



# Genomes of Gut Bacteria from *Nasonia* Wasps Shed Light on Phylosymbiosis and Microbe-Assisted Hybrid Breakdown

 Karissa L. Cross,<sup>a,b</sup> Brittany A. Leigh,<sup>a,b</sup> E. Anne Hatmaker,<sup>a,b</sup> Aram Mikaelyan,<sup>b,c</sup> Asia K. Miller,<sup>a,b</sup>  Seth R. Bordenstein<sup>a,b,d,e</sup>

<sup>a</sup>Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA

<sup>b</sup>Vanderbilt Microbiome Initiative, Vanderbilt University, Nashville, Tennessee, USA

<sup>c</sup>Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, USA

<sup>d</sup>Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

<sup>e</sup>Vanderbilt Institute for Infection, Immunology and Inflammation, Vanderbilt University Medical Center, Nashville, Tennessee, USA

**ABSTRACT** Phylosymbiosis is a cross-system trend whereby microbial community relationships recapitulate the host phylogeny. In *Nasonia* parasitoid wasps, phylosymbiosis occurs throughout development, is distinguishable between sexes, and benefits host development and survival. Moreover, the microbiome shifts in hybrids as a rare *Proteus* bacterium in the microbiome becomes dominant. The larval hybrids then catastrophically succumb to bacterium-assisted lethality and reproductive isolation between the species. Two important questions for understanding phylosymbiosis and bacterium-assisted lethality in hybrids are (i) do the *Nasonia* bacterial genomes differ from other animal isolates and (ii) are the hybrid bacterial genomes the same as those in the parental species? Here, we report the cultivation, whole-genome sequencing, and comparative analyses of the most abundant gut bacteria in *Nasonia* larvae, *Providencia rettgeri* and *Proteus mirabilis*. Characterization of new isolates shows *Proteus mirabilis* forms a more robust biofilm than *Providencia rettgeri* and that, when grown in coculture, *Proteus mirabilis* significantly outcompetes *Providencia rettgeri*. *Providencia rettgeri* genomes from *Nasonia* are similar to each other and more divergent from pathogenic, human associates. *Proteus mirabilis* from *Nasonia vitripennis*, *Nasonia giraulti*, and their hybrid offspring are nearly identical and relatively distinct from human isolates. These results indicate that members of the larval gut microbiome within *Nasonia* are most similar to each other, and the strain of the dominant *Proteus mirabilis* in hybrids is resident in parental species. Holobiont interactions between shared, resident members of the wasp microbiome and the host underpin phylosymbiosis and hybrid breakdown.

**IMPORTANCE** Animal and plant hosts often establish intimate relationships with their microbiomes. In varied environments, closely related host species share more similar microbiomes, a pattern termed phylosymbiosis. When phylosymbiosis is functionally significant and beneficial, microbial transplants between host species and host hybridization can have detrimental consequences on host biology. In the *Nasonia* parasitoid wasp genus, which contains a phylosymbiotic gut community, both effects occur and provide evidence for selective pressures on the holobiont. Here, we show that bacterial genomes in *Nasonia* differ from other environments and harbor genes with unique functions that may regulate phylosymbiotic relationships. Furthermore, the bacteria in hybrids are identical to those in parental species, thus supporting a hologenomic tenet that the same members of the microbiome and the host genome impact phylosymbiosis, hybrid breakdown, and speciation.

**KEYWORDS** *Proteus*, *Providencia*, *Nasonia*, gut bacteria, microbiome, phage

**Citation** Cross KL, Leigh BA, Hatmaker EA, Mikaelyan A, Miller AK, Bordenstein SR. 2021. Genomes of gut bacteria from *Nasonia* wasps shed light on phylosymbiosis and microbe-assisted hybrid breakdown. *mSystems* 6:e01342-20. <https://doi.org/10.1128/mSystems.01342-20>.

**Editor** Jonathan L. Klassen, University of Connecticut

**Copyright** © 2021 Cross et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Seth R. Bordenstein, [s.bordenstein@vanderbilt.edu](mailto:s.bordenstein@vanderbilt.edu).

**Received** 9 February 2021

**Accepted** 3 March 2021

**Published** 6 April 2021

Microbiomes can play pivotal roles in host health and disease (1–6) and frequently establish distinguishable associations with host lineages (7–14). Often, the evolutionary relationships between closely related host species mirror the ecological relationships of their microbial communities in a pattern termed phylosymbiosis (15). Across animal and plant clades, closely related host species harbor more similar microbial communities than divergent host species (9, 15–20). Breakdown of phylosymbiotic relationships can also occur when the host genome and microbiome, or hologenome, are altered, such as in hybrid hosts (21–23) or parental hosts receiving a microbiome transplantation from another species (9, 24). It can lead to detrimental effects including hybrid lethality and intestinal pathology (21, 22, 24, 25). Resultantly, it is pertinent to investigate how the bacterial identities present within parental species are or are not altered as a result of hybridization.

A well-developed animal system for studying phylosymbiosis and hologenomic speciation is the *Nasonia* parasitoid wasp genus. It is comprised of four species whose ancestor arose approximately 1 million years ago (MYA) (26). These species are interfertile in the absence of *Wolbachia* endosymbionts and exhibit strong trends in phylosymbiosis even under identical rearing conditions (7, 9, 27). Crosses between *Nasonia* species long cured of their intracellular *Wolbachia* endosymbionts readily produce fit F<sub>1</sub> hybrid females, but most of the F<sub>2</sub> hybrid males die during larval development due to microbe-assisted hybrid lethality that occurs in association with changes in the microbiome, hypermelanization, and transcriptome-wide upregulation of immune gene expression (21, 28). Moreover, transplants of microbiomes into larvae of each *Nasonia* species elicit reductions in host developmental rates and survival, supporting the premise that selective pressures drive phylosymbiosis in the system (24).

Early in development, *Nasonia* larvae possess a simple microbiome that changes composition throughout development (27). Larvae of the two most divergent species, *Nasonia vitripennis* and *Nasonia giraulti*, possess gut microbiomes dominated by *Providencia* bacteria (81 to 96% of sequencing reads), whereas their F<sub>2</sub> larval hybrid offspring are dominated by *Proteus* bacteria (86% of sequencing reads) (21). Furthermore, these F<sub>2</sub> interspecific hybrids exhibit ~90% hybrid lethality between the third and fourth larval instars (21), and germfree rearing can remarkably rescue the hybrid lethality (21, 29). The rescue of hybrid lethality via germfree rearing and recapitulation of death upon inoculation strongly support the dependency of hybrid lethality on gut bacteria (30). Interspecific microbiome transplants between *Nasonia* species with heat-killed bacterial communities also result in slowed larval growth and decreased pupation and adult survival, demonstrating how host responses play an integral role in interacting with their microbiomes (24). The costs that result from gut microbiome transplants occur in an evolutionarily informed manner, which further suggests that selective pressures can underpin phylosymbiosis and holobiont composition (24).

The genomes of both *N. vitripennis* and *N. giraulti* are published (31, 32), and thus, the imperative turns to the bacterial genomes within these two species and their hybrid offspring to understand the nature of the catastrophic events that result in gut bacterium-assisted hybrid lethality. In particular, are the bacteria in *Nasonia* guts distinguishable from other environmental isolates and thus specific to the species complex? And upon hybridization and phylosymbiotic breakdown, are the gut bacteria in parental hosts identical to or different from the bacteria in hybrid hosts? A key aspect in this system is whether the same bacteria present in parental species contribute to reproductive isolation in hybrids.

In *Nasonia*, an alteration in the abundances of *Proteus* and *Providencia* bacteria in hybrid offspring is associated with lethality (21). *Proteus* and *Providencia* are well-characterized bacterial genera due to their role as opportunistic pathogens in both humans and insects (33–36). *Proteus* spp. are present in low abundance in humans, but their overgrowth is often associated with urinary tract infections (35, 37). In *Drosophila melanogaster*, different strains of *Providencia* display various degrees of virulence toward the host, which develops an immune response to potentially combat bacterial infection

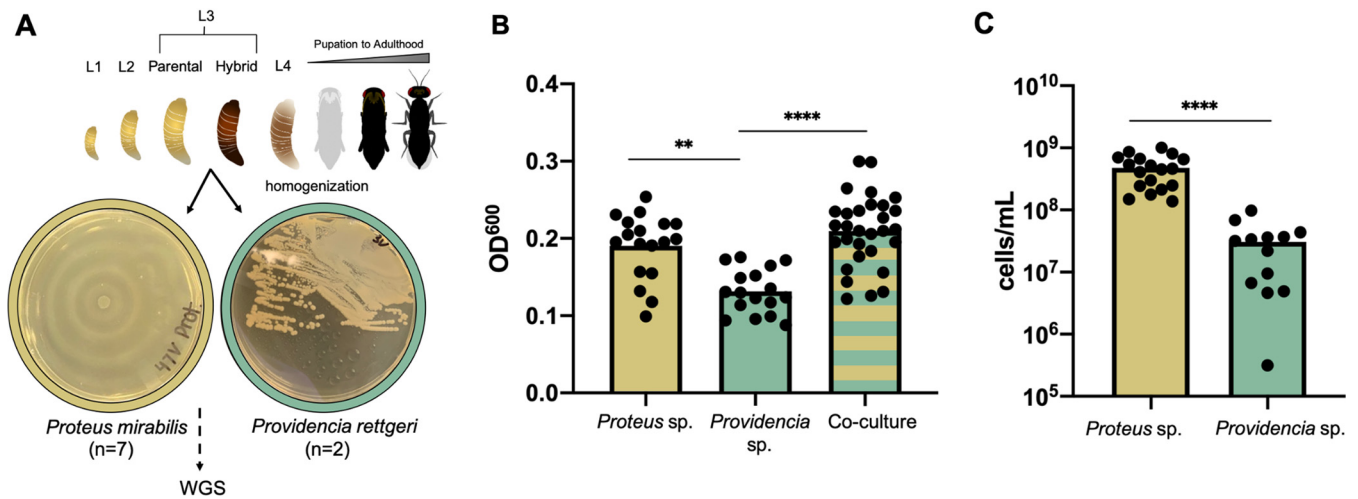
(38–40). Additionally, in *Caenorhabditis elegans*, commensal *Providencia* bacteria in the gut produce a neurotransmitter that promotes fitness of both the host and bacteria (41). The evolution of various host responses to different bacteria demonstrates how intimate interactions between these bacteria and their hosts may mediate pathogenic versus symbiotic relationships.

Since *Proteus* and *Providencia* bacteria are native members of the *Nasonia* gut microbiome across multiple lines (27, 42), there are several questions as to whether or not changes in the strains of gut bacteria mediate hybrid lethality and whether the strains in *Nasonia* are distinct from those in other animals. As *Proteus* and *Providencia* bacteria are readily culturable in the laboratory, we isolated and sequenced the genomes of representative species from *Nasonia* and their hybrid larvae to investigate the genomic diversity of (i) *Proteus* and *Providencia* isolates between *Nasonia* and publicly available genomes from insects and humans, (ii) *Proteus* and *Providencia* isolates between parental *N. vitripennis* and *N. giraulti* species, and (iii) *Proteus* isolates between parental and F<sub>2</sub> hybrid offspring to determine if they are the same. Annotation and evolutionarily guided comparisons of the gene content may inform the nature of what drives phylosymbiosis in *Nasonia* from the microbial side, and what is the nature of bacterium-dependent hybrid lethality.

## RESULTS AND DISCUSSION

**Biofilm formation and genomics of *Providencia* and *Proteus* bacterial isolates from *Nasonia* spp. and their hybrids.** We previously found that the dominant bacterial genus present in the microbiome of the larvae of *N. giraulti* and *N. vitripennis* is the genus *Providencia*, whereas in F<sub>2</sub> recombinant male hybrids of the same developmental stage, the bacterial genus *Proteus* becomes the dominant taxon associated with severe hybrid mortality (21, 27). Both of these bacteria are easily cultivable, are well studied in arthropods, and are opportunistic human pathogens (6, 33–35, 43, 44). Therefore, we sought to isolate and sequence the genomes of representative *Proteus* and *Providencia* species from parental *N. vitripennis* or *N. giraulti* and F<sub>2</sub> hybrids derived from the cross of *N. vitripennis* males × *N. giraulti* females. We concurrently set up intra- and interspecific crosses and collected F<sub>2</sub> third-instar larvae (L3) from virgin F<sub>1</sub> females. F<sub>2</sub> male larvae were surface sterilized with 70% ethanol and sterile water and then homogenized. Homogenate was serially diluted on tryptic soy agar (TSA) plates, and distinct bacterial colonies were randomly chosen, subcultured, and sent for whole-genome sequencing (Fig. 1A).

Bacteria that successfully colonize insect guts often adhere to surfaces and rapidly form biofilms (45). To assess the ability of *Proteus* and *Providencia* to adhere and form biofilms on solid surfaces, we plated representatives from each genus in single culture and coculture, and we measured their resulting biofilms. Over 24 h, *Proteus mirabilis* formed a more robust biofilm than *Providencia rettgeri*, and when the two isolates were plated together in a 1:1 ratio, the biofilm was significantly more abundant than *Providencia rettgeri* alone (Fig. 1B). To determine whether the slight, insignificant biofilm increase in the coculture occurred in an additive manner with the two species contributing equally to the biofilm or whether the increase was principally due to one of the two species, we developed a qPCR assay to quantify their abundance. *Proteus mirabilis* significantly outcompeted *Providencia rettgeri* when grown together ( $P < 0.0001$ ), with *Providencia rettgeri* making up only about 6.4% of the biofilm composition (Fig. 1C). This suggests that *Nasonia* host genetic factors may keep the *Proteus* bacterial abundances under control *in vivo* in parental wasp species; and regulation is then compromised in hybrids where *Proteus* dominates the microbiome, or when the two bacteria are grown in coculture outside the host. *Proteus mirabilis* is a well-characterized human pathogen (46) that has a distinctive swarming behavior which serves to reduce competition for nutrients between bacterial strains (47). When *Nasonia* hybridization produces an F<sub>2</sub> genotypic recombinant that shuffles the genes of the two host species, the swarming behavior of *Proteus mirabilis* coupled with a presumptive breakdown in the ability of the host to



**FIG 1** Characteristics of *Proteus mirabilis* and *Providencia rettgeri* isolates. (A) Workflow for sample collection and examples of unique colony morphologies of *Providencia* (n=2) and *Proteus* (n=7) isolates. *Nasonia* development from larvae to adulthood is depicted on top demonstrating the L3 larval stage in which samples were collected from parental and hybrid lines. Following isolation, samples were prepared for whole-genome sequencing (WGS). (B) Biofilm formation by representative *Proteus* and *Providencia* isolates shows significantly different abilities to form biofilms on solid surfaces. Bars denote significantly different values based on Kruskal-Wallis test with Dunn’s multiple comparison with sample sizes of n=17 to 30 of two biological replicates (\*\*, P=0.0018; \*\*\*\*, P = <0.0001). (C) Composition of the biofilm coculture shows *Proteus mirabilis* significantly outcompetes *Providencia rettgeri* when grown together. *Providencia rettgeri* makes up about 6.4% of the coculture biofilm composition. Bars denote significantly different values calculated by a Mann-Whitney U test with sample size of n=13 to 18 (\*\*\*\*, P = <0.0001).

regulate bacterial abundances may create an environment that permits bacterial overgrowth and ensuing host lethality.

Following isolation of colonies and DNA extraction, whole-genome sequencing was performed using an Illumina MiSeq. Data on genome statistics including number of reads generated, sequencing coverage, and genome size and completion are provided in Tables 1 and 2. For strain nomenclature, Ngir or G, Nvit or V, and Hybrid or H denote *N. giraulti*, *N. vitripennis*, and hybrid, respectively; “L3,” if noted, represents the third larval instar stage used for bacterial isolation and cultivation; numbers at the end of the name are unique designations assigned upon isolation (numerals 1 to 3). Shorthand notation of all genomes presented throughout this study is provided in Table S1 in the supplemental material. Isolate identity was confirmed after sequencing by comparison of the 16S rRNA sequences extracted from each genome to the RefSeq database.

We sequenced *Providencia rettgeri* isolates from *N. giraulti* and *N. vitripennis* parental lines. The genome size for *Providencia* strain Wasp.NgirL3.3 is 4,254,678 bp with 163-fold

**TABLE 1** *Providencia rettgeri* sequencing and genome statistics<sup>c</sup>

Strain reference	Wasp.NgirL3.3	Wasp.NvitL3.3	Wasp.Nvit03	Fly.Dmel1	Human.RB151
NCBI strain name	NgirL3-3G	NvitL3-3V	NVIT03 <sup>a</sup>	Dmel1 <sup>a</sup>	RB151 <sup>b</sup>
Host source	<i>N. giraulti</i> L3 larvae	<i>N. vitripennis</i> L3 larvae	<i>N. vitripennis</i> gut	<i>D. melanogaster</i> hemolymph	Human (urine from UTI)
No. of reads	3,774,124	1,321,295	191,241	680,000	167,518
No. of contigs	16	20	73	9	1
Avg coverage (fold)	163	189	164	30	581
Genome size (bp)	4,254,678	4,314,756	4,397,852	4,251,508	4,780,676
GC content (%)	40.2	40.2	40.3	40.2	40.5
No. of CDS (coding)	3,898	3,897	4,076	3,848	4,484
No. of plasmids	1	1	0	1	1
No. of prophage regions	3	3	6	4	2
% completed	100	99.91	100	98.61	99.46
% contaminated	1.62	1.89	1.08	1.08	1.08
Reference	This study	This study	Wang and Brucker (42)	Galac and Lazzaro (38)	Marquez-Ortiz et al. (44)

<sup>a</sup>Previously published insect isolates (accession numbers: NVIT03, [QUAF01000007](#); Dmel1, [CM001774](#)).

<sup>b</sup>Human isolate reference strain (accession number: RB151, [PRJNA342046](#)).

<sup>c</sup>Abbreviations: UTI, urinary tract infection; CDS, coding sequences.

**TABLE 2** *Proteus mirabilis* sequencing and genome statistics

Strain reference	Wasp.NgirL3.1	Wasp.NgirL3.2	Wasp.Hybl3.1	Wasp.Hybl3.2	Wasp.Hybl3.3	Wasp.NvitL3.1	Wasp.NvitL3.2	Human.HI4320 <sup>a</sup>
NCBI strain name	NgirL3-1G	NgirL3-2G	F <sub>2</sub> Hybrid-1H	F <sub>2</sub> Hybrid-2H	F <sub>2</sub> Hybrid-3H	NvitL3-1V	NvitL3-2V	HI4320 <sup>a</sup>
Host source	<i>N. giraulti</i> L3 larvae	<i>N. giraulti</i> L3 larvae	<i>Nasonia</i> L3 hybrid	<i>Nasonia</i> L3 hybrid	<i>Nasonia</i> L3 hybrid	<i>N. vitripennis</i> L3 larvae	<i>N. vitripennis</i> L3 larvae	Human (urine)
No. of reads	4,180,770	4,215,774	3,928,914	3,823,446	3,926,886	4,193,134	4,059,594	60,288
No. of contigs	57	61	46	62	63	42	62	1
Avg coverage (fold)	399	348	443	323	331	471	377	8
Genome size (bp)	4,024,338	4,023,051	4,076,851	4,023,528	4,024,027	4,070,909	4,026,776	4,063,606
GC content (%)	38.6	38.6	38.6	38.6	38.6	38.6	38.6	38.8
No. of CDS (coding)	3,639	3,632	3,688	3,621	3,628	3,683	3,628	3,611
No. of plasmids	0	0	0	0	0	0	0	1
No. of prophage regions	19	17	19	19	17	18	17	4
% completed	100	100	100	100	100	100	100	100
% contaminated	0	0	0	0	0	0	0	0
Reference	This study	This study	This study	This study	This study	This study	This study	Pearson et al. (35)

<sup>a</sup>Human isolate reference strain (accession number: HI4320, PRJNA12624).

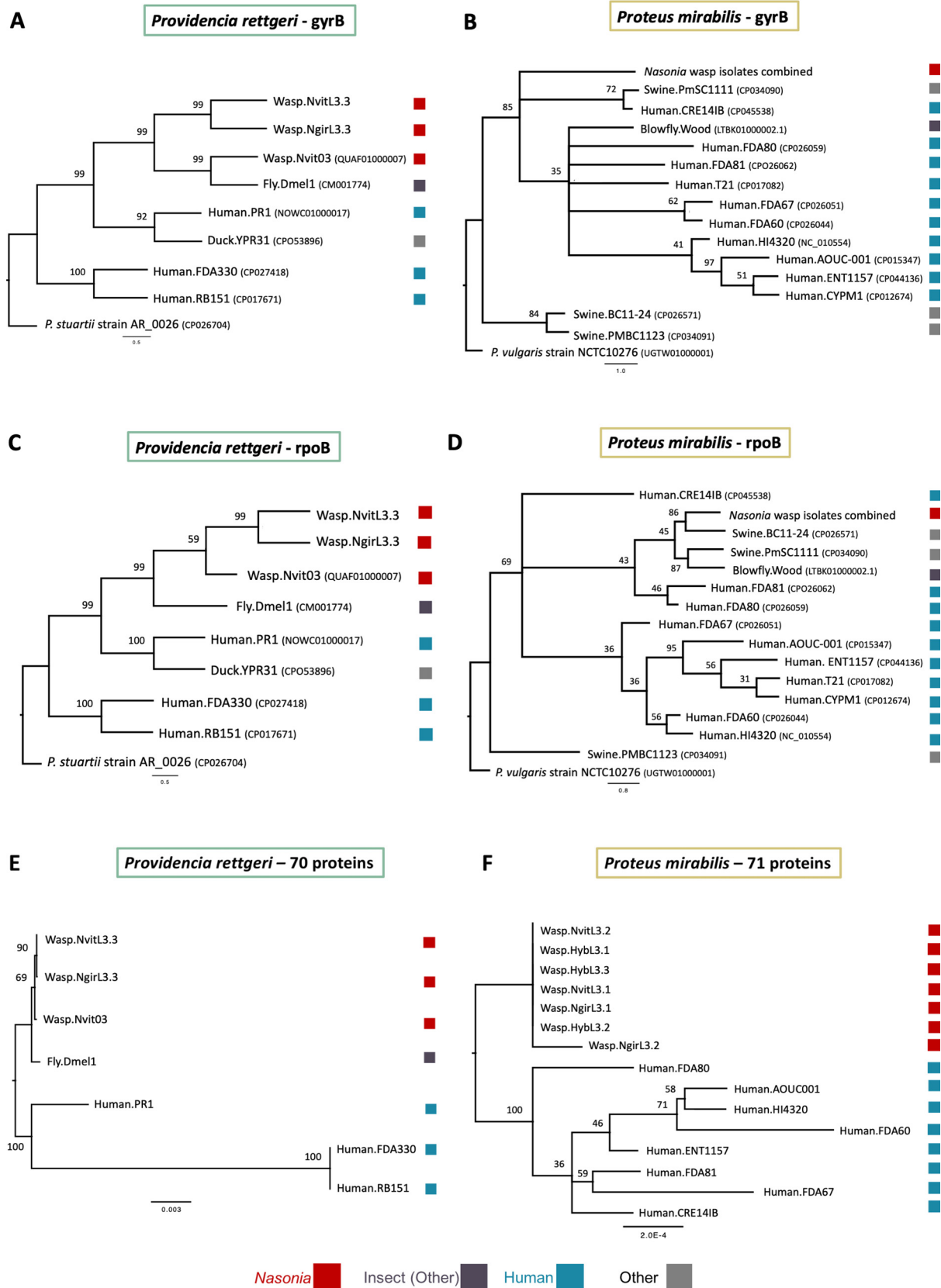
coverage, and Wasp.NvitL3.3 is 4,324,254 bp with 189-fold coverage. Both strains have a G+C content of 40.2%. A single plasmid of 75,229 bp was identified in *Providencia* strain Wasp.NgirL3.3 with 324-fold coverage, and a single plasmid of 75,094 bp was identified in Wasp.NvitL3.3 with 116-fold coverage. The two plasmids have 100% nucleotide identity between them.

The average genome size for the seven *Proteus mirabilis* isolates is 4,046,931 bp ranging between 3,823,446 and 4,215,774 bp with an average coverage of 385-fold. The G+C content of all *Proteus mirabilis* isolates is 38.6%. No circular elements were identified in any isolate. Using CheckM (48) for genome completion and contamination estimates, the *Providencia rettgeri* and *Proteus mirabilis* genomes are predicted to be 99.91 to 100% complete with 0.00 to 1.89% contamination (Tables 1 and 2).

**Phylogenetic placement of *Nasonia* bacterial isolates with other host-associated lineages.** To characterize and compare the genomic diversities of the sequenced *Providencia rettgeri* and *Proteus mirabilis* isolates, we surveyed the National Center for Biotechnology Information (NCBI) databases for representative genomes. *Proteus mirabilis* is a commensal, gastrointestinal bacterium in host-associated environments (46, 49), including in humans (50), tree shrews (51, 52), and insects (21, 27, 53). It is well characterized due to its role as an opportunistic pathogen in different clinical settings (33–36), especially urinary tract infections involving catheters (37), as well as bacteremia and wound infections (46). However, less is known regarding *Providencia rettgeri*, despite also serving as an opportunistic pathogen in both humans and insects (5, 6, 38, 43, 44). We characterized our *Nasonia* isolates based on single-copy gene phylogenies of the housekeeping genes *DNA gyrase B* (*gyrB*, topoisomerase) and *RNA polymerase beta subunit* (*rpoB*, RNA synthesis), a 70- to 71-multiprotein concatenated phylogeny (Fig. 2), and whole-genome average nucleotide identity (ANI) for all-against-all pairwise comparisons with publicly available whole-genome sequences (Fig. S1). We report two key results.

First, the *Providencia rettgeri* insect isolates are distinct from human-associated isolates. In the maximum likelihood (ML) phylogeny for both *rpoB* and *gyrB*, the *Nasonia* isolates from this study form a well-supported clade with previously published, insect-associated *Providencia rettgeri* strains Wasp.Nvit03 (QUAF01000007) from *Nasonia vitripennis* (42) and Fly.Dmel1 (NZ\_CM001774) from *Drosophila melanogaster* (38) (Fig. 2A and C). To explore these relationships further, we built a concatenated protein tree using 70 core bacterial proteins to garner a higher-resolution view of the relatedness between *Providencia rettgeri* strains used in this study (Table S1). A split in the phylogeny of these strains was apparent based on host origin, whereby *Nasonia* isolates grouped together with *D. melanogaster* and apart from human isolates, consistent with the ML phylogenies (Fig. 2E). Lastly, based on whole-genome pairwise average nucleotide identity (ANI) (54), we classified Fly.Dmel1, Wasp.Nvit03, Wasp.NvitL3.3, and Wasp.NgirL3.3 as the same bacterial strain based on their >99% ANI. The closest human reference isolate to the insect *Providencia rettgeri* isolates was strain Human.PR1 with an ANI of ~92%, followed by strains Human.RB151 and Human.FDA330 at an ANI of ~84% (Fig. S1). ANI values of ≥95% are considered the same species (54, 55); therefore, the human isolates of *Providencia rettgeri* may represent different species.

The *Proteus mirabilis* isolates sequenced from *Nasonia* have identical *gyrB* and *rpoB* sequences at the nucleotide level, and a representative sequence was used to build the ML phylogenies (Fig. 2B and D). The *Proteus mirabilis* ML trees (Fig. 2B and D) provide moderate support for the placement of the *Nasonia* isolates, with most publicly available human-associated *Proteus mirabilis* genomes forming a separate clade with low support. The low phylogenetic support is due to the high similarity of these genes between close relatives. Therefore, we built a concatenated amino acid protein tree using 71 core bacterial genes to provide more resolution to the diversity within the *Proteus mirabilis* isolates (Fig. 2F, Table S1). The phylogenomic tree showed that *Nasonia*-associated *Proteus mirabilis* isolates are nearly 100% identical and form a grouping distinct from human-associated strains. Lastly, based on ANI, all *Proteus mirabilis* isolates from *Nasonia* were identified as the same strain with >99% ANI. When comparing human-associated



**FIG 2** Phylogenetic placement of *Nasonia* isolates relative to publicly available genomes. (A and B) Phylogeny of *Providencia rettgeri* and *Proteus mirabilis* strains based on 2,415 nucleotides of the gyrase B subunit reconstructed using RAxML. (C and D) Phylogeny of *Providencia rettgeri* and *Proteus mirabilis* strains based on 2,415 nucleotides of the RNA polymerase B subunit reconstructed using RAxML. (Continued on next page)

whole-genome sequences of *Proteus mirabilis* with our *Nasonia* isolates, all isolates are classified as the same subspecies with an ANI of >98.5% (Fig. 2B and D; Fig. S1). Collectively, there is little to no genetic diversity present among all the *Proteus mirabilis* bacterial isolates from *Nasonia* species and their hybrids.

**Comparative genomics within and between bacteria from *Nasonia* and other animals. (i) *Providencia rettgeri* pangenomics.** For comparative genomic analyses, we utilized five publicly available high-quality whole-genome sequences of *Providencia rettgeri* to investigate the unique genomic changes that arose in *Nasonia*-associated and other animal-associated *Providencia rettgeri* isolates. The *Providencia rettgeri* genomes are provided in Table S1. The public genomes are from *Nasonia vitripennis* (Wasp.Nvit03) (42), *Drosophila melanogaster* (Fly.Dmel1) (38), and three human samples (reference strain Human.RB151 [44], Human.FDA330 [56], and Human.PR1). Combined with our two *Nasonia*-associated isolates of *Providencia rettgeri*, Wasp.NvitL3.3 and Wasp.NgirL3.3, pangenomic analyses in anvio (57, 58) produced a total of 6,789 gene clusters (COGs). We grouped these gene clusters into seven bins based on their presence/absence across the genomes: (i) *Providencia* species Core (2,949 gene clusters), (ii) Human Specific (58 gene clusters), (iii) *Nasonia* Specific (55 gene clusters), (iv) Insect Specific (142 gene clusters), (v) *Drosophila* Specific (197 gene clusters), (vi) *Nasonia*—Nashville (168 gene clusters), and (vii) *Nasonia*—Cambridge (269 gene clusters) (Fig. 3). The city nomenclature between isolates Wasp.NvitL3.3, Wasp.NgirL3.3, and Wasp.Nvit03 is used to denote potential genomic differences between laboratories and geographies as these two lines were derived from the same *N. vitripennis* line less than 6 years ago and reared apart since then. More genomes from these lines will be required to further evaluate the evolution of genomic differences. The core genome represents gene clusters present in all genomes analyzed (human associated and insect associated). Most gene clusters (43.4%) fall within the *Providencia* species Core bin, as expected, where the majority of COG functions relate to [C] energy production and conversion, [E] amino acid transport and metabolism, and [J] translation, ribosomal structure and biogenesis (Tables S2 and S3).

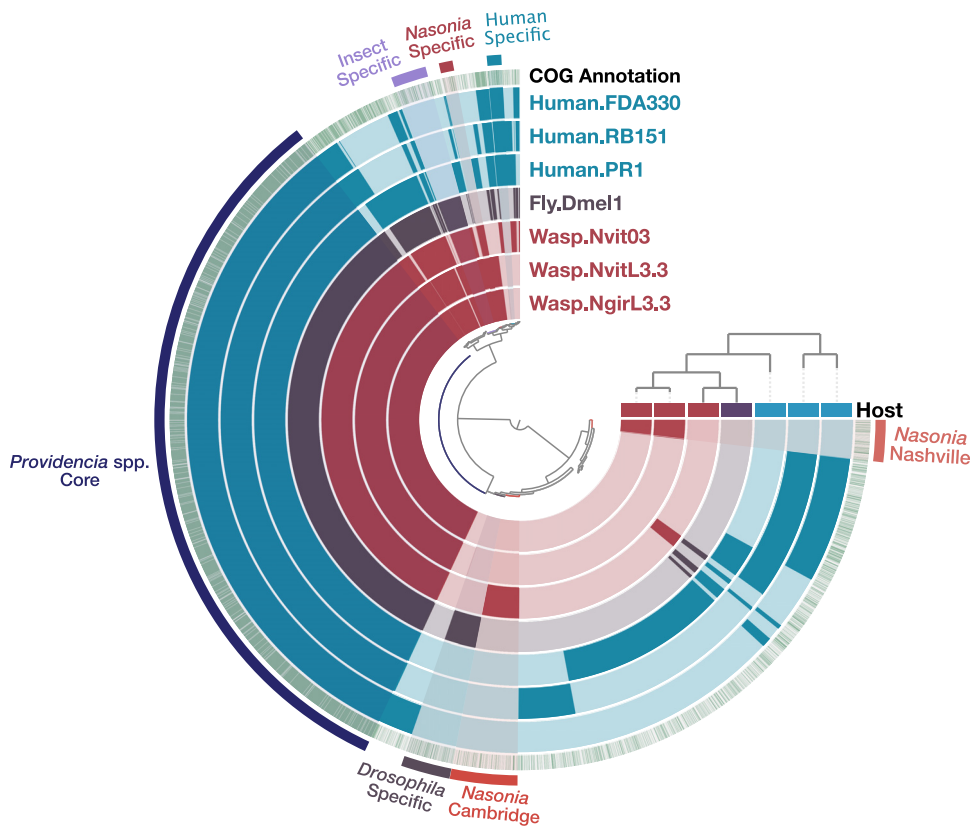
As mentioned above, we determined the phylogenetic relationships among these *Providencia rettgeri* whole-genome sequences and noted a split in the phylogeny of these strains based on host origin, which is consistent with their ANI divergence (Fig. 2 and Fig. S1). As adaptations to specific host environments may drive strain divergence in *Providencia rettgeri*, we sought to investigate host-associated genomic variation in *Providencia rettgeri* between (i) humans and insects, (ii) *Nasonia* and human/*Drosophila*, and (iii) various *Nasonia* isolates to further explore how bacterial genomic variation partitions across host variation.

(a) *Functional differences between *Providencia rettgeri* isolates from insects and humans.* The human and insect bins encompass accessory proteins, also referred to as gene clusters (GCs), present in either all human- or insect-associated genomes (*Nasonia* and *Drosophila* combined) (Fig. 3). We divide the discussion of these unique bins into those with and without shared functions. Distinct gene clusters between human and insect isolates based on the Markov cluster algorithm (MCL) (59, 60) that share homologous functional assignments (defined as distinct proteins having the same COG assignment) span hemolysins, large exoproteins involved in heme utilization or adhesin, and adhesin and type 1 pilus proteins (Table S2). These functions are typically involved in host cell lysis (61), biofilm formation (62), and attachment to mucosal surfaces (62–64). Notably, COG3539 for pilin (type 1 fimbria component protein) is the most abundant functional annotation in all *Providencia rettgeri* genomes, averaging around 58 occurrences in each genome. Different proteins encoding similar functions between human- and insect-associated bacterial genomes could be due to slightly different adaptations of these genes as the bacteria evolved within their

## FIG 2 Legend (Continued)

*Proteus mirabilis* strains based on 4,029 nucleotides of the RNA polymerase beta subunit reconstructed using RAxML. (E and F) Phylogeny of *Providencia rettgeri* and *Proteus mirabilis* strains based on concatenated amino acid alignments of 70 to 71 core bacterial proteins reconstructed using PhyML.





**FIG 3** Pangenome of *Providencia rettgeri* human and insect isolates. The seven inner layers correspond to the seven genomes, including the sequenced *Nasonia* isolates in maroon, previously published *Drosophila* isolate in purple, and human-associated genomes in blue. The bars in the inner circles show the presence of gene clusters (GCs) in a given genome with the outer green circle depicting known COGs (green) versus unknown (white) assignments. The outermost layer of color-coded lines and text highlights groups of GCs that correspond to the genome core or to group-specific GCs. Genomes are ordered based on their gene-cluster distribution across genomes, which is shown at the top right corner (tree). Central dendrograms depict protein cluster hierarchy when displayed as protein cluster frequency. The top horizontal layer underneath the tree represents host source (*Nasonia*, maroon; *Drosophila*, purple; human, blue).

specified niche.

We investigated what functions are present in the human-associated *Providencia rettgeri* genomes and missing or reduced in insect-associated isolates. In the human-associated *Providencia rettgeri* genomes, there are twice as many proteins relating to COG functions for [U] intracellular trafficking, secretion, and vesicular transport (7.05% versus 2.35% of total COG functions) (Tables S2 and S3). This difference seems to be primarily driven by the increased occurrence of proteins relating to some of the same functions above, namely, hemolysins and large exoproteins involved in heme utilization or adhesin. Large exoproteins involved in heme utilization or adhesin (COG3210) are secreted to outer bacterial membranes and serve as attachment factors to host cells (64). For example, *Providencia stuartii* expressing higher levels of MR/K hemagglutinin adhered better to catheters and led to persistence in catheter-associated bacteriuria (65). Similarly, the filamentous hemagglutinin (FHA) protein secreted by *Bordetella pertussis* is used for attachment to the host epithelium early in its pathogenesis (66). Collectively, the increased presence of these genes in human-associated *Providencia rettgeri* suggests an adaptive lifestyle toward human colonization and virulence. Relatedly, *Drosophila* species have developed unique antimicrobial peptides that suppress and resist *Providencia rettgeri* by reducing bacterial burden (39, 40).

Furthermore, CRISPR-Cas system-associated endonucleases Cas1 and Cas3 were strictly in human-associated *Providencia rettgeri* strains Human.FDA330 and Human.

RB151 and absent in all other strains. The adaptive CRISPR-Cas systems are multigenic and include short repeated palindromic regions alongside spacers of phage DNA that serve as recognition factors triggering an “immune” response when phages encoding the spacer attempt to infect the microbe (67). CRISPR-Cas systems are crucial for subversion of phage, and as such, human-associated isolates with CRISPR-Cas have fewer integrated prophages (five total;  $n=2$ ) than the *Nasonia*-associated ones (12 total;  $n=3$ ) (Table 1). The human-associated isolate Human.PR1 without CRISPR-Cas maintains five unique predicted prophages within its genome, similar to the number found in each *Nasonia*-associated isolate, and it also maintains the highest ANI to the *Nasonia* isolates at ~92% compared to the ~84% of all other human isolates. The emphasis of these functions in human-associated *Providencia rettgeri* isolates, as opposed to insect-associated isolates, suggests human isolates are better poised to combat phage invasions.

To further tease apart the annotated functions specific in *Nasonia* and its close relative in *D. melanogaster*, we performed a functional enrichment analysis to statistically identify COGs only in the insect isolates (Fig. 3 and Table S4). The functional enrichment analysis computed in *anvi'o* uses a Generalized Linear Model with the logit linkage function to compute an enrichment score,  $P$  value, and a false-detection rate  $q$  value between the different genome data sets (57, 58). This analysis identified six COG functions unique to insect-associated genomes: L-rhamnose isomerase (COG4806), L-rhamnose mutarotase (COG3254), predicted metal-dependent hydrolase (COG1735), uncharacterized iron-regulated membrane protein (COG3182), putative heme iron utilization protein (COG3721), and  $\text{Ca}^{2+}/\text{H}^{+}$  antiporter (COG0387) (Table S4). L-Rhamnose is a common sugar component of plant and bacterial cell walls (68, 69), serves as a carbon source for some bacteria (70), and may confer protection to host immune proteins (71). For example, adherent-invasive *Escherichia coli* when grown on bile salts (a component of the gut) showed an upregulation of genes involved in sugar degradation, including the rhamnose pathway (72), and host-associated *Listeria monocytogenes* survival increases by glycosylating L-rhamnose and thus decreasing the cell wall permeability to antimicrobial peptides (71). Additionally, as iron is a key metal for biological functions, this poses the question as to why these genes may be enriched in insect-associated environments compared to those in human-associated settings. Iron-regulated membrane proteins and heme iron utilization proteins permit enteric bacteria to readily sense and respond to iron-limiting environments and play a role in iron acquisition (73, 74). Similarly, as mentioned above, the human-associated isolates encode unique heme utilization proteins, emphasizing how these functions are important for specific host-associated environments. Insects differ from mammals in that they secrete ferritin, a protein that contains iron, into their hemolymph at levels 1,000-fold higher than what is found in mammalian blood (75), and it has altered expression in the presence of *Wolbachia* endosymbionts (76). Perhaps the presence of these iron-regulating genes, particularly in insect environments, helps regulate iron homeostasis specific to what is available in insects. Lastly, the  $\text{Ca}^{2+}/\text{H}^{+}$  antiporter may play a role in ion homeostasis to protect bacteria in altered-pH environments and in maintaining internal pH homeostasis (77, 78). It is important to note that although the unadjusted  $P$  value for all genes in the enrichment analysis is significant ( $P=0.008$ ), the adjusted  $q$  value ( $q=0.897$ ) that controls for false-discovery rate does not reach significance due to the sample sizes for these genomes ( $n=7$ ) (Table S4); however, this does not change the presence/absence of these gene clusters across isolates.

(b) *Functional differences between Providencia rettgeri* isolates from *Nasonia* versus those from humans and *Drosophila*. We were next interested in disentangling functions specific in the *Nasonia*-associated isolates. Therefore, we performed the same functional enrichment analysis to identify COGs that occur only in the *Providencia rettgeri* strains from *Nasonia* (Fig. 3 and Table S4). Only two COG categories occurred in the *Nasonia*-associated genomes: phospholipid *N*-methyltransferase (COG3963) and CYTH domain, found in class IV adenylate cyclase and various triphosphatases (COG2954) ( $P=0.03$ ,  $q=1$ ). Phospholipid *N*-methyltransferase synthesizes phosphatidylcholine

(PC), a membrane-forming phospholipid present in only about 15% of bacteria (79, 80). PC may help mediate symbiotic host-microbe interactions as PC is required for virulence in some bacteria and to establish beneficial symbioses with hosts (81). Furthermore, adenylate cyclases are responsible for converting ATP to cyclic AMP (cAMP) and play key regulatory roles such as mediating signal transduction in cells (82), of which the CYTH (CyaB, thiamine triphosphatase) domain of the type IV adenylate cyclases binds organic phosphate (83). In *Pseudomonas aeruginosa*, when the adenylate cyclase gene CyaB was deleted, virulence was attenuated in a mouse model (84). Collectively, these two COGs could play a role in mediating symbiotic relationships, beneficial or pathogenic, in the *Nasonia* host. Lastly, the Wasp.NvitL3.3 and Wasp.NgirL3.3 isolates did not share any mobile elements with any human- or *Drosophila*-associated isolates. However, Fly.Dmel1 from *Drosophila* shares 62% nucleotide similarity with one phage spanning 46 genes from the previously published Wasp.Nvit03 *Nasonia* isolate (phage 1, Fig. S2).

(c) *Functional differences within Providencia rettgeri* isolates from *Nasonia*. We further investigated the differences between *Providencia rettgeri* isolates within *Nasonia* species. Strains Wasp.NvitL3.3 and Wasp.NgirL3.3 isolated from the lab in Nashville, TN, have a 99.99% ANI with each other and 99.3% ANI with Wasp.Nvit03 isolated from a lab located in Cambridge, MA (Fig. S1); therefore, most of their genomic content is similar. No discernible difference could be identified between the two Nashville isolates; however, a small difference exists between these and the isolate from Cambridge derived from the same *N. vitripennis* line, AsymCx, less than 6 years apart (Fig. 3; Table S2). *Providencia rettgeri* from Nashville contains over three times as many genes relating to the COG category for [V] defense mechanisms (10.19% versus 3.03% of total COG functions, Table S3). Although we found no evidence of CRISPR-Cas gene cassettes in *Nasonia*-associated genomes, a major difference is the presence of type I restriction-modification (RM) systems exclusively in the Nashville-associated *Providencia* genomes and absent from *N. vitripennis* strain Wasp.Nvit03 from Cambridge. Type I RM systems cleave away from the recognition site and have three components: a restriction enzyme (*hsdR*), a methyltransferase (*hsdM*), and a specificity subunit (*hsdS*) (85). *Providencia rettgeri* strains Wasp.NvitL3.3 and Wasp.NgirL3.3 are predicted to encode two type I RM systems, one with all three component genes and one missing the restriction enzyme. The Wasp.Nvit03 strain without the type I RM system harbors six putative prophage regions within its genome, while both the newly sequenced Wasp.NvitL3.3 and Wasp.NgirL3.3 strains here contain three prophage regions each. The three phages found within Wasp.NvitL3.3 and Wasp.NgirL3.3 are 99.5 to 100% identical at the nucleotide level across the entire phage regions (Fig. S2), whereas the remaining three phage regions in Wasp.Nvit03 are novel or contain similarity to phage regions in *Providencia rettgeri* from *D. melanogaster*. The absence of both CRISPR-Cas and type I RM systems in the Wasp.Nvit03 strain could be directly correlated with the number of prophages within its genome, but a larger sample size of genomes may be necessary to make any firm predictions.

We also performed functional enrichment analysis to identify genes present only in *Providencia rettgeri* strains isolated from the lab in Nashville and absent in all other strains. We identified seven unique COGs: (i) CDP-glycerol glycerophosphotransferase (COG1887), (ii) uncharacterized conserved protein YBBC (COG3876), (iii) Argonaute homolog implicated in RNA metabolism and viral defense (COG1431), (iv) predicted restriction endonuclease (COG3440), (v) DNA replication protein DnaD (COG3935), (vi) replicative superfamily II helicase (COG1204), and (vii) predicted ATPase archaeal AAA<sup>+</sup> ATPase superfamily (COG1672) ( $P=0.03$ ,  $q=1$ ) (Table S4). Of particular interest is the Argonaute homolog and restriction endonuclease that function as viral and mobile element defense mechanisms (86, 87). Recently, a bacterial Argonaute nuclease from *Clostridium butyricum* was shown to target multicopy genetic elements and suppress the propagation of plasmids and infection by phages via DNA interference (88). In addition to the RM systems, this diverse suite of genes in *Nasonia* specifically from Nashville suggests enhanced protection against viral infection.

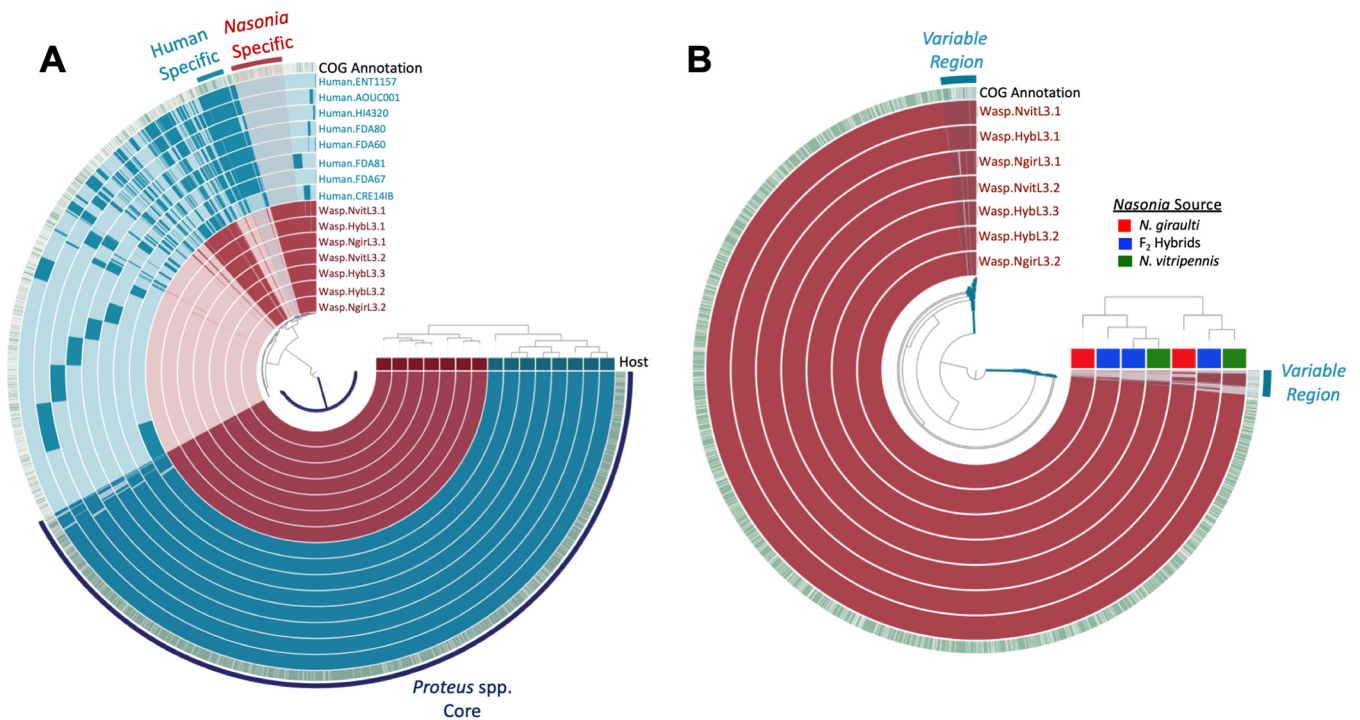
Conversely, strain Wasp.Nvit03 from Cambridge, MA, contains genes relating to

type III secretion systems that are not present in strains from Nashville (COG4791, COG4790, COG4794) ( $P=0.03$ ,  $q=1$ ). Specifically, type III secretion systems serve to deliver effector proteins across bacterial and host membranes that can influence host cell biology (89). For example, this machinery provides efficient protein transfer into eukaryotic cells that could inhibit phagocytosis or downregulate proinflammatory responses of the host (90). Most notably, the biggest difference between the *Nasonia* isolates from Nashville and those from Cambridge is the unique phage machinery. Although two of the three prophages identified in Wasp.NvitL3.3 and Wasp.NgiriL3.3 strains were also present in Wasp.Nvit03 based on amino acid identity, extensive rearrangements were evident (Fig. S2). Notably, one additional phage in Wasp.Nvit03 has 73% nucleotide homology to a phage present in human isolate Human.RB151.

**(ii) *Proteus mirabilis* pangenomics.** We employed the same methodology to compare the *Proteus mirabilis* *Nasonia*-associated genomes from this study ( $n=7$ ) (Table 2) with high-quality, publicly available genomes from close relatives in humans ( $n=8$ ) (Table S1). We were interested in the diversity of *Proteus mirabilis* bacteria across related *Nasonia* species as they diverged 1 MYA, and their microbiomes exhibit strong signs of phyllosymbiosis (24, 27, 30). Importantly, *Proteus* bacteria exhibit low abundance in parental species, but  $F_2$  hybrids exhibit significant breakdown and lethality between the L3 and L4 larval stages; approximately 90% of the  $F_2$  hybrids die in conjunction with *Proteus* becoming the dominant bacterium that causes lethality (21). Consequently, whether the *Proteus mirabilis* bacteria in the hybrids are the same as or different from those in the parents remains a key question in understanding the nature of the bacterium-assisted hybrid breakdown. Our collection of *Proteus mirabilis* isolates encompasses two to three isolates each from *N. vitripennis*, *N. giraulti*, and  $F_2$  hybrid males to control for potential sex effects. Lastly, we also investigated how *Nasonia*-associated isolates compare to those isolated from human-associated environments. All *Proteus mirabilis* genomes are listed in Table S1.

With all genomes combined ( $n=15$ ), pangenomic analyses in *anvi'o* (57, 58) produced a total of 5,421 gene clusters. We grouped these gene clusters into three bins: (i) *Proteus* species Core (3,043 gene clusters), (ii) Human Specific (101 gene clusters), and (iii) *Nasonia* Specific (189 gene clusters) (Fig. 4A). Most gene clusters (56.1%) fall within the *Proteus* species Core bin where the majority of COG functions relate to [C] Energy production and conversion, [E] Amino acid transport and metabolism, and [J] Translation, ribosomal structure and biogenesis, the same three categories as reported for the *Providencia* species (Tables S3 and S5). Of the gene clusters not found in the core, 1.8% and 3.5% are shared just within human-associated and *Nasonia*-associated isolates, respectively, and the 38.5% remaining gene clusters are found within limited subsets of the genomes across isolates. Indeed, human-associated and *Nasonia*-associated *Proteus mirabilis* genomes share  $\geq 98.6\%$  average nucleotide identity (ANI) at the whole-genome level. Moreover, within *Nasonia*-associated isolates, the ANI increases to  $\geq 99.9\%$  identity (Fig. S1). Collectively, this evidence demonstrates that *Nasonia*-associated *Proteus mirabilis* isolates are highly similar to and slightly distinct from human *Proteus mirabilis* isolates. Furthermore, using a set of 71 core bacterial proteins, we determined *Nasonia*-associated *Proteus mirabilis* isolates are 99.9% identical and phylogenetically split from the human-associated strains (Fig. 2). Therefore, we next investigated what differences may exist between the genomes of (i) *Nasonia* species and hybrids and (ii) *Nasonia* and humans.

**(a) Functional differences between *Proteus mirabilis* isolates from *Nasonia*.** We investigated the strain-level diversity within our *Nasonia*-associated *Proteus mirabilis* isolates to distinguish whether or not the *Proteus* strains in hybrid offspring are the same as those in parental *Nasonia*. Pangenomic analyses produced a total of 3,699 gene clusters (Fig. 4B). Functional enrichment analyses found no significant difference between *Proteus mirabilis* isolated from either *N. vitripennis*, *N. giraulti*, or  $F_2$  hybrids. Therefore, the hybrid *Proteus mirabilis* genomes are functionally identical to the parental *Proteus mirabilis* isolates. We identified two regions where there was variability in the



**FIG 4** Pangenome of *Proteus mirabilis* human and *Nasonia* isolates. (A) The 15 inner layers correspond to all 15 genomes, with *Nasonia* isolates shown in maroon and human isolates shown in blue. The top horizontal layer underneath the tree represents host source (*Nasonia*, maroon; human, blue). (B) The seven inner maroon layers correspond to the seven *Nasonia*-associated genomes sequenced as part of this study. The top horizontal layer underneath the tree represents host source (*N. giraulti*, red;  $F_2$  hybrids, blue; *N. vitripennis*, green). For both panels A and B, the solid bars in each of the inner circles show the presence of gene-clusters (GCs) in a given genome while the outer green circle depicts known COGs (green) versus unknown (white) assignments. The outermost layer of solid bars and text highlight groups of GCs that correspond to the genome core or to group-specific GCs. Genomes are ordered based on their gene-cluster distribution across genomes, which is shown at the top right corner (tree). Central dendrograms depict protein cluster hierarchy when displayed as protein cluster frequency.

distribution of gene clusters that totaled  $<0.05\%$  of total gene clusters (Fig. 4B) and grouped them into a bin labeled “Variable Regions” for further investigation. Inspection of this bin identified 92 gene clusters that occurred in subsets of the *Nasonia*-associated *Proteus mirabilis* genomes, of which 68 (74%) are lacking known functional annotations. No consistent trend emerged regarding gene clusters found only within specific *Nasonia*-associated genomes (e.g., gene clusters were not consistently unique in only *N. vitripennis* and hybrid strains or *N. giraulti* and hybrid strains). Hierarchical clustering based on gene cluster frequencies places the  $F_2$  hybrid strains as more similar to paternal *N. vitripennis* strains based on their gene-cluster distribution across genomes, but *N. giraulti* strains are still closely associated within these relationships (Fig. 4B), signifying that these small numbers of differences are not enough to designate the specific parental origin of the  $F_2$  hybrid strains.

For three phages (phages 1 to 3), there is 100% nucleotide similarity and identical gene synteny across all seven *Proteus mirabilis* isolates in *Nasonia*; however, key differences including gene deletions and truncations are apparent in proteins within the five other phages also present (Fig. S3). These five phages have a nucleotide similarity ranging from 93.6 to 99.9% across the genomes, but the most significant dissimilarities are found within the phage tail region, the part of the phage that directly interacts with its bacterial host and often determines host specificity. For example, in phages 4 and 5 from Wasp.NgirL3.1, an early stop codon results in a predicted truncation in the tail spike protein, whereas across all other isolates containing these phages, the tail spike protein is intact. Relative to other isolates, phage 6 found in Wasp.NvitL3.2 and that found in Wasp.HybL3.2 both maintained the same two nucleotide deletions (4-mer and 8-mer) as well as mutations surrounding these deletions in the host specificity protein J gene. Altogether, comparative genomics of the phage coding regions

demonstrate that these isolates are extremely similar, and some of their phages exhibit distinct differences in the tail regions known to evolve fast, namely those that determine host specificity (91).

(b) *Functional differences between *Proteus mirabilis* isolates from *Nasonia* and human.*

Next, comparisons between the *Nasonia* and Human bins (Fig. 4A) revealed that the *Nasonia*-associated *Proteus mirabilis* isolates are particularly enriched for gene clusters with functions for [M] cell wall/membrane/envelope biogenesis (14.58% versus 1.59%) (Table S3). This includes gene functions for glycosyltransferases involved in lipopolysaccharide (LPS) biosynthesis (COG3306), peptidoglycan/LPS O-acetylase OafA/YrhL containing acyltransferase and SGNH-hydrolase domains (COG1835), and UDP-galactopyranose mutase (COG0562). LPS is a major component of Gram-negative bacterial cell walls that can be modified by glycosyltransferases and acyltransferases, and variations in the structure of LPS can provide selective advantages in different environments such as inhibiting bacteriophage binding, temperature tolerance, and antimicrobial resistance (92–95). Functional enrichment analyses in *anvi'o* identified the UDP-galactopyranose mutase COG function as present only in *Nasonia*-associated genomes ( $q=0.0173$ ) and responsible for the biosynthesis of galactofuranose, a sugar structure found in bacterial cell walls that is absent in mammals (96, 97) (Table S6). UDP-galactopyranose was shown to be essential for growth of *Mycobacterium tuberculosis* (98), although it remains unclear what function it may serve specifically in *Nasonia*-associated environments. Additionally, all *Nasonia*-associated genomes and a single human-associated genome (HI4320) encode muramidase (urinary lysozyme) (COG4678;  $q=0.0452$ ), although a similar homologous COG function (COG3772) exists in all human-associated genomes and in three *Nasonia*-associated genomes as well. Muramidase activity has been directly connected to the ability for *Proteus* to differentiate from vegetative to swarmer cells (99), a phenotype that is visible from bacteria isolated here (Fig. 1B).

We also investigated putative adaptive and innate bacterial defense systems within the isolates including CRISPR-Cas, restriction-modification (RM), and Bacteriophage Exclusion (BREX) systems. Although we found no evidence of CRISPR-Cas gene cassettes in any of our isolates, there were predicted type I RM and BREX systems. The specificity subunit (*hdsS*) of the complete type I RM system had an amino acid sequence identical to *hdsS* proteins (>95% nucleotide similarity) found in two other *Proteus mirabilis* strains (PmSDC32 from red junglefowl; CNR20130297 from human, as of September 2020). The putative type I BREX system (100, 101) component genes were intact without frameshifts or premature stop codons. Although the BREX mechanism of foreign DNA removal is not yet understood completely, unlike CRISPR-Cas and RM systems, the defense system does not involve restriction of foreign DNA (100, 101). However, both RM and BREX systems utilize methylation patterns to determine self/nonsel recognition. The absence of CRISPR-Cas systems in *Proteus* has been documented previously (102), although the absence has not been linked to expanded presence of mobile elements. The presence of the newly characterized BREX defense system in *Proteus mirabilis* could prevent phage integration, but the evolutionary history behind either the prophage integration events in *Proteus mirabilis* or the history of the BREX system is unknown. No publicly available *Proteus mirabilis* strains appear to have the same type I RM system (as determined by the conserved methyltransferase protein). *Proteus mirabilis* strains MPE5139 (CP053684.1) and Pm15C1 (KX268685) contain a similar BREX cassette (86% query coverage; 97.75% similarity), but strains placed phylogenetically close to the *Nasonia*-associated isolates such as PmSC1111 (CP034090) and BC11-24 (CP026571) do not. The acquisition of the BREX system after integration of the numerous phage genomes found within *Proteus mirabilis* could explain the apparent lack of phage inhibition, although we cannot say whether this is the case in the current study.

We next investigated the human-associated *Proteus mirabilis* bin and discovered it contains more unique gene clusters for functions relating to [P] inorganic ion transport and metabolism (9.56% versus 0% in *Nasonia* Specific bin) (Table S3). These gene

clusters encompass multiple COG functions relating to outer membrane receptors for iron (COG1629), ferrienterochelin, and colicins (COG4771) as well as copper chaperone proteins (COG2608) and oxidoreductases (NAD-binding) involved in siderophore biosynthesis (COG4693) (Table S5). In addition to exploring the bins, we also performed a functional enrichment analysis in *anvi'o* (57, 58) (Fig. 4A). This analysis identified that siderophore biosynthesis (COG4693), previously noted in the human-associated *Proteus*, occurs in 7 out of 8 of the human-associated genomes and is absent in all *Nasonia*-associated genomes ( $q=0.045$ ) (Table S6). Interestingly, the previously noted outer membrane receptor for ferrienterochelin and colicins (COG4771) has upregulated expression in pathogenic *E. coli* strains compared to commensals (103). Although homologous functions (different proteins with the same functional assignment) for outer membrane receptor proteins (COG1629 and COG4771) are also found in *Nasonia*-associated *Proteus mirabilis* genomes, the presence of oxidoreductase involved in siderophore biosynthesis (COG4693) and increased occurrence and diversity of genes relating to inorganic ion transport and metabolism in human-associated environments suggest that the more pathogenic strains may maintain a variety of antigens that could result in persistent inflammation and infection. When infection ensues, metal homeostasis can change dramatically, and therefore pathogens must be able to compete for the limited metal availability (104). Vertebrate hosts, including humans, can produce proteins such as the iron-sequestering lipocalin 2 to keep trace metals away from bacteria, which in turn creates a selective pressure for bacteria to evolve diverse iron-binding siderophores and a higher affinity toward binding trace metals for thriving in vertebrate environments (104–106). For example, uropathogenic *E. coli* in humans uses the siderophore enterobactin (of which ferrienterochelin is the iron complex) to resist lipocalin 2 (107). Therefore, the value of these proteins in human-associated environments can assist pathogens in resisting nutritional immunity imposed by the host (104, 108, 109).

Human-associated *Proteus mirabilis* isolates also encode  $5\times$  as many transposases as do *Nasonia*-associated *Proteus* isolates. Transposons permit the movement of DNA in and around the bacterial chromosome, which can help facilitate growth and adaptation of bacteria to their host environment (110, 111). Using the same functional enrichment analysis approach described above, we identified just two gene functions that are found in all human-associated *Proteus mirabilis* isolates ( $n=8$ ) and absent in all *Nasonia*-associated isolates ( $q=0.017$ ): (i) the type IV secretory pathway VirB4 component, and (ii) a predicted nuclease of a restriction endonuclease-like (RecB) superfamily, the DUF1016 family (Table S6). Type IV secretion systems secrete virulence factors in which the VirB4 subunit acts as an ATPase and is essential for some bacteria to cause infection (112, 113). Additionally, restriction endonucleases, such as RecB, cleave DNA at specific sites and act as defense mechanisms for bacteria against foreign DNA (114). Altogether, these differences in genomic content in human-associated strains suggest they may be more predisposed to succeed in environments where selective pressure may drive strategies for scavenging host metal nutrients and genome changes related to competition with mobile genetic elements.

**Conclusion.** Host-associated microbial communities often establish intimate and distinguishable relationships that assist host metabolism (115), development (116), behavior (117), and immune maturation (118), among others. Moreover, cross-system trends occur such as phylosymbiosis wherein microbial community ecological relationships recapitulate the host phylogenetic relationships (15). The *Nasonia* parasitoid wasp genus is a model system with phylosymbiosis in both the bacterial and viral compositions, hybrid maladies associated with the microbiome, and bacterial cultivability that permits further investigation of the host-microbe interactions that mediate these processes. Two bacterial genera dominate the larval microbiome of these wasps in pure species and hybrids, *Proteus mirabilis* and *Providencia rettgeri*, and here we have genomically characterized them relative to each other and to isolates from invertebrate and vertebrate animal hosts. There are three important findings from this work. First, this study reports the genome sequencing of bacteria from hybrid hosts. Recent

sequencing of hybrid deep-sea mussels showed that their symbionts were genetically indistinguishable from parental mussels (119), and this premise remains to be evaluated in other hybrid systems. Second, the bacterial species in insect-associated environments are not human contaminants as they differ from other host-associated environments and have adapted unique functions for survival that may more tightly regulate symbiotic relationships in insects. Third, the *Proteus mirabilis* genomic diversity is not unique between *Nasonia* species and hybrids, thus supporting a tenet of hologenomic speciation whereby the dominant bacterium in hybrids is a resident microbial taxon in parental species. Therefore, just as the same alleles of nuclear genes in parental species underpin lethality in hybrids, so do bacteria from parental species in the case of *Proteus mirabilis* and *Nasonia* hybrids. Whole-genome sequencing of both host and microbial constituents of this association now permits a deeper understanding of the multiomic interactions between resident members of the microbiome and the host, which in turn underpin phylosymbiosis and hybrid breakdown in these wasps.

## MATERIALS AND METHODS

**Nasonia rearing.** *Nasonia* species are interfertile in the absence of incompatible *Wolbachia* infections as a result of their recent evolutionary divergence (120), which allows us to take advantage of their haplodiploid sex determination to acquire  $F_2$  recombinant hybrid male offspring from virgin  $F_1$  mothers following parental crosses. We used the *Wolbachia*-uninfected lines *N. vitripennis* AsymCx, *N. giraulti* RV2x(u), and  $F_2$  hybrids from paternal AsymCx  $\times$  maternal RV2x(u) crosses. *Nasonia* wasps were reared under 25°C constant light on *Sarcophaga bullata* pupae. Hybrids were generated as previously described (21). Briefly, we collected virgin females and males from each parental species during early pupal development. Upon eclosion, parental adults were crossed in single-pair matings.  $F_1$  females were collected as virgins in early pupal stages and serially hosted after eclosion every 48 h on two *S. bullata* pupae to generate  $F_2$  haploid recombinant males (collected on third hosting). Parental strains were reared concurrently under identical conditions.

**Isolate cultivation and sequencing.** *Proteus* and *Providencia* bacteria were isolated from L3 larval stages of male *Nasonia giraulti* RV2x(u), *N. vitripennis* AsymCx, and  $F_2$  hybrids [paternal AsymCx  $\times$  maternal RV2x(u)], using tryptic soy agar (TSA) plates (Difco) containing 1.5% agar. For each wasp line, 10 larvae were collected, surface sterilized with 70% ethanol for 1 min, washed with 1 $\times$  phosphate-buffered saline (PBS), and resuspended in 20  $\mu$ l of 1 $\times$  PBS. The resuspended individuals were homogenized using sterile pestles, and serial dilutions were plated on TSA plates. Colonies with distinct morphology were subcultured on fresh TSA plates to ensure isolation and then stored as glycerol stocks (50% glycerol) at  $-80^\circ\text{C}$  for future characterization.

For cultivation and characterization, isolates were maintained on TSA solid agar plates or in TSA broth grown at 37°C, shaking at 130 rpm in liquid culture. For sequencing, multiple distinct isolates were selected from each sample group (*N. giraulti*, *N. vitripennis*, and  $F_2$  hybrids), and genomic DNA was extracted from 3 ml of overnight (tryptic soy broth [TSB]) broth culture using the ZR Duet DNA/RNA MiniPrep Plus kit (Zymo Research) following manufacturer's protocol for the "suspended cells" option. Isolates were designated via the addition of a letter corresponding to the host animal's species: G for *N. giraulti*, V for *N. vitripennis*, and H for hybrid. DNA was sent to the North Carolina State University's Genomic Sciences Laboratory and sequenced using a single flow cell of the Illumina MiSeq platform to produce 2-by-250-bp paired-end reads.

**Biofilm growth.** Colonies from two isolates (NvitL3-1V and NvitL3-3V) from *N. vitripennis* were grown in overnight cultures of LB broth at 37°C and 130 rpm. The next day, cultures were diluted to an  $\text{OD}_{600}$  of 0.05 for each isolate, corresponding to  $5 \times 10^5$  CFU per milliliter. A total of 1 ml was plated onto plastic cell culture 12-well plates in triplicate and incubated in a humid chamber overnight at 37°C. To prepare the coculture inoculate, *Proteus* and *Providencia* were mixed 1:1 before adding a total of 1 ml to the wells. After 24 h, the supernatant was removed and the biofilm was stained and measured using established protocols (121). Briefly, the culture was gently removed, and the plates were gently submerged into distilled water to remove unadhered cells. The plates were allowed to dry within a biosafety hood. The dried plates were then stained with 0.1% crystal violet for 10 min, washed three times in distilled water, and allowed to dry again under the hood. The stain retained was resuspended in 30% acetic acid, and the  $\text{OD}_{600}$  was measured and reported. Values were graphed in GraphPad Prism 8, and statistical significance was determined using a Kruskal-Wallis test with Dunn's correction for multiple comparisons.

To quantitatively determine the coculture composition of the *Proteus mirabilis* and *Providencia rettgeri* biofilm community, we developed a qPCR assay using primers specific for unique single-copy genes in each respective genome for *P. mirabilis* (5'-GGTGAGATTGTATTAATGG and 5'-ATCAGGAAGATGACGAG, annealing temperature 58°C) and *P. rettgeri* (5'-AACTCGGTCAGTCCAAACG and 5'-CTGCATTGTTTCGCTTCTCAC, annealing temperature 66°C). *Proteus* primers were designed for the *ureR* gene using a previously reported forward primer (122) and a new reverse primer. *Providencia* primers were designed based on a phage gene found only within the *Providencia* genomes. The biofilm experiment was repeated as described above except that once the supernatant was removed, the adherent cells in the biofilm were recovered from the 1:1 coculture in 1 ml of LB broth. The cells were then pelleted by spinning at  $10,000 \times g$  for 10 min, and supernatant was removed. DNA from the resulting cell pellet was extracted using the Genra Puregene tissue kit (Qiagen)



according to the manufacturer's protocol and diluted to ~10 nanograms/microliter each. Amplification was with Bio-Rad iTaq Universal SYBR green supermix in a CFX96 real-time C1000 thermal cycler (Bio-Rad, Hercules, CA), and each qPCR was performed at the following thermal profile: 95°C for 3 min, followed by 40 cycles of 95°C for 15 s and the respective annealing temperature for each qPCR primer for 1 min. Samples were calculated using a standard curve generated from dilutions of larger gene products amplified from the same genes for *P. mirabilis* (5'-GCGATTTTACACCGAGTTTC and 5'-ATCCCCATTCTGACATCAA) and *P. rettgeri* (5'-CCGTTGTGTGTTTGGTATCG and 5'-GTAAGCTGCGTGGATTGGTT). Primer specificity was determined *in silico* by BLASTing each primer sequence against each genome and by testing each primer pair on DNA from each bacterial species and observing no amplification either by PCR or by qPCR. Isolates and materials are available upon request.

**Bacterial genome assemblies and annotations.** All reads were trimmed using Trimmomatic v0.32 (123) and quality checked with FastQC (124). Reads were further filtered using the Geneious v. 11.1.5 filter and trim workflow (125). The reads of each isolate were assembled using SPAdes 3.13.0 (126), and the quality of each assembly was determined using QUAST (127); reads under 2 kb were discarded. Genomes were annotated using Prokka (128) and visualized in Geneious v11.0.3. Genome completion and contamination estimates were calculated using CheckM (48) and average nucleotide identity (ANI) using FastANI (54) online in KBase ([www.kbase.us](http://www.kbase.us)) (129). Prophages were identified using VirSorter (130), and plasmids were identified through Prokka annotations and increased numbers of reads mapping through Bowtie 2 (131).

**Genomic comparisons.** For phylogenetic placement of *Proteus* and *Providencia* bacteria isolated in this study, we aligned nucleotide sequences of the gyrase B subunit (*gyrB*) and RNA polymerase B subunit (*rpoB*) separately within Geneious 11.1.5 (125) using the multiple align tool. Using these alignments, we constructed separate unrooted maximum-likelihood (ML) phylogenies with 1,000 replicates for bootstrapping, using the GTR GAMMA model within RAxML for *gyrB* from both *P. rettgeri* and *P. mirabilis*. Evolutionary models were determined with jModelTest (132). Trees were visualized in FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Concatenated amino acid sequences for phylogenomic analyses were computed in *anvi'o* (57) using 'anvi-get-sequences-for-hmm-hits' with the flags `-return-best-hit`, `-get-aa-sequences`, and `-concatenate` with the `-hmm-source` `Bacteria_71`: [type: singlecopy] ADK, AICARFT\_IMPCHas, ATP-synt, ATP-synt\_A, Adenylsucc\_synt, Chorismate\_synt, EF\_TS, Exonuc\_VII\_L, GrpE, Ham1p\_like, IPPT, OSCP, PGK, Pept\_tRNA\_hydro, RBFA, RNA\_pol\_L, RNA\_pol\_Rpb6, RRF, RecO\_C, Ribonuclease\_P, Ribosom\_S12\_S23, Ribosomal\_L1, Ribosomal\_L13, Ribosomal\_L14, Ribosomal\_L16, Ribosomal\_L17, Ribosomal\_L18p, Ribosomal\_L19, Ribosomal\_L2, Ribosomal\_L20, Ribosomal\_L21p, Ribosomal\_L22, Ribosomal\_L23, Ribosomal\_L27, Ribosomal\_L27A, Ribosomal\_L28, Ribosomal\_L29, Ribosomal\_L3, Ribosomal\_L32p, Ribosomal\_L35p, Ribosomal\_L4, Ribosomal\_L5, Ribosomal\_L6, Ribosomal\_L9\_C, Ribosomal\_S10, Ribosomal\_S11, Ribosomal\_S13, Ribosomal\_S15, Ribosomal\_S16, Ribosomal\_S17, Ribosomal\_S19, Ribosomal\_S2, Ribosomal\_S20p, Ribosomal\_S3\_C, Ribosomal\_S6, Ribosomal\_S7, Ribosomal\_S8, Ribosomal\_S9, RsfS, RuvX, SecE, SecG, SecY, SmpB, TsaE, UPF0054, YajC, eIF-1a, ribosomal\_L24, tRNA-synt\_1d, tRNA\_m1G\_MT. Note that *Providencia rettgeri* strain Dmel1 was missing the gene for AICARFT\_IMPCHas, so this was manually removed from the *Providencia* alignment in Geneious resulting in a total of 70 genes. Concatenated alignments were imported into Geneious (Biomatters, Ltd., Auckland, New Zealand) (125) v11.0.3, and protein trees were built with PhyML using an LG substitution model and 100 bootstraps. Evolutionary models were determined with ModelTest-NG for amino acids (132).

We used *anvi'o* (57) v6.2 following the pangenomics workflow (58) to analyze pangenomes of our *Proteus* and *Providencia* isolates with publicly available reference genomes from the National Center for Biotechnology Information (NCBI) (see Table S1 in the supplemental material), respectively. If genomes contained plasmids, plasmids were included within the genome file. Briefly, we used the program 'anvi-script-FASTA-to-contigs-db' to convert genome fasta nucleotide files into contig databases for each genome which uses Prodigal (133) v2.6.2 for gene calling. We then annotated each gene using 'anvi-run-ncbi-contigs'. A genome storage file was created to collect each genome database using 'anvi-gen-genomes-storage', and the pangenome was computed using 'anvi-pan-genome' with flags `-mcl-inflation 10` and `-use-ncbi-blast`, which uses the MCL algorithm (59, 60) to identify clusters in amino acid sequence similarity search results and blastp (134) for the amino acid sequence similarity search. Genomes were classified by host (human versus insect) using 'anvi-import-misc-data', and functional enrichment analyses were performed using 'anvi-get-enriched-functions-per-pan-group' with `-annotation-source COG_FUNCTION`. We defined "core" genes of each species pangenome as gene clusters that were present in every genome and accessory genes as those present in only a subset of genomes (e.g., *Nasonia* specific gene clusters). Figures were visualized in *anvi'o* interactive interface, Inkscape version 1.0 (available from <https://inkscape.org/>), and finalized in Microsoft Office PowerPoint (v16.37).

**Data availability.** All whole-genome sequences were deposited in GenBank under BioProject PRJNA660265. BioSample accession numbers and further metadata are provided in Table S1 in the supplemental material.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, TIF file, 0.8 MB.

**FIG S2**, TIF file, 0.7 MB.

**FIG S3**, TIF file, 1.1 MB.

**TABLE S1**, XLSX file, 0.01 MB.

**TABLE S2**, XLSX file, 0.2 MB.

**TABLE S3**, XLSX file, 0.02 MB.

**TABLE S4**, XLSX file, 0.02 MB.

**TABLE S5**, XLSX file, 0.1 MB.

**TABLE S6**, XLSX file, 0.01 MB.

## ACKNOWLEDGMENTS

This work was supported by the Vanderbilt Microbiome Initiative to S.R.B., National Institutes of Health Ruth Kirschstein Postdoctoral Fellowship to B.A.L., Deutsche Forschungsgemeinschaft (DFG) postdoctoral fellowship (MI 2242/1-1) to A.M., and Vanderbilt University SyBBURE Searle Undergraduate Research Program to A.K.M. Any opinion, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institutes of Health, the National Science Foundation, or Vanderbilt University.

A.M. and S.R.B. conceptualized the study and designed experiments. A.M. performed microbial isolation and genome sequencing. K.L.C., B.A.L., E.A.H., and A.K.M. carried out experiments and performed data analysis. K.L.C. and B.A.L. analyzed the genomes and performed pangenomic analyses. K.L.C. and S.R.B. wrote the manuscript with input from B.A.L. S.R.B. supervised the project.

## REFERENCES

- Shropshire JD, Bordenstein SR. 2016. Speciation by symbiosis: the microbiome and behavior. *mBio* 7:e01785-15. <https://doi.org/10.1128/mBio.01785-15>.
- Rosengaus RB, Zecher CN, Schultheis KF, Brucker RM, Bordenstein SR. 2011. Disruption of the termite gut microbiota and its prolonged consequences for fitness. *Appl Environ Microbiol* 77:4303-4312. <https://doi.org/10.1128/AEM.01886-10>.
- Jing T-Z, Qi F-H, Wang Z-Y. 2020. Most dominant roles of insect gut bacteria: digestion, detoxification, or essential nutrient provision? *Microbiome* 8:38. <https://doi.org/10.1186/s40168-020-00823-y>.
- Duploux A, Minard G, Saastamoinen M. 2020. The gut bacterial community affects immunity but not metabolism in a specialist herbivorous butterfly. *Ecol Evol* 10:8755-8769. <https://doi.org/10.1002/ece3.6573>.
- Wang Y, Rozen DE. 2018. Gut microbiota in the burying beetle, *Nicrophorus vespilloides*, provide colonization resistance against larval bacterial pathogens. *Ecol Evol* 8:1646-1654. <https://doi.org/10.1002/ece3.3589>.
- Msaad Guerfali M, Djobbi W, Charaabi K, Hamden H, Fadhl S, Marzouki W, Dhaouedi F, Chevrier C. 2018. Evaluation of *Providencia rettgeri* pathogenicity against laboratory Mediterranean fruit fly strain (*Ceratitis capitata*). *PLoS One* 13:e0196343. <https://doi.org/10.1371/journal.pone.0196343>.
- Leigh BA, Bordenstein SR, Brooks AW, Mikaelyan A, Bordenstein S. 2018. Finer-scale phylosymbiosis: insights from insect viromes. *mSystems* 3:e00131-18. <https://doi.org/10.1128/mSystems.00131-18>.
- Kohl KD, Dearing MD, Bordenstein SR. 2018. Microbial communities exhibit host species distinguishability and phylosymbiosis along the length of the gastrointestinal tract. *Mol Ecol* 27:1874-1883. <https://doi.org/10.1111/mec.14460>.
- Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. 2016. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biol* 14:e2000225. <https://doi.org/10.1371/journal.pbio.2000225>.
- Bordenstein SR, Theis KR. 2015. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol* 13:e1002226. <https://doi.org/10.1371/journal.pbio.1002226>.
- Díaz-Sánchez S, Estrada-Peña A, Cabezas-Cruz A, de la Fuente J. 2019. Evolutionary insights into the tick hologenome. *Trends Parasitol* 35:725-737. <https://doi.org/10.1016/j.pt.2019.06.014>.
- Dunaj SJ, Bettencourt BR, Garb JE, Brucker RM. 2020. Spider phylosymbiosis: divergence of widow spider species and their tissues' microbiomes. *BMC Evol Biol* 20:104. <https://doi.org/10.1186/s12862-020-01664-x>.
- Song SJ, Sanders JG, Delsuc F, Metcalf J, Amato K, Taylor MW, Mazel F, Lutz HL, Winker K, Graves GR, Humphrey G, Gilbert JA, Hackett SJ, White KP, Skeen HR, Kurtis SM, Withrow J, Braile T, Miller M, McCracken KG, Maley JM, Ezenwa VO, Williams A, Blanton JM, McKenzie VJ, Knight R. 2020. Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. *mBio* 11:e02901-19. <https://doi.org/10.1128/mBio.02901-19>.
- Pollock FJ, McMinds R, Smith S, Bourne DG, Willis BL, Medina M, Thurber RV, Zaneveld JR. 2018. Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nat Commun* 9:4921. <https://doi.org/10.1038/s41467-018-07275-x>.
- Lim SJ, Bordenstein SR. 2020. An introduction to phylosymbiosis. *Proc Biol Sci* 287:20192900. <https://doi.org/10.1098/rspb.2019.2900>.
- Dietrich C, Köhler T, Brune A. 2014. The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl Environ Microbiol* 80:2261-2269. <https://doi.org/10.1128/AEM.04206-13>.
- Amato KR, Sanders JG, Song SJ, Nute M, Metcalf JL, Thompson LR, Morton JT, Amir A, McKenzie VJ, Humphrey G, Gogul G, Gaffney J, Baden AL, Britton GAO, Cuzzo FP, Di Fiore A, Dominy NJ, Goldberg TL, Gomez A, Kowalewski MM, Lewis RJ, Link A, Sauter ML, Tecot S, White BA, Nelson KE, Stumpf RM, Knight R, Leigh SR. 2019. Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *ISME J* 13:576-587. <https://doi.org/10.1038/s41396-018-0175-0>.
- Chiarello M, Auguet J-C, Bettarel Y, Bouvier C, Claverie T, Graham NAJ, Rieuvilleneuve F, Sucre E, Bouvier T, Villéger S. 2018. Skin microbiome of coral reef fish is highly variable and driven by host phylogeny and diet. *Microbiome* 6:147. <https://doi.org/10.1186/s40168-018-0530-4>.
- Doane MP, Morris MM, Papudeshi B, Allen L, Pande D, Haggerty JM, Johri S, Turnlund AC, Peterson M, Kacev D, Nosal A, Ramirez D, Hovel K, Ledbetter J, Alker A, Avalos J, Baker K, Bhide S, Billings E, Byrum S, Clemens M, Demery AJ, Lima LFO, Gomez O, Gutierrez O, Hinton S, Kieu D, Kim A, Loaiza R, Martinez A, McGhee J, Nguyen K, Parlan S, Pham A, Price-Waldman R, Edwards RA, Dinsdale EA. 2020. The skin microbiome of elasmobranchs follows phylosymbiosis, but in teleost fishes, the microbiomes converge. *Microbiome* 8:93. <https://doi.org/10.1186/s40168-020-00840-x>.
- Bouffaud M-L, Poirier M-A, Muller D, Moëgne-Loccoz Y. 2014. Root microbiome relates to plant host evolution in maize and other Poaceae: Poaceae evolution and root bacteria. *Environ Microbiol* 16:2804-2814. <https://doi.org/10.1111/1462-2920.12442>.

21. Brucker RM, Bordenstein SR. 2013. The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus *Nasonia*. *Science* 341:667–669. <https://doi.org/10.1126/science.1240659>.
22. Li Z, Wright A-DG, Si H, Wang X, Qian W, Zhang Z, Li G. 2016. Changes in the rumen microbiome and metabolites reveal the effect of host genetics on hybrid crosses. *Environ Microbiol Rep* 8:1016–1023. <https://doi.org/10.1111/1758-2229.12482>.
23. Edwards JE, Schennink A, Burden F, Long S, van Doorn DA, Pellikaan WF, Dijkstra J, Saccenti E, Smidt H. 2020. Domesticated equine species and their derived hybrids differ in their fecal microbiota. *Anim Microbiome* 2:8. <https://doi.org/10.1186/s42523-020-00027-7>.
24. van Opstal EJ, Bordenstein SR. 2019. Phylosymbiosis impacts adaptive traits in *Nasonia* wasps. *mBio* 10:e00887-19. <https://doi.org/10.1128/mBio.00887-19>.
25. Wang J, Kalyan S, Steck N, Turner LM, Harr B, Künzel S, Vallier M, Häslér R, Franke A, Oberg H-H, Ibrahim SM, Grassl GA, Kabelitz D, Baines JF. 2015. Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. *Nat Commun* 6:6440. <https://doi.org/10.1038/ncomms7440>.
26. Campbell BC, Steffen-Campbell JD, Werren JH. 1993. Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28S rDNA sequences. *Insect Mol Biol* 2:225–237. <https://doi.org/10.1111/j.1365-2583.1994.tb00142.x>.
27. Brucker RM, Bordenstein SR. 2012. The roles of host evolutionary relationships (genus: *Nasonia*) and development in structuring microbial communities. *Evolution* 66:349–362. <https://doi.org/10.1111/j.1558-5646.2011.01454.x>.
28. Beukeboom LW, Koevoets T, Morales HE, Ferber S, van de Zande L. 2015. Hybrid incompatibilities are affected by dominance and dosage in the haplodiploid wasp *Nasonia*. *Front Genet* 6:140. <https://doi.org/10.3389/fgene.2015.00140>.
29. Shropshire JD, van Opstal EJ, Bordenstein SR. 2016. An optimized approach to germ-free rearing in the jewel wasp *Nasonia*. *PeerJ* 4:e2316. <https://doi.org/10.7717/peerj.2316>.
30. Brucker RM, Bordenstein SR. 2012. Speciation by symbiosis. *Trends Ecol Evol* 27:443–451. <https://doi.org/10.1016/j.tree.2012.03.011>.
31. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, *Nasonia* Genome Working Group, Beukeboom LW, Desplan C, Elsik CG, Grimelikhuijzen CJ, Kitts P, Lynch JA, Murphy T, Oliveira DCSG, Smith CD, van de Zande L, Worley KC, Zdobnov EM, Aerts M, Albert S, Anaya VH, Anzola JM, Barchuk AR, Behura SK, Bera AN, Berenbaum MR, Bertossa RC, Bitondi MMG, Bordenstein SR, Bork P, Bornberg-Bauer E, Brunain M, Cazzamali G, Chaboub L, Chacko J, Chavez D, Childers CP, Choi J-H, Clark ME, Claudianos C, Clinton RA, Cree AG, Cristino AS, Dang PM, et al. 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327:343–348. <https://doi.org/10.1126/science.1178028>.
32. Wang X, Kelkar YD, Xiong X, Martinson EO, Lynch J, Zhang C, Werren JH, Wang X. 2020. Genome report: whole genome sequence and annotation of the parasitoid jewel wasp *Nasonia giraulti* laboratory strain RV2X[u]. *G3 (Bethesda)* 10:2565–2572. <https://doi.org/10.1534/g3.120.401200>.
33. Bitar I, Mattioni Marchetti V, Mercato A, Nucleo E, Anesi A, Bracco S, Rognoni V, Hrabak J, Migliavacca R. 2020. Complete genome and plasmid sequences of a clinical *Proteus mirabilis* isolate producing plasmid mediated NDM-1 from Italy. *Microorganisms* 8:339. <https://doi.org/10.3390/microorganisms8030339>.
34. Di Pilato V, Chiarelli A, Boinett CJ, Riccobono E, Harris SR, D'Andrea MM, Thomson NR, Rossolini G, Giani T. 2016. Complete genome sequence of the first KPC-type carbapenemase-positive *Proteus mirabilis* strain from a bloodstream infection. *Genome Announc* 4:e00607-16. <https://doi.org/10.1128/genomeA.00607-16>.
35. Pearson MM, Sebahia M, Churcher C, Quail MA, Seshasayee AS, Luscombe NM, Abdellah Z, Arrosmith C, Atkin B, Chillingworth T, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabinowitsch E, Walker D, Whithead S, Thomson NR, Rather PN, Parkhill J, Mobley HLT. 2008. Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. *J Bacteriol* 190:4027–4037. <https://doi.org/10.1128/JB.01981-07>.
36. Octavia S, Xu W, Ng OT, Marimuthu K, Venkatachalam I, Cheng B, Lin RTP, Teo JWP. 2020. Identification of AbaR4 Acinetobacter baumannii resistance island in clinical isolates of blaOXA-23-positive *Proteus mirabilis*. *J Antimicrob Chemother* 75:521–525. <https://doi.org/10.1093/jac/dkz472>.
37. Schaffer JN, Pearson MM. 2015. *Proteus mirabilis* and urinary tract infections. *Microbiol Spectr* 3(5). <https://doi.org/10.1128/microbiolspec.UTI-0017-2013>.
38. Galac MR, Lazzaro BP. 2012. Comparative genomics of bacteria in the genus *Providencia* isolated from wild *Drosophila melanogaster*. *BMC Genomics* 13:612. <https://doi.org/10.1186/1471-2164-13-612>.
39. Unckless RL, Howick VM, Lazzaro BP. 2016. Convergent balancing selection on an antimicrobial peptide in *Drosophila*. *Curr Biol* 26:257–262. <https://doi.org/10.1016/j.cub.2015.11.063>.
40. Hanson MA, Dostálová A, Ceroni C, Poidevin M, Kondo S, Lemaitre B. 2019. Synergy and remarkable specificity of antimicrobial peptides in vivo using a systematic knockout approach. *Elife* 8:e44341. <https://doi.org/10.7554/eLife.44341>.
41. O'Donnell MP, Fox BW, Chao P-H, Schroeder FC, Sengupta P. 2020. A neurotransmitter produced by gut bacteria modulates host sensory behaviour. *Nature* 583:415–420. <https://doi.org/10.1038/s41586-020-2395-5>.
42. Wang G-H, Brucker RM. 2019. Genome sequence of *Providencia rettgeri* NVIT03, isolated from *Nasonia vitripennis*. *Microbiol Resour Announc* 8:e01157-18. <https://doi.org/10.1128/MRA.01157-18>.
43. Tchuente PLS, Rabenandrasana MAN, Ramparany L, Ratsima E, Enouf V, Randrianirina F, Collard J-M. 2020. Genome-based insights into the resistomes and mobilomes of two *Providencia rettgeri* strains isolated from wound infections in Madagascar. *J Glob Antimicrob Resist* 20:178–182. <https://doi.org/10.1016/j.jgar.2019.07.013>.
44. Marquez-Ortiz RA, Haggerty L, Sim EM, Duarte C, Castro-Cardozo BE, Beltran M, Saavedra S, Vanegas N, Escobar-Perez J, Petty NK. 2017. First complete *Providencia rettgeri* genome sequence, the NDM-1-producing clinical strain RB151. *Genome Announc* 5:e01472-16. <https://doi.org/10.1128/genomeA.01472-16>.
45. Maltz MA, Weiss BL, O'Neill M, Wu Y, Aksoy S. 2012. OmpA-mediated biofilm formation is essential for the commensal bacterium *Sodalis glossinidius* to colonize the tsetse fly gut. *Appl Environ Microbiol* 78:7760–7768. <https://doi.org/10.1128/AEM.01858-12>.
46. Drzewiecka D. 2016. Significance and roles of *Proteus* spp. bacteria in natural environments. *Microb Ecol* 72:741–758. <https://doi.org/10.1007/s00248-015-0720-6>.
47. Budding AE, Ingham CJ, Bitter W, Vandenbroucke-Grauls CM, Schneeberger PM. 2009. The Dienes phenomenon: competition and territoriality in swarming *Proteus mirabilis*. *J Bacteriol* 191:3892–3900. <https://doi.org/10.1128/JB.00975-08>.
48. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
49. O'Hara CM, Brenner FW, Miller JM. 2000. Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. *Clin Microbiol Rev* 13:534–546. <https://doi.org/10.1128/CMR.13.4.534>.
50. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. 2012. Human gut microbiome viewed across age and geography. *Nature* 486:222–227. <https://doi.org/10.1038/nature11053>.
51. Gu W, Wang W, Tong P, Liu C, Jia J, Lu C, Han Y, Sun X, Kuang D, Li N, Dai J. 2020. Comparative genomic analysis of *Proteus* spp. isolated from tree shrews indicated unexpectedly high genetic diversity. *PLoS One* 15:e0229125. <https://doi.org/10.1371/journal.pone.0229125>.
52. Gu W, Tong P, Liu C, Wang W, Lu C, Han Y, Sun X, Kuang DX, Li N, Dai J. 2019. The characteristics of gut microbiota and commensal Enterobacteriaceae isolates in tree shrew (*Tupaia belangeri*). *BMC Microbiol* 19:203. <https://doi.org/10.1186/s12866-019-1581-9>.
53. Ma Q, Fonseca A, Liu W, Fields AT, Pimsler ML, Spindola AF, Tarone AM, Crippen TL, Tomberlin JK, Wood TK. 2012. *Proteus mirabilis* interkingdom swarming signals attract blow flies. *ISME J* 6:1356–1366. <https://doi.org/10.1038/ismej.2011.210>.
54. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.
55. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. 2020. Diversity within species: interpreting strains in microbiomes. *Nat Rev Microbiol* 18:491–506. <https://doi.org/10.1038/s41579-020-0368-1>.
56. Sichtig H, Minogue T, Yan Y, Stefan C, Hall A, Tallon L, Sadzewicz L, Nadendla S, Klimke W, Hatcher E, Shumway M, Aldea DL, Allen J, Koehler J, Slezak T, Lovell S, Schoepp R, Scherf U. 2019. FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and

- regulatory science. *Nat Commun* 10:3313. <https://doi.org/10.1038/s41467-019-11306-6>.
57. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. <https://doi.org/10.7717/peerj.1319>.
  58. Delmont TO, Eren AM. 2018. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 6:e4320. <https://doi.org/10.7717/peerj.4320>.
  59. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584. <https://doi.org/10.1093/nar/30.7.1575>.
  60. Van Dongen S. 2008. Graph clustering via a discrete uncoupling process. *SIAM J Matrix Anal Appl* 30:121–141. <https://doi.org/10.1137/040608635>.
  61. Murthy AMV, Phan M-D, Peters KM, Nhu NTK, Welch RA, Ulett GC, Schembri MA, Sweet MJ. 2018. Regulation of hemolysin in uropathogenic *Escherichia coli* fine-tunes killing of human macrophages. *Virulence* 9:967–980. <https://doi.org/10.1080/21505594.2018.1465786>.
  62. Spaulding CN, Schreiber HL, Zheng W, Dodson KW, Hazen JE, Conover MS, Wang F, Svenmarker P, Luna-Rico A, Francetic O, Andersson M, Hultgren S, Egelman EH. 2018. Functional role of the type 1 pilus rod structure in mediating host-pathogen interactions. *Elife* 7:e31662. <https://doi.org/10.7554/eLife.31662>.
  63. Martinez JJ, Mulvey MA, Schilling JD, Pinkner JS, Hultgren SJ. 2000. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO J* 19:2803–2812. <https://doi.org/10.1093/emboj/19.12.2803>.
  64. Jacob-Dubuisson F, Loch C, Antoine R. 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Mol Microbiol* 40:306–313. <https://doi.org/10.1046/j.1365-2958.2001.02278.x>.
  65. Mobley HLT, Chippendale GR, Tenney JH, Mayer AR, Crisp LJ, Penner JL, Warren JW. 1988. MR/K hemagglutination of *Providencia stuartii* correlates with adherence to catheters and with persistence in catheter-associated bacteriuria. *J Infect Dis* 157:264–271. <https://doi.org/10.1093/infdis/157.2.264>.
  66. Loch C, Geoffroy MC, Renaud G. 1992. Common accessory genes for the *Bordetella pertussis* filamentous hemagglutinin and fimbriae share sequence similarities with the papC and papD gene families. *EMBO J* 11:3175–3183. <https://doi.org/10.1002/j.1460-2075.1992.tb05394.x>.
  67. Barrangou R, Marraffini LA. 2014. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol Cell* 54:234–244. <https://doi.org/10.1016/j.molcel.2014.03.011>.
  68. Giraud M-F, Naismith JH. 2000. The rhamnose pathway. *Curr Opin Struct Biol* 10:687–696. [https://doi.org/10.1016/S0959-440X\(00\)00145-7](https://doi.org/10.1016/S0959-440X(00)00145-7).
  69. Rodionova IA, Li X, Thiel V, Stolyar S, Stanton K, Fredrickson JK, Bryant DA, Osterman AL, Best AA, Rodionov DA. 2013. Comparative genomics and functional analysis of rhamnose catabolic pathways and regulons in bacteria. *Front Microbiol* 4:407. <https://doi.org/10.3389/fmicb.2013.00407>.
  70. Boronat A, Aguilar J. 1981. Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: differences in induction of propanediol oxidoreductase. *J Bacteriol* 147:181–185. <https://doi.org/10.1128/JB.147.1.181-185.1981>.
  71. Carvalho F, Atilano ML, Pombinho R, Covas G, Gallo RL, Filipe SR, Sousa S, Cabanes D. 2015. L-rhamnosylation of *Listeria monocytogenes* wall teichoic acids promotes resistance to antimicrobial peptides by delaying interaction with the membrane. *PLoS Pathog* 11:e1004919. <https://doi.org/10.1371/journal.ppat.1004919>.
  72. Delmas J, Gibold L, Faïs T, Batista S, Lereboure M, Sinel C, Vazeille E, Cattoir V, Buisson A, Barnich N, Dalmaso G, Bonnet R. 2019. Metabolic adaptation of adherent-invasive *Escherichia coli* to exposure to bile salts. *Sci Rep* 9:2175. <https://doi.org/10.1038/s41598-019-38628-1>.
  73. Otto BR, Sparrius M, Verweij-van Vught AM, MacLaren DM. 1990. Iron-regulated outer membrane protein of *Bacteroides fragilis* involved in heme uptake. *Infect Immun* 58:3954–3958. <https://doi.org/10.1128/AI.58.12.3954-3958.1990>.
  74. Klebba PE, McIntosh MA, Neilands JB. 1982. Kinetics of biosynthesis of iron-regulated membrane proteins in *Escherichia coli*. *J Bacteriol* 149:880–888. <https://doi.org/10.1128/JB.149.3.880-888.1982>.
  75. Tang X, Zhou B. 2013. Iron homeostasis in insects: insights from *Drosophila* studies. *IUBMB Life* 65:863–872. <https://doi.org/10.1002/iub.1211>.
  76. Kremer N, Voronin D, Charif D, Mavingui P, Mollereau B, Vavre F. 2009. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. *PLoS Pathog* 5:e1000630. <https://doi.org/10.1371/journal.ppat.1000630>.
  77. Resch CT, Winogrodzki JL, Patterson CT, Lind EJ, Quinn MJ, Dibrov P, Häse CC. 2010. The putative Na<sup>+</sup>/H<sup>+</sup> antiporter of *Vibrio cholerae*, VcNhaP2, mediates the specific K<sup>+</sup>/H<sup>+</sup> exchange in vivo. *Biochemistry* 49:2520–2528. <https://doi.org/10.1021/bi902173y>.
  78. Waditee R, Buaboocha T, Kato M, Hibino T, Suzuki S, Nakamura T, Takabe T. 2006. Carboxyl-terminal hydrophilic tail of a NhaP type Na<sup>+</sup>/H<sup>+</sup> antiporter from cyanobacteria is involved in the apparent affinity for Na<sup>+</sup> and pH sensitivity. *Arch Biochem Biophys* 450:113–121. <https://doi.org/10.1016/j.abb.2006.02.013>.
  79. Geiger O, López-Lara IM, Sohlenkamp C. 2013. Phosphatidylcholine biosynthesis and function in bacteria. *Biochim Biophys Acta* 1831:503–513. <https://doi.org/10.1016/j.bbalip.2012.08.009>.
  80. Sohlenkamp C, López-Lara IM, Geiger O. 2003. Biosynthesis of phosphatidylcholine in bacteria. *Prog Lipid Res* 42:115–162. [https://doi.org/10.1016/S0163-7827\(02\)00050-4](https://doi.org/10.1016/S0163-7827(02)00050-4).
  81. Aktas M, Wessel M, Hacker S, Klüsener S, Gleichenhagen J, Narberhaus F. 2010. Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic cells. *Eur J Cell Biol* 89:888–894. <https://doi.org/10.1016/j.ejcb.2010.06.013>.
  82. Zhang G, Liu Y, Ruoho AE, Hurler JH. 1997. Structure of the adenylyl cyclase catalytic core. *Nature* 386:247–253. <https://doi.org/10.1038/386247a0>.
  83. Iyer LM, Aravind L. 2002. The catalytic domains of thiamine triphosphatase and CyaB-like adenylyl cyclase define a novel superfamily of domains that bind organic phosphates. *BMC Genomics* 3:33. <https://doi.org/10.1186/1471-2164-3-33>.
  84. Smith RS, Wolfgang MC, Lory S. 2004. An adenylate cyclase-controlled signaling network regulates *Pseudomonas aeruginosa* virulence in a mouse model of acute pneumonia. *Infect Immun* 72:1677–1684. <https://doi.org/10.1128/iai.72.3.1677-1684.2004>.
  85. Murray NE. 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* 64:412–434. <https://doi.org/10.1128/mmb.64.2.412-434.2000>.
  86. Makarova KS, Wolf YI, van der Oost J, Koonin EV. 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* 4:29. <https://doi.org/10.1186/1745-6150-4-29>.
  87. Hegge JW, Swarts DC, van der Oost J. 2018. Prokaryotic Argonaute proteins: novel genome-editing tools? *Nat Rev Microbiol* 16:5–11. <https://doi.org/10.1038/nrmicro.2017.73>.
  88. Kuzmenko A, Oguienko A, Esyunina D, Yudin D, Petrova M, Kudinova A, Maslova O, Ninova M, Ryazansky S, Leach D, Aravin AA, Kulbachinskiy A. 2020. DNA targeting and interference by a bacterial Argonaute nuclease. *Nature* 587:632–637. <https://doi.org/10.1038/s41586-020-2605-1>.
  89. Tseng T-T, Tyler BM, Setubal JC. 2009. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* 9:52. <https://doi.org/10.1186/1471-2180-9-S1-52>.
  90. Mota LJ, Cornelis GR. 2005. The bacterial injection kit: type III secretion systems. *Ann Med* 37:234–249. <https://doi.org/10.1080/07853890510037329>.
  91. Holtzman T, Globus R, Molshanski-Mor S, Ben-Shem A, Yosef I, Qimron U. 2020. A continuous evolution system for contracting the host range of bacteriophage T7. *Sci Rep* 10:307. <https://doi.org/10.1038/s41598-019-57221-0>.
  92. Li Y, Powell DA, Shaffer SA, Rasko DA, Pelletier MR, Leszyk JD, Scott AJ, Masoudi A, Goodlett DR, Wang X, Rietz CRH, Ernst RK. 2012. LPS remodeling is an evolved survival strategy for bacteria. *Proc Natl Acad Sci U S A* 109:8716–8721. <https://doi.org/10.1073/pnas.1202908109>.
  93. Maldonado RF, Sá-Correia I, Valvano MA. 2016. Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS Microbiol Rev* 40:480–493. <https://doi.org/10.1093/femsre/fuw007>.
  94. Knirel YA, Prokhorov NS, Shashkov AS, Ovchinnikova OG, Zdorovenko EL, Liu B, Kostryukova ES, Larin AK, Koonidova AK, Letarov AV. 2015. Variations in O-antigen biosynthesis and O-acetylation associated with altered phage sensitivity in *Escherichia coli* 4s. *J Bacteriol* 197:905–912. <https://doi.org/10.1128/JB.02398-14>.
  95. Kulikov EE, Majewska J, Prokhorov NS, Golomidova AK, Tatarskiy EV, Letarov AV. 2017. Effect of O-acetylation of O antigen of *Escherichia coli* lipopolysaccharide on the nonspecific barrier function of the outer membrane. *Microbiology* 86:310–316. <https://doi.org/10.1134/S0026261717030080>.
  96. Tanner JJ, Boechi L, Andrew McCammon J, Sobrado P. 2014. Structure, mechanism, and dynamics of UDP-galactopyranose mutase. *Arch Biochem Biophys* 544:128–141. <https://doi.org/10.1016/j.abb.2013.09.017>.
  97. Soltero-Higgin M, Carlson EE, Gruber TD, Kiessling LL. 2004. A unique catalytic mechanism for UDP-galactopyranose mutase. *Nat Struct Mol Biol* 11:539–543. <https://doi.org/10.1038/nsmb772>.

98. Pan F, Jackson M, Ma Y, McNeil M. 2001. Cell wall core galactofuran synthesis is essential for growth of mycobacteria. *J Bacteriol* 183:3991–3998. <https://doi.org/10.1128/JB.183.13.3991-3998.2001>.
99. Strating H, Vandenende C, Clarke AJ. 2012. Changes in peptidoglycan structure and metabolism during differentiation of *Proteus mirabilis* into swarmer cells. *Can J Microbiol* 58:1183–1194. <https://doi.org/10.1139/w2012-102>.
100. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J* 34:169–183. <https://doi.org/10.15252/embj.201489455>.
101. Chaudhary K. 2018. Bacteriophage EXclusion (BREX): a novel anti-phage mechanism in the arsenal of bacterial defense system. *J Cell Physiol* 233:771–773. <https://doi.org/10.1002/jcp.25973>.
102. Alam MS, Chakraborty S, Rahman T, Hosen MI, Paul A, Hasan AKMM, Hossain MA. 2018. Investigation of the potential association between clustered regularly interspersed short palindromic repeats (CRISPR) and antibiotic resistance pattern of bacterial strains isolated from medical waste and environmental water. *Open J Med Microbiol* 8:13–25. <https://doi.org/10.4236/ojmm.2018.82002>.
103. Sommer U, Petersen J, Pfeiffer M, Schrotz-King P, Morszczek C. 2010. Comparison of surface proteomes of enterotoxigenic (ETEC) and commensal *Escherichia coli* strains. *J Microbiol Methods* 83:13–19. <https://doi.org/10.1016/j.mimet.2010.07.011>.
104. Palmer LD, Skaar EP. 2016. Transition metals and virulence in bacteria. *Annu Rev Genet* 50:67–91. <https://doi.org/10.1146/annurev-genet-120215-035146>.
105. Flo TH, Smith KD, Sato S, Rodriguez DJ, Holmes MA, Strong RK, Akira S, Aderem A. 2004. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestering iron. *Nature* 432:917–921. <https://doi.org/10.1038/nature03104>.
106. Abergel RJ, Wilson MK, Arceneaux JEL, Hoette TM, Strong RK, Byers BR, Raymond KN. 2006. Anthrax pathogen evades the mammalian immune system through stealth siderophore production. *Proc Natl Acad Sci U S A* 103:18499–18503. <https://doi.org/10.1073/pnas.0607055103>.
107. Shields-Cutler RR, Crowley JR, Hung CS, Stapleton AE, Aldrich CC, Marschall J, Henderson JP. 2015. Human urinary composition controls antibacterial activity of siderocalin. *J Biol Chem* 290:15949–15960. <https://doi.org/10.1074/jbc.M115.645812>.
108. Perry WJ, Spraggins JM, Sheldon JR, Grunenwald CM, Heinrichs DE, Cassat JE, Skaar EP, Caprioli RM. 2019. *Staphylococcus aureus* exhibits heterogeneous siderophore production within the vertebrate host. *Proc Natl Acad Sci U S A* 116:21980–21982. <https://doi.org/10.1073/pnas.1913991116>.
109. Tiwari V, Rajeswari MR, Tiwari M. 2019. Proteomic analysis of iron-regulated membrane proteins identify FhuE receptor as a target to inhibit siderophore-mediated iron acquisition in *Acinetobacter baumannii*. *Int J Biol Macromol* 125:1156–1167. <https://doi.org/10.1016/j.ijbiomac.2018.12.173>.
110. Fan C, Wu Y-H, Decker CM, Rohani R, Gesell Salazar M, Ye H, Cui Z, Schmidt F, Huang WE. 2019. Defensive function of transposable elements in bacteria. *ACS Synth Biol* 8:2141–2151. <https://doi.org/10.1021/acssynbio.9b00218>.
111. Hartl DL, Dykhuizen DE, Miller RD, Green L, de Framond J. 1983. Transposable element IS50 improves growth rate of *E. coli* cells without transposition. *Cell* 35:503–510. [https://doi.org/10.1016/0092-8674\(83\)90184-8](https://doi.org/10.1016/0092-8674(83)90184-8).
112. Sgro GG, Oka GU, Souza DP, Cenens W, Bayer-Santos E, Matsuyama BY, Bueno NF, dos Santos TR, Alvarez-Martinez CE, Salinas RK, Farah CS. 2019. Bacteria-killing type IV secretion systems. *Front Microbiol* 10:1078. <https://doi.org/10.3389/fmicb.2019.01078>.
113. Schulein R, Dehio C. 2002. The VirB/VirD4 type IV secretion system of *Bartonella* is essential for establishing intraerythrocytic infection. *Mol Microbiol* 46:1053–1067. <https://doi.org/10.1046/j.1365-2958.2002.03208.x>.
114. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG, Murray NE. 2014. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res* 42:3–19. <https://doi.org/10.1093/nar/gkt990>.
115. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Pettersson S. 2012. Host-gut microbiota metabolic interactions. *Science* 336:1262–1267. <https://doi.org/10.1126/science.1223813>.
116. Sommer F, Bäckhed F. 2013. The gut microbiota—masters of host development and physiology. *Nat Rev Microbiol* 11:227–238. <https://doi.org/10.1038/nrmicro2974>.
117. Heijtz RD, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S. 2011. Normal gut microbiota modulates brain development and behavior. *Proc Natl Acad Sci U S A* 108:3047–3052. <https://doi.org/10.1073/pnas.1010529108>.
118. Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9:313–323. <https://doi.org/10.1038/nri2515>.
119. Ücker M, Ansoorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N. 2020. Deep-sea mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable symbionts. *bioRxiv* <https://doi.org/10.1101/2020.09.11.282681>.
120. Niehuis O, Judson AK, Gadau J. 2008. Cytonuclear genic incompatibilities cause increased mortality in male  $F_2$  hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics* 178:413–426. <https://doi.org/10.1534/genetics.107.080523>.
121. O'Toole GA. 2011. Microtiter dish biofilm formation assay. *J Vis Exp* (47):2437. <https://doi.org/10.3791/2437>.
122. Zhang W, Niu Z, Yin K, Liu P, Chen L. 2013. Quick identification and quantification of *Proteus mirabilis* by polymerase chain reaction (PCR) assays. *Ann Microbiol* 63:683–689. <https://doi.org/10.1007/s13213-012-0520-x>.
123. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
124. Simon A. 2010. FastQC: a quality control tool for high throughput sequence data.
125. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
126. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
127. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
128. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
129. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
130. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. <https://doi.org/10.7717/peerj.985>.
131. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
132. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol Biol Evol* 37:291–294. <https://doi.org/10.1093/molbev/msz189>.
133. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
134. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).