Title: Comparative genomics support reduced-genome *Paraburkholderia* symbionts of *Dictyostelium discoideum* amoebas are ancestrally adapted professional symbionts

Suegene Noh[1*], Benjamin J. Capodanno[1,2], Songtao Xu[1], Marisa C. Hamilton[1,3], Joan E. Strassmann[4], David C. Queller[4]

[1] Department of Biology, Colby College, Waterville, Maine, USA

[2] Brotman Baty Institute for Precision Medicine, Seattle, Washington, USA

[3] University Program in Genetics and Genomics, Duke University, Durham, North Carolina, USA

[4] 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. In addition, there is a large difference in genome size between *P. agricolaris* (8.7 million base pairs) vs. *P. hayleyella* and *P. bonniea* (4.1 Mbp) and in GC content (62% vs. 59%). 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 two reduced genomes of *P. bonniea* and *P. hayleyella* display characteristics that support evolution in a host environment. In addition, all three *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. Ultimately, our combined evidence supports that the reduced-genome *D. discoideum*-symbionts have evolved to be professional symbionts ancestrally adapted to their protist hosts.

Introduction

The social amoeba *Dictyostelium discoideum* (Eumycetozoa; Dictyosteliales) is a predatory soil protist frequently used for studying host-pathogen interactions (Cosson & Soldati 2008; Bozzaro & Eichinger 2011; Dunn et al. 2017). It is also an emerging model for host-microbe symbiosis in the broad sense, which we 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 (Tipton et al. 2019; Hentschel 2021; Drew et al. 2021). A subset of *D. discoideum* strains isolated from soil persistently carry intracellular gram-negative *Paraburkholderia* (Betaproteobacteria; Burkholderiales) (Brock et al. 2011; DiSalvo et al. 2015; Haselkorn et al. 2019). Multi-locus sequence typing analyses and whole genome phylogenies showed that these symbionts comprise 2 independent clades (Haselkorn et al. 2019; Brock et al. 2020). 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 (Brock et al. 2020).

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 (Husnik et al. 2021). The association between *D. discoideum* and its *Paraburkholderia* symbionts appears to be facultative. *D. discoideum*-symbionts are able to simultaneously maintain a free-living and host-associated lifestyle (Haselkorn et al. 2019;

DiSalvo et al. 2015). The fitness outcomes to host and symbiont appear to be context-dependent (Scott et al. in press), as with most facultative host-microbe symbioses (Drew et al. 2021). *D. discoideum* amoeba hosts generally suffer negative fitness consequences of association. When amoeba 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 (Brock et al. 2011; DiSalvo et al. 2015; Shu, Brock, et al. 2018; Miller et al. 2020). However potentially important context-dependent fitness benefits to the host may be (1) increased availability of food bacteria in relatively inhospitable environments as a result of secondary carriage, and (2) improved competitive ability against other *D. discoideum* strains by potentially passing on symbiont infections or releasing *Paraburkholderia* secretions in a defensive manner (Brock et al. 2011, 2013). 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 (Garcia et al. 2019). While not a direct demonstration of any fitness benefits, *P. agricolaris* and *P. hayleyella* show positive chemotaxis toward *D. discoideum* supernatant (Shu, Zhang, et al. 2018).

We present a comparative genomics analysis of the three type 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) vs. *P. hayleyella* and *P. bonniea* (4.1 Mbp) and in GC content (62% vs. 59%) (Brock et al. 2020; Haselkorn et al. 2019). Genome reduction is a pattern associated with long-term host association in many symbiotic bacteria (Moran & Plague 2004; Moran 2002; Maurelli 2007; Bliven & Maurelli 2012; Merhej et

al. 2013, 2009; Toft & Andersson 2010) including pathogenic *Burkholderia* (Nierman et al. 2004). Therefore, we investigated any significant differences in genome characteristics in the genomes of *D. discoideum* symbionts, and particularly in the two reduced genomes of *P. hayleyella* and *P. bonniea*. Based on what we know from the best studied endosymbiont genomes of multicellular animals we look for patterns that frequently accompany host adaptation, such as fewer genes in functional categories related to metabolism, DNA repair, and gene regulation (McCutcheon & Moran 2012; Andersson & Kurland 1998).

Because the ability to infect *D. discoideum* appears to be a shared derived trait among *Paraburkholderia* symbionts of *D. discoideum*, we focus 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 (Brock et al. 2020; Haselkorn et al. 2019), we pay 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 including selection and drift as with the rest of the genome (Abby & Daubin 2007; Arnold et al. 2022; Brockhurst et al. 2019; Liu et al. 2004). In the context of symbiosis, key horizontally transferred genes can enable new symbiotic relationships (e.g. symbiosis islands) (Hacker & Carniel 2001). If host adaptation-induced genome reduction is ongoing, we expect 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) (Ochman & Davalos 2006). 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 (Losada et al. 2010; Manzano-Marín & Latorre 2016), eventually leading to the highly reduced genomes seen in obligate symbionts.

Methods

Paraburkholderia genome selection and gene prediction

Genome sequencing methods were described previously (Brock et al. 2020). 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 (Chin et al. 2013). After an initial round of annotation, we identified the chromosomal replication initiator *dnaA* sequence and *Initiator replication protein* in each assembly's contig and re-oriented each contig from these genes using Circlator (Hunt et al. 2015). We used SMRT analysis software Quiver to repolish each assembly (Chin et al. 2013).

We chose the following *Paraburkholderia* with finished genomes for more detailed comparison: *P. fungorum* strain ATCC BAA-463 (Coenye et al. 2001), originally isolated from the fungus *Phanerochaete chrysosporium* (Seigle-Murandi et al. 1996), *P. sprentiae* strain WSM5005 (De Meyer et al. 2013) isolated from root nodules of the domesticated legume *Lebeckia ambigua*, *P. terrae* strain DSM17804 (Yang et al. 2006) isolated from broad-leaved forest soil, and *P. xenovorans* strain LB400 (Goris et al. 2004) 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 8 additional *Paraburkholderia* genomes that span the clade that includes *P. agricolaris*, *P. hayleyella* and *P. bonniea*. We added 4 plant-associated species genomes (*P. megapolitana* LMG23650, *P. phenoliruptrix* BR3459a, *P. phymatum* STM815, *P. phytofirmans* PsJN) and 4 free-living species genomes (*P. caledonica* PHRS4, *P. phenazinium* LMG2247, *P. sartisoli* LMG24000, and *P. terricola* mHS1) (Vandamme et al. 2007; Coenye et al. 2004; Vandamme et al. 2002; Sessitsch et al. 2005; Viallard et al. 1998; Vanlaere et al. 2008; Goris et al. 2002). All

genomes were downloaded from NCBI and considered complete (Table S1). With the exception of *P. sartisoli* and *P. phenazinium*, all selected genomes are also finished.

We re-annotated each genome with Prokka v1.14.6 (Seemann 2014) 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 (Syberg-Olsen et al. 2021) with DIAMOND v2.0.6.144 (Buchfink et al. 2015) BLAST against the NCBI RefSeq non-redundant 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 (Darling et al. 2010) identifies locally collinear 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 (Krzywinski et al. 2012). We used ggraph v2.0.3 and igraph v1.2.11 in R v3.6.0 (R Core Team 2019) to generate the hive plots.

Horizontally transferred genetic element detection

We used the ISFinder (Siguier et al. 2006) webserver (accessed October 27, 2021) and its nucleotide BLAST to identify putative IS elements to test for their proliferation in each

156 genome. We identified the best hits by comparing overlapping hits by e-value and bit score. We
157 retained hits that were at least 70% coverage of the IS element it matched in the database.
158 Genomic islands are clusters of genes of horizontally transferred origin and have been
159 found in a range of sizes from as small as 5 to as large as 500 kilobases (Dobrindt et al. 2004;
160 Langille et al. 2010; Bertelli et al. 2019). We applied IslandPath-DIMOB and SIGI-HMM as
161 implemented via the IslandViewer 4 webserver (Bertelli et al. 2017). IslandPath-DIMOB uses
162 sequence composition and mobility genes, while SIGI-HMM uses codon usage bias. We then
163 used pairwise reciprocal megablast to determine whether any of the predicted genomic islands
164 were shared among *D. discoideum*-symbiont genomes.
165 Lastly, we looked for individually occurring horizontally transferred genes using
166 DarkHorse2 v2.0_rev09. DarkHorse2 compares individual genes against the NCBI NR database
167 and detects genes with unusual distributions of hits by calculating a lineage probability index
168 (LPI) score (Podell & Gaasterland 2007; Delaye et al. 2020). Vertically inherited genes will have
169 a high LPI score because most high-scoring BLASTP hits will belong to close taxonomic
170 relatives. Horizontally transferred genes are detected because high-scoring BLASTP hits will be
171 taxonomically distant, leading to lower LPI scores. We used DIAMOND to perform BLASTP,
172 then following suggestions from the author excluded self and sister species hits (*P. fungorum* for
173 *P. agricolaris*, *P. bonniea* and *P. hayleyella* from each other), and set the global filter threshold
174 to 0.02 to allow candidate matches to have bit scores up to 2% different from the best non-self
175 match.
176
177 *Gene functional annotation*
178 We performed broad scale functional annotation with COG (Clusters of Orthologous
179 Groups) (Galperin et al. 2015; Tatusov et al. 2000), and KO (Kyoto Encyclopedia of Genes and
180 Genomes (KEGG) Orthology) (Kanehisa, Sato, Kawashima, et al. 2016; Kanehisa et al. 2017).
181 We assigned COG by RPS-BLAST (Altschul et al. 1997) against COG position-specific scoring

matrices downloaded from the NCBI Conserved Domain Database (version July 31, 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 (Huntemann et al. 2015). We assigned KO using the BlastKOALA webserver (http://kegg.jp/blastkoala/; accessed July 13-28, 2020) that performed BLASTP against the KEGG GENES database at the prokaryote Genus and eukaryote Family level (Kanehisa, Sato & Morishima 2016).

We compared functional genome composition in terms of numbers of genes observed in each COG category using agglomerative clustering and non-metric multidimensional scaling (NMDS). Both were implemented in R: agglomerative clustering using cluster v2.1.2, and NMDS using vegan v2.5-7. Both analyses support that 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), proportions of intact genes vs. pseudogenes. To determine which COG categories contribute to this difference, we used a binomial exactTest (McMurdie & Holmes 2014) using edgeR v3.26.8 (Robinson et al. 2010). Because the enrichment of functional categories of genes for the comparison of *P. bonniea* and *P. hayleyella* vs. other *Paraburkholderia* genomes may be due to 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 (post-normalization) and absolute (raw counts) sense, we investigated which specific COGs were contributing to the difference. We used KEGG Mapper (Kanehisa & Sato 2020) and its Reconstruct Pathway tool (https://www.genome.jp/kegg/tool/map_pathway.html; accessed March 14, 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 (Page et al. 2015) 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 (2020) 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 (Yang 2007). Protein sequence multi-fasta files for each core gene were aligned with MUSCLE v3.8.31 (Edgar 2004), then converted into codon alignments using PAL2NAL v14 (Suyama et al. 2006). 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 AIC scores to that of the null hypothesis (H0) that there should be no significant variation in molecular evolution across the species tree. 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 omega (dN/dS) between groups of species.

Essential amino acid biosynthetic repertoire

We used GapMind webserver (http://papers.genomics.lbl.gov/cgi-bin/gapView.cgi; accessed January 26, 2022) to evaluate any loss of essential amino acid biosynthesis pathways in each genome. GapMind detects genes involved in 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 (Price et al. 2020, 2018). It can handle fusion proteins (two enzymes fused into a single protein) and split proteins (multi-domain enzyme split into up to two proteins).

Protein secretion system repertoire and effector prediction

We used TXSScan (Abby et al. 2016) implemented in Galaxy/ Pasteur (accessed October 2, 2020) to identify protein secretion systems in the three focal and 12 additional *Paraburkholderia* genomes. TXSScan identifies protein secretion systems (Types I-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 sub-type, and whether it is contained within a single operon (single locus) or across a few neighboring operons (multi locus). 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 (Winsor et al. 2008).

We classified all T3SS and T6SS found in our 15 genomes. For T3SS, we used the T3Enc database v1.0 (Hu et al. 2017) and downloaded three representative amino acid sequences of thirteen 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 (Edgar 2004) and made gene trees using the Le and Gascuel substitution model with FastTree v2.1.10 (Price et al. 2010). We estimated a species tree from these gene trees using ASTRID v2.2.1 (Vachaspati & Warnow 2015) and ASTRAL v5.7.8 (Zhang et al. 2018). We followed the

same methods for T6SS using the SecReT6 database v3.0 (Li et al. 2015) and the conserved component genes tssB (sheath; COG3516), tssC (sheath; COG3517), and tssF (baseplate; COG3519).

We used VFDB (Virulence factors of Pathogenic Bacteria; accessed January 25, 2022) (Chen et al. 2005) 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 April 20, 2021) to predict secreted effectors. BastionHub (Wang et al. 2021) combines a hidden Markov model based approach and a machine learning approach. Lastly, we used effectiveELD (Jehl et al. 2011; Eichinger et al. 2016) on the effectiveDB server (accessed March 28, 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 (Schmitz-Esser et al. 2010; Gomez-Valero & Buchrieser 2019; Schulz et al. 2016). InterProScan (Jones et al. 2014) webserver (accessed March 29, 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 (Buels et al. 2016; Skinner et al. 2009). 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 (it is currently behind a login while we build out its firewall so please contact S. Noh for access). The GitHub repositories supporting the browser

285 are available at https://github.com/noh-lab/burk-browser and https://github.com/noh-lab/jbrowse-
286 executables.
287
288 *Data and Code Availability*
289 All analyses and figures found in this manuscript can be generated and recreated using input
290 data and code available at the GitHub repository https://github.com/noh-lab/comparative-dicty-
291 symbionts.
292
293 **Results and Discussion**
294 *D. discoideum-symbionts represent two distinct categories, reduced vs. non-reduced size*
295 *genomes*
296 Sequencing using PacBio technology indicated that the genomes of all three *D.*
297 *discoideum*-symbiont species are each comprised of 2 chromosomes, albeit resulting in
298 different total genome sizes. The *P. agricolaris* genome was more than twice the size of both *P.*
299 *bonniea* and *P. hayleyella* (8.7 vs. 4.1 million base pairs). The overall gene content (CDS)
300 comparison was also proportionate, with approximately 7700 genes predicted for *P. agricolaris*
301 as opposed to approximately 3600 genes for the reduced genomes (Table 1). The genome size
302 and gene count of *P. agricolaris* is on par with other the *Paraburkholderia* genomes we
303 examined (Table 1).
304
305 Table 1. Genome statistics of *Paraburkholderia* symbionts of *D. discoideum* and other
306 representative *Paraburkholderia* strains for comparison

	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	18	12	12
tRNA	71	56	63
Locality (host)	*D. discoideum* in Virginia, USA	*D. discoideum* in Virginia, USA	*D. discoideum* in Virginia, USA

307

	***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 Western Australia	Broad-leaved forest soil in South Korea	PCB-contaminated soil in USA

308 *Genes and CDS counts were predicted by Prokka and are inclusive of pseudogenes

309 *Pseudogenes were predicted by Pseudofinder

310

311 Whole genome alignments of all ten finished genomes found 153 locally colinear blocks,
312 ranging in sizes as small as 262 base pairs and as large as 208,252 base pairs in the *P.*
313 *agricolaris* genome. *P. bonniea* and *P. hayleyella* share with each other considerable synteny
314 but also possess a large inverted region relative to each other (Figure 1). Both of these reduced
315 genomes show extensive genome rearrangement compared to the genome of *P. agricolaris*, or
316 any of the other *Paraburkholderia* genomes (Figure 1 & S1). The genome of *P. agricolaris*
317 shares a large degree of synteny with other *Paraburkholderia* genomes in chromosome 1 as
318 indicated by the overall lack of gaps toward the center of each hive plot (Figure S1).

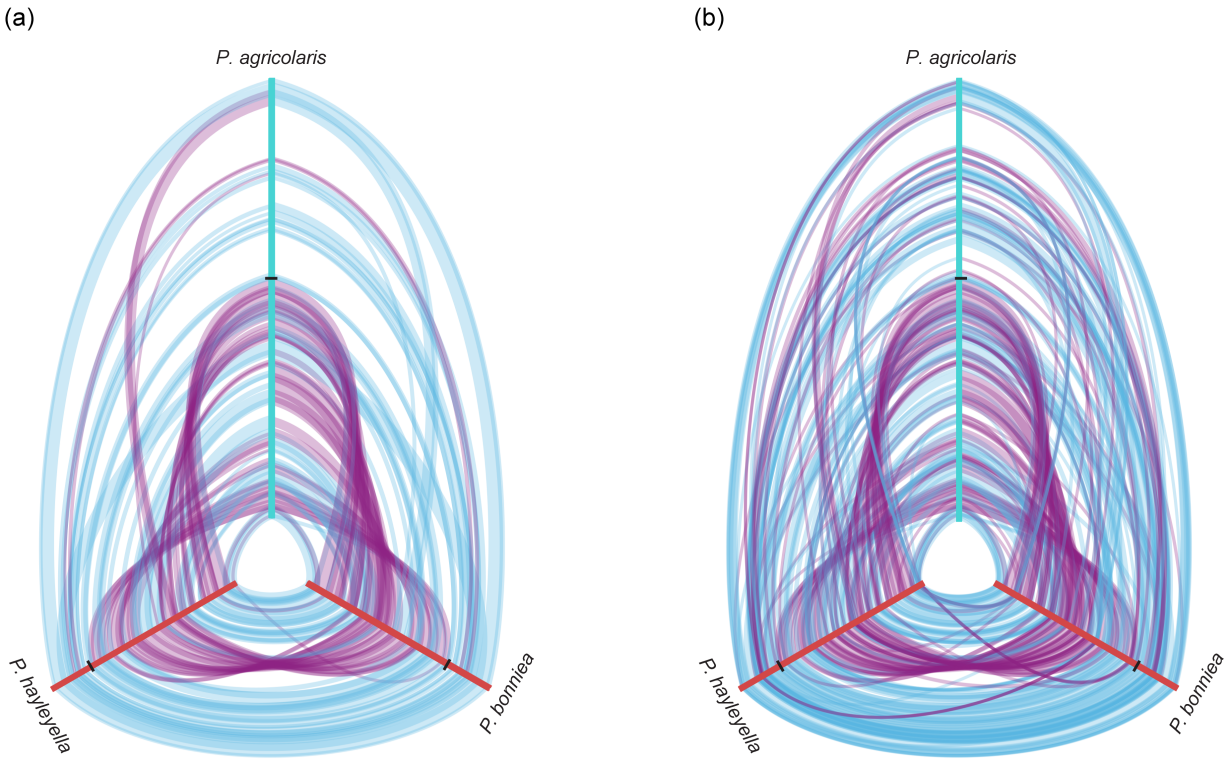


Figure 1. Hive plots of whole genome comparisons of *D. discoideum*-symbiont genomes. Locally colinear blocks between pairs of genomes are shown as bands that connect the axes (genomes). Only blocks above the median size are shown on the left (a) for visual clarity, while all blocks are shown on the right (b). 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.

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) 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 166-218 IS elements, many of which were flanking regions that were randomly lost

among the examined strains in what appears to be ongoing genome reduction (Losada et al. 2010). 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 supports that all three *D. discoideum*-symbiont genomes are relatively stable, and the two reduced genomes are currently not in flux.

The reduced D. discoideum-symbiont genomes show evidence of functional adaptation to the host environment

For each *D. discoideum*-symbiont genome, 65-68 % of genes were annotated with COG (Figure S2), and 53-61 % with KO. The agglomerative clustering and 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* (Figure 2). Further investigation of specific functional differences between the two groups (reduced genomes vs. non-reduced) indicated nine COG categories that were significantly different (exactTest, FDR << 0.01). Of these, four were consistently different in both normalized and raw counts in the same direction (Figure S3). Fewer genes than expected 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). More genes than expected were found in the reduced genomes for Cell motility (N).

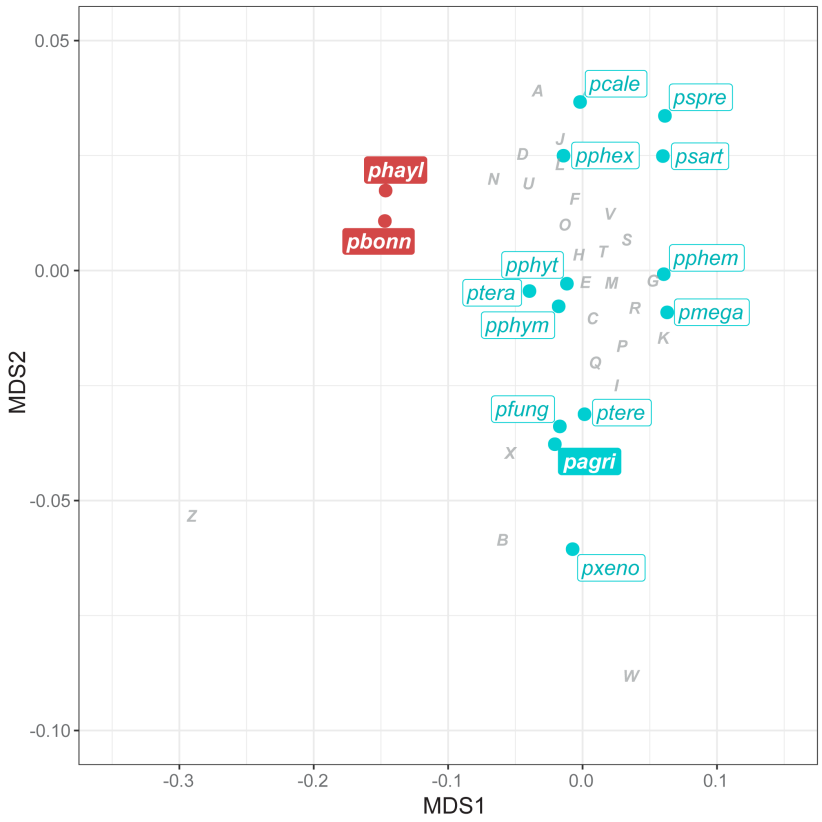


Figure 2. Comparison of reduced (red) and non-reduced (turquoise) genomes in terms of their functional compositions in nonmetric multidimensional space. The contributions of COG categories are projected with minor adjustments to avoid overlap with other features. (pagri = *P. agricolaris*; pbonn = *P. 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*) (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)

We looked within each COG category that was significantly different between reduced and non-reduced 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 (Figure 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 (Ottemann & Miller 1997; Duan et al. 2013). Among *Burkholderia*, *B. pseudomallei* flagella have been shown to be necessary for post-invasion virulence in mice (Chua et al. 2003). *B. pseudomallei* and *B. thailandensis* each have two flagellar clusters, and in *B. thailandensis* the second cryptic cluster is involved in post-invasion intracellular motility (French et al. 2011). We found a second flagellar cluster in *P. bonniea* but not in the other *D. discoideum*-symbiont genomes (see also 'Secretion systems' section).

The other significant COG categories were less abundant in the reduced genomes than expected (Figure 3bcd). Many families of transcriptional regulator COGs were less abundant in the reduced genomes, as is often seen with reduced symbiotic bacterial genomes (Merhej et al. 2013; Wilcox et al. 2003). 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 (Garmory & Titball 2004; Harland et al. 2007).

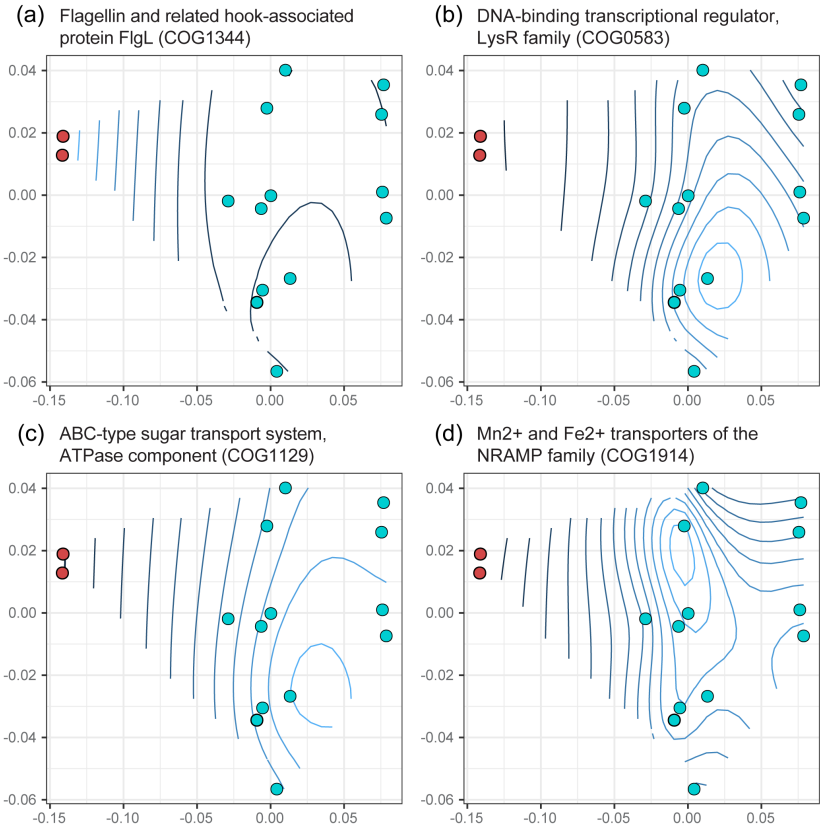


Figure 3. Representative individual COGs belonging to categories (a) Cell motility, (b) Transcription, (c) Carbohydrate transport and metabolism, and (d) Inorganic ion transport and metabolism (P) that were significantly overrepresented or underrepresented in the reduced genomes of *D. discoideum*-symbionts. Contours of abundances are superimposed on the nonmetric multidimensional space from Figure 2. *P. bonniea* and *P. hayleyella* are shown as red points to the left, while *P. agricolaris* is not distinguished from the other genomes in turquoise. Lighter blue contour lines indicate higher abundance compared to darker blue lines.

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 non-reduced

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 (Schaefers 2020). Compared to *P. agricolaris*, the reduced genomes lacked genes encoding nitrate 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 (Mangalea et al. 2017). 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*.

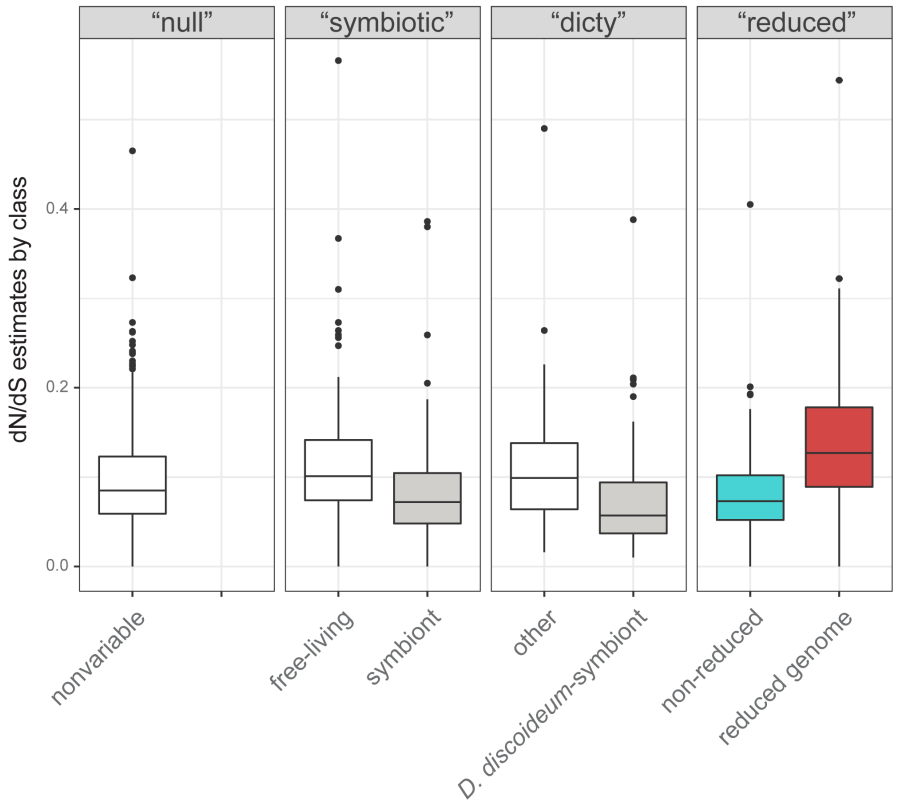
The reduced D. discoideum-symbiont genomes may experience a combination of stronger and relaxed purifying selection relative to other Paraburkholderia genomes

We identified 1673 core genes shared by the 15 *Paraburkholderia* species genomes we investigated (Figure 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 2). 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 $P << 0.01$), and those that show evidence of relaxed selection and significantly higher dN/dS in genomes of reduced size (“reduced”; Wilcoxon test $P << 0.01$) (Figure 4). These results indicate that the reduced genomes of *P. bonniea* and *P. hayleyella* possess a combination of genes experiencing

435 stronger selective constraints and genes under weaker selective constraints relative to the
436 genomes of other *Paraburkholderia*. This pattern is in contrast to genomes of obligate
437 symbionts where the majority of genes are experiencing genetic drift and weaker selection
438 constraints across their entire genomes (Sabater-Muñoz et al. 2017; Wernegreen 2017).
439
440 Table 2. Hypotheses tested regarding molecular evolution in the 1673 core genes shared
441 across 15 Paraburkholderia genomes.

Hypothesis	Number 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 vs. symbiotically associated with a eukaryotic host (2 rate ratios)
"dicty"	137	Selection pressure is different when species are unassociated vs. 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)

442



443

Figure 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").

The three *D. discoideum*-symbiont genomes shared 1977 genes total, including the 1673 core genes (Figure 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 'Secretion systems' section), *bhuRSTUV* genes, and helix-turn-helix motif-containing GntR and LysR transcriptional regulators. *Bordatella* heme utilization (*bhu*) genes are virulence factors in mammalian and avian host infection (Murphy et al. 2002; Vanderpool & Armstrong 2001), and transcriptional regulators with helix-turn-helix motifs have been frequently associated with virulence in pathogens (Finlay & Falkow 1997).

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 (Moran et al. 2008; Lo et al. 2016; McCutcheon et al. 2019). However nutrient-dependent relationships are less likely in protist-prokaryote symbioses because protist host diets tend to be much more diverse compared to multicellular eukaryotes (Husnik et al. 2021). 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 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 suggest it is unlikely that a recent horizontal gene transfer event substantially

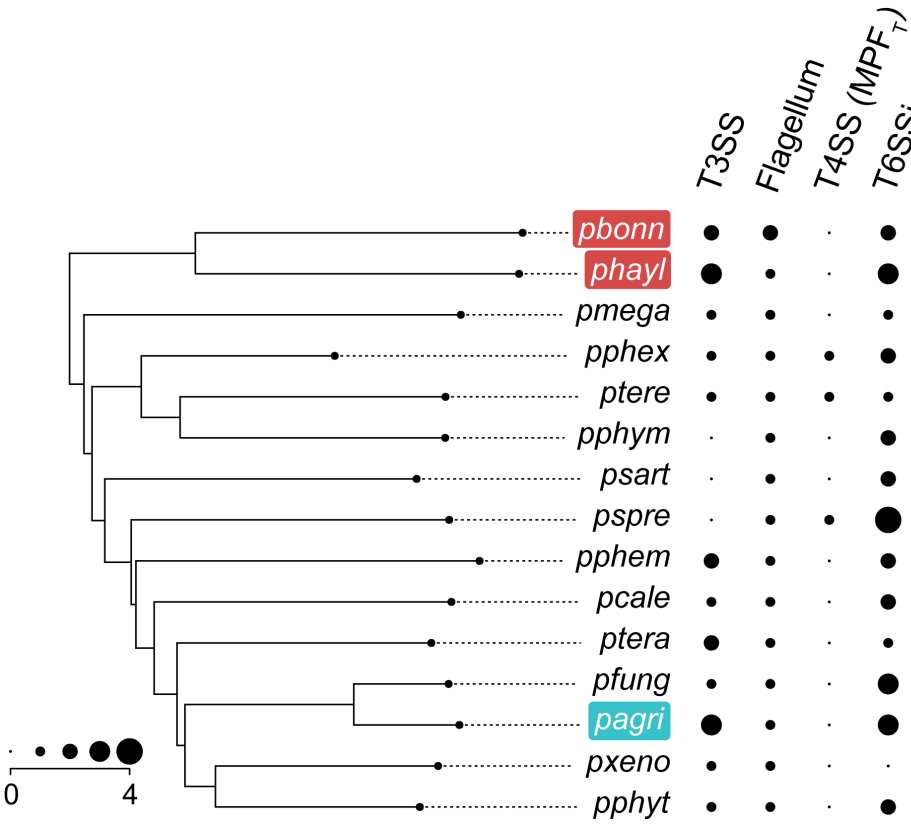
contributed to the shared ability of these symbionts to persistently infect *D. discoideum*. If such an event had occurred, any such genes seem to have experienced amelioration over evolutionary time and cannot easily be distinguished from the rest of the genome (Lawrence & Ochman 1997).

Shared secretion systems may mediate D. discoideum-Paraburkholderia symbiont interactions

Bacterial secretion systems are frequently implicated in host-symbiont interactions (Tseng et al. 2009; Coombes 2009). 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 (Figure 5; Table S4). Classification of T3SS showed that one specific T3SS operon shared among *D. discoideum*-symbionts falls into category 8 T3SS (Figure 6 & S4). This category of T3SS also includes BurBor found in the plant pathogen *Robbsia* (previously *Burkholderia*) *andropogonis* (Mannaa et al. 2019), as well as *Bordatella* species that include mammalian pathogens (Kamanova 2020). In addition, one specific T6SS operon is shared among *D. discoideum*-symbiont genomes and belongs to category i1 (Figure 7 & S5). More importantly, this T6SS operon clusters together with the virulence-causing T6SS-5 operon found in *Burkholderia mallei*, *B. pseudomallei*, and *B. thailandensis* (Lennings et al. 2019). *B. mallei* causes glanders disease and is an obligate pathogen that evolved from an ancestor shared with melioidosis-causing soil bacterium *B. pseudomallei* (Schell et al. 2007; Burtnick et al. 2011; Losada et al. 2010). *B. thailandensis* is sister species to the other two, and is a facultative pathogen similar to *B. pseudomallei* but with much lower clinical virulence (Lennings et al. 2019).

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

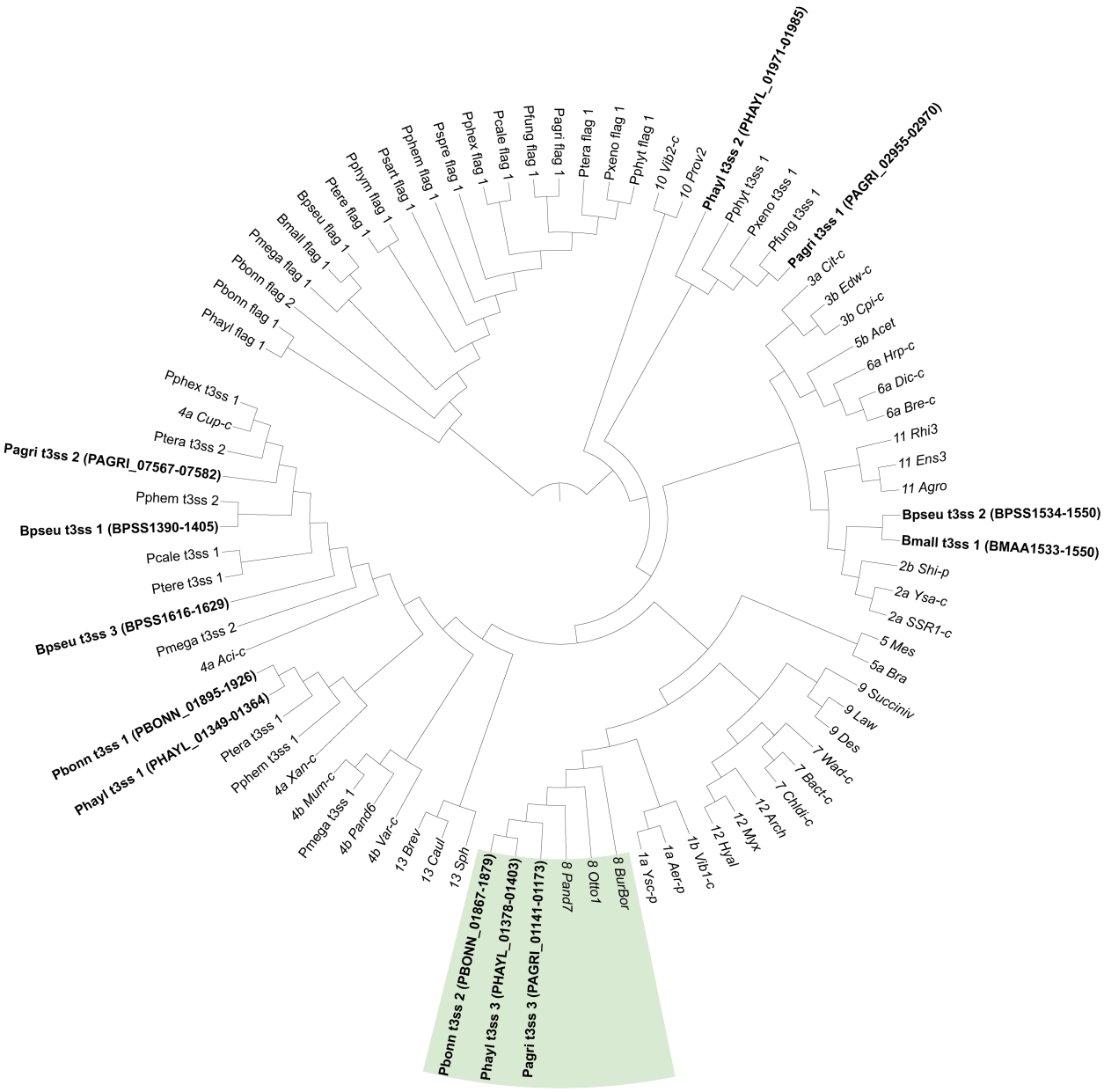
522 BurBor-like T3SS found in the *D. discoideum*-symbiont genomes. It is worth noting that the two
523 adjacent T3SS-3 and T6SS-5 operons in *B. pseudomallei* and *B. thailandensis* have been
524 shown to be functionally linked and necessary for virulence, with the T3SS-3 effectors regulating
525 the expression of the adjacent T6SS-5 (Sun et al. 2010; Chen et al. 2011; Schwarz et al. 2010;
526 French et al. 2011).



527

528 Figure 5. The abundances of secretion systems detected in *D. discoideum*-symbiont genomes
529 and other *Paraburkholderia*. For the Type 4 Secretion System, only protein secretion (as
530 opposed to conjugation-related) T4SS abundances are shown. The phylogeny is a species tree
531 based on Brock et al 2020. (pagri = *P. agricolaris*; pbonn = *P. bonniea*; phayl = *P. hayleyella*;
532 pcale = *P. caledonica*; pfung = *P. fungorum*, pmega = *P. megapolitana*, pphem = *P.*
533 *phenazinium*; pphex = *P. phenoliruptrix*; pphym = *P. phymatum*; pphyt = *P. phytofirmans*; psart
534 = *P. sartisoli*; pspre = *P. sprentiae*; ptera = *P. terricola*; ptere = *P. terrae*; pxeno = *P.*
535 *xenovorans*).

536



537

538 Figure 6. Type 3 secretion systems and flagella categorized using the conserved component
539 genes sctJ (inner membrane ring; IPR003282), sctN (ATPase; IPR005714), and sctV (export
540 apparatus; IPR006302). Branch lengths were ignored to improve readability of the ASTRAL tree
541 topology. T3SS categories precede the name of the operon (e.g. “8 Pand7” is operon Pand7
542 belonging to category 8), downloaded from T3Enc database v1.0 (Hu et al. 2017). Tip labels for
543 T3SS in the three *D. discoideum*-symbiont genomes and in *B. mallei* and *B. pseudomallei* are

shown in bold font face with gene IDs for ease of reference. The clade containing the shared T3SS operon is shaded.

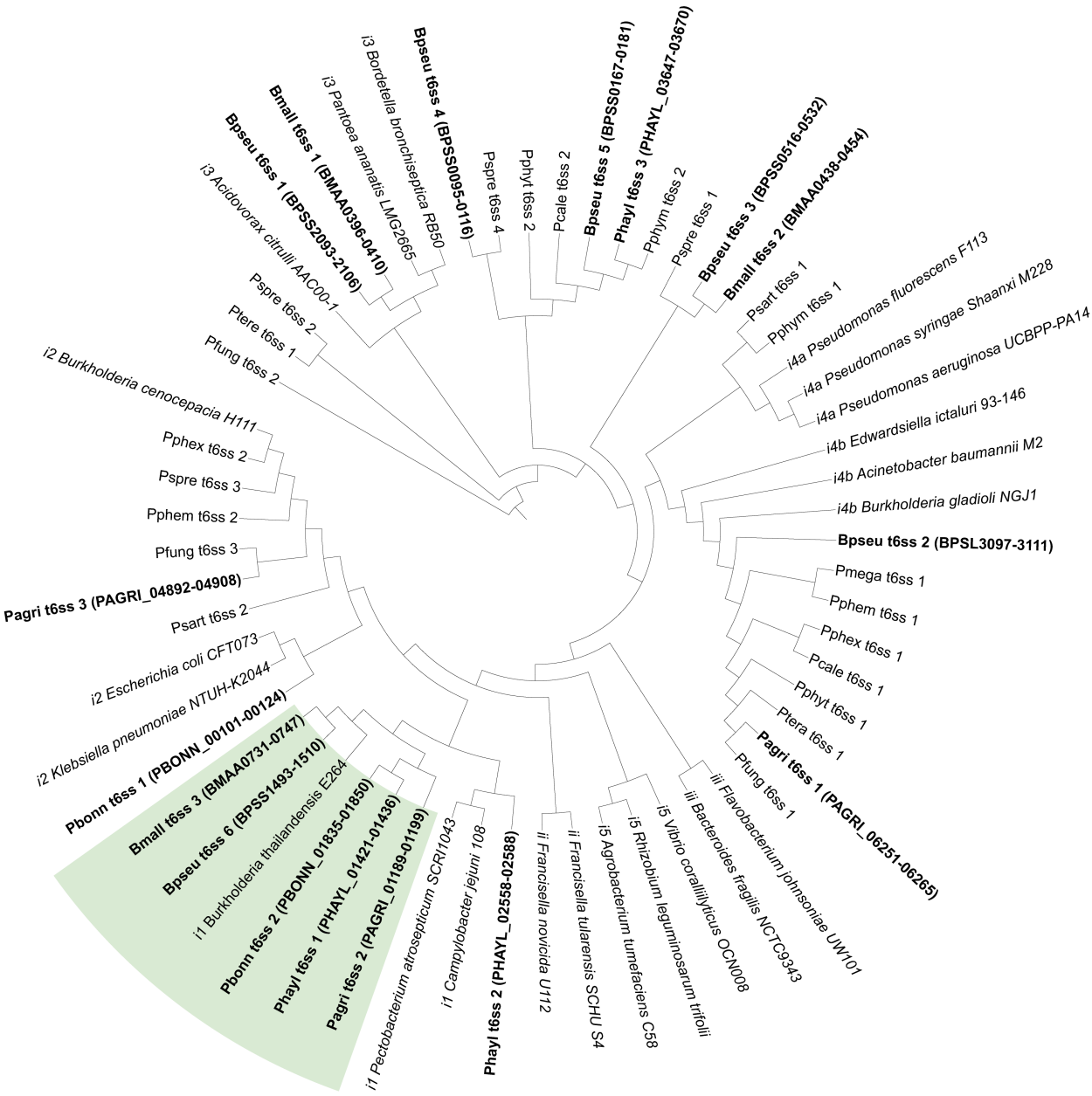


Figure 7. Type 6 secretion systems categorized using the conserved component genes tssB (sheath; COG3516), tssC (sheath; COG3517), and tssF (baseplate; COG3519). Branch lengths were ignored to improve readability of the ASTRAL tree topology. T6SS categories precede the name of the strain to which the operon belongs (e.g. “ii Francisella novicida U112” is belongs to

552 category ii), downloaded from SecReT6 database v3.0 (Li et al. 2015). T6SS in the three *D.*
553 *discoideum*-symbiont genomes and in *B. mallei* and *B. pseudomallei* are shown in bold font face
554 with gene IDs for ease of reference. The clade containing the shared T6SS operon is shaded.
555
556 We attempted to identify effector proteins that might be functionally linked to these *D.*
557 *discoideum*-symbiont secretion systems (Table S5). We identified homologs of the T6SS
558 effector VgrG-5 that would likely be associated with the shared T6SS-5-like operon (Table S5).
559 Unexpectedly, VgrG-5 in *P. agricolaris* (gene ID PAGRI_01155) is a homolog but not an
560 ortholog to VgrG-5 in the two reduced genomes (PBONN_01842 and PHAYL_01429). This
561 suggests the possibility of two independent evolutionary origins of this T6SS effector, and
562 potentially different functional roles. In *Burkholderia thailandensis*, VgrG-5 is necessary for post-
563 infection cell-to-cell spread within mammalian hosts (Schwarz et al. 2014). For each genome,
564 we predicted additional secretion system effectors, including a chaperonin ClpB and a
565 sodium/solute symporter for *P. agricolaris*, and RHS (rearrangement hotspot) proteins that may
566 mediate contact-dependent growth inhibition during bacterial competition for *P. hayleyella*.
567 Lastly, we predicted secreted effectors containing eukaryotic domains specific to our Pfam clans
568 of interest (Ank, TPR, LRR, Pentapeptide, F-box, and RING). Previous investigations of amoeba
569 symbiont genomes have observed enrichment of proteins possessing these domains that
570 hypothetically mediate physiological interactions with a eukaryotic host (Schmitz-Esser et al.
571 2010; Gomez-Valero & Buchrieser 2019; Schulz et al. 2016). Notably, two proteins each directly
572 adjacent to VgrG-5 in *P. agricolaris* (PAGRI_01156-7) and *P. bonniea* (PBONN_01840-1) each
573 contained pentapeptide repeat domains. InterProScan searches indicated that two proteins in *P.*
574 *hayleyella* (PHAYL_01430-1) adjacent to VgrG-5 also contain pentapeptide repeat domains.
575 However, no known functions are predicted for these protein pairs.
576
577 **Conclusion**

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 find evidence that relative to the other *Paraburkholderia* genomes we investigated, all three *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 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 (Wallner et al. 2021; Mannaa et al. 2019). 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 lifecycle of these pathogenic *Burkholderia* (Schwarz et al. 2014). 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 post-infection. We also find 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 any of the other *Paraburkholderia* genomes we compared. It is intriguing to consider the possibility that these genes were transferred among symbiont genomes within the *D. discoideum* amoeba host environment. Different *D. discoideum*-symbiont species have been found coinfecting amoeba hosts (Haselkorn et al. 2019), and diverse *Acanthamoeba* symbionts appear to share genes with each other that are functionally enriched for host interaction (Wang & Wu 2017).

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 two reduced genomes of

604 *P. bonniea* and *P. hayleyella* display characteristics that support their evolution in a host
605 environment. All three species retain the ability to live outside of *D. discoideum*, but the
606 genomes of *P. bonniea* and *P. hayleyella* show fewer transcriptional regulators, as well as fewer
607 carbohydrate and inorganic ion transporters. The reduced genomes possess a combination of
608 genes with molecular evolution patterns that indicate specific responses to the host environment
609 (both stronger and weaker evolutionary constraints) rather than uniform deterioration under
610 genetic drift. In addition, the lack of IS element proliferation and absence of excessive
611 pseudogene accumulation compared to other *Paraburkholderia* genomes indicate that these
612 already reduced genomes are relatively stable.

613 These combined pieces of evidence supports that the reduced genome *D. discoideum*-
614 symbionts are professional symbionts specifically adapted to their protists hosts. We adopt the
615 term “professional symbiont” from Husnik and colleages (2021) to refer to symbiont lineages
616 that are ancestrally adapted to their specific hosts, that possess compact and streamlined
617 genomes. Accordingly, we hypothesize that the symbiotic relationship between *D. discoideum*
618 and the species with reduced genomes is persistent and potentially quite old. However, given
619 the short generation time of protists it is entirely possible that what we call “old” is not as ancient
620 as the symbioses and similarly stable stages of genome reduction observed in microbial
621 symbionts of multicellular eukaryotes. In contrast to *P. bonniea* or *P. hayleyella*, intraspecific
622 genetic variation appears to be larger for *P. agricolaris* (Haselkorn et al. 2019), suggesting that
623 *P. agricolaris* host adaptation may be ongoing and more dynamic. We look forward to
624 expanding these analyses to a larger collection of *D. discoideum*-symbiont genomes in the
625 future, to identify both convergent and divergent host adaptation patterns among *D. discoideum*-
626 symbionts and to continue to add to a growing body of work across diverse protist-prokaryote
627 symbioses.
628
629

Acknowledgements

We thank students and colleagues from the Noh lab (Anna Chen, Kayla Dixon, Laura Drepanos) and Strassmann-Queller lab (Tammy Haselkorn, Clarissa Dzikunu) who contributed to aspects of this project during its development. This work is supported by the National Science Foundation under Grant Numbers IOS 1656756 and DEB 1753743 (JS and DQ), and by the National Institutes of Health and its National Institute of General Medical Sciences by an Institutional Development Award (IDeA) under Grant Numbers P20GM103423 (subaward to SN) and Colby College startup funds (SN).

Reference

Abby S, Daubin V. 2007. Comparative genomics and the evolution of prokaryotes. Trends in Microbiology. 15:135–141. doi: 10.1016/j.tim.2007.01.007.

Abby SS et al. 2016. Identification of protein secretion systems in bacterial genomes. Sci Rep. 6:23080. doi: 10.1038/srep23080.

Altschul SF et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25:3389–3402. doi: 10.1093/nar/25.17.3389.

Andersson SGE, Kurland CG. 1998. Reductive evolution of resident genomes. Trends in Microbiology. 6:263–268. doi: 10.1016/S0966-842X(98)01312-2.

Arnold BJ, Huang I-T, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria. Nat Rev Microbiol. 20:206–218. doi: 10.1038/s41579-021-00650-4.

Bertelli C et al. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Research. 45:W30–W35. doi: 10.1093/nar/gkx343.

Bertelli C, Tilley KE, Brinkman FSL. 2019. Microbial genomic island discovery, visualization and analysis. Brief Bioinform. 20:1685–1698. doi: 10.1093/bib/bby042.

Bliven KA, Maurelli AT. 2012. Antivirulence genes: insights into pathogen evolution through gene loss. Infection and Immunity. 80:4061–4070. doi: 10.1128/IAI.00740-12.

Bozzaro S, Eichinger L. 2011. The professional phagocyte *Dictyostelium discoideum* as a model host for bacterial pathogens. Curr Drug Targets. 12:942–954. doi: 10.2174/138945011795677782.

660 Brock DA et al. 2020. Endosymbiotic adaptations in three new bacterial species associated with
661 *Dictyostelium discoideum*: *Paraburkholderia agricolaris* sp. nov., *Paraburkholderia hayleyella*
662 sp. nov., and *Paraburkholderia bonniea* sp. nov. PeerJ. 8:e9151. doi: 10.7717/peerj.9151.

663 Brock DA, Douglas TE, Queller DC, Strassmann JE. 2011. Primitive agriculture in a social
664 amoeba. Nature. 469:393–396. doi: 10.1038/nature09668.

665 Brock DA, Read S, Bozhchenko A, Queller DC, Strassmann JE. 2013. Social amoeba farmers
666 carry defensive symbionts to protect and privatize their crops. Nature Communications. 4:2385.
667 doi: 10.1038/ncomms3385.

668 Brockhurst MA et al. 2019. The ecology and evolution of pangenomes. Current Biology.
669 29:R1094–R1103. doi: 10.1016/j.cub.2019.08.012.

670 Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND.
671 Nat Methods. 12:59–60. doi: 10.1038/nmeth.3176.

672 Buels R et al. 2016. JBrowse: a dynamic web platform for genome visualization and analysis.
673 Genome Biol. 17:66. doi: 10.1186/s13059-016-0924-1.

674 Burtnick MN et al. 2011. The cluster 1 type VI secretion system Is a major virulence determinant
675 in *Burkholderia pseudomallei*. Infection and Immunity. 79:1512–1525. doi: 10.1128/IAI.01218-
676 10.

677 Chen L et al. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids
678 Research. 33:D325–D328. doi: 10.1093/nar/gki008.

679 Chen Y et al. 2011. Regulation of type VI secretion system during *Burkholderia pseudomallei*
680 infection. Infection and Immunity. 79:3064–3073. doi: 10.1128/IAI.05148-11.

681 Chin C-S et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT
682 sequencing data. Nat Methods. 10:563–569. doi: 10.1038/nmeth.2474.

683 Chua KL, Chan YY, Gan YH. 2003. Flagella are virulence determinants of *Burkholderia*
684 *pseudomallei*. Infection and Immunity. 71:1622–1629. doi: 10.1128/IAI.71.4.1622-1629.2003.

685 Coenye T et al. 2001. *Burkholderia fungorum* sp. nov. and *Burkholderia caledonica* sp. nov.,
686 two new species isolated from the environment, animals and human clinical samples. Int. J. Syst.
687 Evol. Microbiol. 51:1099–1107. doi: 10.1099/00207713-51-3-1099.

688 Coenye T, Henry D, Speert DP, Vandamme P. 2004. *Burkholderia phenoliruptrix* sp. nov., to
689 accommodate the 2,4,5-trichlorophenoxyacetic acid and halophenol-degrading strain AC1100.
690 Systematic and Applied Microbiology. 27:623–627. doi: 10.1078/0723202042369992.

691 Coombes BK. 2009. Type III secretion systems in symbiotic adaptation of pathogenic and non-
692 pathogenic bacteria. Trends in Microbiology. 17:89–94. doi: 10.1016/j.tim.2008.11.006.

693 Cosson P, Soldati T. 2008. Eat, kill or die: when amoeba meets bacteria. Current Opinion in
694 Microbiology. 11:271–276. doi: 10.1016/j.mib.2008.05.005.

695 Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene
696 gain, loss and rearrangement. PLOS ONE. 5:e11147. doi: 10.1371/journal.pone.0011147.

697 De Meyer SE et al. 2013. *Burkholderia sprentiae* sp. nov., isolated from *Lebeckia ambigua* root
698 nodules. International Journal of Systematic and Evolutionary Microbiology. 63:3950–3957. doi:
699 10.1099/ijs.0.048777-0.

700 Delaye L, Vargas C, Latorre A, Moya A. 2020. Inferring horizontal gene transfer with
701 DarkHorse, Phylomizer, and ETE Toolkits. Methods Mol Biol. 2075:355–369. doi: 10.1007/978-
702 1-4939-9877-7_25.

703 DiSalvo S et al. 2015. *Burkholderia* bacteria infectiously induce the proto-farming symbiosis of
704 *Dictyostelium* amoebae and food bacteria. PNAS. 112:E5029–E5037. doi:
705 10.1073/pnas.1511878112.

706 Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and
707 environmental microorganisms. Nat Rev Microbiol. 2:414–424. doi: 10.1038/nrmicro884.

708 Drew GC, Stevens EJ, King KC. 2021. Microbial evolution and transitions along the parasite–
709 mutualist continuum. Nat Rev Microbiol. 19:623–638. doi: 10.1038/s41579-021-00550-7.

710 Duan Q, Zhou M, Zhu L, Zhu G. 2013. Flagella and bacterial pathogenicity. Journal of Basic
711 Microbiology. 53:1–8. doi: 10.1002/jobm.201100335.

712 Dunn JD et al. 2017. Eat prey, live: *Dictyostelium discoideum* as a model for cell-autonomous
713 defenses. Front Immunol. 8:1906. doi: 10.3389/fimmu.2017.01906.

714 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
715 throughput. Nucleic Acids Research. 32:1792–1797. doi: 10.1093/nar/gkh340.

716 Eichinger V et al. 2016. EffectiveDB—updates and novel features for a better annotation of
717 bacterial secreted proteins and Type III, IV, VI secretion systems. Nucleic Acids Research.
718 44:D669–D674. doi: 10.1093/nar/gkv1269.

719 Finlay BB, Falkow S. 1997. Common themes in microbial pathogenicity revisited. Microbiology
720 and Molecular Biology Reviews. 61:136–169. doi: 10.1128/mmbr.61.2.136-169.1997.

721 French CT et al. 2011. Dissection of the *Burkholderia* intracellular life cycle using a
722 photothermal nanoblade. Proceedings of the National Academy of Sciences. 108:12095–12100.
723 doi: 10.1073/pnas.1107183108.

724 Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage
725 and improved protein family annotation in the COG database. Nucleic Acids Research.
726 43:D261–D269. doi: 10.1093/nar/gku1223.

727 Garcia JR, Larsen TJ, Queller DC, Strassmann JE. 2019. Fitness costs and benefits vary for two
728 facultative *Burkholderia* symbionts of the social amoeba, *Dictyostelium discoideum*. Ecology
729 and Evolution. 9:9878–9890. doi: 10.1002/ece3.5529.

730 Garmory HS, Titball RW. 2004. ATP-binding cassette transporters are targets for the
731 development of antibacterial vaccines and therapies. Infection and Immunity. 72:6757–6763.
732 doi: 10.1128/IAI.72.12.6757-6763.2004.

733 Gomez-Valero L, Buchrieser C. 2019. Intracellular parasitism, the driving force of evolution of
734 *Legionella pneumophila* and the genus *Legionella*. Microbes and Infection. 21:230–236. doi:
735 10.1016/j.micinf.2019.06.012.

736 Goris J et al. 2004. Classification of the biphenyl- and polychlorinated biphenyl-degrading strain
737 LB400T and relatives as *Burkholderia xenovorans* sp. nov. International Journal of Systematic
738 and Evolutionary Microbiology. 54:1677–1681. doi: 10.1099/ijs.0.63101-0.

739 Goris J et al. 2002. Diversity of transconjugants that acquired plasmid pJP4 or pEMT1 after
740 inoculation of a donor strain in the A- and B-horizon of an agricultural soil and description of
741 *Burkholderia hospita* sp. nov. and *Burkholderia terricola* sp. nov. Systematic and Applied
742 Microbiology. 25:340–352. doi: 10.1078/0723-2020-00134.

743 Hacker J, Carniel E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity.
744 EMBO reports. 2:376–381. doi: 10.1093/embo-reports/kve097.

745 Harland DN, Dassa E, Titball RW, Brown KA, Atkins HS. 2007. ATP-binding cassette systems
746 in *Burkholderia pseudomallei* and *Burkholderia mallei*. BMC Genomics. 8:83. doi:
747 10.1186/1471-2164-8-83.

748 Haselkorn TS et al. 2019. The specificity of *Burkholderia* symbionts in the social amoeba
749 farming symbiosis: prevalence, species, genetic and phenotypic diversity. Molecular Ecology.
750 28:847–862. doi: 10.1111/mec.14982.

751 Hentschel U. 2021. Harnessing the power of host–microbe symbioses to address grand
752 challenges. Nat Rev Microbiol. 19:615–616. doi: 10.1038/s41579-021-00619-3.

753 Hu Y et al. 2017. A global survey of bacterial type III secretion systems and their effectors.
754 Environmental Microbiology. 19:3879–3895. doi: 10.1111/1462-2920.13755.

755 Hunt M et al. 2015. Circlator: automated circularization of genome assemblies using long
756 sequencing reads. Genome Biology. 16:294. doi: 10.1186/s13059-015-0849-0.

757 Huntemann M et al. 2015. The standard operating procedure of the DOE-JGI Microbial Genome
758 Annotation Pipeline (MGAP v.4). Standards in Genomic Sciences. 10:86. doi: 10.1186/s40793-
759 015-0077-y.

760 Husnik F et al. 2021. Bacterial and archaeal symbioses with protists. Current Biology. 31:R862–
761 R877. doi: 10.1016/j.cub.2021.05.049.

762 Jehl M-A, Arnold R, Rattei T. 2011. Effective—a database of predicted secreted bacterial
763 proteins. Nucleic Acids Research. 39:D591–D595. doi: 10.1093/nar/gkq1154.

764 Jones P et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics.
765 30:1236–1240. doi: 10.1093/bioinformatics/btu031.

766 Kamanova J. 2020. Bordetella type III secretion injectosome and effector proteins. Frontiers in
767 Cellular and Infection Microbiology. 10:466.

768 Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on
769 genomes, pathways, diseases and drugs. Nucleic Acids Research. 45:D353–D361. doi:
770 10.1093/nar/gkw1092.

771 Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from protein
772 sequences. Protein Science. 29:28–35. doi: 10.1002/pro.3711.

773 Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference
774 resource for gene and protein annotation. Nucleic Acids Research. 44:D457–D462. doi:
775 10.1093/nar/gkv1070.

776 Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
777 Functional Characterization of Genome and Metagenome Sequences. Journal of Molecular
778 Biology. 428:726–731. doi: 10.1016/j.jmb.2015.11.006.

779 Krzywinski M, Birol I, Jones SJ, Marra MA. 2012. Hive plots—rational approach to visualizing
780 networks. Briefings in Bioinformatics. 13:627–644. doi: 10.1093/bib/bbr069.

781 Langille MGI, Hsiao WWL, Brinkman FSL. 2010. Detecting genomic islands using
782 bioinformatics approaches. Nat Rev Microbiol. 8:373–382. doi: 10.1038/nrmicro2350.

783 Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and
784 exchange. J Mol Evol. 44:383–397. doi: 10.1007/pl00006158.

785 Lennings J, West TE, Schwarz S. 2019. The *Burkholderia* type VI secretion system 5:
786 composition, regulation and role in virulence. Frontiers in Microbiology. 9:3339.

787 Li J et al. 2015. SecReT6: a web-based resource for type VI secretion systems found in bacteria.
788 Environmental Microbiology. 17:2196–2202. doi: 10.1111/1462-2920.12794.

789 Liu Y, Harrison PM, Kunin V, Gerstein M. 2004. Comprehensive analysis of pseudogenes in
790 prokaryotes: widespread gene decay and failure of putative horizontally transferred genes.
791 Genome Biology. 5:R64. doi: 10.1186/gb-2004-5-9-r64.

792 Lo W-S, Huang Y-Y, Kuo C-H. 2016. Winding paths to simplicity: genome evolution in
793 facultative insect symbionts. FEMS Microbiology Reviews. 40:855–874. doi:
794 10.1093/femsre/fuw028.

795 Losada L et al. 2010. Continuing evolution of *Burkholderia mallei* through genome reduction
796 and large-scale rearrangements. Genome Biology and Evolution. 2:102–116. doi:
797 10.1093/gbe/evq003.

798 Mangalea MR, Plumley BA, Borlee BR. 2017. Nitrate sensing and metabolism inhibit biofilm
799 formation in the opportunistic pathogen *Burkholderia pseudomallei* by reducing the intracellular
800 concentration of c-di-GMP. Frontiers in Microbiology. 8.
801 https://www.frontiersin.org/article/10.3389/fmicb.2017.01353 (Accessed March 17, 2022).

802 Mannaa M, Park I, Seo Y-S. 2019. Genomic features and insights into the taxonomy, virulence,
803 and benevolence of plant-associated *Burkholderia* species. International Journal of Molecular
804 Sciences. 20:121. doi: 10.3390/ijms20010121.

805 Manzano-Marín A, Latorre A. 2016. Snapshots of a shrinking partner: genome reduction in
806 Serratia symbiotica. Scientific Reports. 6:32590. doi: 10.1038/srep32590.

807 Maurelli AT. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of
808 bacterial pathogens. FEMS Microbiology Letters. 267:1–8. doi: 10.1111/j.1574-
809 6968.2006.00526.x.

810 McCutcheon JP, Boyd BM, Dale C. 2019. The life of an insect endosymbiont from the cradle to
811 the grave. Current Biology. 29:R485–R495. doi: 10.1016/j.cub.2019.03.032.

812 McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev
813 Microbiol. 10:13–26. doi: 10.1038/nrmicro2670.

814 McMurdie PJ, Holmes S. 2014. Waste not, want not: why rarefying microbiome data is
815 inadmissible. PLOS Computational Biology. 10:e1003531. doi: 10.1371/journal.pcbi.1003531.

816 Merhej V, Georgiades K, Raoult D. 2013. Postgenomic analysis of bacterial pathogens repertoire
817 reveals genome reduction rather than virulence factors. Briefings in Functional Genomics.
818 12:291–304. doi: 10.1093/bfgp/elt015.

819 Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. 2009. Massive comparative genomic
820 analysis reveals convergent evolution of specialized bacteria. Biology Direct. 4:13. doi:
821 10.1186/1745-6150-4-13.

822 Miller JW, Bocke CR, Tresslar AR, Schniepp EM, DiSalvo S. 2020. *Paraburkholderia*
823 symbionts display variable infection patterns that are not predictive of amoeba host outcomes.
824 Genes. 11:674. doi: 10.3390/genes11060674.

825 Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell.
826 108:583–586. doi: 10.1016/S0092-8674(02)00665-7.

827 Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial
828 symbionts. Annual Review of Genetics. 42:165–190. doi:
829 10.1146/annurev.genet.41.110306.130119.

Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr. Opin. Genet. Dev. 14:627–633. doi: 10.1016/j.gde.2004.09.003.

Murphy ER et al. 2002. BhuR, a virulence-associated outer membrane protein of *Bordetella avium*, is required for the acquisition of iron from heme and hemoproteins. Infection and Immunity. 70:5390–5403. doi: 10.1128/IAI.70.10.5390-5403.2002.

Nierman WC et al. 2004. Structural flexibility in the *Burkholderia mallei* genome. Proceedings of the National Academy of Sciences. 101:14246–14251. doi: 10.1073/pnas.0403306101.

Ochman H, Davalos LM. 2006. The nature and dynamics of bacterial genomes. Science. 311:1730–1733. doi: 10.1126/science.1119966.

Ottemann KM, Miller JF. 1997. Roles for motility in bacterial–host interactions. Molecular Microbiology. 24:1109–1117. doi: 10.1046/j.1365-2958.1997.4281787.x.

Page AJ et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31:3691–3693. doi: 10.1093/bioinformatics/btv421.

Podell S, Gaasterland T. 2007. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. Genome Biology. 8:R16. doi: 10.1186/gb-2007-8-2-r16.

Price MN et al. 2018. Filling gaps in bacterial amino acid biosynthesis pathways with high-throughput genetics. PLOS Genetics. 14:e1007147. doi: 10.1371/journal.pgen.1007147.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE. 5:e9490. doi: 10.1371/journal.pone.0009490.

Price MN, Deutschbauer AM, Arkin AP. 2020. GapMind: automated annotation of amino acid biosynthesis. mSystems. 5:e00291-20. doi: 10.1128/mSystems.00291-20.

R Core Team. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139–140. doi: 10.1093/bioinformatics/btp616.

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. doi: 10.1038/ismej.2017.18.

Schaefers MM. 2020. Regulation of virulence by two-component systems in pathogenic *Burkholderia*. Infection and Immunity. 88:e00927-19. doi: 10.1128/IAI.00927-19.

Schell MA et al. 2007. Type VI secretion is a major virulence determinant in *Burkholderia mallei*. Mol Microbiol. 64:1466–1485. doi: 10.1111/j.1365-2958.2007.05734.x.

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. Journal of Bacteriology. 192:1045–1057. doi: 10.1128/JB.01379-09.

Schulz F et al. 2016. A Rickettsiales symbiont of amoebae with ancient features. Environmental Microbiology. 18:2326–2342. doi: 10.1111/1462-2920.12881.

Schwarz S et al. 2010. Burkholderia type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. PLOS Pathogens. 6:e1001068. doi: 10.1371/journal.ppat.1001068.

Schwarz S et al. 2014. VgrG-5 Is a *Burkholderia* type VI secretion system-exported protein required for multinucleated giant cell formation and virulence. Infection and Immunity. 82:1445–1452. doi: 10.1128/IAI.01368-13.

Scott TJ, Queller DC, Strassmann JE. in press. Context dependence in the symbiosis between *Dictyostelium discoideum* and *Paraburkholderia*. Evolution Letters. n/a. doi: 10.1002/evl3.281.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068–2069. doi: 10.1093/bioinformatics/btu153.

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. doi: 10.1128/aem.62.7.2477-2481.1996.

Sessitsch A et al. 2005. *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. International Journal of Systematic and Evolutionary Microbiology. 55:1187–1192. doi: 10.1099/ijs.0.63149-0.

Shu L, Brock DA, et al. 2018. Symbiont location, host fitness, and possible coadaptation in a symbiosis between social amoebae and bacteria Ruzzini, T, editor. eLife. 7:e42660. doi: 10.7554/eLife.42660.

Shu L, Zhang B, Queller DC, Strassmann JE. 2018. *Burkholderia* bacteria use chemotaxis to find social amoeba *Dictyostelium discoideum* hosts. ISME J. 12:1977–1993. doi: 10.1038/s41396-018-0147-4.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Research. 34:D32–D36. doi: 10.1093/nar/gkj014.

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. 2009. JBrowse: a next-generation genome browser. Genome Res. 19:1630–1638. doi: 10.1101/gr.094607.109.

Sun GW et al. 2010. Identification of a regulatory cascade controlling Type III Secretion System 3 gene expression in *Burkholderia pseudomallei*. Molecular Microbiology. 76:677–689. doi: 10.1111/j.1365-2958.2010.07124.x.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research. 34:W609–W612. doi: 10.1093/nar/gkl315.

Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husnik F. 2021. Pseudofinder: detection of pseudogenes in prokaryotic genomes. 2021.10.07.463580. doi: 10.1101/2021.10.07.463580.

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.

Tipton L, Darcy JL, Hynson NA. 2019. A developing symbiosis: enabling cross-talk between ecologists and microbiome scientists. Front Microbiol. 10:292. doi: 10.3389/fmicb.2019.00292.

Toft C, Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nature Reviews Genetics. 11:465–475. doi: 10.1038/nrg2798.

Tseng T-T, Tyler BM, Setubal JC. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. BMC Microbiology. 9:S2. doi: 10.1186/1471-2180-9-S1-S2.

Vachaspati P, Warnow T. 2015. ASTRID: Accurate Species TRees from Internode Distances. BMC Genomics. 16:S3. doi: 10.1186/1471-2164-16-S10-S3.

Vandamme P et al. 2007. *Burkholderia bryophila* sp. nov. and *Burkholderia megapolitana* sp. nov., moss-associated species with antifungal and plant-growth-promoting properties | Microbiology Society. International Journal of Systematic and Evolutionary Microbiology. 57:2228–2235. doi: https://doi.org/10.1099/ijs.0.65142-0.

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. Systematic and Applied Microbiology. 25:507–512. doi: 10.1078/07232020260517634.

Vanderpool CK, Armstrong SK. 2001. The *Bordetella bhu* locus is required for heme iron utilization. Journal of Bacteriology. 183:4278–4287. doi: 10.1128/JB.183.14.4278-4287.2001.

Vanlaere E et al. 2008. *Burkholderia sartisoli* sp. nov., isolated from a polycyclic aromatic hydrocarbon-contaminated soil. International Journal of Systematic and Evolutionary Microbiology. 58:420–423. doi: 10.1099/ijs.0.65451-0.

Viallard V et al. 1998. *Burkholderia graminis* sp. nov., a rhizospheric Burkholderia species, and reassessment of [*Pseudomonas*] *phenazinium*, [*Pseudomonas*] *pyrrocinia* and [*Pseudomonas*] *glathei* as *Burkholderia*. International Journal of Systematic and Evolutionary Microbiology. 48:549–563. doi: 10.1099/00207713-48-2-549.

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. Frontiers in Microbiology. 12. https://www.frontiersin.org/article/10.3389/fmicb.2021.761215 (Accessed March 31, 2022).

933 Wang J et al. 2021. BastionHub: a universal platform for integrating and analyzing substrates
934 secreted by Gram-negative bacteria. Nucleic Acids Research. 49:D651–D659. doi:
935 10.1093/nar/gkaa899.

936 Wang Z, Wu M. 2017. Comparative genomic analysis of *Acanthamoeba* endosymbionts
937 highlights the role of amoebae as a "melting pot" shaping the *Rickettsiales* evolution. Genome
938 Biology and Evolution. 9:3214–3224. doi: 10.1093/gbe/evx246.

939 Wernegreen JJ. 2017. In it for the long haul: evolutionary consequences of persistent
940 endosymbiosis. Current Opinion in Genetics & Development. 47:83–90. doi:
941 10.1016/j.gde.2017.08.006.

942 Wilcox JL, Dunbar HE, Wolfinger RD, Moran NA. 2003. Consequences of reductive evolution
943 for gene expression in an obligate endosymbiont. Molecular Microbiology. 48:1491–1500. doi:
944 10.1046/j.1365-2958.2003.03522.x.

945 Winsor GL et al. 2008. The Burkholderia Genome Database: facilitating flexible queries and
946 comparative analyses. Bioinformatics. 24:2803–2804. doi: 10.1093/bioinformatics/btn524.

947 Yang H-C, Im W-T, Kim KK, An D-S, Lee S-T. 2006. *Burkholderia terrae* sp. nov., isolated
948 from a forest soil. International Journal of Systematic and Evolutionary Microbiology. 56:453–
949 457. doi: 10.1099/ijs.0.63968-0.

950 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol.
951 24:1586–1591. doi: 10.1093/molbev/msm088.

952 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree
953 reconstruction from partially resolved gene trees. BMC Bioinformatics. 19:153. doi:
954 10.1186/s12859-018-2129-y.

955

956