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

Karissa L. Cross, Brittany A. Leigh, E. Anne Hatmaker, Aram Mikaelyan, Asia K. Miller, Seth R. Bordenstein

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.

IMPORTANT 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


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.

Address correspondence to Seth R. Bordenstein, 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 F1 hybrid females, but most of the F2 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 F1 larval hybrid offspring are dominated by Proteus bacteria (86% of sequencing reads) (21). Furthermore, these F2 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 F₂ 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 phyllosymbiosis 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 F₂ 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 F₂ hybrids derived from the cross of N. vitripennis males × N. giraulti females. We concurrently set up intra- and interspecific crosses and collected F₂ third-instar larvae (L3) from virgin F₁ females. F₂ 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 F₂ 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
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>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NCBI strain name</td>
<td>Ngir.L3-3G</td>
<td>Nvit.L3-3V</td>
<td>Nvit03</td>
<td>Dmel1</td>
<td>RB151</td>
</tr>
<tr>
<td>Host source</td>
<td><em>N. giraulti</em> L3 larvae</td>
<td><em>N. vitripennis</em> L3 larvae</td>
<td><em>N. vitripennis</em> gut</td>
<td><em>D. melanogaster</em> hemolymph</td>
<td>Human (urine from UTI)</td>
</tr>
<tr>
<td>No. of reads</td>
<td>3,774,124</td>
<td>1,321,295</td>
<td>191,241</td>
<td>680,000</td>
<td>167,518</td>
</tr>
<tr>
<td>No. of contigs</td>
<td>16</td>
<td>20</td>
<td>73</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Avg coverage (fold)</td>
<td>163</td>
<td>189</td>
<td>164</td>
<td>30</td>
<td>581</td>
</tr>
<tr>
<td>Genome size (bp)</td>
<td>4,254,678</td>
<td>4,314,756</td>
<td>4,397,852</td>
<td>4,251,508</td>
<td>4,780,676</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>40.2</td>
<td>40.2</td>
<td>40.3</td>
<td>40.2</td>
<td>40.5</td>
</tr>
<tr>
<td>No. of CDS (coding)</td>
<td>3,898</td>
<td>3,897</td>
<td>4,076</td>
<td>3,848</td>
<td>4,484</td>
</tr>
<tr>
<td>No. of plasmids</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>No. of prophage regions</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>% completed</td>
<td>100</td>
<td>99.91</td>
<td>100</td>
<td>98.61</td>
<td>99.46</td>
</tr>
<tr>
<td>% contaminated</td>
<td>1.62</td>
<td>1.89</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>Reference</td>
<td>This study</td>
<td>This study</td>
<td>Wang and Brucker (42)</td>
<td>Galac and Lazzaro (38)</td>
<td>Marquez-Ortiz et al. (44)</td>
</tr>
</tbody>
</table>

*Previously published insect isolates (accession numbers: Nvit03, QUAF01000007; Dmel1, CM001774).

*Human isolate reference strain (accession number: RB151, PRJNA342046).

*Abbreviations: UTI, urinary tract infection; CDS, coding sequences.
### TABLE 2 Proteus mirabilis sequencing and genome statistics

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

*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 "(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.
speciﬁed niche.

We investigated what functions are present in the human-associatedProvidencia rettgerigenomes and missing or reduced in insect-associated isolates. In the human-associatedProvidencia rettgerigenomes, 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 stuartiiexpressing 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-associatedProvidencia rettgerisuggests an adaptive lifestyle toward human colonization and virulence. Relatedly, Drosophila species have developed unique antimicrobial peptides that suppress and resistProvidencia rettgeri by reducing bacterial burden (39, 40).

Furthermore, CRISPR-Cas system-associated endonucleases Cas1 and Cas3 were strictly in human-associatedProvidencia rettgeristrains 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 CRIPSR-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 \( \sim 92\% \) compared to the \( \sim 84\% \) 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 link 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: \( \alpha \)-rhamnose isomerase (COG4806), \( \alpha \)-rhamnose mutarotase (COG3254), predicted metal-dependent hydrolase (COG1735), uncharacterized iron-regulated membrane protein (COG3182), putative heme iron utilization protein (COG3721), and Ca\(^{2+}/\)H\(^{+}\) antiporter (COG0387) (Table S4). \( \alpha \)-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 \( \alpha \)-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 Ca\(^{2+}/\)H\(^{+}\) antiporter may play a role in iron 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.Nvtl3.3 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.Nvtl3.3 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.Nvtl3.3 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.Nvtl3.3 and Wasp.NgirL3.3 strains here contain three prophage regions each. The three phages found within Wasp.Nvtl3.3 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 + 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.NgirL3.3 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 phylosymbiosis (24, 27, 30). Importantly, Proteus bacteria exhibit low abundance in parental species, but F2 hybrids exhibit significant breakdown and lethality between the L3 and L4 larval stages; approximately 90% of the F2 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 F2 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 F2 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
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 F2 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 F2 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

**FIG 4** Pangeneome 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; F2 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.
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 anv’ı’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 (hsdS) of the complete type I RM system had an amino acid sequence identical to hsdS 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/nonself 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× 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 phyllosymbiosis wherein microbial community ecological relationships recapitulate the host phylogenetic relationships (15). The Nasonia parasitoid wasp genus is a model system with phyllosymbiosis 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 multomic interactions between resident members of the microbiome and the host, which in turn underpin phyllosymbiosis 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 the *Wolbachia*-uninfected lines *N. vitripennis* AsymCx, *N. giraulti* RV2x(u), and *F*₂ hybrids from paternal AsymCx × maternal RV2x(u) crosses. *Nasonia* wasps were reared under 25°C constant light on *Sacophaga 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*₂ 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 × 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× phosphate-buffered saline (PBS), and resuspended in 20 μl of 1× 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°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 protocol for the “suspended cells” option.

Isolates were designated via the addition of a letter corresponding to the host animal 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 (NvLtL3-1V and NvLtL3-3V) 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 OD₆₀₀ of 0.05 for each isolate, corresponding to 5 × 10⁷ 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 OD₆₀₀ 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’-GGTGGAGTTTGGATATTG and 5’-ATCGAGAAATGACGAG, annealing temperature 58°C) and *P. rettgeri* (5’-AACCTGCTAGTCCCAAACG and 5’-CTGCAATGTCTCCTCTCAC, 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 × g for 10 min, and supernatant was removed. DNA from the resulting cell pellet was extracted using the Gentra Puregene tissue kit (Qiagen).
according to the manufacturer’s protocol and diluted to ~10 nanograms/microlitre each. Amplification was with Bio-Rad Taq 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′-CGAGGTTATACCCAGGTTC and 5′-ATCCCCATTGTGACATCACA) and *P. rettgeri* (5′-CGTTTGGTGTTGTTGATCG and 5′-GAAGCTGCTGAGATGTT). 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 filtration 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) (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 gyrA 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 each bacterial species and observing no amplification either by PCR or by qPCR. Isolates and materials are available upon request.

**Comparative Genomics of Bacteria from Nasonia Wasps**

**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.

March/April 2021 Volume 6 Issue 2 e01342-20
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., and A.K.M. carried out experiments and performed data analysis. K.L.C. and B.A.L. analyzed the genomes and performed pangenome analyses. K.L.C. and S.R.B. wrote the manuscript with input from B.A.L. S.R.B. supervised the project.

REFERENCES

Comparative Genomics of Bacteria from Nasonia Wasps


regulatory science. Nat Commun 10:3313. https://doi.org/10.1038/s41467-
019-13306-6.

57. Eren AM, Ezen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont

58. Delmont TO, Eren AM. 2018. Linking pan genomes and metagenomes: the

https://doi.org/10.1093/nar/30.7.1575.

60. Van Dongen S. 2008. Graph clustering via a discrete uncoupling process.

61. Murthy AM, Phan M-D, Peters KM, Nhu NTK, Welch RA, Ulett GC, Schembri MA, Sweet MJ. 2018. Regulation of hemolysin in uropatho-
genic Escherichia coli: fine-tuned killing of human macrophages. Viru-


63. Martinez JJ, Mulvey MA, Schilling JD, Pinker JS, Hultgren SJ. 2000. Type
1 pilus-mediated bacterial invasion of bladder epithelial cells. EMBO J

64. Jacob-DuBoisson F, Locht C, Antoine R. 2001. Two-partner secretion in
2001.0278.x.

dis/157.2.264.


67. Boronat A, Aguilar J. 1981. Metabolism of L-fucose and L-rhamnose in
Escherichia coli: differences in induction of propanediol oxidoreductase.


69. Martinez JJ, Mulvey MA, Schilling JD, Pinker JS, Hultgren SJ. 2000. Type
1 pilus-mediated bacterial invasion of bladder epithelial cells. EMBO J

70. Hegge JW, Swarts DC, van der Oost J. 2018. Prokaryotic Argonaute pro-
org/10.1038/nrnmicro.2017.73.

71. Kuzmenko A, Oguienko A, Eysunina D, Yudin D, Petrova M, Kudinova A,

72. Maldonado RF, Sá-Correia I, Valvano MA. 2016. Lipopolysaccharide modi-
fication of O antigen of Escherichia coli lipopolysaccharides. Curr Opin Bio-

73. Otto BR, Sparrius M, Verweij-van Vught AM, MacLaren DM. 1990. Iron-

74. Klebba PE, McIntosh MA, Nellands JB. 1982. Kinetics of biosynthesis of iron-


000630.

77. Restch C, Winogrodzki JL, Patterson CT, Lind J, Quinn MJ, Divpop B, Häse CC. 2010. The putative Na+/H+ antiporter of Vibrio cholerae, V-

T. 2006. Carbonyl-terminal hydrophobic tail of a NhA2P type Na+/H+ anti-
porter from cyanobacteria is involved in the apparent affinity for Na+ and

79. Geiger O, López-Lara IM, Sohlenkamp C. 2013. Phosphatidylcholine bio-
https://doi.org/10.1016/j.bbamem.2012.08.009.

80. Sohlenkamp C, López-Lara IM, Geiger O. 2003. Biosynthesis of phosphat-
diylyceride in bacteria. Prog Lipid Res 42:115–162. https://doi.org/10.1016/S0959-
440X(02)00050-4.


83. Iyer LM, Aravind L. 2002. The catalytic domains of thiamine triphosphate-
org/10.1186/1471-2164-3-33.


