AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Reduced and Nonreduced Genomes in *Paraburkholderia* Symbionts of Social Amoebas

Suegene Noh,[a] Benjamin J. Capodanno,[a,b] Songtao Xu,[a] Marisa C. Hamilton,[a,c] Joan E. Strassmann,[d] David C. Queller[d]

[a]Department of Biology, Colby College, Waterville, Maine, USA
[b]Brotman Baty Institute for Precision Medicine, Seattle, Washington, USA
[c]University Program in Genetics and Genomics, Duke University, Durham, North Carolina, USA
[d]Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA

ABSTRACT The social amoeba *Dictyostelium discoideum* is a predatory soil protist frequently used for studying host-pathogen interactions. A subset of *D. discoideum* strains isolated from soil persistently carry symbiotic *Paraburkholderia*, recently formally described as *P. agricolaris*, *P. bonniea*, and *P. hayleyella*. The three facultative symbiont species of *D. discoideum* present a unique opportunity to study a naturally occurring symbiosis in a laboratory model protist. There is a large difference in genome size between *P. agricolaris* (8.7 million base pairs [Mbp]) versus *P. hayleyella* and *P. bonniea* (4.1 Mbp). We took a comparative genomics approach and compared the three genomes of *D. discoideum* symbionts to 12 additional *Paraburkholderia* genomes to test for genome evolution patterns that frequently accompany host adaptation. Overall, *P. agricolaris* is difficult to distinguish from other *Paraburkholderia* based on its genome size and content, but the reduced genomes of *P. bonniea* and *P. hayleyella* display characteristics indicative of genome streamlining rather than deterioration during adaptation to their protist hosts. In addition, *D. discoideum*-symbiont genomes have increased secretion system and motility genes that may mediate interactions with their host. Specifically, adjacent BurBor-like type 3 and T6SS-5-like type 6 secretion system operons shared among all three *D. discoideum*-symbiont genomes may be important for host interaction. Horizontal transfer of these secretion system operons within the amoeba host environment may have contributed to the unique ability of these symbionts to establish and maintain a symbiotic relationship with *D. discoideum*.

IMPORTANCE Protists are a diverse group of typically single cell eukaryotes. Bacteria and archaea that form long-term symbiotic relationships with protists may evolve in additional ways than those in relationships with multicellular eukaryotes such as plants, animals, or fungi. Social amoebas are a predatory soil protist sometimes found with symbiotic bacteria living inside their cells. They present a unique opportunity to explore a naturally occurring symbiosis in a protist frequently used for studying host-pathogen interactions. We show that one amoeba-symbiont species is similar to other related bacteria in genome size and content, while the two reduced-genome-symbiont species show characteristics of genome streamlining rather than deterioration during adaptation to their host. We also identify sets of genes present in all three amoeba-symbiont genomes that are potentially used for host-symbiont interactions. Because the amoeba symbionts are distantly related, the amoeba host environment may be where these genes were shared among symbionts.

KEYWORDS symbiosis, protist, *Burkholderia*, *Dictyostelium*, genome reduction

The social amoeba *Dictyostelium discoideum* (*Eumycetozoa*; *Dictyosteliales*) is a predatory soil protist frequently used for studying host-pathogen interactions (1–3). It is also an emerging model for host-microbe symbiosis in the broad sense, which we

Editor E. Maggie Sogin, Univ. California Merced

Address correspondence to Suegene Noh, suegene.noh@colby.edu.

The authors declare no conflict of interest.

Received 18 June 2022
Accepted 25 August 2022
Published 13 September 2022

define here as an intimate association between a eukaryotic host and a prokaryote symbiont that can result in positive, neutral, or negative fitness consequences in both parties involved (4–6). A subset of *D. discoideum* strains isolated from soil persistently carry intracellular Gram-negative *Paraburkholderia* (*Betaproteobacteria*; *Burkholderiales*) (7–9). Multilocus sequence typing analyses and whole-genome phylogenies showed that these symbionts comprise two independent clades (9, 10). Subsequently, *P. agricolaris* and the two sister species *P. bonniea* and *P. hayleyella* were formally described as new species sufficiently different from any other previously described *Paraburkholderia* using genetic and phenotypic evidence (10).

The three *Paraburkholderia* symbionts of *D. discoideum* present a unique opportunity to study a naturally occurring symbiosis in a laboratory model protist. They additionally present opportunities for insight into the diversity of protist-prokaryote symbioses, which are understudied compared to the symbiotic relationships of multicellular eukaryotes and their microbial symbionts (11). The association between *D. discoideum* and its *Paraburkholderia* symbionts appears to be facultative, and these symbionts are able to simultaneously maintain a free-living and host-associated lifestyle (8, 9). With *Paraburkholderia* the fitness outcomes to host and symbiont appear to be context dependent (12), as with most facultative host-microbe symbioses (6). *D. discoideum* amoeba hosts generally suffer negative fitness consequences of association. When amoebas are infected with their *Paraburkholderia* symbionts in the lab, the hosts tend to eat less food bacteria during vegetative growth, migrate shorter distances as slugs during their multicellular social cycle, form shorter and smaller volume fruiting bodies, produce fewer spores, and carry other bacteria alive (secondary carriage) into their next vegetative growth cycle (7, 8, 13, 14). However, potentially important context-dependent fitness benefits to the host may be (i) increased availability of food bacteria in relatively inhospitable environments as a result of secondary carriage, and (ii) improved competitive ability against other *D. discoideum* strains by potentially passing on symbiont infections or releasing *Paraburkholderia* secretions in a defensive manner (7, 15). We know less about fitness outcomes of association for *Paraburkholderia* symbionts, but *P. hayleyella* (though not *P. agricolaris*) reaches higher population densities in the presence of *D. discoideum* compared to on its own in soil medium (16). While not a direct demonstration of any fitness benefits, *P. agricolaris* and *P. hayleyella* show positive chemotaxis toward *D. discoideum* supernatant (17). Later, *D. discoideum* was discovered to also associate with unculturable Chlamydiae and *Amoebophilus* symbionts but with no obvious fitness effects of infection or clear patterns of coinfection with *Paraburkholderia* (18).

We present a comparative genomics analysis of the three types of strains of *Paraburkholderia* isolated from *D. discoideum* (*P. agricolaris* BaQS159, *P. hayleyella* BhQS11, and *P. bonniea* BbQS859). We isolated all three strains from *D. discoideum* hosts collected at Mountain Lake Biological Station in Virginia, USA. Notably, there is a large difference in genome size between *P. agricolaris* (8.7 million base pairs [Mbp]) versus *P. hayleyella* and *P. bonniea* (4.1 Mbp) and in GC content (62% versus 59%) (9, 10). Genome reduction is a pattern associated with long-term host association in many symbiotic bacteria (19–25), including pathogenic *Burkholderia* (26) and amoeba-associated *Legionella* (27, 28). Therefore, we investigated any significant differences in genome characteristics in the genomes of *D. discoideum* symbionts and particularly in the reduced genomes of *P. hayleyella* and *P. bonniea*. Based on what we know from the best-studied endosymbiont genomes of multicellular animals, we looked for patterns that frequently accompany host adaptation, such as fewer genes in functional categories related to metabolism, DNA repair, and gene regulation (29, 30).

Because the ability to infect *D. discoideum* appears to be a shared derived trait among *Paraburkholderia* symbionts of *D. discoideum*, we focused several analyses on shared orthologous genes across the three genomes in comparison with other *Paraburkholderia*. Given the estimated large phylogenetic distance between the two *D. discoideum*-symbiont clades (9, 10), we paid particular attention to shared horizontally transferred genetic elements. Horizontal gene transfer generally contributes to an increase in prokaryote genomic repertoires but is subject to evolutionary processes,

TABLE 1 Genome statistics of *Paraburkholderia* symbionts of *D. discoideum*

Genome statistics	*P. agricolaris* BaQS159	*P. bonniea* BbQS859	*P. hayleyella* BhQS11
Scaffold count	2	2	2
Genome size (chr1; chr2)	8,721,420 (4,816,966; 3,904,454)	4,098,182 (3,175,376; 922,806)	4,125,700 (3,295,139; 830,561)
GC content (%)	61.6	58.7	59.2
Genes (total)	7,811	3,600	3,686
CDS (total)	7,721	3,531	3,610
Pseudogene count	579	265	315
rRNA count	18	12	12
tRNA count	71	56	63
Locality (host)	*D. discoideum* in Virginia, USA	*D. discoideum* in Virginia, USA	*D. discoideum* in Virginia, USA

including selection and drift as with the rest of the genome (31–34). In the context of symbiosis, key horizontally transferred genes can enable new symbiotic relationships (e.g., symbiosis islands) (35). If host adaptation-induced genome reduction is ongoing, we expected symbiont genomes to show signs of instability in the form of excess nonfunctional horizontally transferred genetic elements (e.g., insertion sequence [IS] elements or pseudogenes) (36). IS elements in particular connect the themes of genome reduction and horizontally transferred genetic elements. They often proliferate during earlier stages of host adaptation and enable genome rearrangements and deterioration (37, 38), eventually leading to the highly reduced genomes seen in obligate symbionts.

RESULTS

***D. discoideum* symbionts represent two distinct categories, reduced versus nonreduced size genomes.** Sequencing using PacBio technology indicated that the genomes of all three *D. discoideum*-symbiont species are each comprised of two chromosomes, albeit resulting in different total genome sizes. The *P. agricolaris* genome was more than twice the size of both *P. bonniea* and *P. hayleyella* (8.7 versus 4.1 million basepairs). The overall gene content (CDS) comparison was also proportionate, with approximately 7,700 genes predicted for *P. agricolaris* as opposed to approximately 3,600 genes for the reduced genomes (Table 1). The genome size and gene count of *P. agricolaris* are on par with other the *Paraburkholderia* genomes we examined (Table 2).

Whole-genome alignments of all 10 finished genomes found 153 locally colinear blocks, ranging in sizes as small as 262 bp and as large as 208,252 bp in the *P. agricolaris* genome. *P. bonniea* and *P. hayleyella* share with each other considerable synteny but also possess a large inverted region relative to each other (Fig. 1). Both of these reduced genomes show extensive genome rearrangement compared to the genome of *P. agricolaris*, or any of the other *Paraburkholderia* genomes (Fig. 1 and Fig. S1 in the supplemental material). The genome of *P. agricolaris* shares a large degree of synteny with other *Paraburkholderia* genomes in chromosome 1 as indicated by the overall lack of gaps toward the center of each hive plot (Fig. S1).

We found few IS elements in the *D. discoideum*-symbiont genomes. *P. agricolaris* has the IS elements IS1090 (6 copies) and ISBmu21 (1 copy), *P. bonniea* has ISBp1 (2 copies)

TABLE 2 Genome statistics of other representative *Paraburkholderia* strains for comparison

Genome statistics	*P. fungorum* ATCC BAA-463	*P. sprentiae* WSM5005	*P. terrae DSM* 17804	*P. xenovorans* LB400
Scaffold count	4	5	4	3
Genome size (total)	9,058,983	7,829,542	10,062,489	9,702,951
GC content	61.8	63.2	61.9	62.6
Genes (total)	8,260	7,185	9,045	8,760
CDS (total)	8,174	7,087	8,957	8,675
Pseudogene count	746	857	803	845
rRNA count	18	21	18	18
tRNA count	67	76	69	66
Locality (host)	White-rot fungus *Phanerochaete chrysosporium* in Sweden	Domesticated legume *Lebeckia ambigua* in Australia	Broad-leaved forest soil in South Korea	Polychlorinated biphenyl (PCB)-contaminated soil in USA

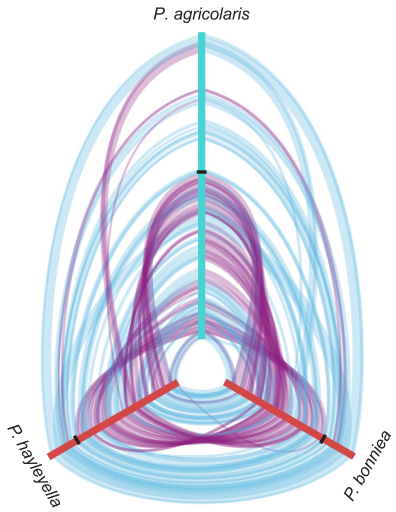

FIG 1 Hive plot of whole-genome comparisons of *D. discoideum*-symbiont genomes. Locally colinear blocks between pairs of genomes are shown as bands that connect the axes (genomes). Only blocks above the median size are shown. Alignment of locally colinear blocks are distinguished between forward (blue) and reverse (purple) orientation. Axes are oriented center out, and boundaries between chromosomes are shown as ticks.

and ISBuph1 (3 copies), and *P. hayleyella* has a single ISPa37 in their genomes (Table S2). Among the other *Paraburkholderia* genomes we examined, the highest number of IS elements was found in *P. xenovorans* LB400 (62 total), while others possessed intermediate numbers ranging from 5 in *P. sprentiae* to 21 in *P. fungorum*. For reference, genomes of *B. mallei* possess between 166 and 218 IS elements, many of which were flanking regions that were randomly lost among the examined strains in what appears to be ongoing genome reduction (37). There was also no evidence of excess pseudogenes in the reduced genomes relative to other *Paraburkholderia* genomes (Table 1). Double-strand break repair pathways (KEGG map03440) were complete in all three *D. discoideum*-symbiont genomes. As genomes with ongoing genome reduction often have numerous IS elements and pseudogenes, and incomplete double-strand break repair pathways, the combined evidence suggests that all three *D. discoideum*-symbiont genomes are relatively stable and the two reduced genomes are currently not in flux.

The reduced genomes of *D. discoideum* symbionts show evidence of functional adaptation to the host environment. For each *D. discoideum*-symbiont genome, 65 to 68% of genes were annotated with Clusters of Orthologous Groups (COG) (Fig. S2), and 53 to 61% with Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO). The agglomerative clustering and nonmetric multidimensional scaling (NMDS) analyses of COG category representation across genomes resulted in *P. bonniea* and *P. hayleyella* clustering with each other and apart from other *Paraburkholderia*, including *P. agricolaris* (Fig. 2). Further investigation of specific functional differences between the two groups (reduced genomes versus nonreduced) indicated nine COG categories that were significantly different (exactTest, false discovery rate $\ll$ 0.01). Of these, four were consistently different in both normalized and raw counts (Fig. S3): fewer genes were detected in the reduced genomes of *P. bonniea* and *P. hayleyella* for Transcription (category K), Carbohydrate transport and metabolism (G), and Inorganic ion transport and metabolism (P), and more genes were found in the reduced genomes for Cell motility (N).

We looked within each COG category that was significantly different between reduced and nonreduced genomes in more detail. First, we found several flagella biosynthesis, basal body, and hook protein COGs that were more abundant in the reduced genomes than expected (Fig. 3a). Flagella are often associated with bacterial virulence, not only through providing motility but also adhesion, invasion, and the secretion and regulation of virulence factors (39, 40). Among *Burkholderia*, *B. pseudomallei* flagella have been shown to be necessary for postinvasion virulence in mice (41). *B.*

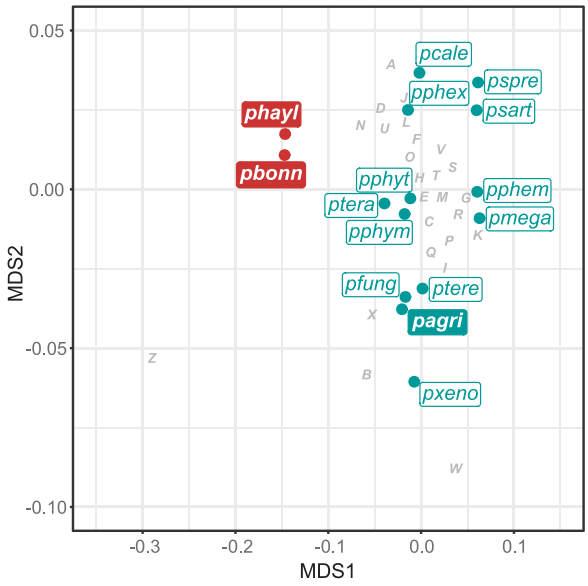

FIG 2 Comparison of reduced (red) and nonreduced (turquoise) genomes in terms of their functional compositions in nonmetric multidimensional space. The contributions of COG categories are projected with minor adjustments to avoid overlap with other features (pagri, *P. agricolaris*; pbonn, *P. bonniea*; phayl, *P. hayleyella*; pcale, *P. caledonica*; pfung, *P. fungorum*, pmega, *P. megapolitana*, pphem, *P. phenazinium*; pphex, *P. phenoliruptrix*; pphym, *P. phymatum*; pphyt, *P. phytofirmans*; psart, *P. sartisoli*; pspre, *P. sprentiae*; ptera, *P. terricola*; ptere, *P. terrae*; pxeno, *P. xenovorans*) (COG categories: J, Translation, ribosomal structure and biogenesis; A, RNA processing and modification; K, Transcription; L, Replication, recombination and repair; B, Chromatin structure and dynamics; D, Cell cycle control, cell division, chromosome partitioning; Y, Nuclear structure; V, Defense mechanisms; T, Signal transduction mechanisms; M, Cell wall/membrane/envelope biogenesis; N, Cell motility; Z, Cytoskeleton; W, Extracellular structures; U, Intracellular trafficking, secretion, and vesicular transport O, Posttranslational modification, protein turnover, chaperones; X, Mobilome: prophages, transposons; C, Energy production and conversion; G, Carbohydrate transport and metabolism; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; P, Inorganic ion transport and metabolism; Q, Secondary metabolites biosynthesis, transport and catabolism; R, General function prediction only; S, Function unknown.

pseudomallei and *B. thailandensis* each have two flagellar clusters, and in *B. thailandensis* the second cryptic cluster is involved in postinvasion intracellular motility (42). We found two flagellar clusters in *P. bonniea* compared to one in the other *D. discoideum*-symbiont genomes (see also "Shared secretion systems may mediate *D. discoideum* - *Paraburkholderia*-symbiont interactions").

The other significant COG categories were less abundant in the reduced genomes than expected (Fig. 3b to d). Many families of transcriptional regulator COGs were less abundant in the reduced genomes, as is often seen with reduced symbiotic bacterial genomes (23, 43). Similarly, several ATP binding cassette (ABC)-type sugar and metal ion transporter COGs were less abundant in the reduced genomes. ABC transporters are often reduced in number in bacteria with intracellular niches compared to extracellular or environmental ones, as intracellular environments are relatively more stable compared to extracellular environments (44, 45).

Analysis with KEGG mapper reconstruction confirmed several missing sugar transport systems in the reduced genomes compared to *P. agricolaris*, including sorbitol/mannitol, L-arabinose, galactofuranose, D-xylose, fructose, and rhamnose. Genes encoding iron (III) transporters were also absent in the two reduced symbiont genomes compared to nonreduced *P. agricolaris*. A similar analysis of ABC transporters also revealed the presence of heme exporter proteins in both reduced genomes but not in *P. agricolaris* and a capsular polysaccharide transport system in *P. bonniea* only. We also examined two-component system (TCS) transporters with KEGG mapper because pathogenic *Burkholderia* have multiple two-component systems related to virulence in plant and animal infection models (46). Compared to *P. agricolaris*, the reduced genomes lacked genes encoding nitrate

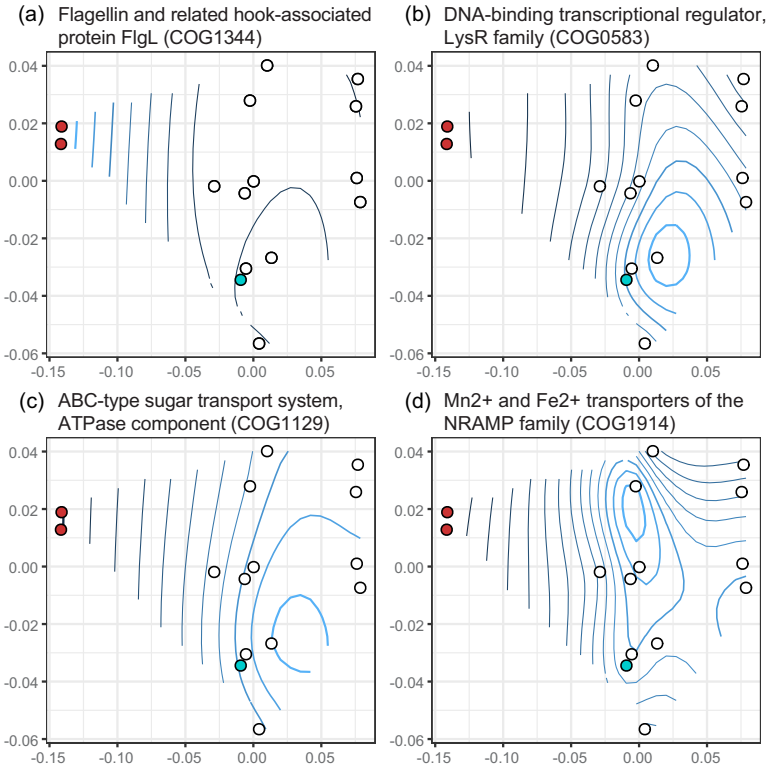

FIG 3 Representative individual COGs belonging to categories Cell motility (a), Transcription (b), Carbohydrate transport and metabolism (c), and Inorganic ion transport and metabolism (d) that were significantly overrepresented (a) or underrepresented (b–d) in the reduced genomes of *D. discoideum* symbionts. Contours of abundances are superimposed on the nonmetric multidimensional space from Fig. 2. *P. bonniea* and *P. hayleyella* are shown as red points to the left, while *P. agricolaris* is distinguished from the other genomes (white) as a turquoise point. Lighter blue contour lines indicate higher abundance compared to darker blue lines.

reductase proteins and chemotaxis proteins typically involved in biofilm formation through cyclic di-GMP regulation. In free-living *B. pseudomallei*, these two two-component systems are linked, as the presence of nitrate has been shown to reduce intracellular cyclic di-GMP levels and inhibit biofilm formation (47). It appears these two-component systems and the aforementioned transporters have not been maintained under selection during host adaptation and genome reduction in *P. bonniea* and *P. hayleyella*.

The reduced *D. discoideum*-symbiont genomes may experience a combination of stronger and relaxed purifying selection relative to other Paraburkholderia genomes. We identified 1,673 core genes shared by the 15 *Paraburkholderia* species genomes we investigated (Fig. S2). When we examined dN/dS as a signature of molecular evolution, the majority of the *Paraburkholderia* core genes showed nonsignificant variation in selection pressure across the species phylogeny. However, a large proportion of core genes (~40%) showed an alternative pattern of molecular evolution (Table 3). These genes show one of two patterns of molecular evolution: those that appear to experience increased selection pressure and significantly lower dN/dS once symbiotically associated with eukaryotes or specifically with *D. discoideum* ("symbiotic" and "dicty"; both Wilcoxon test $P \ll 0.01$), and those that show evidence of relaxed selection and significantly higher dN/dS in genomes of reduced size ("reduced"; Wilcoxon test $P \ll 0.01$) (Fig. 4). These results indicate that the reduced genomes of *P. bonniea* and *P. hayleyella* possess a combination of genes experiencing stronger selective constraints and genes under weaker selective constraints relative to the genomes of other *Paraburkholderia*. This pattern is in contrast to genomes of obligate symbionts where the majority of genes are experiencing genetic drift and weaker selection constraints across their entire genomes (48, 49).

TABLE 3 Hypotheses tested regarding molecular evolution in the 1,673 core genes shared across 15 *Paraburkholderia* genomes

Hypothesis	No. of core genes	Detailed description
"null"	1001	Selection pressure does not vary across the tree
"symbiotic"	163	Selection pressure is different when species are free-living versus symbiotically associated with a eukaryotic host (2 rate ratios)
"dicty"	137	Selection pressure is different when species are unassociated versus symbiotically associated with *D. discoideum* (2 rate ratios)
"reduced"	372	Selection pressure is different in species with reduced genomes (*P. bonniea and P. hayleyella*) (2 rate ratios)

The three *D. discoideum*-symbiont genomes shared 1,977 genes total, including the 1,673 core genes (Fig. S2). Of the 1977 *D. discoideum*-symbiont-shared genes (inclusive of core genes), 120 were not orthologous to genes found in any of the other *Paraburkholderia* genomes we compared. These genes included type 3 and type 6 secretion system component genes (see "Shared secretion systems may mediate *D. discoideum* - *Paraburkholderia*-symbiont interactions"), *bhuRSTUV* genes, and helix-turn-helix motif-containing GntR and LysR transcriptional regulators. *Burkholderia* heme uptake (*bhu*) genes are thought to be important for intracellular iron acquisition in *B. pseudomallei* (50) and are related to *Bordatella* heme utilization (*bhu*) genes that are virulence factors in mammalian and avian host infection (51, 52). Transcriptional regulators with helix-turn-helix motifs have been frequently associated with virulence in pathogens (53).

The relationship between *D. discoideum* and its symbionts is unlikely to be based on amino acid exchange. The gradual loss of essential amino acid biosynthetic ability is a feature of genome reduction in many microbial symbionts that have nutrient exchange relationships with their hosts (54–56). However, nutrient-dependent relationships are less likely in protist-prokaryote symbioses because protist host diets tend

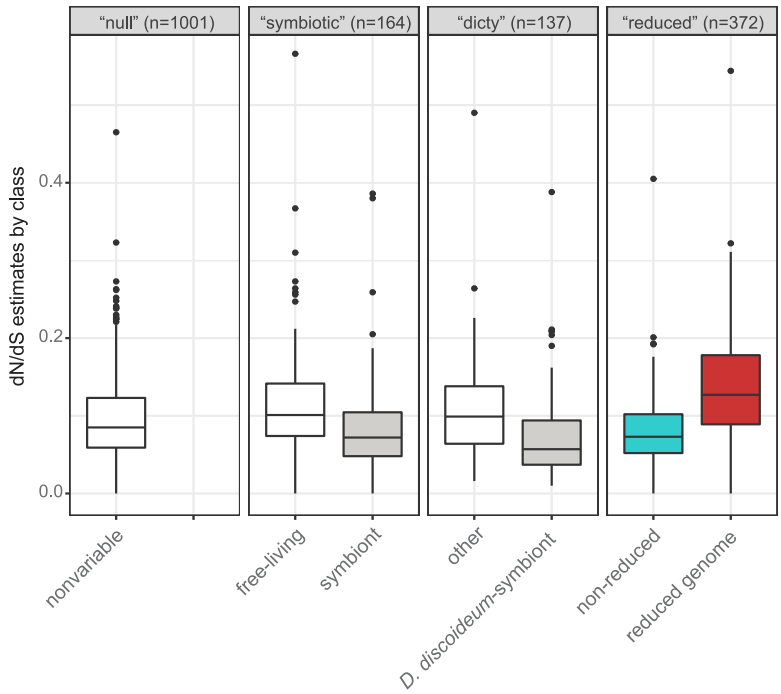

FIG 4 Core genes divided into the hypothesis that best predicts their patterns of molecular evolution. Core genes included genes evolving under stronger selective constraints with significantly lower dN/dS in genomes of symbionts of *D. discoideum* or other eukaryotes ("symbiotic" and "dicty") and genes showing evidence of relaxed selective constraints with significantly higher dN/dS in the reduced genomes of *P. bonniea* and *P. hayleyella* ("reduced"). *P. bonniea* and *P. hayleyella* genes are included in the groups: symbiont, *D. discoideum*-symbiont, and reduced genome within each hypothesis.

to be much more diverse compared to multicellular eukaryotes (11). Accordingly, the three *D. discoideum*-symbiont species are predicted to synthesize all essential amino acids (Table S3), albeit with some variation in degrees of confidence. High-confidence candidates were identified for each of the steps of amino acid biosynthesis in *P. agricolaris*, but some pathways included medium confidence steps in the other two species with reduced genomes. In *P. bonniea*, the L-arginine biosynthesis pathway contained one medium confidence enzyme (Ornithine carbamoyltransferase *argI*) that was a lower coverage match (78%) than the high-confidence threshold (>80%). There is more evidence for a potential breakdown of essential amino acid synthesis in *P. hayleyella*. *P. hayleyella* had four potential gaps in its amino acid biosynthesis pathways. The L-isoleucine, L-leucine, and L-valine pathways shared a single medium-confidence enzyme candidate that is potentially a L-arabonate dehydratase rather than the necessary dihydroxy-acid dehydratase *ilvD* based on ublast bit scores. The L-tryptophan pathway had two medium confidence enzyme candidates for phosphoribosylanthranilate isomerase (*PRAI*), and the better scoring one was a lower coverage match (71%) than the high-confidence threshold. However, given the degree of genome reduction that has already occurred in the reduced-genome *D. discoideum* symbionts, we consider it unlikely that the symbiotic relationship is based on amino acid exchange as essential amino acid synthesis pathways appear largely intact.

***D. discoideum*-symbiont genomes share few recently horizontally transferred genetic elements.** We looked for evidence of shared horizontally transmitted genetic elements. We identified 38, 29, and 27 genomic islands in each *D. discoideum*-symbiont genome (*P. agricolaris*, *P. bonniea*, and *P. hayleyella*), but none of the predicted genomic islands were closely related to a genomic island in another *D. discoideum*-symbiont genome. We found 133, 109, and 120 individually horizontally transferred genes in each *D. discoideum*-symbiont genome. One candidate was shared among all three genomes (type VI secretion system contractile sheath, large subunit) while two additional candidates were shared by *P. bonniea* and P. *hayleyella* (PIN family putative toxin-antitoxin system, toxin component; class I SAM-dependent methyltransferase). The scarcity of easily identified shared horizontally transferred genetic elements suggests it is unlikely that a recent horizontal gene transfer event substantially contributed to the shared ability of these symbionts to persistently infect *D. discoideum*. If such an event had occurred, any such genes seem to have experienced amelioration over evolutionary time and cannot easily be distinguished from the rest of the genome (57).

Shared secretion systems may mediate *D. discoideum* - *Paraburkholderia*-symbiont interactions. Bacterial secretion systems are frequently implicated in host-symbiont interactions (58, 59). All *D. discoideum*-symbiont genomes possessed multiple type III secretions systems (T3SS) and type VI secretion systems (T6SS) in larger numbers than several of the other *Paraburkholderia* genomes examined (Fig. 5; Table S4). Classification of T3SS showed that one specific T3SS operon shared among *D. discoideum* symbionts falls into category 8 T3SS (Fig. 6 and Fig. S4). This category of T3SS also includes BurBor found in the plant pathogen *Robbsia* (previously *Burkholderia*) *andropogonis* (60), as well as *Bordatella* species that include mammalian pathogens (61). In addition, one specific T6SS operon is shared among *D. discoideum*-symbiont genomes and belongs to category i1 (Fig. 7 and Fig. S5). More importantly, this T6SS operon clusters together with the virulence-causing T6SS-5 operon found in *Burkholderia mallei*, *B. pseudomallei*, and *B. thailandensis* (62). *B. mallei* causes glanders disease and is an obligate pathogen that evolved from an ancestor shared with melioidosis-causing soil bacterium *B. pseudomallei* (37, 63, 64). *B. thailandensis* is sister species to the other two and is a facultative pathogen similar to *B. pseudomallei* but with much lower clinical virulence (62).

The T6SS-5-like and BurBor-like T3SS operons shared by the *D. discoideum* symbionts are found directly next to each other on the respective genomes of *P. agricolaris*, *P. bonniea*, and *P. hayleyella*. *B. pseudomallei* and *B. thailandensis* also have a T3SS (T3SS-3 in the literature) adjacent to their T6SS-5, but these T3SS-3 operons appear to be unrelated to the BurBor-like T3SS found in the *D. discoideum*-symbiont genomes. It is worth noting that the two adjacent T3SS-3 and T6SS-5 operons in *B. pseudomallei* and *B.*

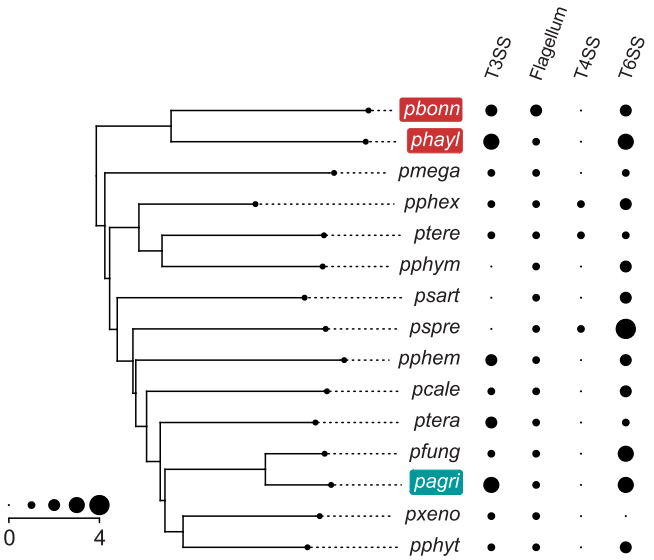

FIG 5 The abundances of secretion systems detected in *D. discoideum*-symbiont genomes and other *Paraburkholderia*. For the type 4 secretion system, only protein secretion (as opposed to conjugation-related) T4SS abundances are shown. The phylogeny is a species tree of the 15 *Paraburkholderia* genomes we examined, reduced from the larger species tree in Brock et al. (10).

thailandensis have been shown to be functionally linked and necessary for virulence, with the T3SS-3 effectors regulating the expression of the adjacent T6SS-5 (42, 65–67).

We attempted to identify effector proteins that might be functionally linked to these *D. discoideum*-symbiont secretion systems (Table S5). We identified homologs of the T6SS effector VgrG-5 that would likely be associated with the shared T6SS-5-like operon (Table S5). Unexpectedly, VgrG-5 in *P. agricolaris* (gene ID PAGRI_01155) is a homolog but not an ortholog to VgrG-5 in the two reduced genomes (PBONN_01842 and PHAYL_01429). This suggests the possibility of two independent evolutionary origins of this T6SS effector and potentially different functional roles. In *Burkholderia thailandensis*, VgrG-5 is necessary for postinfection cell-to-cell spread within mammalian hosts (68). For each genome, we predicted additional secretion system effectors, including a chaperonin ClpB and a sodium/solute symporter for *P. agricolaris* and RHS (rearrangement hot spot) proteins that may mediate contact-dependent growth inhibition during bacterial competition for *P. hayleyella*. Finally, we predicted secreted effectors containing eukaryotic domains specific to our Pfam clans of interest (Ank, TPR, LRR, Pentapeptide, F-box, and RING). Previous investigations of amoeba symbiont genomes have observed enrichment of proteins possessing these domains that hypothetically mediate physiological interactions with a eukaryotic host (69–71). Notably, two proteins each directly adjacent to VgrG-5 in *P. agricolaris* (PAGRI_01156-7) and *P. bonniea* (PBONN_01840-1) each contained pentapeptide repeat domains. InterProScan searches indicated that two proteins in *P. hayleyella* (PHAYL_01430-1) adjacent to VgrG-5 also contain pentapeptide repeat domains. However, no known functions are predicted for these protein pairs.

DISCUSSION

The genomes of *Paraburkholderia* symbionts of *D. discoideum* present a unique opportunity to compare the significantly differently sized genomes of three symbiont species that share the ability to persistently infect *D. discoideum*. We found evidence that relative to the other *Paraburkholderia* genomes we investigated, *D. discoideum*-symbiont genomes have increased secretion system and motility genes that potentially mediate interactions with their host. Specifically, adjacent type 3 and type 6 secretion system operons shared across all three *D. discoideum*-symbiont genomes may have an important role. The BurBor-like

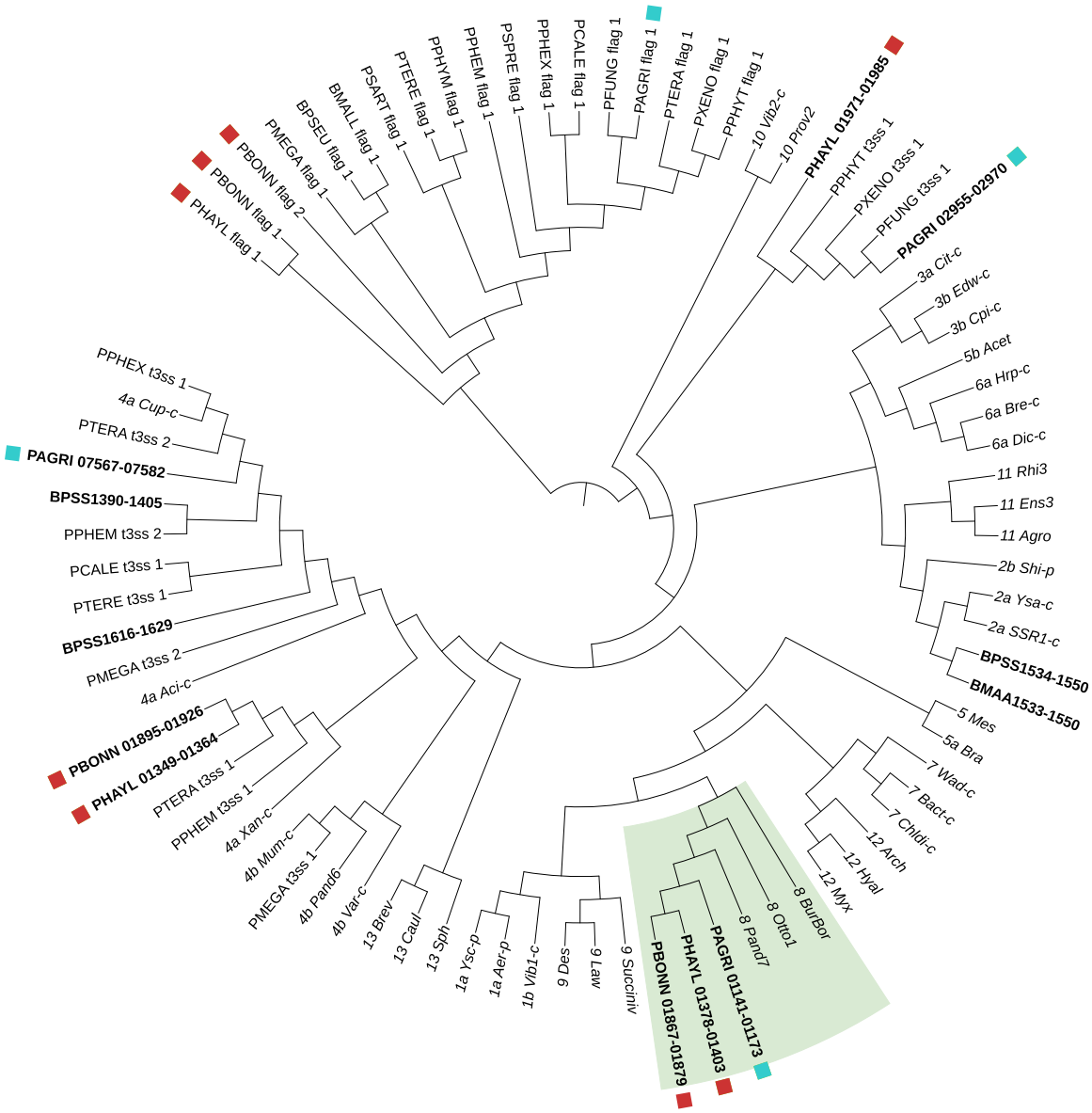

FIG 6 Type 3 secretion systems and flagella categorized using the conserved component genes sctJ (inner membrane ring; IPR003282), sctN (ATPase; IPR005714), and sctV (export apparatus; IPR006302). Branch lengths were ignored to improve readability of the ASTRAL tree topology. T3SS categories precede the name of the operon (e.g., "*8 Pand7*" is operon Pand7 belonging to category 8) downloaded from T3Enc database v1.0 (Hu et al. [135]). Tip labels for T3SS in the three *D. discoideum*-symbiont genomes, *B. mallei* (BMAA), or *B. pseudomallei* (BPSS) are shown in bold font face with gene IDs for ease of cross-reference. *D. discoideum*-symbiont genome T3SS operons are marked with a square symbol, and the clade containing the shared T3SS operon is shaded.

T3SS operon is closely related to one found in the plant pathogen *Robbsia andropogonis*. It includes a needle apparatus uncommon among *Burkholderia* T3SS that is used to inject rhizobitoxine into a wide range of plant hosts (60, 72). The adjacent T6SS operon is closely related to T6SS-5 shared by *B. mallei*, *B. pseudomallei*, and *B. thailandensis*. T6SS-5 is functionally important for the intercellular life cycle of these pathogenic *Burkholderia* (68). We hypothesize that the BurBor-like T3SS operon is used during initial host infection and the T6SS-5-like operon may have a functional role postinfection. We also found orthologs to the T6 effector VgrG-5 specific to T6SS-5, as well as two neighboring potential effectors with eukaryote-like pentapeptide repeat domains in the three *D. discoideum*-symbiont genomes.

Some but not all of the component genes of the shared T6SS-5-like and BurBor-like T3SS operons are among the 120 *D. discoideum*-symbiont-shared genes not found in

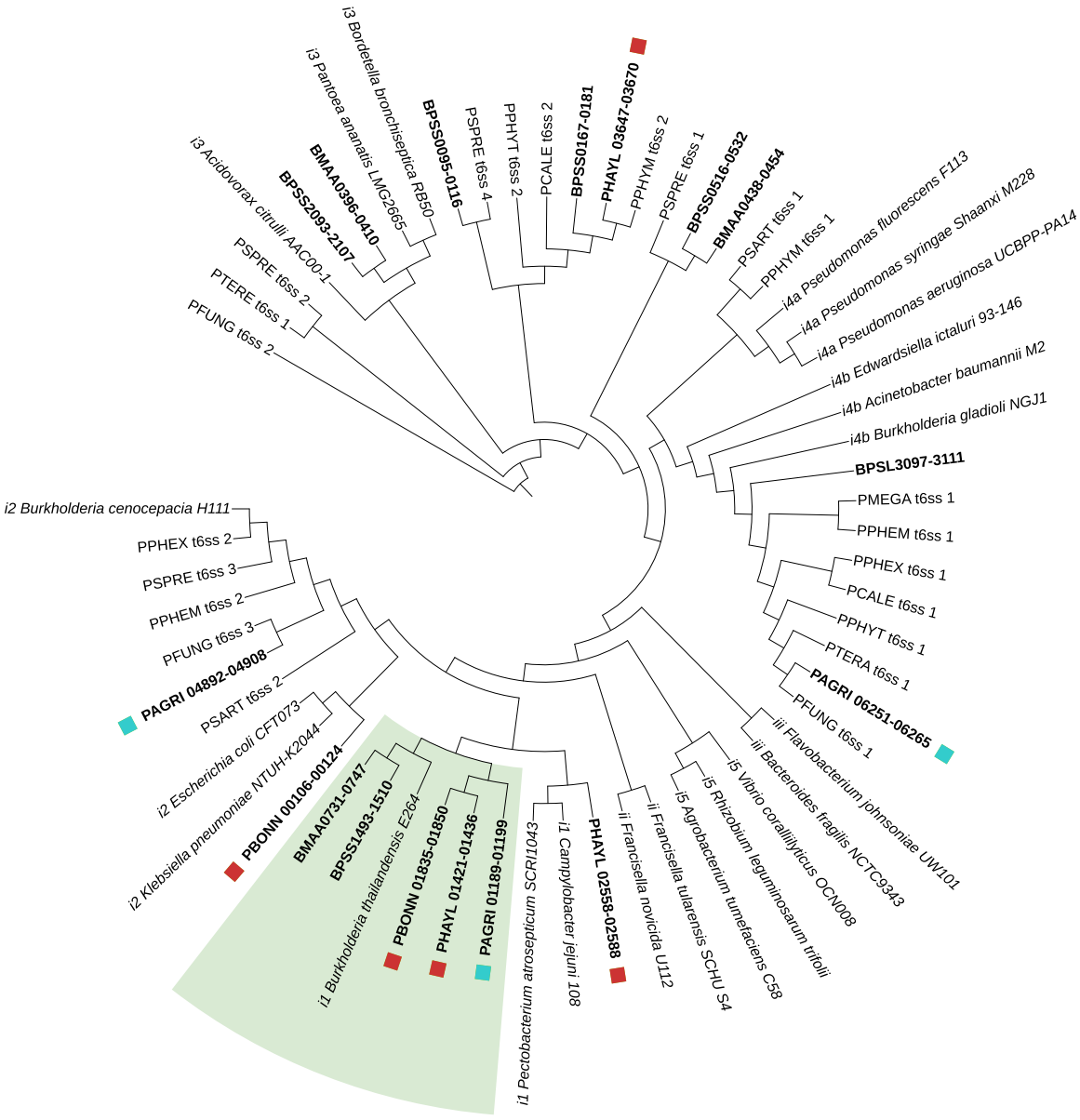

FIG 7 Type 6 secretion systems categorized using the conserved component genes tssB (sheath; COG3516), tssC (sheath; COG3517), and tssF (baseplate; COG3519). Branch lengths were ignored to improve readability of the ASTRAL tree topology. T6SS categories precede the name of the strain to which the operon belongs (e.g., "*ii Francisella novicida U112*" belongs to category ii), downloaded from SecReT6 database v3.0 (Li et al. [139]). Tip labels for T6SS in the three *D. discoideum*-symbiont genomes, *B. mallei* (BMAA), or *B. pseudomallei* (BPSS, BPSL) are shown in bold font face with gene IDs for ease of cross-reference. *D. discoideum*-symbiont genome T3SS operons are marked with a square symbol, and the clade containing the shared T6SS operon is shaded.

any of the other *Paraburkholderia* genomes we compared. It is possible that an unsampled *Paraburkholderia* genome is the source of these secretion systems and other potential symbiosis-mediating factors, including the *bhuRSTUV* iron acquisition genes. However, given the estimated large phylogenetic distance between *P. agricolaris* versus *P. bonniea* and *P. hayleyella*, we hypothesize that one of these clades horizontally acquired the ability to symbiotically associate with *D. discoideum* from the other. Key genes could have been horizontally transferred among symbiont genomes within the *D. discoideum* amoeba host environment, especially since different *D. discoideum*-symbiont species have been found coinfecting amoeba hosts (9). The potential role of protists and amoebas in particular for enabling gene exchange among potential pathogens has been noted before (73, 74) and evidence to support this role continues to accumulate. For

example, diverse *Acanthamoeba* symbionts appear to share genes with each other that are functionally enriched for host interaction (75), and the genomes of *Mycobacterium* that can coinfect amoebas contain several genes with significant homology to other amoeba-resistant bacterial genomes (76). Finally, protist-associated *Parachlamydiaceae* have open pan-genomes that suggest the horizontal acquisition of genes to adapt to different intracellular niches (77). This is in contrast to more specialized obligate intracellular pathogens in the *Chlamydiaceae*.

While the secretion system features shared among *Paraburkholderia* symbionts of *D. discoideum* are striking, *P. agricolaris* is otherwise difficult to distinguish from other *Paraburkholderia* based on its genome size and content. However, the reduced genomes of *P. bonniea* and *P. hayleyella* display characteristics indicative of their evolution in a host environment. All three species retain the ability to live outside *D. discoideum*, but the genomes of *P. bonniea* and *P. hayleyella* show fewer transcriptional regulators, as well as fewer carbohydrate and inorganic ion transporters. The reduced genomes possess a combination of genes with molecular evolution patterns that indicate specific responses to the host environment (both stronger and weaker evolutionary constraints) rather than uniform deterioration under genetic drift. In addition, the lack of IS element proliferation and absence of excessive pseudogene accumulation compared to other *Paraburkholderia* genomes suggest that these already reduced genomes are relatively stable.

These combined pieces of evidence support the view that the reduced-genome *D. discoideum* symbionts are "professional symbionts," symbiont lineages that may be ancestrally adapted and possess reduced genomes that are compact and streamlined rather than haphazardly deteriorated (11). Similar to what we observed in the genomes of *P. bonniea* and *P. hayleyella*, streamlined genomes of professional symbionts are expected to have intact core essential genes, including DNA recombination and repair pathways, and various systems that can mediate interactions with their hosts, including secretion systems and effectors (11). We note that genome streamlining as a theory refers to primarily adaptive genome reduction in free-living microbes that is predicted to be the result of a large effective population size and selection that favors lower genome complexity (78, 79). The streamlined rather than deteriorated genomes of *P. bonniea* and *P. hayleyella* do exhibit characteristics that overlap the predicted outcomes of genome streamlining (e.g., intact core metabolic repertoire, lack of pseudogenes and mobile elements, reduced regulatory factors). However, the underlying cause of genome reduction must be distinct given the facultative nature of *D. discoideum-Paraburkholderia* symbiosis. *P. bonniea* and *P. hayleyella* genomes may have evolved under potentially fluctuating regimes of effective population size and selection pressures depending on whether the symbionts were within an amoeba host versus free living in the soil environment. Intraspecific genetic variation appears to be larger for *P. agricolaris* compared to *P. bonniea* or *P. hayleyella* (9), suggesting that *P. agricolaris* host adaptation may be ongoing and more dynamic. We look forward to expanding this work to a larger collection of *D. discoideum*-symbiont genomes in the future to identify both convergent and divergent host adaptation patterns among *D. discoideum* symbionts and to continue to add to a growing body of work across diverse protist-prokaryote symbioses.

Burkholderia sensu lato include a wide range of species that interact with eukaryotic hosts in symbiotic relationships (80, 81). Our specific results have pointed us mostly toward *Burkholderia* that are mammalian pathogens for clues regarding how *D. discoideum* and their symbiotic *Paraburkholderia* interact. However, many plants, fungi, and particularly insects have close associations with *Burkholderia* (82). With these other eukaryotic hosts, *Burkholderia* symbionts can offer benefits as diverse as nitrogen fixation or metabolism, pesticide or plant secondary compound degradation, or provide bioactive secondary metabolites as defensive compounds (82). We anticipate drawing upon the rich body of work on *Burkholderia* symbioses in follow-up investigations in the *D. discoideum-Paraburkholderia* system. For example, *Burkholderia gladioli* strains include

pathogens of several plants and defensive symbionts of herbivorous *Lagriinae* beetles (83). Two *B. gladioli* strains isolated from *Lagria villosa* beetles include a reduced-genome strain that is the dominant symbiont and a nonreduced strain that occurs more sporadically (84). Both strains produce different antifungals that protect *L. villosa* beetle eggs but can also infect plants and reduce seed production (83, 85, 86). In contrast to *P. bonniea* and *P. hayleyella*, the reduced-genome *B. gladioli* strain is missing genes for several metabolic pathways and DNA repair. The horizontal gain of genes for synthesis of the defensive compound lagriamide is thought to be the key event that led to symbiosis establishment between *L. villosa* beetles and a plant-associated *B. gladioli* ancestor of the reduced-genome strain (84).

Conclusion. Among the three *Paraburkholderia* symbionts of *D. discoideum*, the genome size and content of *P. agricolaris* are similar to other *Paraburkholderia* while the reduced genomes of *P. bonniea* and *P. hayleyella* display characteristics of genome streamlining rather than deterioration during host adaptation. Despite these differences, we found adjacent type 3 and type 6 secretion system operons shared across all three *D. discoideum*-symbiont genomes that may have an important role in host-symbiont interactions. These and other shared features suggest a role for horizontal gene transfer within the amoeba host environment that may contribute to the unique ability of these symbionts to establish and maintain a symbiotic relationship with *D. discoideum*.

MATERIALS AND METHODS

***Paraburkholderia* genome selection and gene prediction.** Genome sequencing methods were described previously (10). Briefly, we prepared high-quality DNA from individual strains grown on SM/5 agar media using Qiagen Genomic tips (20/G). Two genomes (*P. agricolaris* and *P. bonniea*) were sequenced by the University of Washington PacBio Sequencing Services, and *P. hayleyella* was sequenced by the Duke University Center for Genomic and Computational Biology, all on the PacBio SMRT II platform. Reads were assembled via HGAP versions 1.87 and 1.85 (87). After an initial round of annotation, we identified the chromosomal replication initiator *dnaA* sequence and *initiator replication protein* in each assembly's contig and reoriented each contig from these genes using Circlator (88). We used SMRT analysis software Quiver to repolish each assembly (87).

We chose the following *Paraburkholderia* with finished genomes for more detailed comparison: *P. fungorum* strain ATCC BAA-463 (89), originally isolated from the fungus *Phanerochaete chrysosporium* (90); *P. sprentiae* strain WSM5005 (91) isolated from root nodules of the domesticated legume *Lebeckia ambigua*; *P. terrae* strain DSM17804 (92) isolated from broad-leaved forest soil; and *P. xenovorans* strain LB400 (93) isolated from polychlorinated biphenyl-contaminated soil. We refer to these four as our representative *Paraburkholderia* genomes. For broader scale analyses of molecular evolution and comparative genomics, we used eight additional *Paraburkholderia* genomes that span the clade that includes *P. agricolaris*, *P. hayleyella*, and *P. bonniea*. We added four plant-associated species genomes (*P. megapolitana* LMG23650, *P. phenoliruptrix* BR3459a, *P. phymatum* STM815, *P. phytofirmans* PsJN) and four free-living species genomes (*P. caledonica* PHRS4, *P. phenazinium* LMG2247, *P. sartisoli* LMG24000, and *P. terricola* mHS1) (94–100). All genomes were downloaded from NCBI and considered complete (Table S1). While the genomes of *P. sartisoli* and *P. phenazinium* are fragmented into multiple contigs, all remaining selected genomes are finished into full-length chromosomes (101).

We reannotated each genome with Prokka v1.14.6 (102) using the annotation file of *Burkholderia pseudomallei* strain K96243 (downloaded from Burkholderia Genome DB version 9.1) as a source of known proteins. Next, we found putative pseudogenes in each genome using Pseudofinder v1.0 (103) with DIAMOND v2.0.6.144 (104) BLAST against the NCBI RefSeq nonredundant protein database (downloaded August 27, 2021) in Annotate mode. Genes predicted to be pseudogenes due to truncation (less than 65% of average length of similar genes by default) or fragmentation (adjacent predicted reading frames match the same known protein) were removed from further analysis.

Whole-genome alignment. The genome aligner progressiveMauve (105) identifies locally colinear blocks (LCBs), local alignments that occur in the same sequence order and orientation across multiple genomes. We used all 13 finished *Paraburkholderia* genomes (all genomes noted above except *P. sartisoli* and *P. phenazinium*) for the initial whole-genome progressiveMauve alignment in Mauve v2015-02-25. Next we compared the positions and orientations of locally colinear blocks across our three *D. discoideum*-symbiont genomes and each of these against the four representative *Paraburkholderia* genomes to identify large-scale synteny using hive plots (106). We used ggraph v2.0.3 and igraph v1.2.11 in R v3.6.0 (107) to generate the hive plots.

Horizontally transferred genetic element detection. We used the ISFinder (108) webserver (accessed 17 October 2021) and its nucleotide BLAST to identify putative IS elements to test for their proliferation in each genome. We identified the best hits by comparing overlapping hits by E value and bit score. We retained hits that were at least 70% coverage of the IS element it matched in the database.

Genomic islands are clusters of genes of horizontally transferred origin and have been found in a range of sizes from as small as 5 to as large as 500 kb (109–111). We applied IslandPath-DIMOB and SIGI-HMM as implemented via the IslandViewer 4 webserver (112). IslandPath-DIMOB uses sequence

composition and mobility genes, while SIGI-HMM uses codon usage bias. We then used pairwise reciprocal megablast to determine whether any of the predicted genomic islands were shared among *D. discoideum*-symbiont genomes.

Finally, we looked for individually occurring horizontally transferred genes using DarkHorse2 v2.0_rev09. DarkHorse2 compares individual genes against the NCBI NR database and detects genes with unusual distributions of hits by calculating a lineage probability index (LPI) score (113, 114). Vertically inherited genes will have a high-LPI score because most high-scoring BLASTP hits will belong to close taxonomic relatives. Horizontally transferred genes are detected because high-scoring BLASTP hits will be taxonomically distant, leading to lower LPI scores. We used DIAMOND to perform BLASTP, and then following suggestions from the author, excluded self and sister species hits (*P. fungorum* for *P. agricolaris*, *P. bonniea*, and *P. hayleyella* from each other) and set the global filter threshold to 0.02 to allow candidate matches to have bit scores up to 2% different from the best nonself match.

Gene functional annotation. We performed broad-scale functional annotation with Clusters of Orthologous Groups (COG) (115, 116) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) (117, 118). We assigned COG by RPS-BLAST (119) against COG position-specific scoring matrices downloaded from the NCBI Conserved Domain Database (version 31 July 2019). We followed JGI MGAP v4 practices and used an E-value cutoff of 0.01 and query coverage of at least 70% to be considered a valid assignment (120). We assigned KO using the BlastKOALA webserver (http://kegg.jp/blastkoala/; accessed 13 to 28 July 2020) that performed BLASTP against the KEGG GENES database at the prokaryote Genus and eukaryote Family level (121).

We compared functional genome composition in terms of the numbers of genes observed in each COG category using agglomerative clustering and nonmetric multidimensional scaling (NMDS). Both were implemented in R: agglomerative clustering using cluster v2.1.2, and NMDS using vegan v2.5-7. In both analyses, the reduced genomes of *P. bonniea* and *P. hayleyella* comprise their own cluster while all other genomes clustered together. We compared genome statistics by cluster, including genome size, GC% (proportion of GC nucleotides in the genome), and proportions of intact genes versus pseudogenes. To determine which COG categories contribute to this difference, we used a binomial exactTest (122) using edgeR v3.26.8 (123). Because the enrichment of functional categories of genes for the comparison of *P. bonniea* and *P. hayleyella* versus other *Paraburkholderia* genomes may be due to the maintenance of necessary genes despite genome size degradation, we looked at both normalized and raw count comparisons. For each COG category that was significantly differently detected between the two clusters both in the relative (postnormalization) and absolute (raw counts) sense, we investigated which specific COGs were contributing to the difference. We used KEGG Mapper (124) and its Reconstruct Pathway tool (https://www.genome.jp/kegg/tool/map_pathway.html; accessed 14 March 2022) to corroborate differences in pathway components in genomes based on KO annotations.

Core genome molecular evolution. To determine orthologous genes shared among all examined genomes, we performed a pan-genome analysis using Roary v3.13.0 (125) with a 70% identity threshold. To test hypotheses regarding changes in lineage-specific rates of molecular evolution in *Paraburkholderia* symbionts of *D. discoideum*, we used core genes detected in the Roary pan genome analysis. We used the whole-genome species tree from Brock et al. (10) and dropped any additional taxa using the drop.tip() function in phytools v1.0-1 in R. This species tree was used throughout the subsequent molecular evolution analyses using PAML v4.9d (126). Protein sequence multifasta files for each core gene were aligned with MUSCLE v3.8.31 (127) and then converted into codon alignments using PAL2NAL v14 (128). We ran a series of codeml analyses on each codon alignment with proportional branch lengths as recommended by the PAML manual.

We applied three alternative hypotheses that test whether patterns of molecular evolution were altered by a symbiotic lifestyle ("symbiotic"), association specifically with *D. discoideum* ("dicty"), or in the reduced genomes of *P. bonniea* and *P. hayleyella* ("reduced") (Table S1). We compared each of their Akaike Information Criterion (AIC) scores to that of the null hypothesis (H0) that there should be no significant variation in molecular evolution across the species tree. AIC attempts to minimize information loss in order to select the "best" model (129, 130). The hypothesis with the smallest AIC score with at least a 1-point difference from the null hypothesis was considered the best fit. For genes that showed patterns of molecular evolution that best fit an alternative hypothesis, we used Wilcoxon signed-rank tests in R to compare estimates of dN/dS, the ratio of nonsynonymous substitutions per nonsynonymous site over synonymous substitutions per synonymous site, between groups of species.

Essential amino acid biosynthetic repertoire. We used GapMind webserver (http://papers.genomics.lbl.gov/cgi-bin/gapView.cgi; accessed 26 January 2022) to evaluate any loss of essential amino acid biosynthesis pathways in each genome. GapMind detects genes involved in the biosynthesis of 17 amino acids (all standard amino acids excluding alanine, aspartate, and glutamate) and chorismate based on MetaCyc pathways using a combination of sequence similarity and protein family profiles (131, 132). It can handle fusion proteins (two enzymes fused into a single protein) and split proteins (multidomain enzyme split into up to two proteins).

Protein secretion system repertoire and effector prediction. We used TXSScan (133) implemented in Galaxy/Pasteur (accessed 2 October 2020) to identify protein secretion systems in the three focal and 12 additional *Paraburkholderia* genomes. TXSScan identifies protein secretion systems (types I to VI and IX, including type IV and tight adherence [Tad] pili) and flagella based on 204 experimentally studied protein profiles. It also determines whether a secretion system is complete by the presence of mandatory and forbidden component genes by subtype, and whether it is contained within a single operon (single locus) or across a few neighboring operons (multilocus). We used the genomes of *Burkholderia mallei* ATCC 23344 and *Burkholderia pseudomallei* K96243 here to serve as ground truth

because their secretion systems are well studied. A small number of T6SS and one T3SS were classified as incomplete due to misidentifying secretion system component homologs (e.g., TssC as IglB). These were manually corrected and included in the analyses. We verified these manually corrected operons against secretion system databases and Burkholderia Genome DB v9.1 (134).

We classified all T3SS and T6SS found in our 15 genomes. For T3SS, we used the T3Enc database v1.0 (135) and downloaded three representative amino acid sequences of 13 categories of T3SS for the conserved component genes sctJ (inner membrane ring; IPR003282), sctN (ATPase; IPR005714), and sctV (export apparatus; IPR006302). We aligned protein sequences of each component gene using MUSCLE v3.8.31 (127) and made gene trees using the Le and Gascuel substitution model with FastTree v2.1.10 (136). We estimated a species tree from these gene trees using ASTRID v2.2.1 (137) and ASTRAL v5.7.8 (138). We followed the same methods for T6SS using the SecReT6 database v3.0 (139) and the conserved component genes tssB (sheath; COG3516), tssC (sheath; COG3517), and tssF (baseplate; COG3519).

We used VFDB (Virulence factors of Pathogenic Bacteria; accessed 25 January 2022) (140) and downloaded protein sequences of known *Bordatella* T3 Secreted Effectors and *Burkholderia* T3 and T6 Secreted Effectors. We used DIAMOND BLASTP and these proteins as query sequences against the predicted amino acid sequences of each genome. We also used the webserver BastionHub (accessed 20 April 2021) to predict secreted effectors. BastionHub (141) combines a hidden Markov model-based approach and a machine learning approach. Finally, we used effectiveELD (142, 143) on the effectiveDB server (accessed 28 March 2022) to find putative secreted proteins that contain a eukaryotic-like domain. We specifically looked for proteins with domains that belong to Pfam clans for Ank (ankyrin), TPR (tetratricopeptide repeat), LRR (leucine-rich repeat), Pentapeptide, F-box, and RING (including U-box). These domains were selected based on previous reports regarding large numbers of proteins containing eukaryotic domains among amoeba symbionts (69–71). InterProScan (144) webserver (accessed 29 March 2022) was used for additional investigation of secreted effector candidates.

***Paraburkholderia* genome browser.** We built a web Genome Browser for each *D. discoideum*-symbiont genome for convenient browsing of all annotated genomic features mentioned above. We used JBrowse v1 (145, 146). The front-end web application was developed in Centos Steam 8 version of Linux. We used NGINX Web Server v1.14.1 and Java OpenJDK v1.8.0_322. The browser is available at https://burk.colby.edu. The GitHub repositories supporting the browser are available at https://github.com/noh-lab/burk-browser and https://github.com/noh-lab/jbrowse-executables.

Data and code availability. All analyses and figures found in this article can be generated and recreated using input data and code available at the GitHub repository https://github.com/noh-lab/comparative-dicty-symbionts.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 1.1 MB.
FIG S2, EPS file, 2.1 MB.
FIG S3, EPS file, 1.9 MB.
FIG S4, EPS file, 2 MB.
FIG S5, EPS file, 2.6 MB.
TABLE S1, DOCX file, 0.02 MB.
TABLE S2, DOCX file, 0.01 MB.
TABLE S3, DOCX file, 0.03 MB.
TABLE S4, DOCX file, 0.02 MB.
TABLE S5, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank past and current members of the Noh lab (Anna Chen, Kayla Dixon, and Laura Drepanos) and the Strassmann-Queller lab (Tammy Haselkorn and Clarissa Dzikunu) who contributed to aspects of this project during its development. Ron Peck, Laura Runyen-Janecky, and two parties of reviewers provided helpful comments that improved previous drafts.

This work was supported by the National Science Foundation under grant numbers IOS 1656756 and DEB 1753743 (to J.E.S. and D.C.Q.) and National Institutes of Health and its National Institute of General Medical Sciences by an Institutional Development Award (IDeA) under grant number P20GM103423 (subaward to S.N.) and Colby College startup funds (to S.N.).

REFERENCES

1. Cosson P, Soldati T. 2008. Eat, kill or die: when amoeba meets bacteria. Curr Opin Microbiol 11:271–276. https://doi.org/10.1016/j.mib.2008.05.005.
2. Bozzaro S, Eichinger L. 2011. The professional phagocyte *Dictyostelium discoideum* as a model host for bacterial pathogens. Curr Drug Targets 12:942–954. https://doi.org/10.2174/138945011795677782.
3. Dunn JD, Bosmani C, Barisch C, Raykov L, Lefrançois LH, Cardenal-Muñoz E, López-Jiménez AT, Soldati T. 2017. Eat prey, live: *Dictyostelium discoideum* as a model for cell-autonomous defenses. Front Immunol 8:1906. https://doi.org/10.3389/fimmu.2017.01906.
4. Tipton L, Darcy JL, Hynson NA. 2019. A developing symbiosis: enabling cross-talk between ecologists and microbiome scientists. Front Microbiol 10:292. https://doi.org/10.3389/fmicb.2019.00292.
5. Hentschel U. 2021. Harnessing the power of host–microbe symbioses to address grand challenges. Nat Rev Microbiol 19:615–616. https://doi.org/10.1038/s41579-021-00619-3.
6. Drew GC, Stevens EJ, King KC. 2021. Microbial evolution and transitions along the parasite–mutualist continuum. Nat Rev Microbiol 19:623–638. https://doi.org/10.1038/s41579-021-00550-7.
7. Brock DA, Douglas TE, Queller DC, Strassmann JE. 2011. Primitive agriculture in a social amoeba. Nature 469:393–396. https://doi.org/10.1038/nature09668.
8. DiSalvo S, Haselkorn TS, Bashir U, Jimenez D, Brock DA, Queller DC, Strassmann JE. 2015. *Burkholderia* bacteria infectiously induce the proto-farming symbiosis of *Dictyostelium* amoebae and food bacteria. Proc Natl Acad Sci U S A 112:E5029–E5037.
9. Haselkorn TS, DiSalvo S, Miller JW, Bashir U, Brock DA, Queller DC, Strassmann JE. 2019. The specificity of *Burkholderia* symbionts in the social amoeba farming symbiosis: prevalence, species, genetic and phenotypic diversity. Mol Ecol 28:847–862. https://doi.org/10.1111/mec.14982.
10. Brock DA, Noh S, Hubert ANM, Haselkorn TS, DiSalvo S, Suess MK, Bradley AS, Tavakoli-Nezhad M, Geist KS, Queller DC, Strassmann JE. 2020. Endosymbiotic adaptations in three new bacterial species associated with *Dictyostelium discoideum*: *Paraburkholderia agricolaris* sp. nov., *Paraburkholderia hayleyella* sp. nov., and *Paraburkholderia bonniea* sp. nov. PeerJ 8:e9151. https://doi.org/10.7717/peerj.9151.
11. Husnik F, Tashyreva D, Boscaro V, George EE, Lukeš J, Keeling PJ. 2021. Bacterial and archaeal symbioses with protists. Curr Biol 31:R862–R877. https://doi.org/10.1016/j.cub.2021.05.049.
12. Scott TJ, Queller DC, Strassmann JE. 2022. Context dependence in the symbiosis between *Dictyostelium discoideum* and *Paraburkholderia*. Evol Lett 6:245–254. https://doi.org/10.1002/evl3.281.
13. Shu L, Brock DA, Geist KS, Miller JW, Queller DC, Strassmann JE, DiSalvo S. 2018. Symbiont location, host fitness, and possible coadaptation in a symbiosis between social amoebae and bacteria. Elife 7:e42660. https://doi.org/10.7554/eLife.42660.
14. Miller JW, Bocke CR, Tresslar AR, Schniepp EM, DiSalvo S. 2020. *Paraburkholderia* symbionts display variable infection patterns that are not predictive of amoeba host outcomes. Genes 11:674. https://doi.org/10.3390/genes11060674.
15. Brock DA, Read S, Bozhchenko A, Queller DC, Strassmann JE. 2013. Social amoeba farmers carry defensive symbionts to protect and privatize their crops. Nat Commun 4:2385. https://doi.org/10.1038/ncomms3385.
16. Garcia JR, Larsen TJ, Queller DC, Strassmann JE. 2019. Fitness costs and benefits vary for two facultative *Burkholderia* symbionts of the social amoeba, *Dictyostelium discoideum*. Ecol Evol 9:9878–9890. https://doi.org/10.1002/ece3.5529.
17. Shu L, Zhang B, Queller DC, Strassmann JE. 2018. *Burkholderia* bacteria use chemotaxis to find social amoeba *Dictyostelium discoideum* hosts. 8. ISME J 12:1977–1993. https://doi.org/10.1038/s41396-018-0147-4.
18. Haselkorn TS, Jimenez D, Bashir U, Sallinger E, Queller DC, Strassmann JE, DiSalvo S. 2021. Novel Chlamydiae and *Amoebophilus* endosymbionts are prevalent in wild isolates of the model social amoeba *Dictyostelium discoideum*. Environ Microbiol Rep 13:708–719. https://doi.org/10.1111/1758-2229.12985.
19. Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14:627–633. https://doi.org/10.1016/j.gde.2004.09.003.
20. Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108:583–586. https://doi.org/10.1016/S0092-8674(02)00665-7.
21. Maurelli AT. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. FEMS Microbiol Lett 267:1–8. https://doi.org/10.1111/j.1574-6968.2006.00526.x.
22. Bliven KA, Maurelli AT. 2012. Antivirulence genes: insights into pathogen evolution through gene loss. Infect Immun 80:4061–4070. https://doi.org/10.1128/IAI.00740-12.
23. Merhej V, Georgiades K, Raoult D. 2013. Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors. Brief Funct Genomics 12:291–304. https://doi.org/10.1093/bfgp/elt015.
24. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol Direct 4:13. https://doi.org/10.1186/1745-6150-4-13.
25. Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet 11:465–475. https://doi.org/10.1038/nrg2798.
26. Nierman WC, DeShazer D, Kim HS, Tettelin H, Nelson KE, Feldblyum T, Ulrich RL, Ronning CM, Brinkac LM, Daugherty SC, Davidsen TD, Deboy RT, Dimitrov G, Dodson RJ, Durkin AS, Gwinn ML, Haft DH, Khouri H, Kolonay JF, Madupu R, Mohammoud Y, Nelson WC, Radune D, Romero CM, Sarria S, Selengut J, Shamblin C, Sullivan SA, White O, Yu Y, Zafar N, Zhou L, Fraser CM. 2004. Structural flexibility in the *Burkholderia mallei* genome. Proc Natl Acad Sci U S A 101:14246–14251. https://doi.org/10.1073/pnas.0403306101.
27. Gomez-Valero L, Rusniok C, Cazalet C, Buchrieser C. 2011. Comparative and functional genomics of *Legionella* identified eukaryotic like proteins as key players in host–pathogen interactions. Front Microbiol 2:208. https://doi.org/10.3389/fmicb.2011.00208.
28. Cazalet C, Rusniok C, Brüggemann H, Zidane N, Magnier A, Ma L, Tichit M, Jarraud S, Bouchier C, Vandenesch F, Kunst F, Etienne J, Glaser P, Buchrieser C. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. Nat Genet 36:1165–1173. https://doi.org/10.1038/ng1447.
29. McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. 1. Nat Rev Microbiol 10:13–26. https://doi.org/10.1038/nrmicro2670.
30. Andersson SGE, Kurland CG. 1998. Reductive evolution of resident genomes. Trends Microbiol 6:263–268. https://doi.org/10.1016/S0966-842X(98)01312-2.
31. Abby S, Daubin V. 2007. Comparative genomics and the evolution of prokaryotes. Trends Microbiol 15:135–141. https://doi.org/10.1016/j.tim.2007.01.007.
32. Arnold BJ, Huang I-T, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria. 4. Nat Rev Microbiol 20:206–218. https://doi.org/10.1038/s41579-021-00650-4.
33. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. 2019. The ecology and evolution of pangenomes. Curr Biol 29:R1094–R1103. https://doi.org/10.1016/j.cub.2019.08.012.
34. Liu Y, Harrison PM, Kunin V, Gerstein M. 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol 5:R64. https://doi.org/10.1186/gb-2004-5-9-r64.
35. Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. EMBO Rep 2:376–381. https://doi.org/10.1093/embo-reports/kve097.
36. Ochman H, Davalos LM. 2006. The nature and dynamics of bacterial genomes. Science 311:1730–1733. https://doi.org/10.1126/science.1119966.
37. Losada L, Ronning CM, DeShazer D, Woods D, Fedorova N, Stanley Kim H, Shabalina SA, Pearson TR, Brinkac L, Tan P, Nandi T, Crabtree J, Badger J, Beckstrom-Sternberg S, Saqib M, Schutzer SE, Keim P, Nierman WC. 2010. Continuing evolution of *Burkholderia mallei* through genome reduction and large-scale rearrangements. Genome Biol Evol 2:102–116. https://doi.org/10.1093/gbe/evq003.
38. Manzano-Marín A, Latorre A. 2016. Snapshots of a shrinking partner: genome reduction in *Serratia symbiotica*. Sci Rep 6:32590. https://doi.org/10.1038/srep32590.
39. Ottemann KM, Miller JF. 1997. Roles for motility in bacterial–host interactions. Mol Microbiol 24:1109–1117. https://doi.org/10.1046/j.1365-2958.1997.4281787.x.
40. Duan Q, Zhou M, Zhu L, Zhu G. 2013. Flagella and bacterial pathogenicity. J Basic Microbiol 53:1–8. https://doi.org/10.1002/jobm.201100335.
41. Chua KL, Chan YY, Gan YH. 2003. Flagella are virulence determinants of *Burkholderia pseudomallei*. Infect Immun 71:1622–1629. https://doi.org/10.1128/IAI.71.4.1622-1629.2003.

42. French CT, Toesca IJ, Wu T-H, Teslaa T, Beaty SM, Wong W, Liu M, Schröder I, Chiou P-Y, Teitell MA, Miller JF. 2011. Dissection of the *Burkholderia* intracellular life cycle using a photothermal nanoblade. Proc Natl Acad Sci U S A 108:12095–12100. https://doi.org/10.1073/pnas.1107183108.
43. Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. Mol Microbiol 48:1491–1500. https://doi.org/10.1046/j.1365-2958.2003.03522.x.
44. Garmory HS, Titball RW. 2004. ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. Infect Immun 72:6757–6763. https://doi.org/10.1128/IAI.72.12.6757-6763.2004.
45. Harland DN, Dassa E, Titball RW, Brown KA, Atkins HS. 2007. ATP-binding cassette systems in *Burkholderia pseudomallei* and *Burkholderia mallei*. BMC Genomics 8:83. https://doi.org/10.1186/1471-2164-8-83.
46. Schaefers MM. 2020. Regulation of virulence by two-component systems in pathogenic *Burkholderia*. Infect Immun 88:e00927-19. https://doi.org/10.1128/IAI.00927-19.
47. Mangalea MR, Plumley BA, Borlee BR. 2017. Nitrate sensing and metabolism inhibit biofilm formation in the opportunistic pathogen *Burkholderia pseudomallei* by reducing the intracellular concentration of c-di-GMP. Front Microbiol 8:1353. https://doi.org/10.3389/fmicb.2017.01353.
48. Sabater-Muñoz B, Toft C, Alvarez-Ponce D, Fares MA. 2017. Chance and necessity in the genome evolution of endosymbiotic bacteria of insects. ISME J 11:1291–1304. https://doi.org/10.1038/ismej.2017.18.
49. Wernegreen JJ. 2017. In it for the long haul: evolutionary consequences of persistent endosymbiosis. Curr Opin Genet Dev 47:83–90. https://doi.org/10.1016/j.gde.2017.08.006.
50. Butt AT, Thomas MS. 2017. Iron acquisition mechanisms and their role in the virulence of *Burkholderia* species. Front Cell Infect Microbiol 7:460. https://doi.org/10.3389/fcimb.2017.00460.
51. Murphy ER, Sacco RE, Dickenson A, Metzger DJ, Hu Y, Orndorff PE, Connell TD. 2002. BhuR, a virulence-associated outer membrane protein of *Bordetella avium*, is required for the acquisition of iron from heme and hemoproteins. Infect Immun 70:5390–5403. https://doi.org/10.1128/IAI.70.10.5390-5403.2002.
52. Vanderpool CK, Armstrong SK. 2001. The *Bordetella bhu* locus is required for heme iron utilization. J Bacteriol 183:4278–4287. https://doi.org/10.1128/JB.183.14.4278-4287.2001.
53. Finlay BB, Falkow S. 1997. Common themes in microbial pathogenicity revisited. Microbiol Mol Biol Rev 61:136–169. https://doi.org/10.1128/mmbr.61.2.136-169.1997.
54. Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42:165–190. https://doi.org/10.1146/annurev.genet.41.110306.130119.
55. Lo W-S, Huang Y-Y, Kuo C-H. 2016. Winding paths to simplicity: genome evolution in facultative insect symbionts. FEMS Microbiol Rev 40:855–874. https://doi.org/10.1093/femsre/fuw028.
56. McCutcheon JP, Boyd BM, Dale C. 2019. The life of an insect endosymbiont from the cradle to the grave. Curr Biol 29:R485–R495. https://doi.org/10.1016/j.cub.2019.03.032.
57. Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol 44:383–397. https://doi.org/10.1007/PL00006158.
58. Tseng T-T, Tyler BM, Setubal JC. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiol 9:S2. https://doi.org/10.1186/1471-2180-9-S1-S2.
59. Coombes BK. 2009. Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. Trends Microbiol 17:89–94. https://doi.org/10.1016/j.tim.2008.11.006.
60. Mannaa M, Park I, Seo Y-S. 2018. Genomic features and insights into the taxonomy, virulence, and benevolence of plant-associated *Burkholderia* species. Int J Mol Sci 20:121. https://doi.org/10.3390/ijms20010121.
61. Kamanova J. 2020. *Bordetella* type III secretion injectosome and effector proteins. Front Cell Infect Microbiol 10:466. https://doi.org/10.3389/fcimb.2020.00466.
62. Lennings J, West TE, Schwarz S. 2018. The *Burkholderia* type VI secretion system 5: composition, regulation and role in virulence. Front Microbiol 9:3339. https://doi.org/10.3389/fmicb.2018.03339.
63. Schell MA, Ulrich RL, Ribot WJ, Brueggemann EE, Hines HB, Chen D, Lipscomb L, Kim HS, Mrázek J, Nierman WC, Deshazer D. 2007. Type VI secretion is a major virulence determinant in *Burkholderia mallei*. Mol Microbiol 64:1466–1485. https://doi.org/10.1111/j.1365-2958.2007.05734.x.
64. Burtnick MN, Brett PJ, Harding SV, Ngugi SA, Ribot WJ, Chantratita N, Scorpio A, Milne TS, Dean RE, Fritz DL, Peacock SJ, Prior JL, Atkins TP, DeShazer D. 2011. The cluster 1 type VI secretion system is a major virulence determinant in *Burkholderia pseudomallei*. Infect Immun 79:1512–1525. https://doi.org/10.1128/IAI.01218-10.
65. Sun GW, Chen Y, Liu Y, Tan G-YG, Ong C, Tan P, Gan Y-H. 2010. Identification of a regulatory cascade controlling Type III Secretion System 3 gene expression in *Burkholderia pseudomallei*. Mol Microbiol 76:677–689. https://doi.org/10.1111/j.1365-2958.2010.07124.x.
66. Chen Y, Wong J, Sun GW, Liu Y, Tan G-YG, Gan Y-H. 2011. Regulation of type VI secretion system during *Burkholderia pseudomallei* infection. Infect Immun 79:3064–3073. https://doi.org/10.1128/IAI.05148-11.
67. Schwarz S, West TE, Boyer F, Chiang W-C, Carl MA, Hood RD, Rohmer L, Tolker-Nielsen T, Skerrett SJ, Mougous JD. 2010. *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. PLoS Pathog 6:e1001068. https://doi.org/10.1371/journal.ppat.1001068.
68. Schwarz S, Singh P, Robertson JD, LeRoux M, Skerrett SJ, Goodlett DR, West TE, Mougous JD. 2014. VgrG-5 Is a *Burkholderia* type VI secretion system-exported protein required for multinucleated giant cell formation and virulence. Infect Immun 82:1445–1452. https://doi.org/10.1128/IAI.01368-13.
69. Schmitz-Esser S, Tischler P, Arnold R, Montanaro J, Wagner M, Rattei T, Horn M. 2010. The genome of the amoeba symbiont "Candidatus *Amoebophilus asiaticus*" reveals common mechanisms for host cell interaction among amoeba-associated bacteria. J Bacteriol 192:1045–1057. https://doi.org/10.1128/JB.01379-09.
70. Gomez-Valero L, Buchrieser C. 2019. Intracellular parasitism, the driving force of evolution of *Legionella pneumophila* and the genus *Legionella*. Microbes Infect 21:230–236. https://doi.org/10.1016/j.micinf.2019.06.012.
71. Schulz F, Martijn J, Wascher F, Lagkouvardos I, Kostanjšek R, Ettema TJG, Horn M. 2016. A Rickettsiales symbiont of amoebae with ancient features. Environ Microbiol 18:2326–2342. https://doi.org/10.1111/1462-2920.12881.
72. Wallner A, Moulin L, Busset N, Rimbault I, Béna G. 2021. Genetic diversity of type 3 secretion system in *Burkholderia s.l.* and links with plant host adaptation. Front Microbiol 12:761215. https://doi.org/10.3389/fmicb.2021.761215.
73. Bertelli C, Greub G. 2012. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. Front Cell Infect Microbiol 2:110. https://doi.org/10.3389/fcimb.2012.00110.
74. Moliner C, Fournier P-E, Raoult D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS Microbiol Rev 34:281–294. https://doi.org/10.1111/j.1574-6976.2009.00209.x.
75. Wang Z, Wu M. 2017. Comparative genomic analysis of *Acanthamoeba* endosymbionts highlights the role of amoebae as a "melting pot" shaping the *Rickettsiales* evolution. Genome Biol Evol 9:3214–3224. https://doi.org/10.1093/gbe/evx246.
76. Lamrabet O, Merhej V, Pontarotti P, Raoult D, Drancourt M. 2012. The genealogic tree of mycobacteria reveals a long-standing sympatric life into free-living protozoa. PLoS One 7:e34754. https://doi.org/10.1371/journal.pone.0034754.
77. Köstlbacher S, Collingro A, Halter T, Schulz F, Jungbluth SP, Horn M. 2021. Pangenomics reveals alternative environmental lifestyles among chlamydiae. Nat Commun 12:4021. https://doi.org/10.1038/s41467-021-24294-3.
78. Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. ISME J 8:1553–1565. https://doi.org/10.1038/ismej.2014.60.
79. Lynch M. 2006. Streamlining and simplification of microbial genome architecture. Annu Rev Microbiol 60:327–349. https://doi.org/10.1146/annurev.micro.60.080805.142300.
80. Coenye T, Vandamme P. 2003. Diversity and significance of *Burkholderia* species occupying diverse ecological niches. Environ Microbiol 5:719–729. https://doi.org/10.1046/j.1462-2920.2003.00471.x.
81. Eberl L, Vandamme P. 2016. Members of the genus *Burkholderia*: good and bad guys. F1000Res 5:1007. https://doi.org/10.12688/f1000research.8221.1.
82. Kaltenpoth M, Flórez LV. 2020. Versatile and dynamic symbioses between insects and *Burkholderia* bacteria. Annu Rev Entomol 65:145–170. https://doi.org/10.1146/annurev-ento-011019-025025.
83. Wierz JC, Gaube P, Klebsch D, Kaltenpoth M, Flórez LV. 2021. Transmission of bacterial symbionts with and without genome erosion between a beetle host and the plant environment. Front Microbiol 12:715601. https://doi.org/10.3389/fmicb.2021.715601.
84. Waterworth SC, Flórez LV, Rees ER, Hertweck C, Kaltenpoth M, Kwan JC. 2020. Horizontal gene transfer to a defensive symbiont with a reduced

genome in a multipartite beetle microbiome. mBio 11:e02430-19. https://doi.org/10.1128/mBio.02430-19.
85. Flórez LV, Scherlach K, Gaube P, Ross C, Sitte E, Hermes C, Rodrigues A, Hertweck C, Kaltenpoth M. 2017. Antibiotic-producing symbionts dynamically transition between plant pathogenicity and insect-defensive mutualism. Nat Commun 8:15172. https://doi.org/10.1038/ncomms15172.
86. Flórez LV, Scherlach K, Miller IJ, Rodrigues A, Kwan JC, Hertweck C, Kaltenpoth M. 2018. An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in *Lagria villosa* beetles. Nat Commun 9:2478. https://doi.org/10.1038/s41467-018-04955-6.
87. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
88. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
89. Coenye T, Laevens S, Willems A, Ohlén M, Hannant W, Govan JR, Gillis M, Falsen E, Vandamme P. 2001. *Burkholderia fungorum* sp. nov. and *Burkholderia caledonica* sp. nov., two new species isolated from the environment, animals and human clinical samples. Int J Syst Evol Microbiol 51:1099–1107. https://doi.org/10.1099/00207713-51-3-1099.
90. Seigle-Murandi F, Guiraud P, Croize J, Falsen E, Eriksson KL. 1996. Bacteria are omnipresent on *Phanerochaete chrysosporium* Burdsall. Appl Environ Microbiol 62:2477–2481. https://doi.org/10.1128/aem.62.7.2477-2481.1996.
91. De Meyer SE, Cnockaert M, Ardley JK, Maker G, Yates R, Howieson JG, Vandamme P. 2013. *Burkholderia sprentiae* sp. nov., isolated from *Lebeckia ambigua* root nodules. Int J Syst Evol Microbiol 63:3950–3957. https://doi.org/10.1099/ijs.0.048777-0.
92. Yang H-C, Im W-T, Kim KK, An D-S, Lee S-T. 2006. *Burkholderia terrae* sp. nov., isolated from a forest soil. Int J Syst Evol Microbiol 56:453–457. https://doi.org/10.1099/ijs.0.63968-0.
93. Goris J, De Vos P, Caballero-Mellado J, Park J, Falsen E, Quensen JF, Tiedje JM, Vandamme P. 2004. Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400T and relatives as *Burkholderia xenovorans* sp. nov. Int J Syst Evol Microbiol 54:1677–1681. https://doi.org/10.1099/ijs.0.63101-0.
94. Vandamme P, Opelt K, Knöchel N, Berg C, Schönmann S, De Brandt E, Eberl L, Falsen E, Berg G. 2007. *Burkholderia bryophila* sp. nov. and *Burkholderia megapolitana* sp. nov., moss-associated species with antifungal and plant-growth-promoting properties. Int J Syst Evol Microbiol 57:2228–2235. https://doi.org/10.1099/ijs.0.65142-0.
95. Coenye T, Henry D, Speert DP, Vandamme P. 2004. *Burkholderia phenoliruptrix* sp. nov., to accommodate the 2,4,5-trichlorophenoxyacetic acid and halophenol-degrading strain AC1100. Syst Appl Microbiol 27:623–627. https://doi.org/10.1078/0723202042369992.
96. Vandamme P, Goris J, Chen W-M, de Vos P, Willems A. 2002. *Burkholderia tuberum* sp. nov. and *Burkholderia phymatum* sp. nov., nodulate the roots of tropical legumes. Syst Appl Microbiol 25:507–512. https://doi.org/10.1078/07232020260517634.
97. Sessitsch A, Coenye T, Sturz AV, Vandamme P, Barka EA, Salles JF, Van Elsas JD, Faure D, Reiter B, Glick BR, Wang-Pruski G, Nowak JY. 2005. *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. Int J Syst Evol Microbiol 55:1187–1192. https://doi.org/10.1099/ijs.0.63149-0.
98. Viallard V, Poirier I, Cournoyer B, Haurat J, Wiebkin S, Ophel-Keller K, Balandreau J. 1998. 1998. Burkholderia graminis sp. nov., a rhizospheric Burkholderia species, and reassessment of [Pseudomonas] phenazinium, [Pseudomonas] pyrrocinia and [Pseudomonas] glathei as Burkholderia. Int J Syst Evol Microbiol 48:549–563.
99. Vanlaere E, van der Meer JR, Falsen E, Salles JF, de Brandt E, Vandamme P. 2008. *Burkholderia sartisoli* sp. nov., isolated from a polycyclic aromatic hydrocarbon-contaminated soil. Int J Syst Evol Microbiol 58:420–423. https://doi.org/10.1099/ijs.0.65451-0.
100. Goris J, Dejonghe W, Falsen E, De Clerck E, Geeraerts B, Willems A, Top EM, Vandamme P, De Vos P. 2002. Diversity of transconjugants that acquired plasmid pJP4 or pEMT1 after inoculation of a donor strain in the A- and B-horizon of an agricultural soil and description of *Burkholderia hospita* sp. nov. and *Burkholderia terricola* sp. nov. Syst Appl Microbiol 25:340–352. https://doi.org/10.1078/0723-2020-00134.
101. Mardis E, McPherson J, Martienssen R, Wilson RK, McCombie WR. 2002. What is finished, and why does it matter. Genome Res 12:669–671. https://doi.org/10.1101/gr.032102.
102. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
103. Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husnik F. 2022. Pseudofinder: detection of pseudogenes in prokaryotic genomes. Mol Biol Evol 39:msac153. https://doi.org/10.1093/molbev/msac153.
104. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
105. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.
106. Krzywinski M, Birol I, Jones SJ, Marra MA. 2012. Hive plots—rational approach to visualizing networks. Brief Bioinform 13:627–644. https://doi.org/10.1093/bib/bbr069.
107. R Core Team. 2019. R: a language and environment for statistical computing (3.6.0). R Foundation for Statistical Computing, Vienna, Austria.
108. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32–D36. https://doi.org/10.1093/nar/gkj014.
109. Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol 2:414–424. https://doi.org/10.1038/nrmicro884.
110. Langille MGI, Hsiao WWL, Brinkman FSL. 2010. Detecting genomic islands using bioinformatics approaches. Nat Rev Microbiol 8:373–382. https://doi.org/10.1038/nrmicro2350.
111. Bertelli C, Tilley KE, Brinkman FSL. 2019. Microbial genomic island discovery, visualization and analysis. Brief Bioinform 20:1685–1698. https://doi.org/10.1093/bib/bby042.
112. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman FS, Simon Fraser University Research Computing Group. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45:W30–W35. https://doi.org/10.1093/nar/gkx343.
113. Podell S, Gaasterland T. 2007. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. Genome Biol 8:R16. https://doi.org/10.1186/gb-2007-8-2-r16.
114. Delaye L, Vargas C, Latorre A, Moya A. 2020. Inferring horizontal gene transfer with DarkHorse, Phylomizer, and ETE Toolkits. Methods Mol Biol 2075:355–369. https://doi.org/10.1007/978-1-4939-9877-7_25.
115. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269. https://doi.org/10.1093/nar/gku1223.
116. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36. https://doi.org/10.1093/nar/28.1.33.
117. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070.
118. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092.
119. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
120. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). Stand Genomic Sci 10:86. https://doi.org/10.1186/s40793-015-0077-y.
121. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
122. McMurdie PJ, Holmes S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol 10:e1003531. https://doi.org/10.1371/journal.pcbi.1003531.
123. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616.

124. Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci 29:28–35. https://doi.org/10.1002/pro.3711.
125. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421.
126. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. https://doi.org/10.1093/molbev/msm088.
127. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
128. Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612. https://doi.org/10.1093/nar/gkl315.
129. Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723. https://doi.org/10.1109/TAC.1974.1100705.
130. Symonds MRE, Moussalli A. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav Ecol Sociobiol 65:13–21. https://doi.org/10.1007/s00265-010-1037-6.
131. Price MN, Deutschbauer AM, Arkin AP. 2020. GapMind: automated annotation of amino acid biosynthesis. mSystems 5:e00291-20. https://doi.org/10.1128/mSystems.00291-20.
132. Price MN, Zane GM, Kuehl JV, Melnyk RA, Wall JD, Deutschbauer AM, Arkin AP. 2018. Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. PLoS Genet 14:e1007147. https://doi.org/10.1371/journal.pgen.1007147.
133. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EPC. 2016. Identification of protein secretion systems in bacterial genomes. Sci Rep 6:23080. https://doi.org/10.1038/srep23080.
134. Winsor GL, Khaira B, Van Rossum T, Lo R, Whiteside MD, Brinkman FSL. 2008. The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. Bioinformatics 24:2803–2804. https://doi.org/10.1093/bioinformatics/btn524.
135. Hu Y, Huang H, Cheng X, Shu X, White AP, Stavrinides J, Köster W, Zhu G, Zhao Z, Wang Y. 2017. A global survey of bacterial type III secretion systems and their effectors. Environ Microbiol 19:3879–3895. https://doi.org/10.1111/1462-2920.13755.
136. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
137. Vachaspati P, Warnow T. 2015. ASTRID: Accurate Species TRees from Internode Distances. BMC Genomics 16:S3. https://doi.org/10.1186/1471-2164-16-S10-S3.
138. Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153. https://doi.org/10.1186/s12859-018-2129-y.
139. Li J, Yao Y, Xu HH, Hao L, Deng Z, Rajakumar K, Ou H-Y. 2015. SecReT6: a web-based resource for type VI secretion systems found in bacteria. Environ Microbiol 17:2196–2202. https://doi.org/10.1111/1462-2920.12794.
140. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:D325–D328. https://doi.org/10.1093/nar/gki008.
141. Wang J, Li J, Hou Y, Dai W, Xie R, Marquez-Lago TT, Leier A, Zhou T, Torres V, Hay I, Stubenrauch C, Zhang Y, Song J, Lithgow T. 2021. BastionHub: a universal platform for integrating and analyzing substrates secreted by Gram-negative bacteria. Nucleic Acids Res 49:D651–D659. https://doi.org/10.1093/nar/gkaa899.
142. Jehl M-A, Arnold R, Rattei T. 2011. Effective—a database of predicted secreted bacterial proteins. Nucleic Acids Res 39:D591–D595. https://doi.org/10.1093/nar/gkq1154.
143. Eichinger V, Nussbaumer T, Platzer A, Jehl M-A, Arnold R, Rattei T. 2016. EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. Nucleic Acids Res 44:D669–D674. https://doi.org/10.1093/nar/gkv1269.
144. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
145. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH. 2016. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17:66. https://doi.org/10.1186/s13059-016-0924-1.
146. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. 2009. JBrowse: a next-generation genome browser. Genome Res 19:1630–1638. https://doi.org/10.1101/gr.094607.109.