Sato *et al. Microbiome* (2022) 10:178
https://doi.org/10.1186/s40168-022-01372-2

Microbiome

RESEARCH

Open Access

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

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

Abstract

Background: Many animals live in intimate associations with a species-rich microbiome. A key factor in maintaining these beneficial associations is fidelity, defined as the stability of associations between hosts and their microbiota over multiple host generations. Fidelity has been well studied in terrestrial hosts, particularly insects, over longer macroevolutionary time. In contrast, little is known about fidelity in marine animals with species-rich microbiomes at short microevolutionary time scales, that is at the level of a single host population. Given that natural selection acts most directly on local populations, studies of microevolutionary partner fidelity are important for revealing the ecological and evolutionary processes that drive intimate beneficial associations within animal species.

Results: In this study on the obligate symbiosis between the gutless marine annelid *Olavius algarvensis* and its consortium of seven co-occurring bacterial symbionts, we show that partner fidelity varies across symbiont species from strict to absent over short microevolutionary time. Using a low-coverage sequencing approach that has not yet been applied to microbial community analyses, we analysed the metagenomes of 80 *O. algarvensis* individuals from the Mediterranean and compared host mitochondrial and symbiont phylogenies based on single-nucleotide polymorphisms across genomes. Fidelity was highest for the two chemoautotrophic, sulphur-oxidizing symbionts that dominated the microbial consortium of all *O. algarvensis* individuals. In contrast, fidelity was only intermediate to absent in the sulphate-reducing and spirochaetal symbionts with lower abundance. These differences in fidelity are likely driven by both selective and stochastic forces acting on the consistency with which symbionts are vertically transmitted.

Conclusions: We hypothesize that variable degrees of fidelity are advantageous for *O. algarvensis* by allowing the faithful transmission of their nutritionally most important symbionts and flexibility in the acquisition of other symbionts that promote ecological plasticity in the acquisition of environmental resources.

Keywords: Microbiome, Animal-bacterial symbiosis, Symbiont transmission, Phylosymbiosis, Intraspecific genetic variation

Background

Beneficial associations between eukaryotic hosts and bacterial partners are ubiquitous, but how these persist stably over evolutionary time remains a source of debate [1–3]. One of the factors that plays a central role in maintaining beneficial symbioses is partner fidelity, defined as the stability of the association between a host and its symbiont over multiple host generations [4] (see

*Correspondence: ysato@mpi-bremen.de; ndubilier@mpi-bremen.de; manuel_kleiner@ncsu.edu

[1] Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany
[3] Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695, USA
Full list of author information is available at the end of the article

Table 1). In associations with strict fidelity, genetic variants of hosts and symbionts show phylogenetic concordance. Strict fidelity is favoured in associations in which the symbionts are transmitted vertically, that is directly from hosts to their offspring. However, fidelity in vertically transmitted symbioses can be disrupted by host switching, symbiont displacement, acquisition of novel symbionts from free-living microorganisms and symbiont loss. Fidelity is generally weaker in associations with horizontal symbiont transmission in which the symbionts are acquired from free-living microbial populations or co-occurring hosts [4, 5]. However, strong fidelity can also occur in symbioses with horizontal transmission if genotype-dependent partner choice ensures the faithfulness of the association [4].

Partner fidelity has been well studied in obligate associations with only one or a few symbionts, such as aphids [12–14], tsetse flies [15], *Riftia* tube worms [16], Vesicomyidae clams [17–19], *Solemya* clams [20] and the Hawaiian bobtail squid [21]. However, as the number and diversity of symbiont species in a host increase, analysing partner fidelity over multiple host generations and across hundreds to thousands of microbial species and strains that often evolve rapidly and are in continuous flux is highly challenging [22, 23]. In hosts with highly diverse microbiota, such as sponges [24–26], ascidians [26], corals [26, 27], some insects [28–30] and mammals [31–33], phylosymbiosis is therefore the approach most often used (see Table 1). In phylosymbiosis studies, analyses of microbial metagenomes, microbial marker genes, or parts of them, like the 16S rRNA gene, are used to examine if the composition of a given group of host-associated microbiomes reflects the phylogeny of these hosts [28, 29, 31, 34–36]. However, phylosymbiosis is different from partner fidelity. In phylosymbiosis studies, the similarity of microbial communities across hosts is analysed, while in fidelity studies, the coinheritance of symbiont and host genotypes is investigated (see Table 1). By characterizing patterns of host-symbiont coinheritance, partner fidelity can provide more detailed insights into the ecological and evolutionary processes that affect the acquisition and persistence of each single member of complex symbiont communities.

Previous studies of partner fidelity in associations with complex symbiont communities have focussed on terrestrial animals, such as insects [37–41] and humans

Table 1 Definition of key terms used in this study: As many of the terms below are used inconsistently in the literature, we explain here how we interpret them.

Term	Definition
Vertical transmission	The direct transmission of a symbiont from a parent to its offspring. In most symbioses, the transmission is from mother to offspring (maternal), but there are cases of paternal transmission [6, 7].
Horizontal transmission	The transmission of a symbiont to a host from the environment or a co-occurring host [6].
Mixed-mode transmission	The transmission of a symbiont by vertical transmission mixed with occasional or frequent events of horizontal transmission over evolutionary time [5]. Note that the transmission of a symbiont community from one generation to the next in which some members are transmitted vertically and others horizontally is not meant here when using this term.
Partner fidelity	The stability of the association between host and symbiont genotypes over multiple host generations [8]. Partner fidelity is generated by vertical symbiont transmission or genotype-dependent partner choice in horizontal symbiont transmission [4]. Note that studies on mechanism of how microbiomes are transmitted across 1–2 generations in host individuals (e.g. parents to offspring) are not the same as partner fidelity studies, which examine fidelity across many generations in multiple individuals. In this study, we used congruent phylogenies of host mitochondrial genomes and symbiont genomes at microevolutionary scales as an indicator of partner fidelity.
Partner choice	The ability of hosts, symbionts or both to preferentially choose their partner. Partner choice describes interactions between individual partners within their lifetime and is distinct from partner fidelity requiring repeated interactions over evolutionary time [8, 9].
Partner specificity	The taxonomic range of partners in an association [10]. Symbiont specificity is defined as the range of symbionts with which a host associates, while host specificity is defined as the range of hosts with which a symbiont associates. In this study, we distinguish partner specificity from partner fidelity, as the former measures the possible diversity of host-symbiont associations, but not the stability of each association.
Coinheritance	The transmission of two or more traits from a host parent to its offspring. Traits can include any combination of phenotypes, genes, alleles, organelles and symbionts. In this study, we use the term to describe the coinheritance of mitochondria and symbionts from parents to their offspring.
Phylosymbiosis	Microbial community relationships that recapitulate the phylogeny of their hosts [11]. Phylosymbiosis tests how similar the composition of microbial communities is to the phylogeny of host species and can arise through ecological or evolutionary forces. Phylosymbiosis differs from partner fidelity in that the structure of the microbial community is analysed, not the phylogeny of each symbiont taxon.
Microevolution	Evolutionary change in a population over short time scales and generally applied to evolution within a species or conspecific populations.
Macroevolution	Evolutionary change over longer time scales and generally applied to evolution across species and higher taxonomic groups.

[42–45], but the little that is known about partner fidelity in marine animals with a diverse and species-rich microbiome largely stems from studies on sponges and corals (e.g. [46, 47]). Importantly, almost nothing is known about partner fidelity at microevolutionary time scales in wild animals, as past studies on both terrestrial and aquatic animals with species-rich microbiomes have investigated fidelity on larger evolutionary time scales, namely, across populations, species and geographic regions. Microevolutionary studies on single host populations are, however, important because natural selection acts most directly on single populations. Microevolutionary studies can isolate ecological and evolutionary factors that affect the stability and persistence of symbiotic communities from other larger-scale factors such as geographic distance. Moreover, such short time scales matter because strict partner fidelity decreases over evolutionary time in most symbiotic associations [5, 48] and because opportunities for host switching, symbiont displacement and uptake of new symbiont genotypes from the environment increase with time. Furthermore, deleterious genome decay in vertically transmitted symbionts increases with time, making it advantageous for hosts to acquire new symbionts horizontally from the environment or other hosts [49, 50]. In summary, assumptions about partner fidelity based on macroevolutionary scales may not reflect ecologically relevant, microevolutionary interactions between hosts and their microbial communities.

Here we investigated microevolutionary partner fidelity in the marine annelid *Olavius algarvensis* (Fig. 1a). This gutless marine worm is an excellent model for investigating partner fidelity in hosts with multimember symbiont communities, as it lives in an intimate, beneficial association with six to seven symbiont species. Its microbiota is therefore complex and diverse, yet simple enough for investigating the fidelity of each single symbiont species. All gutless annelid species (Clitellata, Naididae, Phallodrilinae, genera

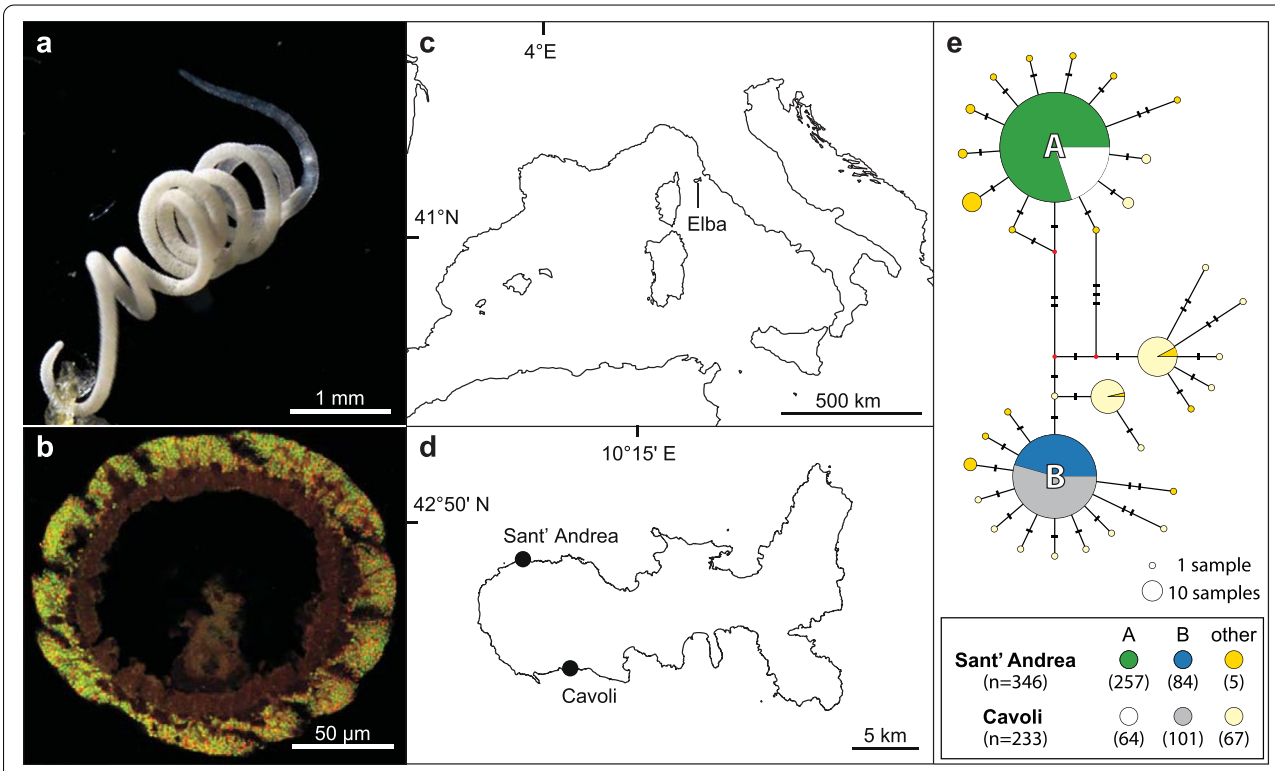

Fig. 1 The *O. algarvensis* population in two bays off the island of Elba was dominated by two mitochondrial haplotypes. **a** Light microscopy image of *Olavius algarvensis*. **b** Fluorescence in situ hybridization image of an *O. algarvensis* cross section, highlighting the symbionts just below the cuticle of the host (gammaproteobacterial symbionts in green and deltaproteobacterial symbionts in red, using general probes for these two phyla). Reproduced with permission from Kleiner et al. [67]. **c** and **d** Location of the two collection sites, Sant' Andrea and Cavoli, two bays off the island of Elba in the Mediterranean. **e** Haplotype network of mitochondrial cytochrome c oxidase subunit 1 (COI) gene sequences of *O. algarvensis* individuals from the two collection sites. The two dominant COI haplotypes A and B co-occurred in both bays. The size of the pie charts corresponds to COI haplotype frequencies. Hatch marks correspond to the number of point mutations between COI haplotypes. Nodes depicted by small red points indicate unobserved intermediates predicted by the algorithm in the haplotype network software. The number of individuals identified as COI haplotype A or B in each bay is in parentheses in the box below the network.

Olavius and *Inanidrilus*; *sensu* Erséus et al. [51]) are regularly associated with at least three to seven symbiont species from different genera and phyla that co-occur within single host individuals [52–59]. All symbionts are harboured in an extracellular region immediately under the outer cuticle of the host (Fig. 1b). Over the estimated 50 million years these hosts have evolved from their gut-bearing ancestors [60], they have become so fully dependent on their symbionts for both nutrition and recycling of their waste compounds that they no longer have a mouth, digestive tract and excretory system [52–54, 61]. Symbiont transmission occurs vertically through smearing when the parents deposit their eggs in the sediment, based on transmission electron microscopy and fluorescence in situ hybridization (FISH) studies of three host species (*Inanidrilus leukodermatus*, *Olavius planus* and *Olavius algarvensis*) [59, 62–64]. However, given the morphological similarity of several members of the bacterial symbiont community, these studies could not resolve if all symbionts are inherited through egg smearing or if some are acquired horizontally from the environment. Furthermore, there is evidence for host switching and displacement in the dominant, sulphur-oxidizing symbiont *Candidatus* Thiosymbion over longer evolutionary time [65, 66].

In the best-studied gutless marine annelid, *O. algarvensis*, seven symbiont species have been identified — two sulphur-oxidizing *Gammaproteobacteria*: *Ca.* Thiosymbion and Gamma3; four sulphate-reducing Deltaproteobacteria: Delta1a, Delta1b, Delta3 and Delta4; and a spirochete [52–55, 68, 69]. Of these seven symbionts, not all consistently co-occur in all host individuals, but all hosts harbour the sulphur-oxidizing symbiont *Ca.* Thiosymbion, the sulphate-reducing Delta1a or Delta1b symbiont and the spirochete [55]. The gammaproteobacterial, and possibly the deltaproteobacterial symbionts, autotrophically fix CO_2 and engage in syntrophic cycling of oxidized and reduced sulphur compounds [52, 53, 67]. Nutrient transfer to the host occurs via phagolysosomal digestion of the symbionts in the epidermal cells underneath the symbiont layer [70, 71]. Evidence from genomic, proteomic and stable isotope analyses indicates resource partitioning, with different symbiont species favouring different energy and carbon sources [53, 54, 61, 70]. This metabolic niche differentiation among these co-occurring symbionts, together with the variability in their abundances across host individuals, indicates different levels of selective pressure, which could be reflected in varying degrees of partner fidelity [72].

In this study, we examined partner fidelity in *O. algarvensis* collected off the island of Elba, Italy, by analysing single-nucleotide polymorphisms (SNPs) across 80 metagenomes. For the hosts, we analysed the phylogeny of their mitochondrial genomes as indicators of vertical symbiont transmission. We then compared mitochondrial phylogeny with that of each symbiont to identify levels of congruence, and correspondingly fidelity, within the microbial consortium. We used a low-coverage sequencing approach for non-model organisms that has not yet been applied to host-microbe associations [73, 74], allowing the analysis of the entire microbial consortium, including low-abundance symbionts.

Results

Two mitochondrial haplotypes dominated the host population

To identify mitochondrial genetic diversity in the *O. algarvensis* population, we sequenced the mitochondrial cytochrome c oxidase subunit I (COI) gene of 579 *O. algarvensis* individuals collected over 4 years from two bays approximately 16 km apart on the island of Elba, Italy (Sant' Andrea and Cavoli; Fig. 1 c and d). A haplotype network, based on 579 COI sequences of 525 bp, revealed two dominant mitochondrial COI haplotypes, here termed A and B, which co-occurred at both locations and were distinct from each other by five nucleotides (Fig. 1e). We sequenced the metagenomes of 20 individuals from the two dominant A and B haplotypes from both locations (in total 80 metagenomes) and assembled complete circular genomes of host mitochondria (mtDNA) from an arbitrarily selected metagenome for A and B haplotypes. The mtDNAs of A and B haplotypes shared 99.3% average nucleotide identity (ANI) and encoded 13 protein-coding genes, 2 ribosomal RNAs and 21 transfer RNAs ($A = 15{,}715$ bp and $B = 15{,}730$ bp). The close phylogenetic relatedness between these two host mitochondrial haplotypes that co-occurred in both bays allowed us to examine the effects of both genetics and geographic location on partner fidelity in *O. algarvensis*.

Intraspecific symbiont diversity was low across and within host individuals

We assembled metagenome-assembled genomes (MAGs) for each of the seven symbionts in *O. algarvensis* and used them as references for all further analyses (Supplementary Table S1; Supplementary Figs. S1 and S2). To ensure that each of the seven symbionts belonged to the same species across all *O. algarvensis* individuals, we examined average nucleotide identity (ANI) of all MAGs that could be recovered from each symbiont species (Supplementary Fig. S3). MAGs from the same species shared more than 95% pairwise ANI, and their 16S rRNA genes were also more than 99% identical (Supplementary Figs. S3 and S4, the only exception was one *Ca.* Thiosymbion pair that had a slightly lower ANI of 94.7%). These sequence

similarities are widely accepted as thresholds for identifying bacterial species (95% ANI for MAGs [75–77] and 98.5 or 99% for 16S rRNA genes [78, 79]).

Symbiont strain diversity within single host individuals was low based on SNP densities in the seven symbiont genomes (0.02–0.49 SNP/kbp; Supplementary Table S2b). These values are lower than or comparable to SNP densities reported for endosymbionts transmitted vertically with repeated events of horizontal transmission, such as the shallow water bivalve *Solemya velum* and the deep-sea scaly-foot snail *Chrysomallon squamiferum* (0.1–1 SNP/kbp and 0.004–3.5 SNP/kbp, respectively) [20, 80], and considerably lower than SNP densities in horizontally transmitted symbionts of the giant tubeworm *Riftia pachyptila* (2.9 SNP/kbp) [81] and deep-sea *Bathymodiolus* mussels (5–11 SNP/kbp) [82]. Since strain diversity was low within host individuals, we treated each symbiont within an *O. algarvensis* individual as a single genotype in the analyses described below.

Sulphur-oxidizing, sulphate-reducing and spirochete symbionts were found in all host individuals

We assessed the relative abundance of symbionts within each of the 80 *O. algarvensis* individuals by quantifying sequencing read abundances for single-copy genes specific to each symbiont species (Fig. 2; Supplementary text 1.1; 162 to 431 single-copy genes per symbiont species). The sulphur-oxidizing symbionts, *Ca.* Thiosymbion and Gamma3, as well as the spirochete, were present in all individuals (Fig. 2a). All host individuals also always had sulphate-reducing symbionts (Delta1a, Delta1b, Delta3 and Delta4), but these varied across host individuals, and no individuals hosted all of them. Delta3 was detected in only six host individuals, making statistical tests meaningless and therefore excluded from subsequent phylogenetic analyses.

Based on relative read abundances for single-copy genes, *Ca.* Thiosymbion was the most abundant symbiont across host individuals (41.6 ± 11.6%; mean ± standard deviation; Fig. 2b), while the abundance of Gamma3 symbiont reads varied considerably in host individuals from 0.1 to 61.2% (28.3 ± 11.3%). The summed relative abundances of the sulphur-oxidizing gammaproteobacteria (69.9 ± 7.4%) and sulphate-reducing deltaproteobacteria (26.7 ± 7.2%) showed consistent ratios across the 80 host individuals, regardless of the location, host COI haplotype, or the combination of these two factors (Supplementary Table S3). Read abundances of the spirochete symbiont were consistently low in host individuals (3.4 ± 2.7%), with one exception of 17.3%.

Probabilistic SNP identification increased the number of hosts and SNP sites for phylogenetic reconstruction of low-abundance symbionts

Conventionally, SNPs in symbiont genomes are identified using a deterministic genotype-calling approach that requires a minimum read coverage (e.g. at least fivefold [83]; Supplementary text 1.4). In our dataset, this coverage requirement excluded symbionts that occurred at low relative abundances (Supplementary Table 4a). In addition, the deterministic genotype calling limits the number of available SNP sites for phylogenetic inference, in our case to a maximum symbiont SNP density of 0.102 SNP/kbp (ranging from 0.006 to 0.102 SNP/kbp; Supplementary Table 4b). To circumvent these limitations, we calculated genotype probabilities and inferred genetic distances from the probability matrices for phylogenetic reconstruction. This allowed us to reconstruct phylogenies of symbionts from more *O. algarvensis* individuals, up to twice as many individuals than using the deterministic approach (between 3 and 131% increase; Supplementary Table S4a). Moreover, the probabilistic approach increased the robustness of our phylogenetic analyses, as it led to a substantial increase in SNP sites (ranging from 0.035 to 0.752 SNP/kbp; Supplementary Table S4b; see Supplementary text 1.4). Our comparison of phylogenies based on deterministic and probabilistic SNP identification showed consistent phylogenetic clustering for the samples that could be analysed using both methods (Fig. 3; Supplementary Fig. S6; Supplementary text 1.4). Therefore, we based our analyses on the probabilistic approach that allowed us to (i) include significantly more samples and (ii) identify more SNP sites to infer more robust phylogenies (Supplementary Table S4).

Congruence of symbiont and mitochondrial phylogenies varied from high to absent

To examine partner fidelity in the *O. algarvensis* symbiosis, we compared the phylogenies of the six most widespread symbionts with that of their hosts' mitochondrial genomes. The degree of congruence between symbiont and host phylogenies reflects the degree of fidelity between partners, with high congruence indicating strong fidelity and vice versa [12, 13, 20]. For the host population, their mtDNA phylogeny revealed a clear divergence between two mitochondrial lineages (termed A- and B-hosts), corresponding to the two COI haplotypes A and B (Fig. 3a). In contrast, the two locations appeared to play a smaller role in shaping mitochondrial phylogeny, with only B-hosts from Sant' Andrea forming a well-supported clade. However, the phylogenetic relationships within the mitochondrial lineages could not be fully resolved due to their limited genetic divergence (Supplementary Fig. S7).

a

COI-haplotype, location	*Ca.*Thiosym.	Gamma3	Delta1a	Delta1b	Delta3	Delta4	Spirochete
A, Sant' Andrea	20/20 (100%)	20/20 (100%)	20/20 (100%)	0/20 (0%)	1/20 (5%)	20/20 (100%)	20/20 (100%)
A, Cavoli	20/20 (100%)	20/20 (100%)	20/20 (100%)	13/20 (65%)	3/20 (15%)	20/20 (100%)	20/20 (100%)
B, Sant' Andrea	20/20 (100%)	20/20 (100%)	9/20 (45%)	16/20 (80%)	2/20 (10%)	17/20 (85%)	20/20 (100%)
B, Cavoli	20/20 (100%)	20/20 (100%)	6/20 (30%)	20/20 (100%)	0/20 (0%)	20/20 (100%)	20/20 (100%)
Total	80/80 (100%)	80/80 (100%)	55/80 (69%)	49/80 (61%)	6/80 (8%)	77/80 (96%)	80/80 (100%)

b

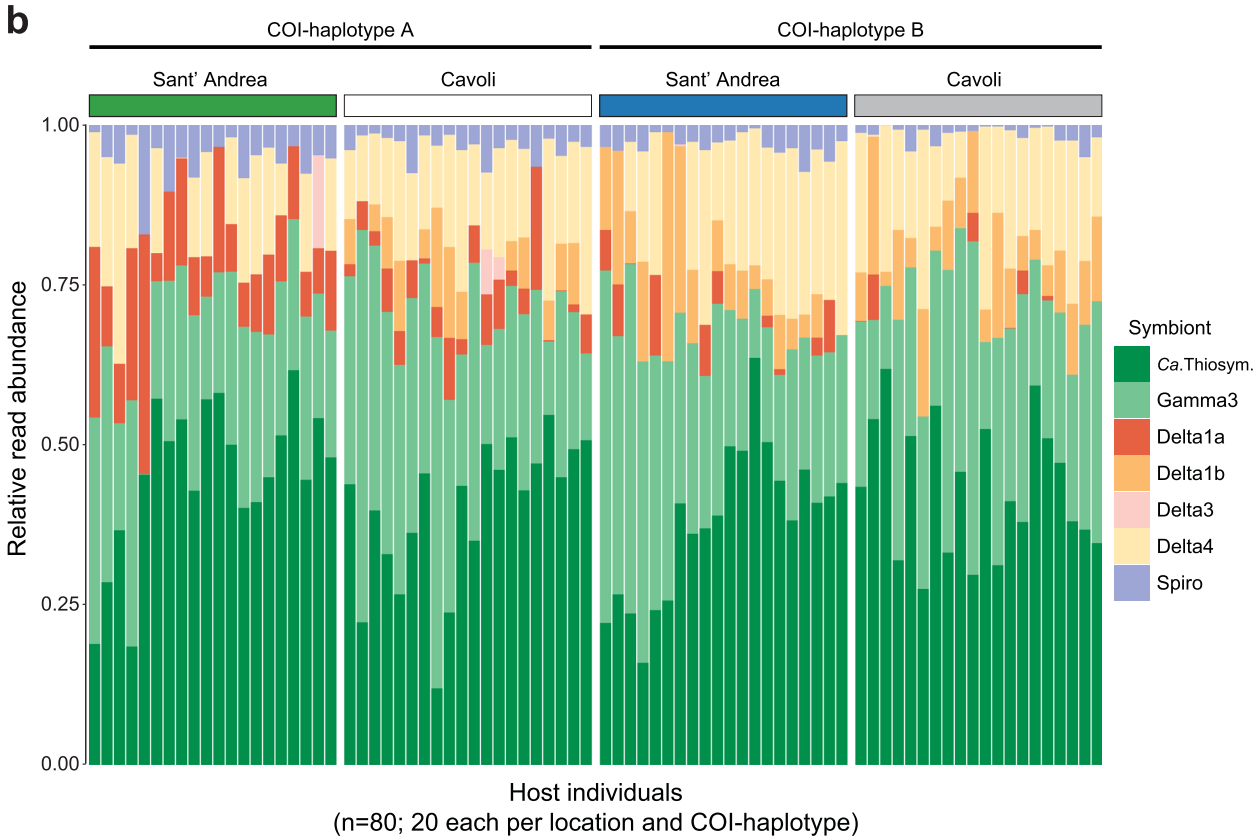

Fig. 2 The composition of the symbiont community in 80 *O. algarvensis* individuals. **a** The number of *O. algarvensis* host individuals from two mitochondrial lineages (COI haplotypes A and B) and two locations (Sant' Andrea and Cavoli) in which the respective symbiont species was detected ($n = 80$ in total; 20 replicates per location and COI haplotype; Supplementary text 1.1). **b** Relative read abundances of symbionts in the 80 *O. algarvensis* individuals. Each column shows the reads from a single host individual. The sulphur-oxidizing symbionts *Ca.* Thiosymbion (*Ca.* Thiosym.) and Gamma3 were the most abundant across host individuals, while the abundances of the sulphate-reducing symbionts (Delta1a, Delta1b, Delta3, Delta4) and the spirochete symbiont (Spiro) were consistently lower. Relative abundances of each symbiont were estimated based on metagenomic sequencing reads that mapped to the single-copy genes of each symbiont (Supplementary text 1.1). Relative symbiont abundances based on 16S rRNA gene sequences in the metagenomes were similar (Supplementary text 1.2; Supplementary Fig. S5)

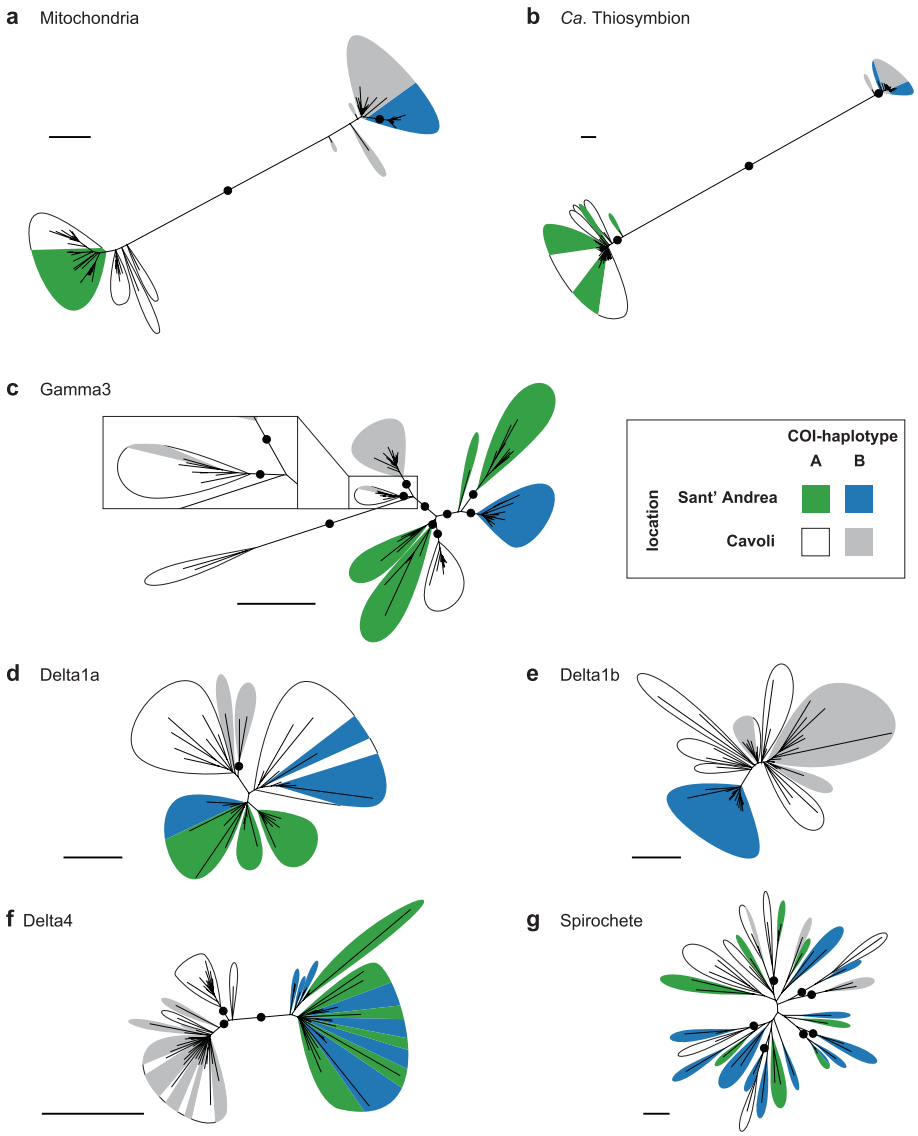

Fig. 3 Comparative phylogenetic analyses of *O. algarvensis* and its microbial consortium members revealed variable patterns of congruence across the six symbionts. Phylogenetic trees based on SNPs across genomes of **a** host mitochondria (166 SNPs, $n = 80$), **b** *Ca.* Thiosymbion (2872 SNPs, $n = 80$), **c** Gamma3 symbiont (618 SNPs, $n = 80$), **d** Delta1a symbiont (375 SNPs, $n = 37$), **e** Delta1b symbiont (624 SNPs, $n = 46$), **f** Delta4 symbiont (675 SNPs, $n = 67$) and **g** spirochete symbiont (88 SNPs, $n = 41$). Phylogenies were inferred from genetic distances calculated from posterior genotype probabilities. Scale bars indicate 0.1 substitution per SNP site. Bootstrap support values > 95% are shown in black circles. Support for branches internal to each coloured leaf was omitted for visibility

For the six symbionts of *O. algarvensis*, we found marked differences in the congruence of their phylogenies with that of their hosts' mitochondrial genomes (Fig. 3 b–g). Congruence was highest in *Ca.* Thiosymbion, which mirrored the mitochondrial phylogeny of their hosts, with symbionts from A- and B-hosts falling into two separate clades (Fig. 3b). The congruence between the two mitochondrial lineages and the two *Ca.* Thiosymbion clades was 100% at both locations. That is, *O. algarvensis* A-hosts always had *Ca.* Thiosymbion A, and *O. algarvensis* B-hosts always had *Ca.* Thiosymbion B, without any exceptions. The slight differences in clade topology in Fig. 3 a and b are due to the lack of sufficient genetic divergence between individual mitochondrial haplotypes and *Ca.* Thiosymbion genotypes (Supplementary Fig. S7b). The Gamma3 symbionts formed nine groups, most of which were statistically supported clades, with each group containing symbionts from either A- or B-hosts (Fig. 3c).

The exception was a Gamma3 symbiont of a B-host from Cavoli that fell in a clade of A-host symbionts from Cavoli (magnified panel in Fig. 3c). Location also appeared to affect the phylogeny of Gamma3 symbionts, with symbionts from the same bay forming distinct groups.

For all other symbionts besides the *Ca.* Thiosymbion and Gamma3, there was no phylogenetic divergence between symbionts from A- and B-hosts (Fig. 3 d–g). Location affected the phylogeny of the Delta4 symbionts, with two well-supported clades separating symbionts from Cavoli and Sant' Andrea (Fig. 3f). The Delta1b symbionts also clustered based on their location, but the clades were not statistically supported (Fig. 3e). For the Delta1a and spirochete, neither host mitochondrial lineage nor location affected their phylogenetic clustering (Fig. 3 d–e, g).

We further examined congruence between the phylogenies of *O. algarvensis* mitochondria and its symbionts using two additional approaches. First, we compared the pairwise genetic distances of hosts and symbionts using three categories: within A- or B-hosts from the same location (within), between A- and B-hosts (mito), and between the two locations Sant' Andrea and Cavoli (location). We tested if genetic distances were explained by host mitochondrial lineage or by location, by analysing pairwise genetic distances in 'within' vs. 'mito' and 'within' vs. 'location' (Fig. 4, Table 2, Supplementary Table S5). For the host, both the mitochondrial lineage and the location had a significant effect on genetic distances, as observed in our phylogenetic SNP analyses. For the symbionts, there was a significant effect of the mitochondrial lineage on genetic distances in *Ca.* Thiosymbion and Gamma3 symbionts, while the effect of location was well supported for the Gamma3 and Delta4 symbionts, again confirming our phylogenetic analyses. An effect of location on genetic distances was also significant for the Delta1b symbionts, but only for B-host symbionts, as they were not detected in A-hosts from Sant' Andrea.

We next examined whether the genetic distance of mtDNA between pairs of *O. algarvensis* individuals corresponded to that of their symbionts. To test for this pattern suggesting genetic co-divergence, correlations of the pairwise genetic distances were calculated between mtDNA and each of the six symbionts (Supplementary Fig. S8, Supplementary Table S6a). The genetic distances of the *Ca.* Thiosymbion symbionts had the strongest positive correlation with mtDNA genetic distances (Table 2, Supplementary Fig. S8a, Supplementary Table S6a). For the Gamma3, Delta1a, Delta1b and Delta4 symbionts, we only observed weak positive correlations (Supplementary Fig. S8 b–e). Genetic co-divergence between the spirochete symbionts and host mitochondria was not detectable (Supplementary Fig. S8f, Supplementary Table S6a). Correlation coefficients (Mantel's R) calculated above for the six symbionts showed a positive correlation with the symbionts' relative read abundances (Fig. 2b; Supplementary Table S6b). In other words, the higher the relative abundance of a symbiont species in host individuals, the greater the degree of genetic co-divergence with *O. algarvensis*.

Discussion

Our analyses revealed that fidelity between the gutless annelid *O. algarvensis* and its endosymbiotic microbial consortium varied from strict to absent. This variability in partner fidelity likely occurred over a short, microevolutionary period, as we analysed individuals within a population of *O. algarvensis* from two very closely related mitochondrial lineages (0.7% divergence) that co-occurred in two bays separated by only 16 km. Our study highlights the importance of examining fidelity over microevolutionary timescales, as it was central to revealing the broad range of fidelity across the members of the *O. algarvensis* symbiont community from strict over intermediate to absent. Over macroevolutionary timescales, strict fidelity is disrupted in *Ca.* Thiosymbion (see below), and it is unlikely that we would have detected the strong to moderate levels of fidelity in other members of the *O. algarvensis* symbiont community over longer evolutionary time.

(See figure on next page.)

Fig. 4 Host mitochondrial lineage and geographic location had a significant effect on the genetic divergence of some but not all symbionts. Host mitochondrial lineages explained genetic divergence in *Ca.* Thiosymbion and Gamma3, while geographic location explained divergence in the Gamma3, Delta1b and Delta4 symbionts. Pairwise genetic distances in *O. algarvensis* mitochondrial genomes and symbionts were calculated from pairs of *O. algarvensis* individuals within the same combination of host lineage (A- or B-host) and location ("Within"), between individuals of A- and B-hosts from the same location ("Between A- and B-hosts") and between individuals from the two locations, Sant' Andrea and Cavoli, but from the same host lineage ("Between locations"). Pairwise genetic distances were compared among these three categories for **a** mitochondria, **b** *Ca.* Thiosymbion, **c** Gamma3 symbiont, **d** Delta1a symbiont, **e** Delta1b symbiont, **f** Delta4 symbiont and **g** spirochete symbiont. 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 (*$p < 0.05$, **$p < 0.01$, see Supplementary Table S5). Orange and blue brackets highlight a significant effect on genetic divergence by the mitochondrial lineage and location, respectively. "N.S." indicates no significant differences among categories ($p > 0.05$)

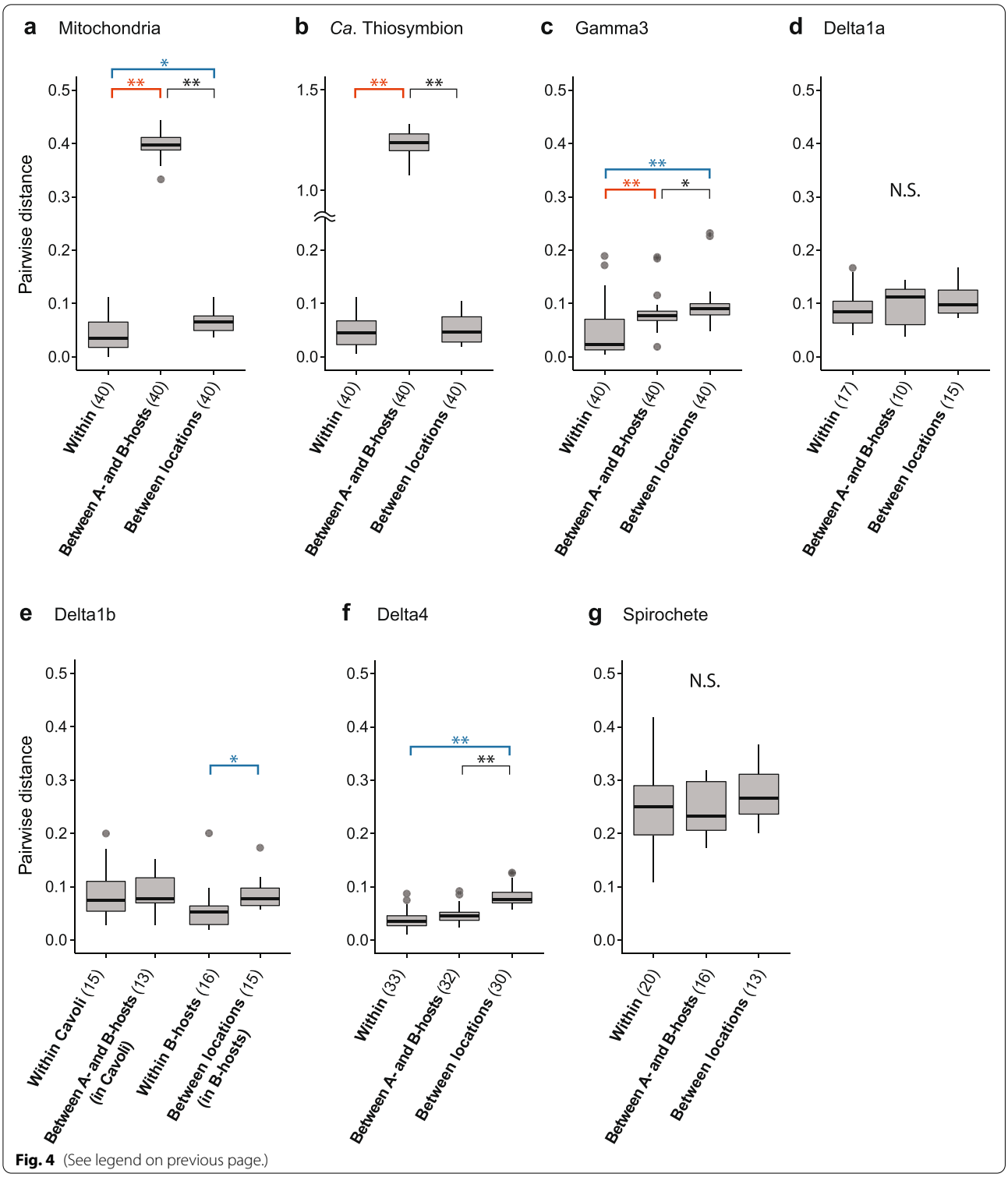

Fig. 4 (See legend on previous page.)

Table 2 Levels of partner fidelity between *O. algarvensis* and its symbiotic consortia based on analyses in this study

Symbiont	Observations					Indication
	Significant effect on genetic distance		Mitochondria-symbiont co-divergence pattern[c]	Relative abundance[d]	Other	Fidelity
***Ca.*Thiosymbion**	Mitochondria[a]		Strong	41.6%	Detected in all host individuals. High genomic divergence between host maternal lineages A and B at both locations	**Strict**
Gamma3	Mitochondria[a]	Location[b]	Present	28.3%	Detected in all host individuals. A single switch from an A- to a B-host observed	**Strong**
Delta1a			Weak	5.3%	Detected in all A-hosts from both locations but in < 50% of B-hosts from both locations	**Moderate**
Delta1b		Location[b]	Present	6.1%	Not detected in A-hosts from Sant'Andrea but detected in 65–100% of other host individuals	**Moderate**
Delta3	NA[e]	NA[e]	NA[e]	0.3%	Detected in only 0–15% of host individuals	NA[e]
Delta4		Location[b]	Weak	14.9%	Detected in 85–100% of host individuals	**Weak**
Spirochete			Absent	3.4%	Detected in all host individuals	**Absent**

[a] Pairwise genetic distances of symbionts were explained by their host mitochondrial lineages A or B (A- or B-hosts)

[b] Pairwise genetic distances of symbionts were explained by locations

[c] Refer Supplementary Fig. S8 and Supplementary Table S6

[d] Mean relative read abundances across all host individuals

[e] Not assessed due to the very low abundance of this symbiont in only six host individuals

Varying degrees of partner fidelity indicate a spectrum of mixed modes between vertical and horizontal transmission

Different degrees of partner fidelity across the microbial consortium of *O. algarvensis* reflect the faithfulness with which the symbionts are transmitted from one generation to the next. The symbionts of marine gutless annelids are transmitted vertically through egg smearing, during which the egg passes through symbiont-rich tissues termed genital pads [62–64]. The egg is then encased in a cocoon and deposited in the sediment. This process offers opportunities for horizontal transmission of bacteria from the surrounding sediment or co-occurring hosts. Indeed, previous studies have shown that over macroevolutionary time, host switching and displacement disrupt fidelity in *Ca.* Thiosymbion, based on comparative phylogenetic analyses of ribosomal genes in 23 gutless annelid species [65, 66]. At microevolutionary time scales, however, our analyses of *O. algarvensis* revealed that transmission is strictly vertical for *Ca.* Thiosymbion, given the strong phylogenetic congruence and genetic co-divergence between these symbionts and their hosts' mitochondria, independent of their collection site (Supplementary text 1.6). We also observed strong fidelity in the Gamma3 symbiont, with only a single switching event from an A- to a B-host in Cavoli (Fig. 3c). Interestingly, such strong fidelity was not observed in the monospecific association between the marine clam *Solemya velum* and its vertically transmitted sulphur-oxidizing symbionts. In these clams, repeated uptake of symbionts from the environment or contemporary hosts occurs within single host populations [20]. This indicates a lesser degree of fidelity between *S. velum* and its symbionts than between *O. algarvensis* and its sulphur-oxidizing symbionts, despite the fact that the clam symbionts are likely transmitted via the ovaries versus the presumed less-restrictive mode of egg smearing in *O. algarvensis*.

Partner fidelity was intermediate to absent for the deltaproteobacterial and spirochete symbionts of *O. algarvensis*. It is therefore likely that these symbionts are regularly acquired horizontally from a free-living population or other co-occurring host individuals. For the Gamma3, Delta1b and Delta4 symbionts, we also observed an effect of their geographic location on their genetic distances (Fig. 4, Table 2). Explanations for this effect include differences in the geographic distribution of genotypes for these symbionts in the environment, exchange of symbionts between contemporary hosts and isolation-by-distance effects on partner choice, but cannot be resolved without additional in-depth analyses of the free-living symbiont population (Supplementary text 1.6).

Combinations of vertical and horizontal transmission modes in co-occurring symbionts have been shown previously, for example in humans [42, 43, 45], sponges [46] and corals [47]. However, these observations were based on comparisons of microbiota in parents and their immediate offspring, and did not examine partner fidelity over multiple generations. As discussed above, time scales matter, because partner fidelity decreases over

evolutionary time in most symbiotic associations [5, 48]. In other words, for any given symbiont that is transmitted vertically across one or a few generations, over longer evolutionary time, horizontal transmission events are likely to disrupt partner fidelity [49, 50]. To our knowledge, the wide range of vertical and horizontal transmission modes of the *O. algarvensis* co-occurring symbionts has not been previously shown at fine microevolutionary time scales.

How can we explain such different degrees of partner fidelity in the *O. algarvensis* symbiosis?

Strong partner fidelity is widespread in associations in which the symbionts are critical for the host's survival and fitness [2, 84]. In *O. algarvensis*, the relative abundance of symbionts was positively correlated with fidelity, with the highest fidelity detected for their sulphur-oxidizing symbionts *Ca*. Thiosymbion, followed by Gamma3 (42% and 28% relative read abundance, respectively; Fig. 2b). These symbionts are the primary producers for their gutless hosts by autotrophically fixing CO_2 into organic compounds [52, 54, 61], and all *O. algarvensis* individuals from both bays harboured these symbionts. Given that *O. algarvensis* gains all of its nutrition by digesting its endosymbionts via endocytosis [70, 71], those symbionts that provide most of its nutrition are likely most strongly selected for, as they most directly affect the fitness of their host.

In contrast, selective pressures on maintaining the symbiotic association are likely relaxed for symbionts that are less critical for their host's fitness or are functionally redundant. The deltaproteobacterial, sulphate-reducing symbionts play an important role by producing reduced sulphur compounds as energy sources for the sulphur-oxidizing bacteria, particularly when these compounds are limiting in the sediment environment [52]. However, all four deltaproteobacterial symbionts reduce sulphate to sulphide [52–54, 68, 69], making this trait functionally redundant among these symbionts. Correspondingly, the presence and abundance of the four sulphate-reducing symbionts varied across host individuals, with different combinations of one to three of these symbionts in each host. This pick and mix pattern indicates that *O. algarvensis* fitness is not dependent on any particular deltaproteobacterial symbiont species but on the presence of functional roles that they share. The lower relative abundance of reads from deltaproteobacterial symbionts (summed average 27%) compared to those of the sulphur oxidizers (70%), as well as their limited biomass [54, 55], further indicates that the deltaproteobacterial symbionts are not as critical for their host's nutrition as *Ca.* Thiosymbion and Gamma3.

The absence of fidelity in the spirochete symbionts was unexpected, given their regular presence in *O. algarvensis* individuals and in other gutless annelids from around the world [56, 85] (Supplementary Fig. S2). While the metabolism of this symbiont is not yet known, it was by far the least abundant symbiont in *O. algarvensis* in this study (3% relative read abundance) as well as previous ones [53–55, 67, 70], indicating a limited role of these symbionts in their hosts' nutrition.

In addition to selective forces, less stringent fidelity in low-abundance symbionts could also be caused by stochastic processes, for example poorer chances of faithful vertical transmission during egg smearing. Assuming that the success with which a symbiont is transmitted to the egg depends on its abundance in the parent, symbionts present in low abundance might face greater chances of not being transmitted than high abundance symbionts. Horizontal reacquisition of these symbionts from the sediment or co-occurring hosts would then explain the disruption of genetic congruence between these symbionts and *O. algarvensis*.

Having your cake and eating it too: the advantage of flexibility in partner fidelity

The lack of stringent partner fidelity between *O. algarvensis* and its deltaproteobacterial and spirochete symbionts indicates that these hosts regularly acquire novel intraspecific genotypes from the environment or from co-occurring host individuals. These newly acquired intraspecific genotypes, together with the interspecific variability in the deltaproteobacterial community across host individuals, could expand the ecological niche of *O. algarvensis* and enable the population to adapt to fluctuating environments. For example, new symbiont genotypes could be better adapted to the environment, provide greater flexibility in the use of resources from the environment and enable resilience to seasonal and long-term temperature changes [5, 86–88]. On the other hand, weak partner fidelity can be costly for the host, including the failure to find suitable symbionts, acquisition of harmful bacteria, or association with 'cheaters' that do not provide mutualistic benefits [89]. However, these costs are likely to be minimal for the deltaproteobacterial symbionts, as their partner quality may depend largely on their ability to produce reduced sulphur compounds, a trait that is critically linked to the energy metabolism of these sulphate-reducing bacteria [52–54], making cheating in this trait unviable for the sulphate-reducing bacteria.

For *O. algarvensis*, fluctuating environmental conditions may be more critical selective forces than the costs of weak partner fidelity. *O. algarvensis*, like other marine annelids, does not have a pelagic life stage and can likely not disperse as widely as many other infaunal

invertebrates [63]. These hosts therefore face the risk of local extinction if environmental conditions become unsuitable for their symbionts. Free-living bacteria rapidly adapt to new environmental conditions [90], and horizontal acquisition of symbionts from the environment can increase the host's potential to adapt to environmental challenges rapidly [91]. The benefit of long-term survival via an evolutionary bet hedging [92, 93] may therefore outweigh the cost of recruiting less-favourable symbionts, as predicted in a recent theoretical study modelling the selective advantages of imperfect vertical transmission of symbionts in variable environments [72].

Probabilistic approaches to SNP identification expand ecological and evolutionary studies of microbial communities

Metagenomic analyses of intraspecific genetic variability in microbial communities typically rely on deep sequencing of each genotype to obtain sufficient SNPs across their genomes [77], incurring substantial sequencing costs. In this study, we reconstructed phylogenies of symbionts directly from genotype probabilities across their genomes, rather than calling genotypes. This probabilistic approach accurately identifies SNPs from low-coverage sequencing data by accounting for the uncertainties of genotyping [73]. This approach has several advantages, including (i) lower sequencing costs compared to high-coverage sequencing, meaning that for the same price, better population coverage can be achieved by sequencing more individuals, (ii) the ability to analyse the genetic diversity of low-abundance symbionts that are present in only some hosts and (iii) an increased robustness of phylogenetic analyses through the recovery of higher SNP numbers. Our study highlights how approaches using genotype probabilities can be applied to the population genetics of host-associated microbiota with variable abundances. Furthermore, our results indicate that probabilistic approaches can also be used to study evolutionary dynamics in free-living microbial communities, thus greatly expanding our toolbox for understanding non-cultivable microorganisms.

Conclusions

We showed that partner fidelity varies from strict to absent in the association between the gutless marine annelid *O. algarvensis* and its microbial consortium. This variability in fidelity was unexpected given that these hosts transmit their symbionts vertically via egg smearing. Our results highlight the importance of examining partner fidelity at microevolutionary scales, as over longer evolutionary time, strict vertical transmission is rare in most symbioses [48, 88]. Understanding the processes that drive fidelity within associations over short to long evolutionary time will help identify the benefits and costs in maintaining symbiotic associations. Such efforts should encompass increasing geographic and taxonomic scales, beginning with local host populations and expanding to regional intraspecific analyses to large-scale global analyses across host species (e.g., [39, 44, 94, 95]). Rapid advances in high-throughput sequencing combined with substantial reductions in sequencing costs using probabilistic SNP calling now make such studies feasible and will contribute to revealing the driving forces that shape the complex and fluid nature of multimember symbioses [87, 96–98].

Methods

Specimen collection and host mitochondrial lineage screening

A total of 579 *O. algarvensis* individuals from two locations on Elba, Italy (Sant' Andrea; 42°48′31″N/10°08′33″E, and Cavoli, 42°44′05″N/10°11′12″E; Fig. 1 c and d; $n = 346$ and 233, respectively) were screened for their mitochondrial lineages based on their mitochondrial COI gene sequences (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 [54]. 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 was individually extracted from single worms using the DNeasy Blood & Tissue Kit (Qiagen, Hilden) according to the manufacturer's instructions. A region of 670 bp of the 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 [99]. PCR-amplicons were 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) according to the manufacturer's instructions. COI sequences were quality filtered with a maximum error rate of 0.5% and aligned using MAFFT v7.45 [100] in Geneious software v11.0.3 (Biomatters, Auckland, New Zealand). A COI haplotype network was built on a 525-bp core alignment using the TCS statistical parsimony algorithm [101] implemented in PopART v1.7 [102]. 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 randomly selected for metagenomic sequencing ($n = 80$ individuals total).

Metagenome sequencing

Sequencing libraries were constructed from the DNA extracted from single worms using a Tn5 transposase purification and tagmentation protocol [103]. The Tn5 transposase was provided by the Protein Science Facility at Karolinska Institutet SciLifeLab (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 for 400–500 bp using the AMPure XP (Beckman Coulter, Indianapolis, IN, USA). Paired-end 150-bp sequences were generated using the Illumina HiSeq3000 System (Illumina, San Diego, CA, USA) with an average total yield of 2.5 Gbp per sample (2548 ± 715 Mbp (mean ± SD)). Construction and quality control of libraries and sequencing were performed at the Max Planck Genome Centre (Cologne, Germany).

Assembly of reference genomes of endosymbionts and *O. algarvensis* 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; Supplementary Table S1). In addition, MAGs of Delta1a, Delta1b and Delta3 were obtained from the public database deposited under the ENA project accession number PRJEB28157 [68, 69]. For the de-novo assembly, 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 [104] implemented in SPAdes v3.9.1 [105]. Clean reads were assembled using MEGAHIT v1.0.6 [106], and symbiont genome bins were identified with MetaBAT v0.26.3 [107]. Bins were assigned to the *Ca.* Thiosymbion, Gamma3, Delta4, and spirochete symbionts based on 99–100% sequence matches with their reference 16S rRNA gene sequences (NCBI accession numbers AF328856, AJ620496, AJ620497, AJ620502 [52, 55], respectively). MAGs were further refined using Bandage v0.08.1 [108] by identifying and inspecting connected contigs on assembly graphs. Completeness and contamination of the MAGs were estimated with CheckM version 1.0.7 [109], and assembly statistics of symbiont genomes were calculated with QUAST v5.0.2 [110] (Supplementary Table S1). For the assessment of overall genetic divergence of symbionts in the 80 *O. algarvensis* individuals, MAGs of all symbionts were binned as above, wherever their read coverage allowed. Average nucleotide identity (ANI) of these MAGs was calculated using FastANI v1.33 [75]. To ensure robust ANI comparisons, MAGs that had 25% or less homologous genes than those of the reference genome (Supplementary Table S1) were excluded from the analysis. Heatmaps of pairwise ANIs within each of the symbionts were generated with *pheatmap* 1.0.12 in R [111].

Complete mitochondrial genomes (mtDNA) were assembled from two metagenomes of *O. algarvensis* from Sant' Andrea representing the two COI haplotypes A and B (specimen IDs: 'OalgSANT_A04' and 'OalgSANT_B04,' respectively). Duplicated sequences due to PCR-amplification during the library preparation were first removed from raw sequences using FastUniq v1.1 [112] prior to adapter clipping and quality trimming with Phred quality score ≥ 2 using Trimmomatic v0.36 [113]. A preliminary 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 [114]. Mitochondrial reads were then identified by mapping to the mtDNA scaffold using *bbmap* in BBTools, and mtDNA was assembled from the identified reads with SPAdes. A circular mtDNA was identified on assembly graphs in Bandage and annotated using MITOS2 webserver (http://mitos2.bioinf.uni-leipzig.de) to confirm the completeness [115].

Characterization of symbiont community composition

The taxonomic composition of the *O. algarvensis* metagenomes was first screened by identifying 16S rRNA gene sequences using phyloFlash v3.3-beta1 [116] with the SILVA SSU database release 132 [117] as reference. The 16S rRNA gene sequences of *O. algarvensis* symbionts were assembled with SPAdes implemented in phyloFlash and aligned using MAFFT in Geneious to identify SNP sites. Chimeric and incomplete (< 1100 bp) SSU sequences were identified in the alignment and excluded.

Relative abundances of *O. algarvensis* symbionts were estimated by mapping metagenome reads to a collection of symbiont-specific sequences of single-copy genes extracted from the genome bins. Orthologous single-copy gene 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 [118]; 8 cases

identified between the Delta1a and Delta1b symbionts) were removed from the final reference sequences of single-copy genes. Metagenomic reads (quality filtered with the same processes as for the mtDNA assembly above) matching to the singe-copy gene sequences were quantified using Kallisto v0.44.0 [119] (Supplementary text 1.1). Symbiont composition estimates were plotted in R using ggplot2 package v3.2.1 [120].

To examine whether our phylogenetic analysis of each *O. algarvensis* symbiont reflects a single-dominant genotype per host individual, levels of symbiont genotype diversity within a host individual were assessed by calculating SNP densities (the number of SNP sites per kbp of reference genome) with previously established procedures [82]. This analysis was performed using a set of publicly available deeply sequenced metagenomes of *O. algarvensis* (listed in Supplementary Table S2a) to ensure that SNP densities were estimated using sufficient symbiont read coverages, and that these estimates could be compared to those in other studies performing similar assessments [20, 80–82].

Identification of single-nucleotide polymorphisms and phylogenetic reconstruction

For identification of SNPs in the genomes of symbionts and mitochondria, quality-controlled reads were first unambiguously split into different symbiont species using *bbsplit* of BBTools, using the reference symbiont MAGs described 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 [121] while ensuring successful mapping of mtDNA reads in the analysis for both A- and B-hosts given the high nucleotide identity of mtDNAs between these two lineages (> 99%; see the "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 [122] and BamUtil v1.0.14 [123], deduplicated with *MarkDuplicates* of Picard Toolkit v2.9.2 (https://github.com/broadinstitute/picard) and realigned around indels with Genome Analysis Toolkit v3.7 [124]. Posterior genotype probabilities were calculated with ANGSD v0.929 [125]. ANGSD is widely used for studies of diploid organisms to infer genotypes from low-coverage sequencing data while taking sequencing errors into account [125–127]. To deal with haploid genotypes in ANGSD, all genotypes were assumed to be 'homozygous' by setting an inbreeding coefficient F of '1,' and a uniform prior were specified for posterior probability calculation (T.S. Korneliussen, 2018, pers. comm.). SNP sites were identified as reference nucleotide positions that were covered $\geq 1\times$ by all samples and showed statistically significant support (SNP p-value < 0.01). When no SNP site was found due to very low or lacking reads from a symbiont, these samples were excluded based on a cut-off of lateral coverages (i.e. % reference genetic sites covered by reads; Supplementary Table S4). For symbionts and mitochondria, their pairwise genetic distances in host individuals were obtained from the matrix of genotype probabilities using NGSdist v1.0.2 [74]. Phylogenetic trees with bootstrap support were computed from the resulting distant matrix using FastME v2.1.5.1 [128] and RAxML v8.2.11 [129].

For the SNP identification by deterministic genotyping (ii), the same symbiont and mitochondrial reads were analysed with the SNIPPY pipeline v3.2 (https://github.com/tseemann/snippy), with the same reference genomes of mtDNA and symbionts as above (Supplementary text 1.4).

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

To examine which factors drive the patterns of genetic divergence in mitochondria and symbionts, their pairwise genetic distances were statistically compared in 3 categories of host pairs: (a) within the same combination of location plus A- or B-hosts, (b) between A- and B-hosts (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 data independence. For the Delta1b symbionts, we separately compared (a) vs. (b) within Cavoli and (a) vs. (c) within B-hosts, because Delta1b did not occur in A-hosts in Sant' Andrea. 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 [130] and FSA package v0.8.25 [131]. 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 [132]. A correlation between Mantel's Rs and relative abundances of symbionts based on reads mapped to single-copy genes was tested with Kendall's rank correlation tau using the function *test.cor* in R's core package.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-022-01372-2.

Additional file 1: Supplementary text: 1.1. Quantification and detection of symbionts based on single-copy marker genes. **1.2**. Assessment of symbiont community compositions based on 16S rRNA genes. **1.3**. Symbiont 16S rRNA gene sequences indicated linkage between host mitochondrial haplotypes and *Candidatus* Thiosymbion. **1.4**. Reconstruction of mitochondria and symbiont phylogenies using a deterministic genotyping approach to SNP-identification. **1.5**. Estimation of the effective population size of symbionts within an *Olavius algarvensis* individual based on genome-wide SNP abundance. **1.6**. Transmission modes of symbionts in *O. algarvensis*. **Supplementary Figure S1**. Phylogenomic tree of symbionts in *O. algarvensis* in relation to reference bacterial genomes. **Supplementary Figure S2**. Phylogeny of 16S ribosomal RNA gene sequences for symbionts in *O. algarvensis* in relation to reference bacterial sequences. **Supplementary Figure S3**. Pairwise average nucleotide identities between metagenome-assembled genomes of *O. algarvensis* symbionts. **Supplementary Figure S4**. Sequence alignments of 16S ribosomal RNA genes of *O. algarvensis* symbionts. **Supplementary Figure S5**. Symbiont composition of individual *O. algarvensis* samples of two COI haplotypes (A and B) from two locations (Sant'Andrea and Cavoli) based on 16S rRNA gene. **Supplementary Figure S6**. Core SNP-trees based on genotypes using a deterministic approach to SNP-identification. **Supplementary Figure S7**. Phylogenies of mitochondria and *Candidatus* Thiosymbion within the major mitochondrial lineages A and B. **Supplementary Figure S8**. Correlation between mitochondrial pairwise genetic distance and pairwise genetic distance of symbionts. **Supplementary Figure S9**. Relative proportions of mapped reads to each of single-copy genes within a single host per symbiont species. **Supplementary Figure S10**. Effective population size estimates of the symbiont per *O. algarvensis* individual. **Supplementary Table S1**. Summary statistics of the reference metagenome-assembled genomes of *O. algarvensis* symbionts. **Supplementary Table S2**. Assessment of strain diversity of symbionts within *O. algarvensis* individuals. **Supplementary Table S3**. Statistical comparisons of ratios between summed relative abundances of deltaproteobacterial and gammaproteobacterial symbionts. **Supplementary Table S4**. Comparison of SNP-identification methods. **Supplementary Table S5**. Statistical comparisons of pairwise genetic distances for the host mitochondria and symbionts. **Supplementary Table S6**. Statistical tests on genetic co-divergence patterns of host mitochondria and symbionts.

Acknowledgements
The authors thank the Hydra Institute for logistical support during sample collection and Toby Kiers (VU University Amsterdam) and members of the Department of Symbiosis for valuable discussions.

Authors' 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 analysed data. YS, JW, RA, ND and MK interpreted results. YS wrote the manuscript with support from ND and MK, and revisions are from all other co-authors. The authors read and approved the final manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Max Planck Society, a Gordon and Betty Moore Foundation Marine Microbial Initiative Investigator Award to ND (Grant GBMF3811), a US 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 programme under the Marie Skłodowska-Curie grant agreement no. 660280 (CW).

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

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany. [2]Gut Microbes and Health Programme, Quadram Institute Bioscience, Norwich NR4 7UQ, UK. [3]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695, USA.

Received: 21 June 2022 Accepted: 15 September 2022
Published online: 22 October 2022

References

1. Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. The evolution of host-symbiont dependence. Nat Commun. 2017;8:15973.
2. Sachs JL, Skophammer RG, Regus JU. Evolutionary transitions in bacterial symbiosis. Proc Natl Acad Sci. 2011;108:10800–7.
3. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. Nat Rev Microbiol. 2008;6:725–40.
4. Douglas AE, Werren JH. Holes in the hologenome: why host-microbe symbioses are not holobionts. mBio. 2016;7:e02099–15.
5. Ebert D. The epidemiology and evolution of symbionts with mixed-mode transmission. Annu Rev Ecol Evol Syst. 2013;44:623–43.
6. Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. Nat Rev Microbiol. 2010;8:218–30.
7. Moran NA, Dunbar HE. Sexual acquisition of beneficial symbionts in aphids. Proc Natl Acad Sci. 2006;103:12803–6.
8. Bull JJ, Rice WR. Distinguishing mechanisms for the evolution of co-operation. J Theor Biol. 1991;149:63–74.
9. Simms EL, Taylor DL. Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. Integr Comp Biol. 2002;42:369–80.
10. Wang D, Yang S, Tang F, Zhu H. Symbiosis specificity in the legume – rhizobial mutualism. Cell Microbiol. 2012;14:334–42.
11. Lim SJ, Bordenstein SR. An introduction to phylosymbiosis. Proc R Soc B Biol Sci. 2020;287:20192900.
12. Liu L, Huang X, Zhang R, Jiang L, Qiao G. Phylogenetic congruence between *Mollitrichosiphum* (Aphididae: Greenideinae) and *Buchnera* indicates insect–bacteria parallel evolution. Syst Entomol. 2013;38:81–92.
13. Funk DJ, Helbling L, Wernegreen JJ, Moran NA. Intraspecific phylogenetic congruence among multiple symbiont genomes. Proc R Soc Lond Ser B Biol Sci. 2000;267:2517–21.
14. Manzano-Marín A, Coeur d'acier A, Clamens A-L, Orvain C, Cruaud C, Barbe V, et al. Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. ISME J. 2020;14:259–73.
15. Symula RE, Marpuri I, Bjornson RD, Okedi L, Beadell J, Alam U, et al. Influence of host phylogeographic patterns and incomplete lineage sorting on within-species genetic variability in *Wigglesworthia* species, obligate symbionts of tsetse flies. Appl Environ Microbiol. 2011;77:8400–8.
16. Polzin J, Arevalo P, Nussbaumer T, Polz Martin F, Bright M. Polyclonal symbiont populations in hydrothermal vent tubeworms and the environment. Proc R Soc B Biol Sci. 2019;286:20181281.
17. Hurtado LA, Mateos M, Lutz RA, Vrijenhoek RC. Coupling of bacterial endosymbiont and host mitochondrial genomes in the hydrothermal vent clam *Calyptogena magnifica*. Appl Environ Microbiol. 2003;69:2058–64.
18. Stewart FJ, Young CR, Cavanaugh CM. Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis. Mol Biol Evol. 2008;25:673–87.

19. Stewart FJ, Cavanaugh CM. Pyrosequencing analysis of endosymbiont population structure: co-occurrence of divergent symbiont lineages in a single vesicomyid host clam. Environ Microbiol. 2009;11:2136–47.
20. Russell SL, Corbett-Detig RB, Cavanaugh CM. Mixed transmission modes and dynamic genome evolution in an obligate animal-bacterial symbiosis. ISME J. 2017;11:1359–71.
21. Koehler S, Gaedeke R, Thompson C, Bongrand C, Visick KL, Ruby E, et al. The model squid–vibrio symbiosis provides a window into the impact of strain- and species-level differences during the initial stages of symbiont engagement. Environ Microbiol. 2019;21:3269–83.
22. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. Nature. 2017;548:43–51.
23. Hawkes CV, Bull JJ, Lau JA. Symbiosis and stress: how plant microbiomes affect host evolution. Philos Transact R Soc B Biol Sci. 2020;375:20190590.
24. Pankey MS, Plachetzki DC, Macartney KJ, Gastaldi M, Slattery M, Gochfeld DJ, et al. Cophylogeny and convergence shape holobiont evolution in sponge–microbe symbioses. Nat Ecol Evol. 2022;6:750–62.
25. Easson CG, Thacker RW. Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. Front Microbiol. 2014;5:532.
26. O'Brien PA, Tan S, Yang C, Frade PR, Andreakis N, Smith HA, et al. Diverse coral reef invertebrates exhibit patterns of phylosymbiosis. ISME J. 2020;14:2211–22.
27. Pollock FJ, McMinds R, Smith S, Bourne DG, Willis BL, Medina M, et al. Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. Nat Commun. 2018;9:4921.
28. Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. PLoS Biol. 2016;14:e2000225.
29. Leigh BA, Bordenstein SR, Brooks AW, Mikaelyan A, Bordenstein SR, Heck M. Finer-scale phylosymbiosis: insights from insect viromes. mSystems. 2018;3:e00131–18.
30. Tinker KA, Ottesen EA. Phylosymbiosis across deeply diverging lineages of omnivorous cockroaches (Order Blattodea). Appl Environ Microbiol. 2020;86:e02513–9.
31. Gaulke CA, Arnold HK, Humphreys IR, Kembel SW, O'Dwyer JP, Sharpton TJ, et al. Ecophylogenetics clarifies the evolutionary association between mammals and their gut microbiota. mBio. 2018;9:e01348–18.
32. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. Science. 2008;320:1647–51.
33. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, et al. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nat Commun. 2017;8:14319.
34. Kohl KD, Varner J, Wilkening JL, Dearing MD. Gut microbial communities of American pikas (Ochotona princeps): evidence for phylosymbiosis and adaptations to novel diets. J Anim Ecol. 2018;87:323–30.
35. Cross KL, Leigh BA, Hatmaker EA, Mikaelyan A, Miller AK, Bordenstein SR, et al. Genomes of gut bacteria from *Nasonia* wasps shed light on phylosymbiosis and microbe-assisted hybrid breakdown. mSystems. 2021;6:e01342–20.
36. Qin M, Chen J, Xu S, Jiang L, Qiao G. Microbiota associated with Mollitrichosiphum aphids (Hemiptera: Aphididae: Greenideinae): diversity, host species specificity and phylosymbiosis. Environ Microbiol. 2021;23:2184–98.
37. Guyomar C, Legeai F, Jousselin E, Mougel C, Lemaitre C, Simon J-C. Multi-scale characterization of symbiont diversity in the pea aphid complex through metagenomic approaches. Microbiome. 2018;6:181.
38. Rock DI, Smith AH, Joffe J, Albertus A, Wong N, O'Connor M, et al. Context-dependent vertical transmission shapes strong endosymbiont community structure in the pea aphid, *Acyrthosiphon pisum*. Mol Ecol. 2018;27:2039–56.
39. Bobay L-M, Wissel EF, Raymann K. Strain structure and dynamics revealed by targeted deep sequencing of the honey bee gut microbiome. mSphere. 2020;5:e00694–20.
40. Powell JE, Martinson VG, Urban-Mead K, Moran NA, Goodrich-Blair H. Routes of acquisition of the gut microbiota of the honey bee *Apis mellifera*. Appl Environ Microbiol. 2014;80:7378–87.
41. Guo W-P, Tian J-H, Lin X-D, Ni X-B, Chen X-P, Liao Y, et al. Extensive genetic diversity of *Rickettsiales* bacteria in multiple mosquito species. Sci Rep. 2016;6:38770.
42. Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, et al. Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. mSystems. 2017;2:e00164–16.
43. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, et al. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. Cell Host Microbe. 2018;24:133–145.e135.
44. Hildebrand F, Gossmann TI, Frioux C, Özkurt E, Myers PN, Ferretti P, et al. Dispersal strategies shape persistence and evolution of human gut bacteria. Cell Host Microbe. 2021;29:1167–1176.e1169.
45. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. Genome Res. 2016;26:1612–25.
46. Björk JR, Díez-Vives C, Astudillo-García C, Archie EA, Montoya JM. Vertical transmission of sponge microbiota is inconsistent and unfaithful. Nat Ecol Evol. 2019;3:1172–83.
47. Quigley KM, Warner PA, Bay LK, Willis BL. Unexpected mixed-mode transmission and moderate genetic regulation of Symbiodinium communities in a brooding coral. Heredity (Edinb). 2018;121:524–36.
48. Russell SL. Transmission mode is associated with environment type and taxa across bacteria-eukaryote symbioses: a systematic review and meta-analysis. FEMS Microbiol Lett. 2019;366:fnz013.
49. Russell SL, Pepper-Tunick E, Svedberg J, Byrne A, Ruelas Castillo J, Vollmers C, et al. Horizontal transmission and recombination maintain forever young bacterial symbiont genomes. PLoS Genet. 2020;16:e1008935.
50. Bennett GM, Moran NA. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proc Natl Acad Sci. 2015;112:10169–76.
51. Erséus C, Wetzel MJ, Gustavsson L. ICZN rules–a farewell to Tubificidae (Annelida, Clitellata). Zootaxa. 2008;1744:66–8.
52. Dubilier N, Mulders C, Ferdelman T, de Beer D, Pernthaler A, Klein M, et al. Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. Nature. 2001;411:298–302.
53. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, et al. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 2006;443:950–5.
54. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. Proc Natl Acad Sci. 2012;109:E1173–82.
55. Ruehland C, Blazejak A, Lott C, Loy A, Erséus C, Dubilier N. Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. Environ Microbiol. 2008;10:3404–16.
56. 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. Appl Environ Microbiol. 2005;71:1553–61.
57. 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. Appl Environ Microbiol. 2006;72:5527–36.
58. Giere O. The gutless marine oligochaete *Phallodrilus leukodermatus*. Structural studies on an aberrant tubificid associated with bacteria. Mar Ecol Prog Ser. 1981;5:353–7.
59. Bright M, Giere O. Microbial symbiosis in Annelida. Symbiosis. 2005;38:1–45.
60. Erséus C, Williams BW, Horn KM, Halanych KM, Santos SR, James SW, et al. Phylogenomic analyses reveal a Palaeozoic radiation and support a freshwater origin for clitellate annelids. Zool Scr. 2020;49:614–40.
61. Kleiner M, Wentrup C, Holler T, Lavik G, Harder J, Lott C, et al. Use of carbon monoxide and hydrogen by a bacteria–animal symbiosis from seagrass sediments. Environ Microbiol. 2015;17:5023–35.
62. Giere O, Langheld C. Structural organisation, transfer and biological fate of endosymbiotic bacteria in gutless oligochaetes. Mar Biol. 1987;93:641–50.

63. Giere O. Ecology and biology of marine Oligochaeta – an inventory rather than another review. Hydrobiologia. 2006;564:103–16.
64. Schimak M-P. Transmission of bacterial symbionts in the gutless oligochaete *Olavius algarvensis*. PhD thesis. University of Bremen. 2016. http://nbn-resolving.de/urn:nbn:de:gbv:46-00105047-16.
65. Zimmermann J, Wentrup C, Sadowski M, Blazejak A, Gruber-Vodicka HR, Kleiner M, et al. Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. Mol Ecol. 2016;25:3203–23.
66. Bergin C, Wentrup C, Brewig N, Blazejak A, Erseus C, Giere O, et al. Acquisition of a novel sulfur-oxidizing symbiont in the gutless marine worm *Inanidrilus exumae*. Appl Environ Microbiol. 2018;84:e02267–17.
67. Kleiner M, Woyke T, Ruehland C, Dubilier N. The *Olavius algarvensis* metagenome revisited: Lessons learned from the analysis of the low-diversity microbial consortium of a gutless marine worm. In: de Bruijn FJ, editor. Handbook of Molecular Microbial Ecology II. Hoboken: Wiley-Blackwell; 2011. p. 319–33.
68. 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*. Microbiol Resourc Announce. 2020;9:e00276–20.
69. Sato Y, Wippler J, Wentrup C, Woyke T, Dubilier N, Kleiner M. High-quality draft genome sequences of the uncultured Delta3 endosymbiont (*Deltaproteobacteria*) assembled from metagenomes of the gutless marine worm *Olavius algarvensis*. Microbiol Resourc Announce. 2020;9:e00704–20.
70. Kleiner M, Dong X, Hinzke T, Wippler J, Thorson E, Mayer B, et al. Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. Proc Natl Acad Sci. 2018;115:E5576–84.
71. Wippler J, Kleiner M, Lott C, Gruhl A, Abraham PE, Giannone RJ, et al. Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*. BMC Genomics. 2016;17:942.
72. Bruijning M, Henry LP, Forsberg SKG, Metcalf CJE, Ayroles JF. Natural selection for imprecise vertical transmission in host–microbiota systems. Nat Ecol Evol. 2022;6:77–87.
73. Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS One. 2012;7:e37558.
74. Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M. Improving the estimation of genetic distances from next-generation sequencing data. Biol J Linn Soc. 2016;117:139–49.
75. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9:5114.
76. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Carnevali PBM, Banfield JF, et al. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. mSystems. 2020;5:e00731–19.
77. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol. 2020;18:491–506.
78. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 2014;64:346–51.
79. Edgar RC. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics. 2018;34:2371–5.
80. Lan Y, Sun J, Chen C, Wang H, Xiao Y, Perez M, et al. Endosymbiont population genomics sheds light on transmission mode, partner specificity, and stability of the scaly-foot snail holobiont. ISME J. 2022. https://doi.org/10.1038/s41396-41022-01261-41394.
81. Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T, et al. Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. Environ Microbiol. 2008;10:727–37.
82. Ansorge R, Romano S, Sayavedra L, Porras MÁG, Kupczok A, Tegetmeyer HE, et al. Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. Nat Microbiol. 2019;4:2487–97.
83. Andreu-Sánchez S, Chen L, Wang D, Augustijn HE, Zhernakova A, Fu J. A benchmark of genetic variant calling pipelines using metagenomic short-read sequencing. Front Genet. 2021;12:648229.
84. Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 2008;42:165–90.
85. Dubilier N, Amann R, Erseus C, Muyzer G, Park S, Giere O, et al. Phylogenetic diversity of bacterial endosymbionts in the gutless marine oligochete *Olavius loisae* (Annelida). Mar Ecol Prog Ser. 1999;178:271–80.
86. Herrera M, Klein SG, Campana S, Chen JE, Prasanna A, Duarte CM, et al. Temperature transcends partner specificity in the symbiosis establishment of a cnidarian. ISME J. 2021;15:141–53.
87. Sudakaran S, Kost C, Kaltenpoth M. Symbiont acquisition and replacement as a source of ecological innovation. Trends Microbiol. 2017;25:375–90.
88. Leftwich PT, Edgington MP, Chapman T. Transmission efficiency drives host-microbe associations. Proc R Soc B Biol Sci. 2020;287:20200820.
89. Salem H, Florez L, Gerardo N, Kaltenpoth M. An out-of-body experience: the extracellular dimension for the transmission of mutualistic bacteria in insects. Proc R Soc B Biol Sci. 2015;282:20142957.
90. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet. 2003;4:457–69.
91. Henry LP, Bruijning M, Forsberg SKG, Ayroles JF. The microbiome extends host evolutionary potential. Nat Commun. 2021;12:5141.
92. Olofsson H, Ripa J, Jonzén N. Bet-hedging as an evolutionary game: the trade-off between egg size and number. Proc R Soc B Biol Sci. 2009;276:2963–9.
93. Childs DZ, Metcalf CJE, Rees M. Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants. Proc R Soc B Biol Sci. 2010;277:3055–64.
94. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, et al. Diversity, structure and convergent evolution of the global sponge microbiome. Nat Commun. 2016;7:11870.
95. Campbell TP, Sun X, Patel VH, Sanz C, Morgan D, Dantas G. The microbiome and resistome of chimpanzees, gorillas, and humans across host lifestyle and geography. ISME J. 2020;14:1584–99.
96. Ferrari J, Vavre F. Bacterial symbionts in insects or the story of communities affecting communities. Philos Transact R Soc B Biol Sci. 2011;366:1389–400.
97. Drew GC, Budge GE, Frost CL, Neumann P, Siozios S, Yañez O, et al. Transitions in symbiosis: evidence for environmental acquisition and social transmission within a clade of heritable symbionts. ISME J. 2021;15:2956–68.
98. Monnin D, Jackson R, Kiers ET, Bunker M, Ellers J, Henry LM. Parallel evolution in the integration of a co-obligate aphid symbiosis. Curr Biol. 2020;30:1949–1957.e1946.
99. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994;3:294–9.
100. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
101. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. Mol Ecol. 2000;9:1657–9.
102. Leigh JW, Bryant D. Popart: full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6:1110–6.
103. Hennig BP, Velten L, Racke I, Tu CS, Thoms M, Rybin V, et al. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. G3 (Bethesda). 2018;8:79–89.
104. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics. 2013;14:S7.
105. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
106. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016;102:3–11.
107. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.
108. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.

109. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
110. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
111. Kolde R. Pheatmap: pretty heatmaps. R package version 1.0.12. 2019. https://cran.r-project.org/package=pheatmap.
112. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, et al. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One. 2012;7:e52249.
113. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30:2114–20.
114. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 2013;41:e129.
115. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69:313–9.
116. Gruber-Vodicka HR, Seah BKB, Pruesse E. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. mSystems. 2020;5:e00920.
117. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012;41:D590–6.
118. Li W, Godzik A. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.
119. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525.
120. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
121. Zischler H, Geisert H, von Haeseler A, Pääbo S. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature. 1995;378:489–92.
122. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
123. Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. 2015;25:918–25.
124. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
125. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics. 2014;15:356.
126. Fuentes-Pardo AP, Ruzzante DE. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. Mol Ecol. 2017;26:5369–406.
127. Fumagalli M, Vieira FG, Linderoth T, Nielsen R. ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics. 2014;30:1486–7.
128. Lefort V, Desper R, Gascuel O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol. 2015;32:2798–800.
129. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
130. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016.
131. Ogle D. Introductory Fisheries Analyses with R. New York: Chapman and Hall/CRC; 2016.
132. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. Vegan: Community ecology package. R package version 2.4-4. 2017. https://CRAN.R-project.org/package=vegan.

Publisher's Note