Polynucleobacter necessarius, a model for genome reduction in both free-living and symbiotic bacteria

Vittorio Boscaro[a,1], Michele Felletti[a,1,2], Claudia Vannini[a], Matthew S. Ackerman[b], Patrick S. G. Chain[c], Stephanie Malfatti[d], Lisa M. Vergez[e], Maria Shin[f], Thomas G. Doak[b], Michael Lynch[b], and Giulio Petroni[a,3]

[a]Department of Biology, Pisa University, 56126 Pisa, Italy; [b]Department of Biology, Indiana University, Bloomington, IN 47401; [c]Los Alamos National Laboratory, Los Alamos, NM 87545; [d]Joint Genome Institute, Walnut Creek, CA 94598; [e]Lawrence Livermore National Laboratory, Livermore, CA 94550; and [f]Eureka Genomics, Hercules, CA 94547

Edited by Nancy A. Moran, University of Texas at Austin, Austin, TX, and approved October 1, 2013 (received for review September 6, 2013)

We present the complete genomic sequence of the essential symbiont *Polynucleobacter necessarius* (*Betaproteobacteria*), which is a valuable case study for several reasons. First, it is hosted by a ciliated protist, *Euplotes*; bacterial symbionts of ciliates are still poorly known because of a lack of extensive molecular data. Second, the single species *P. necessarius* contains both symbiotic and free-living strains, allowing for a comparison between closely related organisms with different ecologies. Third, free-living *P. necessarius* strains are exceptional by themselves because of their small genome size, reduced metabolic flexibility, and high worldwide abundance in freshwater systems. We provide a comparative analysis of *P. necessarius* metabolism and explore the peculiar features of a genome reduction that occurred on an already streamlined genome. We compare this unusual system with current hypotheses for genome erosion in symbionts and free-living bacteria, propose modifications to the presently accepted model, and discuss the potential consequences of translesion DNA polymerase loss.

symbiosis | nonsynonymous mutation rates | *Burkholderiales* | protozoa | genome streamlining

Symbiosis, defined as a close relationship between organisms belonging to different species (1), is a ubiquitous, diverse, and important mechanism in ecology and evolution (e.g., refs. 2–4). In extreme cases, through the establishment of symbiotic relationships, quite unrelated lineages can functionally combine their genomes and generate advantageous emergent features or initiate parasite/host arms races. Ciliates, common unicellular protists of the phylum Ciliophora, are extraordinary receptacles for prokaryotic ecto- and endosymbionts (5, 6) that provide varied examples of biodiversity and ecological roles (6). Nevertheless, most of these symbionts are understudied, partially owing to the scarcity of available molecular data and the absence of sequenced genomes. Yet, thanks to their various biologies and the ease of sampling and cultivating their protist hosts, they are excellent potential models for symbioses between bacteria and heterotrophic eukaryotes. Until recently this field was dominated by studies on endosymbionts of invertebrates, especially insects (e.g., ref. 7), although unicellular systems like amoebas (e.g., refs. 8 and 9) have been shown to be suitable models.

Polynucleobacter necessarius was first described as a cytoplasmic endosymbiont of the ciliate *Euplotes aediculatus* (10, 11). Further surveys detected its presence in a monophyletic group of fresh and brackish water *Euplotes* species (12, 13). All of the investigated strains of these species die soon after being cured of the endosymbiont (10, 12, 13). In the few cases in which *P. necessarius* is not present, a different and rarer bacterium apparently supplies the same function (12, 14). No attempt to grow symbiotic *P. necessarius* outside their hosts has yet been successful (15), strongly suggesting that the relationship is obligate for both partners, in contrast to most other known prokaryote/ciliate symbioses (6).

Thus, the findings of many environmental 16S rRNA gene sequences similar to that of the symbiotic *P. necessarius* (16) but belonging to free-living freshwater bacteria came as a surprise. These free-living strains, which have been isolated and cultivated (17), are ubiquitous and abundant in the plankton of lentic environments (17, 18). They are smaller and do not show the most prominent morphological feature of the symbiotic form: the presence of multiple nucleoids, each containing one copy of the genome (10, 11). It is clear that free-living and endosymbiotic *P. necessarius* are not different life stages of the same organism (15). Nevertheless, these strikingly different bacteria, occupying separate ecological niches, exhibit >99% 16S rRNA gene sequence identity, and phylogenetic analyses fail to separate them into two distinct groups (15). Rather, several lines of evidence point to multiple, recent origins of symbiotic strains from the free-living bacterial pool (14, 15).

Thus, the *Euplotes–Polynucleobacter* symbiosis provides a promising system for the study of changes promoting or caused by the shift to an intracellular lifestyle. The remarkably small (2.16 Mbp) genome of the free-living strain QLW-P1DMWA-1 has been sequenced and studied, especially for features that would explain the success of this lineage in freshwater systems worldwide (19, 20). Phylogenies based on the 16S rRNA gene (13, 14) and multiple-gene analyses (19, 21, 22) consistently cluster *Polynucleobacter* with bacteria of the family *Burkholderiaceae* (*Betaproteobacteria*), either in a basal position or as the sister group of *Ralstonia* and *Cupriavidus*.

Significance

We have investigated multiple aspects of the *Euplotes-Polynucleobacter* system, which provides a unique opportunity for the study of an obligate symbiont with a closely related free-living organism that itself possesses a peculiarly reduced genome and metabolism. We confirmed the robustness and generality of patterns in the evolution of bacterial symbionts' genome, adding at the same time new elements and hypotheses concerning genome reduction in both symbiotic and free-living bacteria. We argue that this system will provide an exceptionally useful model for investigations on symbiosis, because of its peculiarities and the commonness and ease of handling of the ciliate hosts. Genome sequences for independently derived *Polynucleobacter* symbionts will be particularly telling.

Author contributions: C.V., M.L., and G.P. designed research; C.V., P.S.G.C., S.M., L.M.V., and M.S. performed research; V.B., M.F., M.S.A., P.S.G.C., S.M., L.M.V., M.S., and T.G.D. analyzed data; V.B. and M.F. wrote the paper; and G.P. coordinated research.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. NC_010531.1).

[1]V.B. and M.F. contributed equally to this work.

[2]Present address: Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany.

[3]To whom correspondence should be addressed. E-mail: gpetroni@biologia.unipi.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1316687110/-/DCSupplemental.

Here we provide the complete genomic sequence of a symbiotic *P. necessarius* harbored in the cytoplasm of *E. aediculatus* and present a comparative analysis of the two sequenced *Polynucleobacter* genomes, addressing the possible biological basis of the *Euplotes–Polynucleobacter* symbiosis. We also provide insights into the evolution of the unique two-step genome reduction in this bacterial species: the first step involving streamlining in a free-living ancestor and the second a more recent period of genome erosion confined to the symbiotic lineage.

Results and Discussion

General Features of the Genome. The circular chromosome (Fig. S1) of the symbiotic *P. necessarius* strain is 1.56 Mbp long and contains approximately 1,279 protein-coding genes (*SI Results and Discussion, Genome Composition Analysis* and Table S1). The reduction in genome size that has occurred since the establishment of symbiosis can be estimated by comparing this genome with that of the *P. necessarius* free-living strain QLW-P1DMWA-1 (see also Fig. S2), an acceptable procedure given that the symbiont's gene set is largely a subset of that of the free-living strain (*SI Results and Discussion, Genome Composition Analysis* and Fig. S3). This reduction is more apparent as a decrease in the amount of coding DNA (42.3%) than in genome size (27.7%) because of the massive number of pseudogenes in the symbiont (Fig. 1). There are virtually no genes related to mobile elements and extremely few recently duplicated genes in the symbiotic isolate. Horizontally transmitted genes, mostly originated before the split between the two strains, are present (*SI Results and Discussion, Genome Composition Analysis*), but there is no sign of DNA exchange between the symbiont and its host *Euplotes*, nor any other eukaryote. A total of 277 genes in the symbiont genome (105 of which are not shared with the free-living relative) have unknown functions (*SI Results and Discussion, Genome Composition Analysis*).

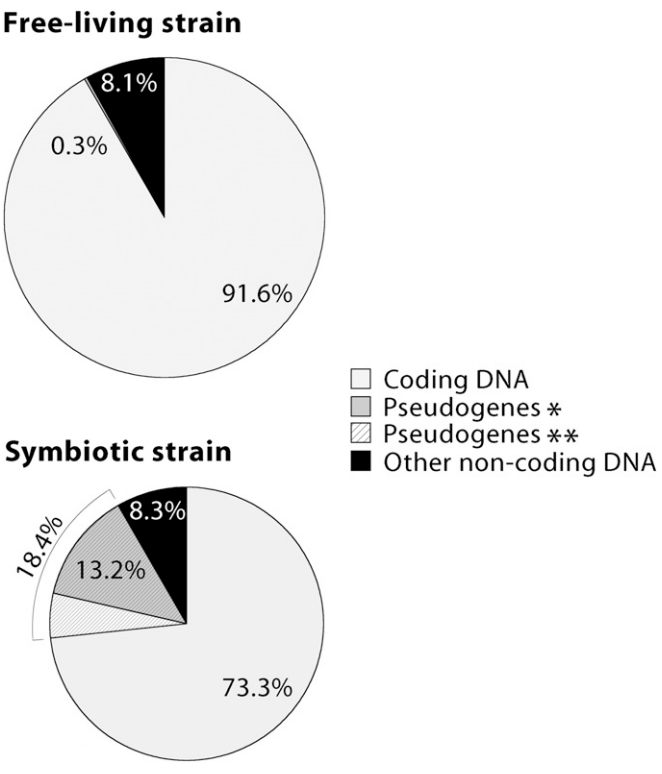

Fig. 1. Percentage allocation of coding and noncoding DNA in the free-living and symbiotic *Polynucleobacter* genomes. The graph areas are proportional to genome sizes (2.16 and 1.56 Mbp, respectively). Pseudogenes were identified with a more conservative (*) and a more permissive (**) approach, as detailed in *SI Results and Discussion, Genome Composition Analysis*.

Metabolism. ***Central metabolism and carbon sources.*** Both the symbiont and the free-living *P. necessarius* lack a glycolytic pathway. They do not possess the central regulatory enzyme of the Embden-Meyerhof pathway (6-phosphofructokinase) nor enzymes specific to the Entner-Doudoroff variant, which is used by most *Burkholderiaceae* bacteria. They also lack any enzyme that could phosphorylate glucose to glucose-6-phosphate or be involved in the assimilation of other monosaccharides. The inability to exploit sugars as carbon or energy sources reflects a general poverty in catabolic pathways.

Genomic analysis (Fig. 2 and *SI Results and Discussion, Details on Metabolic Analysis*; see also ref. 19) suggests that the principal carbon sources for the free-living strain are pyruvate, acetate, carboxylic acids, and probably compounds convertible to them. These can be metabolized to acetyl-CoA, which is the key intermediate of both energy production and anabolism, thanks to a complete glyoxylate cycle, tricarboxylic acids cycle (TCA), and gluconeogenesis pathways. Acetyl-CoA can additionally be directed to the synthesis of polyhydroxybutyrate (PHB), a storage polymer whose production has been investigated in the bacterium *Ralstonia eutropha*, also a member of *Burkholderiaceae* (23). A total of eight amino acids and all TCA intermediates can also be converted to glucose.

The symbiont possesses all of the aforementioned enzymatic paths with the exception of the glyoxylate cycle, the metabolic link between acetyl-CoA and biosynthetic pathways. Notably, genes involved in the polymerization, depolymerization, and metabolic regulation of PHB are still present. Thus, non-TCA carboxylic acids are not exploitable as sole carbon sources, and most enzymes acting on related compounds are missing. Only three amino acids can potentially be converted to glucose in the symbiont (*SI Results and Discussion, Details on Metabolic Analysis*). Thus, it either relies on a very small range of compounds as sole carbon sources, or directly imports various metabolic precursors from its host.

Nitrogen and sulfur metabolism. The free-living *P. necessarius* can perform the assimilatory reduction of nitrate imported from the environment, but the symbiont has lost this ability. An operon including all necessary enzymes for import and catabolism of urea is present only in the free-living strain. Both the free-living and symbiotic *Polynucleobacter* do not possess enzymes involved in nitrification, denitrification, or nitrogen fixation.

The free-living strain can assimilate elemental sulfur and sulfate, whereas the pathway is absent in the symbiont. Moreover, as discussed by Hahn and colleagues (19), the free-living strain possesses an entire set of *sox* genes and hence can probably obtain electrons from hydrogen sulfide (chemolithoheterotrophy). Many of these genes remain in the symbiont, but it is unclear whether the pathway is still functional.

Electron transport chain. Both genomes encode the entire electron transport chain complex and an F-type ATPase. Electrons must come mostly from carboxylic acids and, in the free-living strain, also from hydrogen sulfide. In addition to the most widespread cytochrome *c* oxidase complex, there is a variant in both—the cbb_3 complex—and additionally a *bd* complex in the free-living strain. The cbb_3 complex is present in the genera *Ralstonia* and *Cupriavidus*, and the *bd* complex is present also in *Burkholderia*; both are used in microaerophilic respiration. The symbiont apparently does not posses enzymes able to exploit terminal electron acceptors other than oxygen, whereas the free-living strain possesses a set of alcohol dehydrogenases, suggesting the possibility of energy production through fermentation under anaerobic conditions.

Amino acid and cofactor metabolism. Experimental evidence finds that the free-living isolate can grow on single carbon sources (like acetate) with a few cofactors (19), so we must assume that all amino acid biosynthetic pathways are somehow present in the genome, although it seems that in certain cases the bacterium

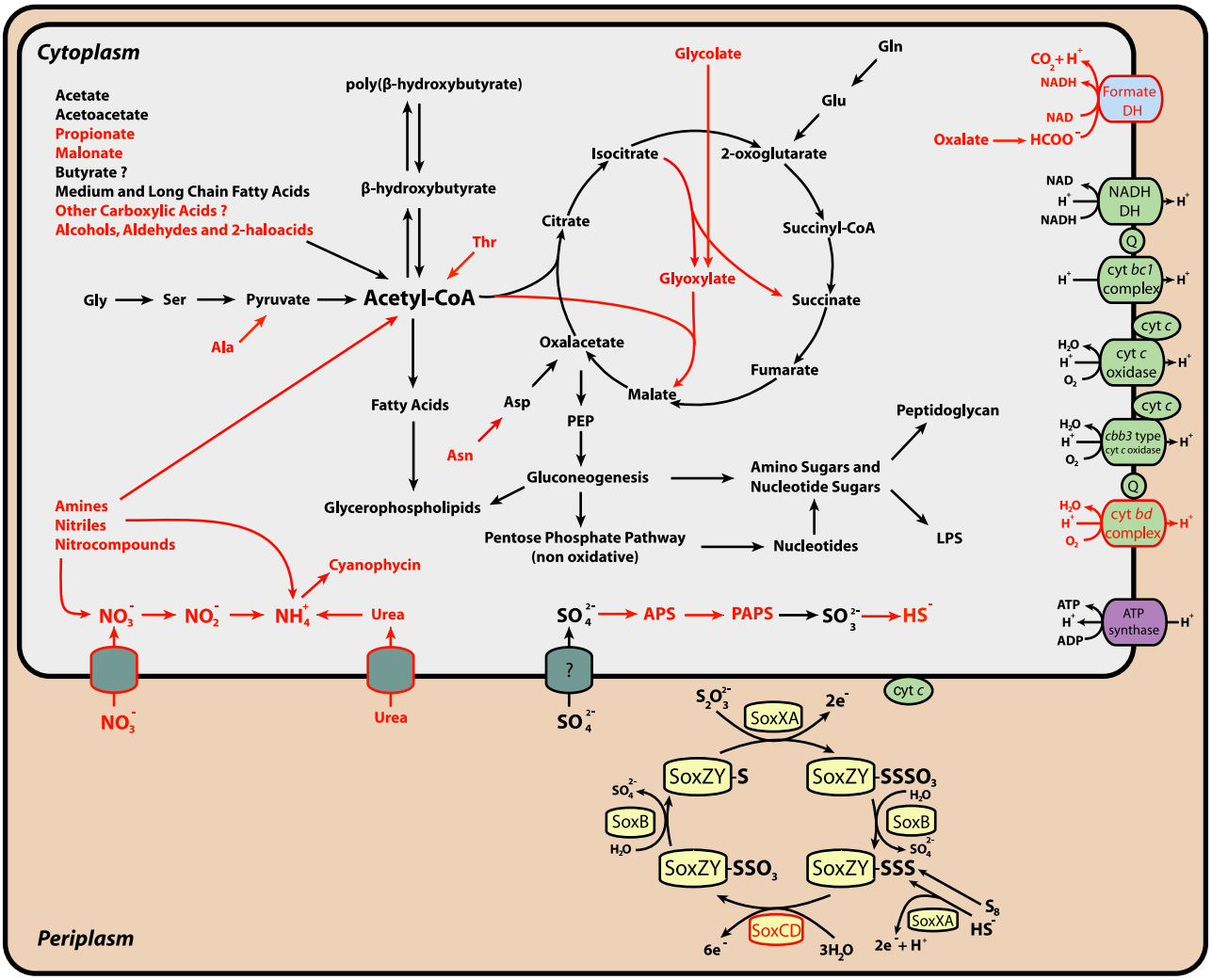

Fig. 2. Schematic drawing of selected pathways of *Polynucleobacter necessarius* metabolism as inferred by genomic analysis. Elements in red are exclusive to the free-living strain's genome. DH, dehydrogenase; cyt, cytochrome.

uses unconventional synthetic strategies (*SI Results and Discussion*, *Details on Metabolic Analysis*). The biosynthesis of four amino acids (alanine, aspartate, serine, and cysteine) has been lost in the symbiont. In a similar way, the symbiont cannot produce many cofactors, although the free-living strain lacks the ability to synthesize some of them as well (*SI Results and Discussion*, *Details on Metabolic Analysis*).

Other biosynthetic pathways. The major biosynthetic pathways stemming from TCA and gluconeogenesis (fatty acids biosynthesis, de novo and salvage nucleotide biosynthesis, and all those providing the necessary intermediates, e.g., the nonoxidative part of the pentose phosphate pathway) are present in both bacteria. As suggested by encoded genes, *P. necessarius* can synthesize phosphatidylserine, phosphatidylethanolamine, phosphatidylglycerol, and cardiolipin, but not phosphatidylinositol or lecithin, as membrane phospholipids. Typical bacterial structures of the cell wall (peptidoglycan and lipopolysaccharide) can be synthesized as well, although the symbiont apparently has a reduced set of genes involved in the recycling and modification of both.

Other Genomic Features. ***Gene classification.*** We classified the genes of both genomes in broad functional categories (*SI Results and Discussion*, *Genome Composition Analysis*), highlighting the differential amounts of gene loss in the symbiont compared with the free-living strain. Genes involved in core metabolism, protein metabolism, and growth are relatively more retained, whereas those related to membranes and transport, sensing and regulation, and other/unknown functions are much more reduced (Fig. S3).

DNA repair systems. DNA repair systems are considerably underrepresented in both *Polynucleobacter* strains. Although base and nucleotide excision repair pathways are intact, the enzymes for mismatch repair (MMR) are missing. Of the DNA polymerases able to perform translesion synthesis and rescue replication from arrest at points of chromosomal lesions (Pol II, Pol IV, and Pol V), only the error-prone Pol V is present in the free-living strain, and none in the symbiont.

No intact homologous recombination pathway is present, either: the *recBCD* system is completely missing, whereas the *recFOR* system lacks the essential gene *recF*. The Holliday junction resolution system (*ruvABC* and *recG*), however, is present. This suggests that in *Polynucleobacter* illegitimate recombination mechanisms are present. Indeed, the genes *recJ*, *recO*, *recR*, and *recA*, which are up-regulators of illegitimate recombination (24), are maintained despite the absence of *recF*.

Membranes and transport. There are fewer than 100 putatively functional genes for transport-related proteins in the symbiotic *P. necessarius* [mostly belonging to large families like ATP-binding cassette (ABC) transporters, tripartite ATP-independent periplasmic transporters (TRAP-T), and major facilitators]. In both strains the phosphotransferase system is reduced to nonspecific

components and a few putative cytoplasmic enzymes. The specificity of most transporters is difficult to assess (see also ref. 19).

Fewer differences between the strains were found in genes related to protein export. Both a Sec and Sec-independent pathways were found, as well as enzymes involved in the recognition of signal peptides, but no complete bacterial secretion system complex is encoded in either strain. Other membrane-related proteins are limited to lipoproteins and very conserved complexes involved in outer membrane integrity (Tol/Par operon), or outer membrane proteins' correct folding and targeting (YaeT complex). A huge putative membrane protein (10,429 aa) of unknown function is encoded by the free-living strain's genome (19) but is absent in the symbiont.

Cell-cycle, sensing, and stress resistance. Whereas genes involved in DNA metabolism (including replication) and cell cycle are relatively conserved between the *P. necessarius* strains, those involved in sensing and resistance were largely lost in the symbiont and are already few in number in the free-living isolate. Many genes of the two-component and TolB systems are likely nonfunctional in the symbiont. This suggests a strongly reduced ability to react to changes in the environment—the ciliate cytoplasm, which is probably more stable than the water column. Among environmental defenses, oxidative stress response systems are particularly reduced; the symbiont lacks, for example, glutathione synthetase, notwithstanding the presence of enzymes requiring glutathione (e.g., glutathione S-transferases).

Physiological Bases of the *Polynucleobacter–Euplotes* Symbiosis. The metabolic profile provides a clear explanation for the inability to grow symbiotic *P. necessarius* strains outside their hosts (15). The symbiont relies on the ciliate at least for carbon sources, organic nitrogen and sulfur, and other essential molecules, including many cofactors. In contrast to ancient or extremely specialized symbionts (e.g., ref. 21), though, it can still perform its own basic anabolic processes and energy production. This condition is similar to that of the other symbiotic member of the family *Burkholderiaceae* with a severely reduced genome: "*Candidatus* Glomeribacter gigasporarum" (22), a beneficial endosymbiont of arbuscular mychorrhizal fungi. *P. necessarius* and "*Ca.* G. gigasporarum" underwent independent events of genome reduction [see also the phylogenetic tree of Ghignone et al. (22)] that produced the loss of different metabolic pathways. For example, "*Ca.* G. gigasporarum" has conserved more amino acid degradation pathways than *P. necessarius* (*SI Results and Discussion*, *Details on Metabolic Analysis*) but lacks instead the β-oxidation pathway (22). Both share the loss of glycolytic pathways.

The symbiotic *P. necessarius* is probably specialized for the intracellular environment in other aspects too, as suggested by its reduced set of genes for sensing and stress resistance. Defensive and regulation mechanisms, as well as membrane and cell wall plasticity, seem to be weakened, providing other plausible reasons for the inability of the bacterium to grow outside its host.

It is more difficult to understand why the symbiont is essential for *Euplotes* survival (12, 15). Many obligate symbionts of eukaryotes described as mutualists serve as a source of essential molecules (7, 25). The possibility that *P. necessarius* provides at least some metabolites to its host cannot be completely ruled out, but we consider it unlikely. *Euplotes* are heterotrophic algal and bacterial feeders and can probably obtain all required amino acids and cofactors from their diet, unlike specialized feeders like sap-feeding insects. More likely, the ciliate host requires *Polynucleobacter* to fix a catabolic deficiency (e.g., in compound degradation) in a pathway usually conserved in both bacteria and eukaryotes but lost in the clade of *Euplotes* species harboring *Polynucleobacter*. Vannini et al. (26) worked on a similar premise and provided evidence for a possible role of these bacteria in glycogen depolymerization; nevertheless, we found nothing in the genome supporting this hypothesis, so the real catabolic pathway involved remains uncertain. The genome of a *Polynucleobacter*-harboring *Euplotes* and further experimental investigation will be able to better address the matter, now that the genomic bases of *Polynucleobacter* biology are established.

It has been pointed out (27) that obligate bacterial symbionts generally do not have "symbiotic genes" coding for exotic functions. Their genomes are fundamentally a subset of those of free-living relatives, and the functional role of symbiosis is better explained by a metabolic cooperation between partners (28, 29). An interesting exception are bacterial secretion systems (BSSs), which are probably involved in the ancestral invasion process and are often found in an active or degraded form in symbiont genomes (25). The symbiotic *P. necessarius* strain does not encode a complete BSS, contrary to some more parasitic-like bacteria that infect amoebas (8). Nevertheless, we found a region similar to a pathogenicity island (positions 625,503–639,686) including 16 ORFs, 10 of which show similarities with type II and type IV secretion system assembly protein genes (the other 6 ORFs do not share significant similarities with any available sequence). This region is of horizontal origin (*SI Results and Discussion*, *Genome Composition Analysis*) and is absent in the free-living strain genome. Most of the genes are considerably shorter (range, 13–78%) than their closest homologs and are possibly nonfunctional. An ancestral free-living strain may have acquired these "invasion genes," which allowed it to survive ingestion and digestion in a predatory *Euplotes*, and was then trapped in the cytoplasm. The ubiquity of free-living *P. necessarius* and the high frequency of BSS genes horizontal transmission, together with no apparent requirement for the evolution of novel functions (at least at the genomic level), may have facilitated the multiple origins and complex pattern of replacement inferred for these essential *Euplotes* symbionts (14).

Genomic Reduction in the Symbiont. The symbiotic *P. necessarius* isolate possesses one of the smallest genomes observed so far in *Betaproteobacteria*, being surpassed only by two exceptional cases that blur the distinction between organism and organelle: "*Candidatus* Tremblaya princeps" (0.14 Mbp, 148 genes) (30) and "*Candidatus* Zinderia insecticola" (0.21 Mbp, 232 genes) (29).

Progressive genome reduction is the rule for obligate symbionts (30, 31). This process has been explained by nonselective mechanisms in insects' symbionts: either relaxed selection and enhanced genetic drift (32, 33) or an increase in mutation rates (34). The first hypothesis stems from the decreased number of essential functions in symbionts and their small population sizes, reduced gene exchange, and frequent occurrence of bottlenecks. The prediction is for a higher synonymous/nonsynonymous (dN/dS) ratio in symbiotic lineages, which has often been reported (32, 33), although a potential bias resulting from synonymous-site saturation has been suggested (34, 35). In *P. necessarius*, dN is slightly higher in the symbiotic than in the free-living strain (Fig. 3), in accordance with expectations. Nevertheless, it was impossible to obtain confident estimates of dS in the two lineages despite their unparalleled level of sequence similarity, because synonymous sites were still saturated on the branch leading to the closest outgroup. When more *P. necessarius* genomes of both free-living and symbiotic strains become available, this problem will almost certainly be circumvented with the help of an intraspecific outgroup.

Moran et al. (31) and McCutcheon and Moran (36) described the steps of genome erosion: the crucial turning point is the loss of DNA repair mechanisms, which brings increased mutation rates, an A+T bias, and massive gene inactivation driven by the spread of mobile elements (MEs). Under these conditions, elimination of DNA in noncoding regions is expected to be due to an intrinsic deletion bias (30, 37). Most of our findings are compatible with this scenario, notwithstanding the differences between the systems studied: *P. necessarius* does not experience a bottleneck during the host asexual division, and ciliate effective population sizes are arguably larger than those of insects. Where our system stands in striking contrast to the prior model is the virtual absence of MEs, paired with the abundance of pseudogenes. MEs and pseudogenes should be a signature of the first

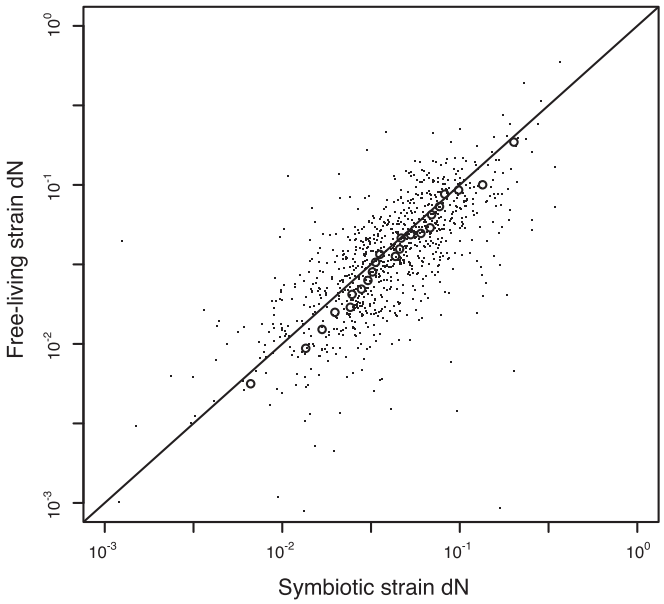

Fig. 3. Nonsynonymous divergence between 904 homolog functional genes in the free-living and symbiotic *Polynucleobacter necessarius*. The rate of divergence from the outgroup is higher in the symbiont (averages: 0.0556, SE: 0.0049 vs. 0.0479, SE: 0.0035; $P = 0.006$, two-tailed t test). Axes are in logarithmic scale, and each circle represents the average position of 40 gene-windows projected onto the diagonal.

stages of genome erosion and until now have always been found together and in large numbers in recently evolved symbionts (30, 35, 36), but in very low quantity, or entirely missing, in the extremely reduced genomes of more ancient symbionts (36). In this regard, the *P. necessarius* case is an exception in recent symbionts and may be explained by the preexisting small size of the ancestral genome or by a relatively large effective population size. It demonstrates, however, that the multiplication of MEs is not a required step of genome erosion.

An aspect that has been less investigated is the role of translesion DNA polymerases (TLPs), and we propose here a scenario that may deserve attention in future studies. The symbiotic *P. necessarius* lacks all TLPs. Although the loss of MMR and homologous recombination may increase the rate of mildly deleterious mutations, the loss of TLPs introduces the risk that a single damaged nucleotide, even in noncoding regions, can entirely block replication. This expectation is supported by the severe reduction in survival of bacteria with artificially impaired translesion synthesis under DNA-damaging conditions (38, 39) and reduced fitness under nonstress conditions (40). A small genome, however, provides less target sites for damage. If the last TLP was lost passively during the early stages of gene inactivation, this would have exerted an additional pressure toward deletions in sequences whose function is not selected for maintenance. The *umuC* and *umuD* genes (encoding the two subunits of Pol V, the only TLP present in the free-living strain) are not even recognizable as pseudogenes in the genome of the symbiotic *Polynucleobacter*, suggesting that the loss of translesion synthesis happened relatively early during the genome erosion process.

Not all proximate causes of genome reduction are known, but among them there is illegitimate recombination, apparently more so when coupled with MMR loss (24). Although the molecular pathways involved are not entirely characterized, TLPs actually participate in deletion accumulation during illegitimate recombination in *Salmonella* (41). However, other mechanisms have been proposed in different systems (e.g., ref. 42). Illegitimate recombination and absence of MMR likely constitute the main source of deletions in *Polynucleobacter* through TLP-independent mechanisms.

The probability that an aborted chromosome replication leads to cell death could be alleviated by multiple copies of the genome. It is thus intriguing to also link the loss of TLPs with the presence of multiple nucleoids in *Polynucleobacter*. Polyploidy has been observed in other intracellular symbionts with reduced genomes, such as *Buchnera* (27), and this explanation might apply to them too (and is not in contradiction with the possibility that a multiple-copy genome is advantageous because of increased gene expression). An alternative hypothesis is that multiple nucleoids arose in *P. necessarius* before the loss of the last remaining TLP, perhaps as byproduct of an unbalanced cell cycle, and paved the way for reduced selection on translesion synthesis maintenance. However, the proximal causes of the polyploidy in the symbiont are not immediately apparent from genomic sequences. The symbiont and the free-living *P. necessarius* share a very similar set of genes involved in chromosome segregation and cell division [including the ParAB-*parS*, structural maintenance of chromosome (SMC), and filamentous temperature-sensitive proteins (Fts) systems (43)].

Genomic Reduction in the Free-Living Progenitor. As stressed above, the genome of the symbiotic *P. necessarius* strain is largely a subset of that of its conspecific relative. However, the free-living *P. necessarius* strain QLW-P1DMWA-1 already possesses a remarkably small genome (19), comparable in size to that of beta-proteobacterial obligate pathogens like *Neisseria* (2.09–2.28 Mbp) and "*Ca.* G. gigasporarum" (approximately 1.72 Mbp) and much smaller than those of other free-living or facultative symbiotic *Burkholderiaceae* (range, 3.75–9.73; Table S2). Thus, a first event of gene loss in the *Polynucleobacter* lineage occurred before the establishment of the symbiosis, presumably in a free-living ancestor.

Genome streamlining in free-living bacteria is less understood than genome erosion in symbionts but has drawn attention in *Prochlorococcus* (44) and "*Candidatus* Pelagibacter ubiquae" (45), two marine taxa with huge global populations. Most authors have argued for adaptive explanations of genome streamlining in free-living bacteria (44, 46), and selection-driven gene loss has been reported for experimental populations (47, 48). In particular, mechanisms requiring relaxed selection and increased drift are generally considered unrealistic because of the huge population sizes of these bacteria.

Nevertheless, the reduced genomes of symbionts and the streamlined genomes of free-living bacteria share many analogies, like a higher AT content and reduced DNA repair mechanisms, in particular MMR loss [reported here for *P. necessarius* and also observed in strains of *Prochlorococcus* (44) and "*Candidatus* Pelagibacter" (49)]. This suggests that nonadaptive mechanisms may have shaped the genomes of these free-living lineages too, perhaps acting in the past when the population sizes were smaller. Although likely detrimental during the first stages, a genome erosion driven by the same mechanisms invoked for symbionts may produce a compact, specialized, and less costly metabolism that could be very successful in the right environment.

Materials and Methods

Purification of the Endosymbiont DNA. Cultures of the ciliate host *E. aediculatus* strain STIR1 were starved, filtered, and treated with chloramphenicol to minimize the amount of contaminating bacteria in the culture medium. The cells were mechanically ruptured to release the symbionts, which were then separated from the eukaryote's cellular fragments through centrifugations at increasing accelerations. Total genomic DNA was extracted from the supernatant after the final centrifugation (protocols and quality controls are detailed in *SI Materials and Methods*).

Sequencing, Assembling, and Annotation. The genome of this symbiotic strain of *P. necessarius* was selected for sequencing on the basis of the US Department of Energy Joint Joint Genome Institute Community Sequencing Program 2006 and is publicly available (accession number NC_010531.1). Sequencing, assembling, and annotation were performed as described elsewhere for the conspecific free-living strain QLW-P1DMWA-1 genome (20). Pseudogenes were identified on the base of reading frame interruptions and/or substantially shorter sequence length with respect to

orthologs, using one more conservative and one more permissive threshold (details in *SI Results and Discussion, Genome Composition Analysis* and Table S1).

Functional Analysis. Predicted ORFs from symbiotic and free-living *P. necessarius* were compared against the nonredundant protein sequences database using BlastP (50). A list of best results was produced for each ORF, and the putative protein product inferred. The KAAS–KEGG Automatic Annotation Server (51) was used as an aid in the interpretation of metabolic pathways in *Polynucleobacter* and to comparatively screen a set of complete genomes from other *Burkholderiaceae* bacteria (Table S2). Data on available genomes were obtained from the National Center for Biotechnology Information microbial genomes webpage (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html).

Substitution Rates. Homologous sequences were defined by closed groups of reciprocal best hits determined from BlastP (50) among the pooled genes belonging to the genomes listed in Table S2. Amino acid sequences were aligned with MUSCLE (52), and nonsynonymous substitution rates were inferred from using PAML (53).

ACKNOWLEDGMENTS. We thank Martin W. Hahn for his fundamental contribution in supporting, as main applicant, the *Polynucleobacter* genome project, performed by the US Department of Energy Joint Genome Institute with support by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231; Matteo Meliani and Clizia Gennai for help in DNA isolation; and Simone Gabrielli for the technical assistance in graphic artwork. This work was supported by European Commission FP7-PEOPLE-2009-IRSES project CINAR PATHOBACTER (Project 247658) and by European Cooperation in Science and Technology (COST) action BM1102.

1. de Bary A (1879) *Die Erscheinung der Symbiose*, ed Trübner KJ (Verlag von Karl, Strassburg).
2. Dyall SD, Brown MT, Johnson PJ (2004) Ancient invasions: From endosymbionts to organelles. *Science* 304(5668):253–257.
3. Gilbert SF, Sapp J, Tauber AI (2012) A symbiotic view of life: We have never been individuals. *Q Rev Biol* 87(4):325–341.
4. Moran NA (2007) Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci USA* 104(Suppl 1):8627–8633.
5. Fokin SI (2012) Frequency and biodiversity of symbionts in representatives of the main classes of Ciliophora. *Eur J Protistol* 48(2):138–148.
6. Görtz H-D (2006) *The Prokaryotes*, eds Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (Springer, New York), pp 364–402.
7. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190.
8. Horn M, et al. (2004) Illuminating the evolutionary history of chlamydiae. *Science* 304(5671):728–730.
9. Schmitz-Esser S, et al. (2010) The genome of the amoeba symbiont "*Candidatus* Amoebophilus asiaticus" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. *J Bacteriol* 192(4):1045–1057.
10. Heckmann K (1975) *Omikron*, ein essentieller Endosymbiont von *Euplotes aediculatus*. *J Protozool* 22(1):97–104.
11. Heckmann K, Schmidt HJ (1987) *Polynucleobacter necessarius* gen. nov., sp. nov., an obligately endosymbiotic bacterium living in the cytoplasm of *Euplotes aediculatus*. *Int J Syst Bacteriol* 37(4):456–457.
12. Heckmann K, Ten Hagen R, Görtz H-D (1983) Freshwater *Euplotes* species with a 9 type 1 cirrus pattern depend upon endosymbionts. *J Protozool* 30(2):284–289.
13. Vannini C, Petroni G, Verni F, Rosati G (2005) *Polynucleobacter* bacteria in the brackish-water species *Euplotes harpa* (Ciliata Hypotrichia). *J Eukaryot Microbiol* 52(2):116–122.
14. Vannini C, Ferrantini F, Ristori A, Verni F, Petroni G (2012) Betaproteobacterial symbionts of the ciliate *Euplotes*: Origin and tangled evolutionary path of an obligate microbial association. *Environ Microbiol* 14(9):2553–2563.
15. Vannini C, et al. (2007) Endosymbiosis in statu nascendi: Close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*). *Environ Microbiol* 9(2):347–359.
16. Zwart G, Crump BC, Kamst-van Agterveld MP, Hagen F, Han SK (2002) Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat Microb Ecol* 28(2):141–155.
17. Hahn MW (2003) Isolation of strains belonging to the cosmopolitan *Polynucleobacter necessarius* cluster from freshwater habitats located in three climatic zones. *Appl Environ Microbiol* 69(9):5248–5254.
18. Jezberová J, et al. (2010) Ubiquity of *Polynucleobacter necessarius* ssp. asymbioticus in lentic freshwater habitats of a heterogeneous 2000 km^2 area. *Environ Microbiol* 12(3):658–669.
19. Hahn MW, et al. (2012) The passive yet successful way of planktonic life: Genomic and experimental analysis of the ecology of a free-living *Polynucleobacter* population. *PLoS ONE* 7(3):e32772.
20. Meincke L, et al. (2012) Complete genome sequence of *Polynucleobacter necessarius* subsp. *asymbioticus* type strain (QLW-P1DMWA-1^T). *Stand Genomic Sci* 6(1):74–83.
21. McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* 2:708–718.
22. Ghignone S, et al. (2012) The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *ISME J* 6(1):136–145.
23. Pohlmann A, et al. (2006) Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. *Nat Biotechnol* 24(10):1257–1262.
24. Nilsson AI, et al. (2005) Bacterial genome size reduction by experimental evolution. *Proc Natl Acad Sci USA* 102(34):12112–12116.
25. Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. *Cell* 126(3):453–465.
26. Vannini C, Lucchesi S, Rosati G (2007) *Polynucleobacter*: Symbiotic bacteria in ciliates compensate for a genetic disorder in glycogenolysis. *Symbiosis* 44(1-3):85–91.
27. Moran NA, Degnan PH (2006) Functional genomics of *Buchnera* and the ecology of aphid hosts. *Mol Ecol* 15(5):1251–1261.
28. Hansen AK, Moran NA (2011) Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci USA* 108(7):2849–2854.
29. McCutcheon JP, von Dohlen CD (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 21(16):1366–1372.
30. Moran NA (2003) Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr Opin Microbiol* 6(5):512–518.
31. Moran NA, McLaughlin HJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* 323(5912):379–382.
32. Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 93(7):2873–2878.
33. Wernegreen JJ, Moran NA (1999) Evidence for genetic drift in endosymbionts (*Buchnera*): Analyses of protein-coding genes. *Mol Biol Evol* 16(1):83–97.
34. Itoh T, Martin W, Nei M (2002) Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci USA* 99(20):12944–12948.
35. Burke GR, Moran NA (2011) Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol* 3:195–208.
36. McCutcheon JP, Moran NA (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10(1):13–26.
37. Kuo C-H, Ochman H (2009) Deletional bias across the three domains of life. *Genome Biol Evol* 1:145–152.
38. Jarosz DF, Cohen SE, Delaney JC, Essigmann JM, Walker GC (2009) A DinB variant reveals diverse physiological consequences of incomplete TLS extension by a Y-family DNA polymerase. *Proc Natl Acad Sci USA* 106(50):21137–21142.
39. Kuban W, et al. (2012) *Escherichia coli* UmuC active site mutants: Effects on translesion DNA synthesis, mutagenesis and cell survival. *DNA Repair (Amst)* 11(9):726–732.
40. Yeiser B, Pepper ED, Goodman MF, Finkel SE (2002) SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proc Natl Acad Sci USA* 99(13):8737–8741.
41. Koskiniemi S, Andersson DI (2009) Translesion DNA polymerases are required for spontaneous deletion formation in *Salmonella typhimurium*. *Proc Natl Acad Sci USA* 106(25):10248–10253.
42. Ikeda H, Shiraishi K, Ogata Y (2004) Illegitimate recombination mediated by double-strand break and end-joining in *Escherichia coli*. *Adv Biophys* 38:3–20.
43. Reyes-Lamothe R, Nicolas E, Sherratt DJ (2012) Chromosome replication and segregation in bacteria. *Annu Rev Genet* 46:121–143.
44. Dufresne A, et al. (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci USA* 100(17):10020–10025.
45. Giovannoni SJ, et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309(5738):1242–1245.
46. Morris JJ, Lenski RE, Zinser ER (2012) The Black Queen Hypothesis: Evolution of dependencies through adaptive gene loss. *mBio* 3(2):e00036–e12.
47. Lee M-C, Marx CJ (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genet* 8(5):e1002651.
48. Koskiniemi S, Sun S, Berg OG, Andersson DI (2012) Selection-driven gene loss in bacteria. *PLoS Genet* 8(6):e1002787.
49. Viklund J, Ettema TJ, Andersson SG (2012) Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol Biol Evol* 29(2):599–615.
50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
51. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35(Web Server issue):W182–185.
52. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
53. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.

Supporting Information

Boscaro et al. 10.1073/pnas.1316687110

SI Results and Discussion

Genome Composition Analysis. Both *Polynucleobacter necessarius* genomes are encoded by a single circular chromosome, 2.16 Mbp long in the free-living strain QLW-P1DMWA-1 and 1.56 Mbp long in the symbiotic strain harbored by *Euplotes aediculatus* STIR1.

Pseudogenes were identified with the software GenePRIMP (1) according to two criteria: interruptions of the reading frame (stop codons, frameshift mutations) and/or severely reduced length (less than 50% of the five best hit homologs, defined as nonfused genes with at least 40% sequence identity with the query gene). In addition to this, we considered as likely nonfunctional those genes whose length in the symbiont genome is less than 75% that of the ortholog in the free-living strain genome (and vice versa). All of the genes in this category encode single enzymes in otherwise missing multiple-gene pathways or transporter operons (that are instead completely encoded in the free-living *Polynucleobacter* genome) and thus likely represent nonfunctional remnants of the process of genome erosion.

The remaining, putatively functional protein-coding genes were divided into eight categories:

i) Metabolism, coding for known enzymes performing reactions involved in core metabolism, energy production, nitrogen and sulfur metabolism, and biosynthetic/catabolic processes.
ii) Transcription, translation, and posttranslational protein processing (including protein degradation).
iii) DNA metabolism and cell cycle.
iv) Membranes and transport.
v) Sensing, homeostasis, and regulation.
vi) Mobile elements-related, coding for transposases, integrases, and/or of phage- or plasmid-origin.
vii) Others, coding for enzymes belonging to known families but whose precise role in cell physiology is not known.
viii) Unknown, putative protein genes that could not be unambiguously assigned.

Table S1 reports comparative data, whereas Fig. S3 visually depicts the degree of gene loss in the symbiont's genome for each protein-coding gene category.

The most apparent feature of the symbiont's genome is its huge number of pseudogenes compared with that of the free-living strains. This strongly reduces the percentage of coding DNA, which is high in the free-living *Polynucleobacter* genome (91.6%). The reduction in the rest of noncoding DNA is instead similar to the genome size reduction (26.4%).

Genes related to metabolic processes (categories 1-3) are relatively more conserved, with a high percentage of shared genes and a comparatively lower number of genes exclusive to the free-living strain. All other categories are strongly underrepresented in the symbiont. Genes included in category 8 ("Unknown") are generally shorter than those in other categories, often less than 300 bp long in both genomes (average in the symbiont: 542 bp). One hundred five predicted genes in the symbiont are not shared with the free-living strain, but these are even shorter (average: 384 bp), and 75 do not share significant sequence similarity with any other available gene. It is plausible that many of them are not actually active, thus reducing the number of proteins expressed by the bacteria and consequently the probability that important features went unnoticed in our analysis.

In all categories the percentage of genes exclusive to the symbiotic strain is negligible, thus allowing use of the related free-living strain's genome as an approximation of that of the common ancestor. Orthologous gene sequences tend to be shorter in the symbiont (average: 93.4%) too.

Horizontally transmitted genes mostly originated before the split between the two strains: the best BlastP hit still is the ortholog in the conspecific genome, but all of the others belong to non-betaproteobacterial prokaryotes. There are approximately 80 of these genes in the symbiont's genome; approximately half have no known function. All genes in the putative pathogenicity island discussed in the main text are of horizontal origin (BlastP best hits include unrelated bacteria like *Pseudomonas*, "*Candidatus* Pelagibacter," and *Legionella*) and do not share an ortholog in the free-living strain's genome.

Whole-genome alignment performed with Mauve software (2) revealed one major inversion event and several minor ones since the divergence of the two strains (Fig. S2).

Details on Metabolic Analysis. *Carbon sources.* According to the analysis of coded enzymes, the principal carbon sources for the free-living strain are pyruvate, tricarboxylic acid cycle (TCA) intermediates, acetate, carboxylic acids, and carboxylic acid derivatives. Some of those compounds constitute the low-molecular-weight photooxidation products of humic substances on which Hahn et al. (3) observed growth of bacterial cultures. In such a metabolism acetyl-CoA represents the key intermediate of both energy production and anabolism, thanks to complete glyoxylate cycle, TCA, and gluconeogenesis pathways. No enzymes involved in mono- and polysaccharide catabolism were found, with the notable exception of a cellulase, which in the free living could be involved in the regulation of biofilm formation in antagonism of cellulose synthase (see below).

Pyruvate and gluconeogenesis. Pyruvate can be directly converted to phosphoenolpyruvate (PEP) with PEP synthetase only in the free-living *Polynucleobacter*. Thus, because of the lack of a complete glyoxylate cycle, pyruvate cannot be used as a sole carbon source in the symbiont. The same is true for all of the carbon sources that are converted to acetyl-CoA as intermediate. Oxalacetate can be used as a substrate for gluconeogenesis through phosphoenolpyruvate carboxykinase in both strains, thus TCA intermediates are glucogenic.

Acetate and acetoacetate derivatives. In both symbiotic and free-living *Polynucleobacter*, acetate can be directly assimilated through acetyl-CoA synthetase. In many bacteria this enzyme was shown to act on propionate and butyrate as well (4). In both strains, acetoacetate can be transferred on CoA and metabolized thanks to a 3-oxoacid-CoA transferase, whereas 3-hydroxybutyrate, coming from the degradation of poly-β-hydroxybutyrate, can be used as a further carbon source after oxidation to acetoacetate. Chloroacetate and other 2-halocids could be metabolized by the 2-haloacid dehalogenase present in both bacteria.

Propionate. In the free-living *Polynucleobacter*, propionate can be directly assimilated through a propionyl-CoA synthetase. Apparently the methylmalonyl-CoA pathway for the propionate assimilation is not complete owing to the lack of methylmalonyl-CoA epimerase; however, propionate could be metabolized, at least in the free-living strain, through a modified β-oxidation pathway involving malonic semialdehyde or 3-oxopropionyl-CoA/malonyl-CoA as intermediates. The ability to catabolize propionate is uncertain in the symbiont because it lacks propionyl-CoA synthetase (although the reaction could be performed by acetyl-CoA synthetase). The β-oxidation pathway variant involving malonic semialdehyde is not functional, whereas the variant involving 3-oxopropionyl-CoA/malonyl-CoA as intermediates might be

present but is not well characterized from the enzymatic and genetic point of view.

Fatty acids and lipids. A complete β-oxidation pathway is present in both the free-living and symbiotic strain, even though they do not seem to be able to metabolize short chain fatty acids (C6, C8, C10) because they lack the short chain acyl-CoA synthetase. This is in accordance with the assimilation tests performed by Hahn et al. (3, 5) on the free-living bacterium. On the other hand medium and long chain fatty acids can be transferred on the CoA by a synthetase. More complex lipids, such as triacylglicerids and some phospholipids, can be metabolized as well owing to the presence of three secretory lipases and three putative patatin-like phospholipases.

Other carboxylic acids. In the free-living strain's genome the presence of a gene encoding for a CoA transferase, which in *Ralstonia* catalyzes reversible transfer reactions of a CoA group from CoA-thioesters to free acids using acetyl-CoA as a CoA donor (6), could be responsible for the assimilation of other carboxylic acids. Malonate can be assimilated thanks to the presence of a specific malonyl-CoA transferase. Only in the free-living strain, three organic acids (pyruvate, malate, and L-lactate) can be oxidized by specific oxidoreductases that direct the reducing equivalents directly into the electron transport chain: pyruvate and malate oxidoreductases reduce the quinone, and L-lactate dehydrogenase reduces the cytochrome.

Glycolate and tartrate. As previously shown by Hahn et al. (3), the free-living strain has the genetic potential to assimilate glycolate and phospho-glycolate (products of algal photorespiration), which are important components of algal exudates. Furthermore, the free-living bacterium seems to have the enzymatic activities (tartrate dehydrogenase/decarboxylase) for the conversion of tartrate, another important component of vegetal exudates, to D-glycerate. It has also the enzyme for the conversion of hydroxypyruvate to D-glycerate. Surprisingly it was not possible to find any gene encoding for a glycerate kinase, which catalyzes the synthesis of the glycolysis intermediate 3-phosphoglycerate. No functional gene for glycolate and tartrate degradation was found in the symbiont.

Oxalate and formate. Only in the free-living bacterium, oxalate can be used as an additional source of electrons (but not as a carbon source). Oxalate can be converted to formate through the concurrent action of the formyl-CoA transferase and oxalyl-CoA decarboxylase. Formate can be further oxidized to CO_2 with the production of NADH, owing to the presence of at least one complete set of genes, present only in the free-living bacterium, encoding for a functional formate dehydrogenase.

Alcohols and aldehydes. Apparently only the free-living *Polynucleobacter* strain has the genetic potential to assimilate organic compounds, such as alcohols and aldehydes, which can be converted in carboxylic acids. Indeed we have found different alcohol dehydrogenases (Fe-dependent, Zn-dependent, short chain dehydrogenase) and an aldehyde dehydrogenase.

Nitrogen-containing compounds. Some nitrogen-containing compounds can be used as additional carbon and nitrogen sources. The free-living bacterium has the possibility to degrade amines and nitroalkanes through an amine oxidase and a nitronate monooxygenase, respectively. These two enzymes remove the nitrogen-containing group producing the relative aldehyde. The hydrolysis of acetamide and formamide is also possible only in the free-living *Polynucleobacter*, thanks to an acetamidase/formamidase.

Aromatic compounds. Finally, a series of enzymes involved in the aromatic compounds degradation pathways were found only in the free-living strain's genome. In particular, we identified several enzymes of the 3-oxoadipate pathway, especially of the ortho-cleavage of catechole, protocatechuate, and chlorobenzoate, three putative nitroreductase genes (responsible of the reduction and the degradation of nitroaromatic compounds), an arylesterase, and a benzoylformate decarboxylase (involved in the benzoate degradation pathway). The 3-oxoacid-CoA transferase discussed above could be involved in the late steps of this pathway. Nevertheless, it was not possible to identify a complete degradation pathway.

Amino acids. *Biosynthesis.* Ammonia assimilation is achieved through the concerted activity of glutamate synthase (GOGAT) and glutamine synthase in both the symbiotic and free-living strain. A monomeric ferredoxin-dependent GOGAT is present in both strains, whereas a dimeric NADPH-dependent glutamate synthase is present only in the free-living strain.

Aspartate transaminase is apparently missing in *P. necessarius*, a feature shared by other bacteria in the *Burkholderiaceae* family (in our screening of the available genomes we found this enzyme only in *Burkholderia phenoliruptrix* BR3459a, *Burkholderia phytofirmans* PsJN, and *Ralstonia solanacearum* GMI1000); one possibility is that aspartate ammonia-lyase, usually a degradation enzyme, is used anabolically as in *Rhizobium lupini* (7). The symbiont lacks even this enzyme. Alternatively, aspartate could be produced through aromatic-amino acid aminotrasferase, which has a significant role in aspartate formation in many bacteria (8). Asparagine synthetase is missing, but the asparagine required for protein synthesis is available through modification of aspartyl-$tRNA^{Asn}$ owing to the presence of the genes *gatA* and *gatB*. This indirect pathway is found in most bacteria, but it usually coexists with the standard one (9). Instead, in *Polynucleobacter* it seems to be the only possibility for the synthesis of the asparaginyl-$tRNA^{Asn}$, a feature also observed in *Deinococcus radiodurans* and *Thermus thermophilus* (10). Methionine biosynthesis through the transsulfuration pathway is not possible because of the lack of one required enzyme in the free-living strain (cystathionine γ-synthase) and two in the symbiont (missing cystathionine β-lyase too). The presence of all of the genes for the biosynthetic pathway of homoserine, plus a homolog of O-succinylhomoserine sulfhydrylase (*metZ*), suggests the possibility that homocysteine is produced through direct sulfhydrilation of O-acetyl-L-homoserine, as seen in *Pseudomonas putida* (11).

The differences between the symbiont and the free-living strains' genomes are that the symbiont has lost one enzyme necessary for cysteine biosynthesis (serine O-acetyltransferase) and one for serine biosynthesis (D-3-phosphoglycerate dehydrogenase). It also lacks glutathione synthase but apparently maintains other enzymes necessary for glutathione biosynthesis and those related to glutathione-dependent detoxification.

Degradation. Compared with the other analyzed members of the *Burkholderiaceae* family, both bacteria have extremely reduced amino acid degradation pathways. A common feature shared with the bacteria belonging to *Burkholderiaceae* is the loss of lysine, tryptophan, and methionine degradation pathways (with the only exception of *B. phenoliruptrix* BR3459a, which has a cystathionine γ-lyase for methionine degradation). Because of the lack of the branched-chain α-ketoacid dehydrogenase complex, the degradation pathways of valine, leucine, and isoleucine seem to be absent in many bacteria of the *Burkholderiaceae* family, as well as in *P. necessarius* (complete degradation pathways were found only in *Burkholderia pseudomallei* K96243, *Burkholderia thailandensis* E264, *Burkholderia mallei* ATCC 23344, *Burkholderia cenocepacia* HI2424, *Burkholderia ambifaria* AMMD, *Burkholderia glumae* BGR1, and *Burkholderia gladioli* BSR3). A few enzymes catalyzing the last reactions of valine, leucine, isoleucine, and lysine degradation pathways are retained in the free-living strain, probably because they are also involved in carboxylic acid degradation and assimilation processes.

In addition, the degradation pathways for cysteine, proline, arginine, histidine, phenylalanine, and tyrosine are missing in both *P. necessarius* strains. Only in a few cases those degradation pathways are lost in other members of the *Burkholderiaceae* family (*Burkholderia rhizoxinica* HKI 454 cannot degrade arginine, histidine, phenylalanine, and tyrosine, and representatives of the genera *Cupriavidus* and *Ralstonia* cannot degrade arginine).

Alanine can be catabolized to pyruvate in the free-living strain through alanine-glyoxylate transaminase, alanine transaminase, and alanine dehydrogenase. The symbiotic bacterium lacks all three enzymes.

Aspartate can be oxidized to oxalacetate in both the free-living strain and the symbiont thanks to aspartate oxidase (NadB), an enzyme that is involved mainly in NAD^+ biosynthesis and that is encoded by all of the *Burkholderiaceae* genomes analyzed. Moreover, only in the free-living strain, the presence of the aspartate ammonia-lyase enables the degradation of aspartate to fumarate. The symbiont has lost asparaginase, thus it cannot degrade asparagine.

Because both bacteria lack glutaminase, glutamine can be converted to glutamate only through the reaction of glutamate synthase. Furthermore, because they are both missing glutamate dehydrogenase, glutamate can be converted to 2-oxoglutarate only through reactions involving glutamate as an amino-group donor (e.g., aromatic amino acid aminotransferase, acetyl-ornithine aminotransferase, phosphoserine aminotransferase, branched-chain amino acid aminotransferase).

Both bacteria lack a complete glycine cleavage system, but glycine can be converted to serine thanks to a glycine hydroxymethyltransferase. Threonine dehydratase, which is present in both strains, can degrade serine and threonine to pyruvate and 2-oxobutyrate, respectively. The following oxidation of 2-oxobutyrate to propionyl-CoA may be performed in the free-living strain but probably not in the symbiont because the latter lacks a 2-oxoacid dehydrogenase activity. Moreover, as explained before, the ability of catabolize propionyl-CoA is uncertain in the symbiont. Because of the lack of a complete glyoxylate cycle in the symbiont, the carbon skeletons of glycine and serine cannot be used for gluconeogenesis and, therefore, as sole carbon source. Nevertheless these two amino acids might be used as additional energy and carbon sources.

In summary, although only aspartate, glutamate, and glutamine could be considered glucogenic in the symbiont, it is unlikely that they are effectively used by the bacterium as carbon and energy sources because their conversions to oxalacetate and 2-oxoglutarate, respectively, seem to be side reactions in other biosynthetic pathways. On this point, the situation seems very different from that of "*Candidatus* Glomeribacter gigasporarum" (*Burkholderiaceae*), the obligate endobacterium of an arbuscolar mycorrhizal fungus, in which serine, glutamate, aspartate, and arginine can be used by the symbiont as carbon and energy sources (12). On the other hand, the amino acids that have the potential to be used as sole carbon sources in the free-living *P. necessarius* are alanine, glycine, serine, threonine, aspartate, asparagine, glutamate, and glutamine.

Cofactors. Both the free-living and symbiotic strain have all of the genes necessary for S-adenosylmethionine, polyamines, haem, NAD^+, and $NADP^+$ biosynthesis. Putrescine is probably synthesized from ornithine after decarboxylation through a putative ornithine/lysine/arginine decarboxylase. Differently from the symbiont, the free-living strain can additionally synthesize thiamine pyrophostate, FAD, CoA, molibdopterin, and folate. The symbiont lacks only the last two enzymes for FAD biosynthesis (riboflavine kinase and FAD synthetase), it needs pantotenate or pantoic acid to make CoA [this feature is shared with "*Ca.* G. gigasporarum" (12)], and it retains only the enzymes for the folate-mediated one-carbon metabolism.

Apparently, both bacteria are unable to synthesize vitamin B12 and siroheme even if they still possess a few genes belonging to this biosynthetic pathway. The free-living strain's genome lacks two key genes (*epd* and *pdxB*) and the symbiont four (*epd*, *pdxB*, *pdxA*, and *pdxJ*) involved in the biosynthesis of B6 vitamers. The biotin biosynthetic pathway is not complete in the free-living strain and is completely lost in the symbiont.

Our genomic analysis seems to be partially in contrast with the data of Hahn et al. (5), which show that the free-living strain QLW-P1DMWA-1 can grow on acetate as sole carbon and energy source in an inorganic medium exclusively supplemented with vitamin B12.

Additional notes. As expected, there are no genes involved in flagellar assembly or motility in either genome. On the other hand, the free-living *P. necessarius* has genes usually related to aggregation, challenging the notion that this organism is strictly planktonic. These include a putative cellulose synthase—whose role is not entirely understood in bacteria but whose presence correlates with the formation of biofilms (13); and diguanylate cyclase, producing the messenger molecule c-di-GMP, which is often involved in triggering and maintaining aggregation (14). These genes were lost in the symbiont.

Another exclusive, and intriguing, feature of the free-living strain is the presence of cyanophycin synthetase. Cyanophycin is a polymer of aspartate and arginine used for nitrogen storage, typical of cyanobacteria (15).

SI Materials and Methods

Purification of the Endosymbiont DNA. The ciliate host *E. aediculatus* strain STIR1 was collected in the freshwater Stirone river (Italy) (16). It was cultured in artificial brackish water (5‰ salinity), in an incubator with constant temperature (19–20 °C) and 12:12 h irradiance of 200 μmol photons m^{-2} s^{-1}, and fed with the green alga *Dunaliella salina*.

The ciliates were starved for at least 1 wk before performing DNA extraction. The day before extraction, approximately 5 L of culture were washed and treated to eliminate possible contaminants. Ciliates were placed in a cylindric filter with pores of 10 μm and rinsed with ∼2.5 L of flowing sterile culture medium, thus concentrating the *Euplotes* cells and diluting algal and bacterial contaminants. The remaining volume (approximately 200 mL) was added with chloramphenicol to a final concentration of 0.2 mg/mL and maintained at 15 °C overnight. Pilot experiments, followed by in situ hybridizations (performed as in ref. 17), showed that such a treatment was efficient in removing bacteria in the medium without affecting *Polynucleobacter* symbionts. After this step, ciliate cells were washed again through passage in the cylindric filter and then through three centrifuges (250 × *g*, 10 min) and the pelleted cells transferred to sterile culture medium. After the third wash, the pellet was suspended in a volume of 2 mL and transferred to a glass grinder for mechanical lysis on ice. Successful rupture of the cellular membranes and release of the symbionts were checked by microscope.

The bacteria were separated from most of the eukaryote's cellular fragments through centrifugations at increasing acceleration. The homogenate was loaded at the top of a PBS solution and centrifuged at 50, 100, 200, and 300 × *g* (10 min), discarding the pellet after each step. A centrifugation at 11,300 × *g* (10 min) was performed to collect the purified bacteria, followed by a lysis step in 1% SDS. A final centrifugation, again at 11,300 × *g*, separated the *Polynucleobacter* DNA, in the supernatant, from DNA-containing residue of eukaryotic organelles in the pellet. Total genomic DNA was extracted from this last supernatant with a modified cetyltrimethylammonium bromide (CTAB) method (see below). To check and optimize this method, DNA extraction from different pellets and supernatants were evaluated through electrophoresis on a 1% agarose gel stained with ethidium bromide. The host DNA appears as a smear of high-to-low size fragments. In the final supernatant, the smear is not visible on the gel. DNA extracted from a total volume of ∼30 L of ciliate cultures was used for genome sequencing.

CTAB Method Protocol.

i) Add lysozyme (100 mg/mL) and mix
ii) Incubate for 5 min at RT

iii) Add 10% SDS and mix
iv) Add Proteinase K (10 mg/mL) and mix
v) Incubate for 1 h at 37 °C
vi) Add 5 M NaCl and mix
vii) Add CTAB/NaCl (4.1 g NaCl and 10 g CTAB dissolved in 100 mL water heated to 65 °C) and mix
viii) Incubate 65 °C for 10 min
ix) Add chloroform:isoamyl alcohol (24:1) and mix
x) Spin at max speed for 10 min at room temperature
xi) Transfer aqueous phase to clean microcentrifuge tubes
xii) Add phenol:chloroform:isoamyl alcohol (25:24:1) and mix
xiii) Spin at maximum speed for 10 min at room temperature
xiv) Transfer aqueous phase and add 0.6 vol isopropanol (–20 °C)
xv) Incubate at room temperature for 30 min
xvi) Spin at maximum speed for 15 min
xvii) Wash pellet with 70% ethanol, spin at max speed for 5 min
xviii) Discard the supernatant and let pellet dry for 5–10 min at room temperature
xix) Resuspend in TE plus RNase A (concentration 10 mg/mL) 99:1
xx) Transfer to sterile microcentrifuge tubes
xxi) Incubate at 37 °C for 20 min

1. Pati A, et al. (2010) GenePRIMP: A gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 7(6):455–457.
2. Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14(7):1394–1403.
3. Hahn MW, et al. (2012) The passive yet successful way of planktonic life: Genomic and experimental analysis of the ecology of a free-living *Polynucleobacter* population. *PLoS ONE* 7(3):e32772.
4. Arias-Barrau E, Olivera ER, Sandoval A, Naharro G, Luengo JM (2006) Acetyl-CoA synthetase from *Pseudomonas putida* U is the only acyl-CoA activating enzyme induced by acetate in this bacterium. *FEMS Microbiol Lett* 260(1):36–46.
5. Hahn MW, Lang E, Brandt U, Wu QL, Scheuerl T (2009) Emended description of the genus *Polynucleobacter* and the species *Polynucleobacter necessarius* and proposal of two subspecies, *P. necessarius* subsp. *necessarius* subsp. nov. and *P. necessarius* subsp. *asymbioticus* subsp. nov. *Int J Syst Evol Microbiol* 59(Pt 8): 2002–2009.
6. Lindenkamp N, Schürmann M, Steinbüchel A (2013) A propionate CoA-transferase of *Ralstonia eutropha* H16 with broad substrate specificity catalyzing the CoA thioester formation of various carboxylic acids. *Appl Microbiol Biotechnol* 97(17): 7699–7709.
7. Kretovich WL, Kariakina TI, Weinova MK, Sidelnikova LI, Kazakova OW (1981) The synthesis of aspartic acid in *Rhizobium lupini* bacteroids. *Plant Soil* 61(1-2):145–156.
8. Umbarger HE (1978) Amino acid biosynthesis and its regulation. *Annu Rev Biochem* 47:533–606.
9. Sheppard K, et al. (2008) From one amino acid to another: tRNA-dependent amino acid biosynthesis. *Nucleic Acids Res* 36(6):1813–1825.
10. Min B, Pelaschier JT, Graham DE, Tumbula-Hansen D, Söll D (2002) Transfer RNA-dependent amino acid biosynthesis: An essential route to asparagine formation. *Proc Natl Acad Sci USA* 99(5):2678–2683.
11. Alaminos M, Ramos JL (2001) The methionine biosynthetic pathway from homoserine in *Pseudomonas putida* involves the *metW, metX, metZ, metH* and *metE* gene products. *Arch Microbiol* 176(1-2):151–154.
12. Ghignone S, et al. (2012) The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *ISME J* 6(1):136–145.
13. Römling U (2002) Molecular biology of cellulose production in bacteria. *Res Microbiol* 153(4):205–212.
14. Tamayo R, Pratt JT, Camilli A (2007) Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu Rev Microbiol* 61:131–148.
15. Oppermann-Sanio FB, Steinbüchel A (2002) Occurrence, functions and biosynthesis of polyamides in microorganisms and biotechnological production. *Naturwissenschaften* 89(1):11–22.
16. Petroni G, Dini F, Verni F, Rosati G (2002) A molecular approach to the tangled intrageneric relationships underlying phylogeny in *Euplotes* (Ciliophora, Spirotrichea). *Mol Phylogenet Evol* 22(1):118–130.
17. Vannini C, et al. (2007) Endosymbiosis in statu nascendi: Close phylogenetic relationship between obligately endosymbiotic and obligately free-living *Polynucleobacter* strains (*Betaproteobacteria*). *Environ Microbiol* 9(2):347–359.

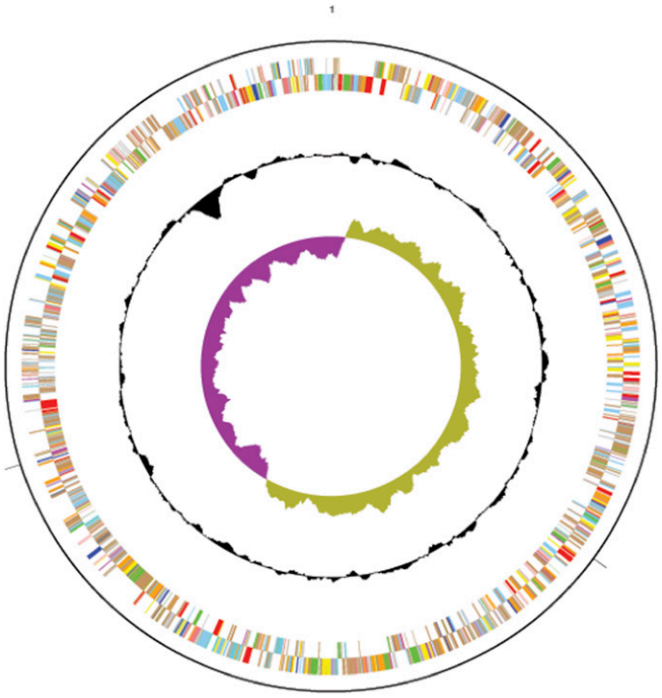

Fig. S1. Circular map of the chromosome. From outside to the center: genes on forward strand, genes on reverse strand, GC content, GC skew. Gene colors reflect COG categories.

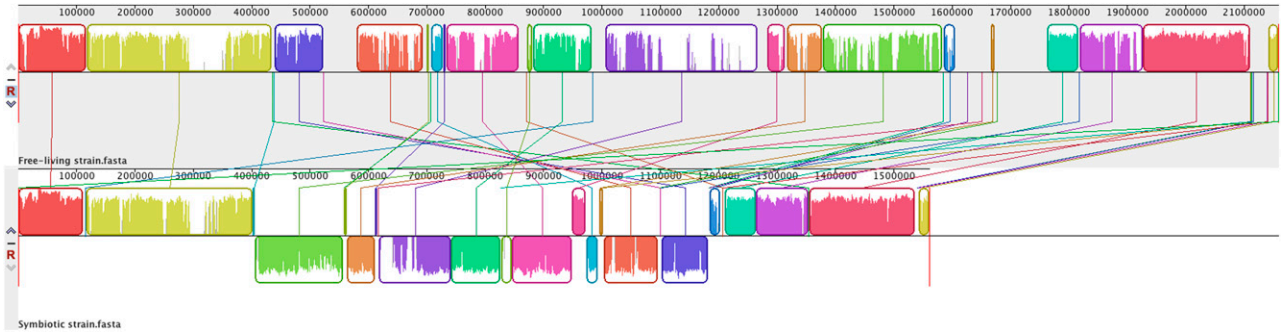

Fig. S2. Free-living/symbiotic *Polynucleobacter* whole-genome alignment performed with Mauve software, showing a major inversion event and several minor ones.

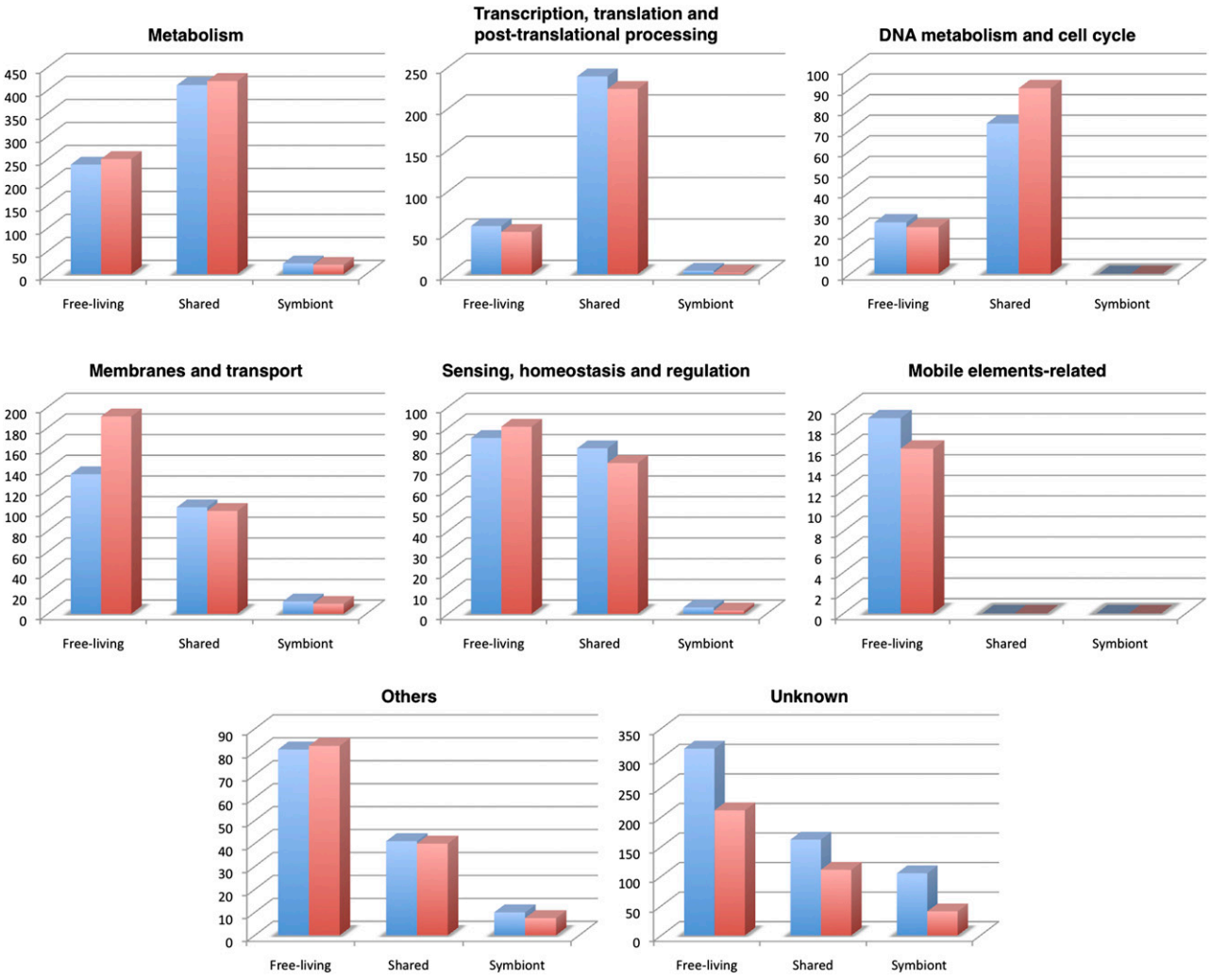

Fig. S3. Bar graphs showing the putative functional genes exclusive to the free-living strain, the symbiont, and shared (with a close ortholog in each genome). Blue bars indicate gene number, red bars indicate the total length of the genes (in kbp, averaged for the shared genes).

Table S1. Genomic features

Parameter	Free-living strain	Symbiotic strain	Ratio (%)
No. of chromosomes	1	1	
Genome size, bp	2,159,490	1,560,469	72.3
GC content, %	44.8	45.6	
Coding size, bp	1,977,881	1,142,322	57.8
Protein-coding genes, *n*	2,075	1,279	61.6
Average protein genes length, bp	949	886	93.4
RNA-coding genes, *n*	44	44	
Pseudogenes, *n**	11	231	
Pseudogenes, *n*†	13	460	
Rest of noncoding DNA size, bp	176,194	129,615	73.6

*Conservative estimate (*SI Results and Discussion*, *Genome Composition Analysis*).
†Permissive estimate (*SI Results and Discussion*, *Genome Composition Analysis*).

Table S2. Non-*Polynucleobacter* reference genomes

Organism	Genome accession no.	Chromosome and plasmid no.	Genome size, Mbp	GC content, %	Protein coding genes, *n*
Burkholderia ambifaria AMMD	NC_010551-3, NC_010557	4	7.53	66.8	6,610
Burkholderia cenocepacia HI2424	NC_08542-5	4	7.70	66.8	6,919
Burkholderia gladioli BSR3	NC_015376-8, NC_015381-3	6	9.05	67.4	7,411
Burkholderia glumae BGR1	NC_012718, NC_012720-1, NC_012723-5	6	7.28	67.9	5,773
Burkholderia mallei ATCC 23344	NC_006348-9	2	5.84	68.5	5,023
Burkholderia multivorans ATCC17616	NC_010801-2, NC_010804-5	4	7.01	66.7	6,258
Burkholderia phenoliruptrix BR3459a	NC_018672, NC_018695-6	3	7.65	63.1	6,496
Burkholderia phymatum STM815	NC_010622-3, NC_010625, NC_010627	4	8.68	62.3	7,496
Burkholderia phytofirmans PsJN	NC_010676, NC_010679, NC_010681	3	8.21	62.3	7,241
Burkholderia pseudomallei K96243	NC_006350-1	2	7.25	68.1	5,728
Burkholderia rhizoxinica HKI 454	NC_014718, NC_014722-3	3	3.75	60.7	3,870
Burkholderia thailandensis E264	NC_007650-1	2	6.72	67.6	5,632
Burkholderia vietnamiensis G4	NC_009254-6, NC_009226-30	8	8.39	65.7	7,617
Burkholderia xenovorans LB400	NC_007951-3	3	9.73	62.6	8,702
Ralstonia eutropha JMP134	NC0073367, NC007347-8	4	7.26	64.5	6,446
Ralstonia solanacearum GMI1000	NC_003295-6	2	5.81	67.0	5,113
Ralstonia pickettii 12D	NC_012849, NC_012851, NC_012855-7	5	5.69	63.3	5,361
Cupriavidus metallidurans CH34	NC007971-4	4	6.91	63.5	6,477
Cupriavidus necator H16	NC_015723-4, NC_015726-7	3	7.42	66.4	6,626
Cupriavidus taiwanensis LMG19424	NC_010528-30	3	6.48	67.0	5,896

All data come from the National Center for Biotechnology Information Microbial Genome Web site (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html).