbio.1001783 (2014).

11 Douglas, A. E. & Werren, J. H. Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts. *mBio* **7**, e02099-02015, doi:10.1128/mBio.02099-15 (2016).

730 12 Koga, R., Meng, X.-Y., Tsuchida, T. & Fukatsu, T. Cellular mechanism for selective vertical
731 transmission of an obligate insect symbiont at the bacteriocyte–embryo interface.
732 *Proceedings of the National Academy of Sciences* **109**, E1230-E1237,
733 doi:10.1073/pnas.1119212109 (2012).

734 13 Cary, S. C. & Giovannoni, S. J. Transovarial inheritance of endosymbiotic bacteria in clams
735 inhabiting deep-sea hydrothermal vents and cold seeps. *Proceedings of the National*
736 *Academy of Sciences* **90**, 5695-5699, doi:10.1073/pnas.90.12.5695 (1993).

737 14 Hosokawa, T., Kikuchi, Y., Nikoh, N., Shimada, M. & Fukatsu, T. Strict Host-Symbiont
738 Cospeciation and Reductive Genome Evolution in Insect Gut Bacteria. *PLOS Biology* **4**, e337,
739 doi:10.1371/journal.pbio.0040337 (2006).

740 15 Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nature*
741 *Reviews Microbiology* **8**, 218-230, doi:10.1038/nrmicro2262 (2010).

742 16 Won, Y.-J. *et al.* Environmental acquisition of thiotrophic endosymbionts by deep-sea
743 mussels of the genus *Bathymodiolus*. *Applied and Environmental Microbiology* **69**, 6785-6792,
744 doi:10.1128/aem.69.11.6785-6792.2003 (2003).

745 17 Cary, S. C., Warren, W., Anderson, E. & Giovannoni, S. J. Identification and localization of
746 bacterial endosymbionts in hydrothermal vent taxa with symbiont-specific polymerase chain
747 reaction amplification and in situ hybridization techniques. *Molecular marine biology and*
748 *biotechnology* **2**, 51-62 (1993).

749 18 Broughton, W. J. & Perret, X. Genealogy of legume-Rhizobium symbioses. *Current Opinion in*
750 *Plant Biology* **2**, 305-311, doi:10.1016/s1369-5266(99)80054-5 (1999).

751 19 Madsen, E. B. *et al.* A receptor kinase gene of the LysM type is involved in legume perception
752 of rhizobial signals. *Nature* **425**, 637-640, doi:10.1038/nature02045 (2003).

753 20 Kaltenpoth, M. *et al.* Partner choice and fidelity stabilize coevolution in a Cretaceous-age
754 defensive symbiosis. *Proceedings of the National Academy of Sciences* **111**, 6359-6364,
755 doi:10.1073/pnas.1400457111 (2014).

756 21 Hurtado, L. A., Mateos, M., Lutz, R. A. & Vrijenhoek, R. C. Coupling of Bacterial Endosymbiont
757 and Host Mitochondrial Genomes in the Hydrothermal Vent Clam *Calyptogena magnifica*.
758 *Applied and Environmental Microbiology* **69**, 2058-2064, doi:10.1128/aem.69.4.2058-
759 2064.2003 (2003).

760 22 Liu, L., HUANG, X., ZHANG, R., JIANG, L. & QIAO, G. Phylogenetic congruence between
761 Mollitrichosiphum (Aphididae: Greenideinae) and Buchnera indicates insect–bacteria parallel
762 evolution. *Systematic Entomology* **38**, 81-92, doi:10.1111/j.1365-3113.2012.00647.x (2013).

763 23 Funk, D. J., Helbling, L., Wernegreen, J. J. & Moran, N. A. Intraspecific phylogenetic
764 congruence among multiple symbiont genomes. *Proceedings of the Royal Society of London.*
765 *Series B: Biological Sciences* **267**, 2517-2521, doi:10.1098/rspb.2000.1314 (2000).

766 24 Symula, R. E. *et al.* Influence of Host Phylogeographic Patterns and Incomplete Lineage
767 Sorting on Within-Species Genetic Variability in *Wigglesworthia* Species, Obligate Symbionts
768 of Tsetse Flies. *Applied and Environmental Microbiology* **77**, 8400-8408,
769 doi:10.1128/aem.05688-11 (2011).

770 25 Peek, A. S., Feldman, R. A., Lutz, R. A. & Vrijenhoek, R. C. Cospeciation of chemoautotrophic
771 bacteria and deep sea clams. *Proceedings of the National Academy of Sciences of the United*
772 *States of America* **95**, 9962-9966, doi:10.1073/pnas.95.17.9962 (1998).

773 26 Bandi, C., Anderson, T. J., Genchi, C. & Blaxter, M. L. Phylogeny of *Wolbachia* in filarial
774 nematodes. *Proceedings of the Royal Society B: Biological Sciences* **265**, 2407-2413 (1998).

775 27 Allen, J. M., Reed, D. L., Perotti, M. A. & Braig, H. R. Evolutionary Relationships of
776 "*Candidatus* Riesia spp.," Endosymbiotic *Enterobacteriaceae* Living within Hematophagous
777 Primate Lice. *Applied and Environmental Microbiology* **73**, 1659-1664,
778 doi:10.1128/aem.01877-06 (2007).

779 28 Stewart, F. J., Young, C. R. & Cavanaugh, C. M. Lateral Symbiont Acquisition in a Maternally
780 Transmitted Chemosynthetic Clam Endosymbiosis. *Molecular Biology and Evolution* **25**, 673-
781 687, doi:10.1093/molbev/msn010 (2008).

782 29 Russell, S. L., Corbett-Detig, R. B. & Cavanaugh, C. M. Mixed transmission modes and dynamic
783 genome evolution in an obligate animal-bacterial symbiosis. *ISME J*,
784 doi:10.1038/ismej.2017.10 (2017).

785 30 Merckx, V. & Bidartondo, M. I. Breakdown and delayed cospeciation in the arbuscular
786 mycorrhizal mutualism. *Proceedings. Biological sciences* **275**, 1029-1035,
787 doi:10.1098/rspb.2007.1622 (2008).

788 31 Won, Y.-J., Jones, W. J. & Vrijenhoek, R. C. Absence of Cospeciation Between Deep-Sea
789 Mytilids and Their Thiotrophic Endosymbionts. *Journal of Shellfish Research* **27**, 129-138, 110
790 (2008).

791 32 Bennett, G. M. & Moran, N. A. Heritable symbiosis: The advantages and perils of an
792 evolutionary rabbit hole. *Proceedings of the National Academy of Sciences* **112**, 10169-10176,
793 doi:10.1073/pnas.1421388112 (2015).

794 33 Moran, N. A. & Plague, G. R. Genomic changes following host restriction in bacteria. *Current*
795 *Opinion in Genetics & Development* **14**, 627-633, doi:10.1016/j.gde.2004.09.003 (2004).

796 34 Stewart, F. J., Young, C. R. & Cavanaugh, C. M. Evidence for Homologous Recombination in
797 Intracellular Chemosynthetic Clam Symbionts. *Molecular Biology and Evolution* **26**, 1391-
798 1404, doi:10.1093/molbev/msp049 (2009).

799 35 Dubilier, N. *et al.* Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an
800 oligochaete worm. *Nature* **411**, 298-302, doi:10.1038/35077067 (2001).

801 36 Moran, N. A., McCutcheon, J. P. & Nakabachi, A. Genomics and Evolution of Heritable
802 Bacterial Symbionts. *Annual Review of Genetics* **42**, 165-190,
803 doi:10.1146/annurev.genet.41.110306.130119 (2008).

804 37 Manzano-Marín, A. *et al.* Serial horizontal transfer of vitamin-biosynthetic genes enables the
805 establishment of new nutritional symbionts in aphids' di-symbiotic systems. *The ISME Journal*
806 **14**, 259-273, doi:10.1038/s41396-019-0533-6 (2020).

807 38 Ponnudurai, R. *et al.* Metabolic and physiological interdependencies in the *Bathymodiolus*
808 *azoricus* symbiosis. *ISME Journal*, doi:10.1038/ismej.2016.124 (2016).

809 39 Lamelas, A. *et al.* Serratia symbiotica from the Aphid Cinara cedri: A Missing Link from
810 Facultative to Obligate Insect Endosymbiont. *PLOS Genetics* **7**, e1002357,
811 doi:10.1371/journal.pgen.1002357 (2011).

812 40 McCutcheon, John P. & von Dohlen, Carol D. An Interdependent Metabolic Patchwork in the
813 Nested Symbiosis of Mealybugs. *Current Biology* **21**, 1366-1372,
814 doi:10.1016/j.cub.2011.06.051 (2011).

815 41 Toenshoff, E. R., Gruber, D. & Horn, M. Co-evolution and symbiont replacement shaped the
816 symbiosis between adelgids (Hemiptera: Adelgidae) and their bacterial symbionts.
817 *Environmental Microbiology* **14**, 1284-1295, doi:10.1111/j.1462-2920.2012.02712.x (2012).

818 42 Koga, R. & Moran, N. A. Swapping symbionts in spittlebugs: evolutionary replacement of a
819 reduced genome symbiont. *The ISME Journal* **8**, 1237-1246, doi:10.1038/ismej.2013.235
820 (2014).

43 Husnik, F. & McCutcheon, J. P. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proceedings of the National Academy of Sciences* **113**, E5416-E5424, doi:10.1073/pnas.1603910113 (2016).

44 Erséus, C., Wetzel, M. J. & Gustavsson, L. ICZN rules–a farewell to Tubificidae (Annelida, Clitellata). *Zootaxa* **1744**, 66-68 (2008).

45 Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950-955, doi:10.1038/nature05192 (2006).

46 Kleiner, M. *et al.* Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proceedings of the National Academy of Sciences* **109**, E1173–E1182, doi:10.1073/pnas.1121198109 (2012).

47 Ruehland, C. *et al.* Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. *Environmental Microbiology* **10**, 3404-3416, doi:10.1111/j.1462-2920.2008.01728.x (2008).

48 Bergin, C. *et al.* Acquisition of a Novel Sulfur-Oxidizing Symbiont in the Gutless Marine Worm *Inanidrilus exumae*. *Applied and Environmental Microbiology* **84**, doi:10.1128/aem.02267-17 (2018).

49 Blazejak, A., Kuever, J., Erséus, C., Amann, R. & Dubilier, N. Phylogeny of 16S rRNA, Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase, and Adenosine 5′ -Phosphosulfate Reductase Genes from Gamma- and Alphaproteobacterial Symbionts in Gutless Marine Worms (Oligochaeta) from Bermuda and the Bahamas. *Applied and Environmental Microbiology* **72**, 5527-5536, doi:10.1128/aem.02441-05 (2006).

50 Blazejak, A., Erséus, C., Amann, R. & Dubilier, N. Coexistence of Bacterial Sulfide Oxidizers, Sulfate Reducers, and Spirochetes in a Gutless Worm (Oligochaeta) from the Peru Margin. *Applied and Environmental Microbiology* **71**, 1553-1561, doi:10.1128/aem.71.3.1553-1561.2005 (2005).

51 Dubilier, N. *et al.* Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). *Marine Ecology Progress Series* **178**, 271-280 (1999).

52 Sato, Y., Wippler, J., Wentrup, C., Dubilier, N. & Kleiner, M. High-Quality Draft Genome Sequences of Two Deltaproteobacterial Endosymbionts, Delta1a and Delta1b, from the Uncultured Sva0081 Clade, Assembled from Metagenomes of the Gutless Marine Worm *Olavius algarvensis*. *Microbiology Resource Announcements* **9**, e00276-00220, doi:10.1128/mra.00276-20 (2020).

53 Sato, Y. *et al.* High-Quality Draft Genome Sequences of the Uncultured Delta3 Endosymbiont (*Deltaproteobacteria*) Assembled from Metagenomes of the Gutless Marine Worm *Olavius algarvensis*. *Microbiology Resource Announcements* **9**, e00704-00720, doi:10.1128/mra.00704-20 (2020).

54 Giere, O. & Langheld, C. Structural organisation, transfer and biological fate of endosymbiotic bacteria in gutless oligochaetes. *Marine Biology* **93**, 641-650, doi:10.1007/BF00392801 (1987).

55 Giere, O. Ecology and Biology of Marine Oligochaeta – an Inventory Rather than another Review. *Hydrobiologia* **564**, 103-116, doi:10.1007/s10750-005-1712-1 (2006).

56 Bright, M. & Giere, O. Microbial symbiosis in Annelida. *Symbiosis* **38**, 1-45 (2005).

57 Krueger, D. M., Gustafson, R. G. & Cavanaugh, C. M. Vertical Transmission of Chemoautotrophic Symbionts in the Bivalve *Solemya velum* (Bivalvia: Protobranchia). *The Biological Bulletin* **190**, 195-202, doi:10.2307/1542539 (1996).

866 58 Robidart, J. C. *et al.* Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed
867 through metagenomics. *Environmental Microbiology* **10**, 727-737, doi:10.1111/j.1462-
868 2920.2007.01496.x (2008).

869 59 Ansorge, R. *et al.* Functional diversity enables multiple symbiont strains to coexist in deep-
870 sea mussels. *Nature Microbiology* **4**, 2487-2497, doi:10.1038/s41564-019-0572-9 (2019).

871 60 Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. & Wang, J. SNP Calling, Genotype Calling,
872 and Sample Allele Frequency Estimation from New-Generation Sequencing Data. *PLOS ONE* **7**,
873 e37558, doi:10.1371/journal.pone.0037558 (2012).

874 61 Vieira, F. G., Lassalle, F., Korneliussen, T. S. & Fumagalli, M. Improving the estimation of
875 genetic distances from Next-Generation Sequencing data. *Biological Journal of the Linnean*
876 *Society* **117**, 139-149, doi:10.1111/bij.12511 (2016).

877 62 Duperron, S. *et al.* Diversity, relative abundance and metabolic potential of bacterial
878 endosymbionts in three Bathymodiolus mussel species from cold seeps in the Gulf of Mexico.
879 *Environmental Microbiology* **9**, 1423-1438, doi:10.1111/j.1462-2920.2007.01259.x (2007).

880 63 Tsuchida, T., Koga, R., Shibao, H., Matsumoto, T. & Fukatsu, T. Diversity and geographic
881 distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid,
882 *Acyrthosiphon pisum*. *Molecular Ecology* **11**, 2123-2135, doi:10.1046/j.1365-
883 294X.2002.01606.x (2002).

884 64 Rock, D. I. *et al.* Context-dependent vertical transmission shapes strong endosymbiont
885 community structure in the pea aphid, *Acyrthosiphon pisum*. *Molecular Ecology* **27**, 2039-
886 2056, doi:10.1111/mec.14449 (2018).

887 65 Kleiner, M. *et al.* Metaproteomics method to determine carbon sources and assimilation
888 pathways of species in microbial communities. *Proceedings of the National Academy of*
889 *Sciences*, doi:10.1073/pnas.1722325115 (2018).

890 66 Kleiner, M. *et al.* Use of carbon monoxide and hydrogen by a bacteria–animal symbiosis from
891 seagrass sediments. *Environmental Microbiology*, DOI: 10.1111/1462-2920.12912,
892 doi:10.1111/1462-2920.12912 (2015).

893 67 Russell, S. L. *et al.* Horizontal transmission and recombination maintain forever young
894 bacterial symbiont genomes. *PLOS Genetics* **16**, e1008935,
895 doi:10.1371/journal.pgen.1008935 (2020).

896 68 Giska, I., Sechi, P. & Babik, W. Deeply divergent sympatric mitochondrial lineages of the
897 earthworm *Lumbricus rubellus* are not reproductively isolated. *BMC Evolutionary Biology* **15**,
898 217, doi:10.1186/s12862-015-0488-9 (2015).

899 69 Donnelly, R. K. *et al.* Nuclear DNA recapitulates the cryptic mitochondrial lineages of
900 Lumbricus rubellus and suggests the existence of cryptic species in an ecotoxological soil
901 sentinel. *Biological Journal of the Linnean Society* **110**, 780-795, doi:10.1111/bij.12171 (2013).

902 70 Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria.
903 *Proceedings of the National Academy of Sciences of the United States of America* **93**, 2873-
904 2878 (1996).

905 71 Kleiner, M., Young, J. C., Shah, M., VerBerkmoes, N. C. & Dubilier, N. Metaproteomics Reveals
906 Abundant Transposase Expression in Mutualistic Endosymbionts. *mBio* **4**, e00223-00213,
907 doi:10.1128/mBio.00223-13 (2013).

908 72 Klepac-Ceraj, V. *et al.* High overall diversity and dominance of microdiverse relationships in
909 salt marsh sulphate-reducing bacteria. *Environmental Microbiology* **6**, 686-698,
910 doi:10.1111/j.1462-2920.2004.00600.x (2004).

911 73 Paz, L.-C., Schramm, A. & Lund, M. B. Biparental transmission of *Verminephrobacter*
912 symbionts in the earthworm *Aporrectodea tuberculata* (Lumbricidae). *FEMS Microbiology*
913 *Ecology* **93**, doi:10.1093/femsec/fix025 (2017).

914 74 Moran, N. A. & Dunbar, H. E. Sexual acquisition of beneficial symbionts in aphids.
915 *Proceedings of the National Academy of Sciences* **103**, 12803-12806,
916 doi:10.1073/pnas.0605772103 (2006).

917 75 De Vooght, L., Caljon, G., Van Hees, J. & Van Den Abbeele, J. Paternal Transmission of a
918 Secondary Symbiont during Mating in the Viviparous Tsetse Fly. *Molecular Biology and*
919 *Evolution* **32**, 1977-1980, doi:10.1093/molbev/msv077 (2015).

920 76 Sudakaran, S., Kost, C. & Kaltenpoth, M. Symbiont Acquisition and Replacement as a Source
921 of Ecological Innovation. *Trends in Microbiology* **25**, 375-390, doi:10.1016/j.tim.2017.02.014
922 (2017).

923 77 Tsuchida, T., Koga, R. & Fukatsu, T. Host Plant Specialization Governed by Facultative
924 Symbiont. *Science* **303**, 1989 (2004).

925 78 Ferrari, J. & Vavre, F. Bacterial symbionts in insects or the story of communities affecting
926 communities. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**,
927 1389-1400, doi:10.1098/rstb.2010.0226 (2011).

928 79 Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of
929 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.
930 *Molecular marine biology and biotechnology* **3**, 294-299 (1994).

931 80 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
932 Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-780,
933 doi:10.1093/molbev/mst010 (2013).

934 81 Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate gene
935 genealogies. *Mol Ecol* **9**, 1657-1659 (2000).

936 82 Leigh, J. W. & Bryant, D. popart: full-feature software for haplotype network construction.
937 *Methods in Ecology and Evolution* **6**, 1110-1116, doi:doi:10.1111/2041-210X.12410 (2015).

938 83 Hennig, B. P. *et al.* Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5
939 Purification and Tagmentation Protocol. *G3 (Bethesda, Md.)* **8**, 79-89,
940 doi:10.1534/g3.117.300257 (2018).

941 84 Therkildsen, N. O. & Palumbi, S. R. Practical low-coverage genomewide sequencing of
942 hundreds of individually barcoded samples for population and evolutionary genomics in
943 nonmodel species. *Molecular Ecology Resources* **17**, 194-208, doi:10.1111/1755-0998.12593
944 (2016).

945 85 Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for
946 error correction in single-cell sequencing. *BMC Genomics* **14**, S7, doi:10.1186/1471-2164-14-
947 s1-s7 (2013).

948 86 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-
949 cell sequencing. *Journal of computational biology : a journal of computational molecular cell*
950 *biology* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

951 87 Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced
952 methodologies and community practices. *Methods* **102**, 3-11,
953 doi:10.1016/j.ymeth.2016.02.020 (2016).

954 88 Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately
955 reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165-e1165,
956 doi:10.7717/peerj.1165 (2015).

957 89 Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo
958 genome assemblies. *Bioinformatics* **31**, 3350-3352, doi:10.1093/bioinformatics/btv383
959 (2015).

960 90 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing
961 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
962 *Genome Research* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

963 91 Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome
964 assemblies. *Bioinformatics* **29**, 1072-1075, doi:10.1093/bioinformatics/btt086 (2013).

965 92 Xu, H. *et al.* FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. *PLOS*
966 *ONE* **7**, e52249, doi:10.1371/journal.pone.0052249 (2012).

967 93 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
968 data. *Bioinformatics (Oxford, England)* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170
969 (2014).

970 94 Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from
971 genomic next-generation sequencing reads—a baiting and iterative mapping approach.
972 *Nucleic Acids Research* **41**, e129-e129, doi:10.1093/nar/gkt371 (2013).

973 95 Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation.
974 *Molecular Phylogenetics and Evolution* **69**, 313-319, doi:10.1016/j.ympev.2012.08.023 (2013).

975 96 Gruber-Vodicka, H. R., Seah, B. K. B. & Pruesse, E. phyloFlash: Rapid Small-Subunit rRNA
976 Profiling and Targeted Assembly from Metagenomes. *mSystems* **5**, e00920-00920,
977 doi:10.1128/mSystems.00920-20 (2020).

978 97 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing
979 and web-based tools. *Nucleic Acids Research* **41**, D590-D596, doi:10.1093/nar/gks1219
980 (2012).

981 98 Li, W. & Godzik, A. CD-HIT: A fast program for clustering and comparing large sets of protein
982 or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).

983 99 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
984 quantification. *Nature Biotechnology* **34**, 525, doi:10.1038/nbt.3519 (2016).

985 100 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).

986 101 Zischler, H., Geisert, H., von Haeseler, A. & Pääbo, S. A nuclear 'fossil' of the mitochondrial D-
987 loop and the origin of modern humans. *Nature* **378**, 489-492, doi:10.1038/378489a0 (1995).

988 102 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079,
989 doi:10.1093/bioinformatics/btp352 (2009).

990 103 Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis
991 framework for variant extraction and refinement from population-scale DNA sequence data.
992 *Genome Research* **25**, 918-925, doi:10.1101/gr.176552.114 (2015).

993 104 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
994 generation DNA sequencing data. *Genome Research* **20**, 1297-1303,
995 doi:10.1101/gr.107524.110 (2010).

996 105 Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation
997 Sequencing Data. *BMC Bioinformatics* **15**, 356, doi:10.1186/s12859-014-0356-4 (2014).

998 106 Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: A Comprehensive, Accurate, and Fast
999 Distance-Based Phylogeny Inference Program. *Molecular Biology and Evolution* **32**, 2798-
1000 2800, doi:10.1093/molbev/msv150 (2015).

1001 107 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1002 phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

1003 108 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective
1004 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and*
1005 *Evolution* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

1006 109 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments.
1007 *Nucleic Acids Research* **47**, W256-W259, doi:10.1093/nar/gkz239 (2019).

1008 110 vegan: Community Ecology Package. R package version 2.4-4. (2017).

1009 111 R-Core-Team. *R: A language and environment for statistical computing.*, (R Foundation for
1010 Statistical Computing, 2016).

1011 112 Ogle, D. *Introductory Fisheries Analyses with R*. 1-337 (Chapman and Hall/CRC, 2016).

1012 113 Watterson, G. A. On the number of segregating sites in genetical models without
1013 recombination. *Theor Popul Biol* **7**, 256-276, doi:10.1016/0040-5809(75)90020-9 (1975).

1014 114 Paradis, E. pegas: an R package for population genetics with an integrated–modular
1015 approach. *Bioinformatics* **26**, 419-420, doi:10.1093/bioinformatics/btp696 (2010).