Genomes of gut bacteria from *Nasonia* wasps shed light on phylosymbiosis and microbe-assisted hybrid breakdown

Karissa L. Cross[a,b], Brittany A. Leigh[a,b], E. Anne Hatmaker[a,b], Aram Mikaelyan[b,c], Asia K. Miller[a,b], Seth R. Bordenstein[a,b,d,e,#]

[a] Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA
[b] Vanderbilt Microbiome Initiative, Vanderbilt University, Nashville, Tennessee, USA
[c] Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, USA
[d] Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[e] Vanderbilt Institute for Infection, Immunology and Inflammation, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Keywords: Proteus, Providencia, Nasonia, gut bacteria, microbiome, phage

Running title: **Comparative genomics of bacteria from *Nasonia* wasps**

ORCID iD: Karissa L. Cross (http://orcid.org/0000-0001-5618-7642),
Brittany A. Leigh (https://orcid.org/0000-0001-5668-9925),
E. Anne Hatmaker (https://orcid.org/0000-0002-7821-8160)
Aram Mikaelyan (https://orcid.org/0000-0002-7276-0194),
Asia K. Miller (https://orcid.org/0000-0003-0378-232X),
Seth R. Bordenstein (https://orcid.org/0000-0001-7346-0954)

[#] Address correspondence to Seth R. Bordenstein, s.bordenstein@vanderbilt.edu

Abstract: 239 words

Text: 6,633 words

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* bacteria in the microbiome becomes dominant. The larval hybrids then catastrophically succumb to bacterial-assisted lethality and reproductive isolation between the species. Two important questions for understanding phylosymbiosis and bacterial-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 when grown in co-culture, *Proteus mirabilis* significantly outcompetes *Providencia rettgeri*. *Providencia rettgeri* genomes from *Nasonia* are similar to each other and more divergent to pathogenic, human-associated strains. *Proteus mirabilis* from *N. vitripennis*, *N. 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 or host hybridization can have detrimental consequences on host biology. In the *Nasonia* parasitoid wasp genus that 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 parental species, thus supporting a hologenomic tenet that the same members of the microbiome and the host genome impact phylosymbiosis, hybrid breakdown, and speciation.

INTRODUCTION

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, is 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 one 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_1 hybrid females, but most of the F_2 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 elicits 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, *N. vitripennis* and *N. giraulti*, possess gut microbiomes dominated by *Providencia* spp. bacteria (81-96% of sequencing reads), whereas their F_2 larval hybrid offspring are dominated by *Proteus* spp. bacteria (86% of sequencing reads) [21]. Furthermore, these F_2 interspecific hybrids exhibit ~90% hybrid lethality between the third and fourth larval instars [21], and germ-free rearing can remarkably rescue the hybrid lethality [21, 29]. The rescue of hybrid lethality via germ-free rearing

and recapitulation of death upon inoculation strongly supports the dependency of hybrid lethality on gut bacteria [30]. Interspecific microbiome transplants between *Nasonia* species with heat-killed bacterial communities also results 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 evolutionary-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 bacteria-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 or different to 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 associates 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 varying degrees of virulence towards the host whereby the host 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 varying 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 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 their 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_2 hybrid offspring to determine if they are the same. Annotation and evolutionary-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 bacterial-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_2 recombinant male hybrids of the same developmental stage, the bacterial genus *Proteus* becomes the dominant taxa associated with severe hybrid mortality [21, 27]. Both of these bacteria are easily cultivable, well-studied in arthropods, and 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_2 hybrids derived from the cross of *N. vitripennis* males x *N. giraulti* females. We concurrently set up intra- and interspecific crosses and collected F_2 third instar larvae (L3) from virgin F_1 females. F_2 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, sub-cultured, and sent for whole genome sequencing **(Figure 1a)**.

Bacteria that successfully colonize insect guts often adhere to surfaces and rapidly form biofilms [45]. To assess the ability for *Proteus* and *Providencia* to adhere and form biofilms on solid surfaces, we plated representatives from each genus in single culture and co-culture, and we measured their resulting biofilms. Over 24 hours, *Proteus mirabilis* formed a more robust biofilm than *Providencia rettgeri*, and when both isolates were plated together in a 1:1 ratio, the biofilm was significantly more abundant than *Providencia rettgeri* alone **(Figure 1b)**. To determine whether the slight, insignificant biofilm increase in the co-culture occurred in an additive manner

with both species contributing equally to the biofilm or if 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* only making up about 6.4% of the biofilm composition **(Figure 1c)**. This suggests that *Nasonia* host genetic factors may keep the *Proteus* bacterial abundances under control *in vivo* in parental wasp species; that regulation is then compromised in hybrids, where *Proteus* dominates the microbiome or when the two bacteria are grown in coculture outside of 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 a F_2 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 (#1-3). Shorthand notation of all genomes presented throughout this study are provided in Supplemental Table 1. 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 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 a 116-fold coverage. Both 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 to 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 99.91-100% complete with 0.00-1.89% contamination (Tables 1 and 2).

Phylogenetic placement of *Nasonia* bacterial isolates with other host-associated lineages

To characterize and compare the genomic diversity 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 bacteria 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, 43, 44, 54]. 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 70-71 multi-protein concatenated phylogeny **(Figure 2),** and whole genome average nucleotide identity (ANI) for all-against-all pairwise comparisons with publicly available whole genome sequences **(Supplemental Figure 1)**. 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* [54] **(Figure 2a,c)**. To explore these relationships further, we also 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 **(Supplemental Table 1)**. 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 **(Figure 2e)**. Lastly, based on whole genome pairwise average nucleotide identity (ANI) [55], 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% **(Supplemental Figure 1)**. ANI values >=95% are considered the same species [55, 56]; 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 **(Figure 2b,d)**. The *Proteus mirabilis* ML trees **(Figure 2b,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 **(Figure 2f, Supplemental Table 1)**. The phylogenomic tree showed that *Nasonia*-associated *Proteus mirabilis* are nearly 100% identical and form a distinct grouping 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 whole genome sequences of *Proteus mirabilis* with our *Nasonia* isolates, all isolates are classified as the same subspecies with an ANI >98.5% (**Figure 2b,d**; **Supplemental Figure 1**). Collectively, there is little to no genetic diversity present amongst all the *Proteus mirabilis* bacterial isolates from *Nasonia* species and their hybrids.

Comparative genomics within and between bacteria from *Nasonia* and other animals

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 Supplemental Table 1. The public genomes are from *Nasonia vitripennis* (Wasp.Nvit03) [42], *Drosophila melanogaster* (Fly.Dmel1) [54], and three human samples (reference strain Human.RB151 [44], Human.FDA330 [57], and Human.PR1). Combined with our two *Nasonia*-associated isolates of *Providencia rettgeri*, Wasp.NvitL3.3 and Wasp.NgirL3.3, pangenomic analyses in anvi'o [58, 59] 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* spp. 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) **(Figure 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 six years apart 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* spp. 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 **(Supplemental Table 2, 3)**.

As aforementioned, 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 **(Figure 2, Supplemental Figures 1)**. 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.

Functional differences between Providencia rettgeri 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) **(Figure 3)**. We divide the discussion of these unique bins to those with shared functions and then those without shared functions. Distinct gene clusters between human and insect isolates based on the Markov Cluster Algorithm (MCL) [60, 61] 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 **(Supplemental Table 2)** These functions are typically involved in host cell lysis [62], biofilm formation [63], and attachment to mucosal surfaces [63–65]. 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 specified niche.

Next, 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% vs. 2.35% of total COG functions) **(Supplemental Table 2, 3)**. 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 [65]. For example, *Providencia stuartii* expressing higher levels of MR/K hemagglutinin adhered better to catheters and led to persistence in catheter-associated bacteriuria [66]. Similarly, the filamentous hemagglutinin (FHA) protein secreted by *Bordetella pertussis* is used for attachment to the host epithelium early in its pathogenesis [67]. Collectively, the increased presence of these genes in human-associated *Providencia rettgeri* suggests an adaptive lifestyle towards human colonization and virulence. Related, *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 [68]. 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 ~92% as 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 **(Figure 3, Supplemental Table 4)**. 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 datasets [58, 59]. 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 Ca2+/H+ antiporter (COG0387) **(Supplemental Table 4)**. L-rhamnose is a common sugar component of plant and bacterial cell walls [69, 70], serves as a carbon source for some bacteria [71], and may confer protection to host immune proteins [72]. For example, adherent-invasive *E. coli* when grown on bile salts (a component of the gut) showed an upregulation of genes involved in sugar degradation, including the rhamnose pathway [73], and host-associated *L. monocytogenes* survival increases by glycosylating L-rhamnose and thus decreasing the cell wall permeability to antimicrobial peptides [72]. 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 as 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 [74, 75]. Similarly, as mentioned above, the human-associated isolates encoded 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 1,000-fold higher than what is found in mammalian blood [76] and has altered expression in the presence of *Wolbachia* endosymbionts [77]. Perhaps the presence of these iron-regulating genes, particularly in insect-environments, helps to regulate iron homeostasis specific to what is available in insects. Lastly, the Ca2+/H+ antiporter may play a role in ion homeostasis to protect bacteria in altered pH environments and in maintaining internal pH homeostasis [78, 79]. 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) **(Supplemental Table 4)**; however, this doesn't change the presence/absence of these gene clusters across isolates.

Functional differences between Providencia rettgeri from Nasonia isolates versus 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* **(Figure 3, Supplemental Table 4)**. 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 [80, 81]. PC may help mediate symbiotic host-microbe interactions as PC is required for virulence in some bacteria and to establish beneficial symbioses with hosts [82]. Furthermore, adenylate cyclases are responsible for converting adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP) and play key regulatory roles such as mediating signal transduction in cells [83], of which the CYTH (CyaB, thiamine triphosphatase) domain of the type IV adenylate cyclases binds organic phosphate [84]. In *Pseudomonas aeruginosa*, when the adenylate cyclase gene CyaB was deleted, virulence was attenuated in a mouse model [85]. 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* does share 62% nucleotide similarity with one phage spanning 46 genes from the previously published Wasp.Nvit03 *Nasonia* isolate (Phage 1; **Supplemental Figure 2**).

Functional differences within Providencia rettgeri Nasonia isolates

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 **(Supplemental Figure 1)**; therefore, most of their genomic content is similar. No discernable 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 six years apart **(Figure 3; Supplemental Table 2)**. *Providencia rettgeri* from Nashville encodes over three times as many genes relating to the COG category for [V] defense mechanisms (10.19% vs. 3.03% of total COG functions, **Supplemental Table 3).** 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*) [86]. *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 99.5-100% identical at the nucleotide level across the entire phage regions (**Supplemental Figure 2**), 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 R-M systems in the Wasp.Nvit03 strain could be directly correlated to 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) **(Supplemental Table 4)**. Of particular interest is the Argonaute homolog and restriction endonuclease that function as viral and mobile element defense mechanisms [87, 88]. 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 [89]. 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 encodes 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 [90]. For example, this machinery provides efficient protein transfer into eukaryotic cells that could inhibit phagocytosis or downregulate pro-inflammatory responses of the host [91]. Most notably, the biggest difference between the *Nasonia* isolates from Nashville vs. 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 **(Supplemental Figure 2).** Notably, one additional phage in Wasp.Nvit03 has 73% nucleotide homology to a phage present in human isolate Human.RB151.

***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) **(Supplemental Table 1)**. We were interested in the diversity of *Proteus mirabilis* bacteria across related *Nasonia* species as they diverged one MYA, and their microbiomes exhibit strong signs of phylosymbiosis [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 bacteria that causes lethality [21]. Consequently, whether the *Proteus mirabilis* bacteria in the hybrids are the same or different from the parents remains a key question to understanding the nature of the bacteria-assisted hybrid breakdown. Our collection of *Proteus mirabilis* isolates encompasses two-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 Supplemental Table 1.

With all genomes combined (n=15), pangenomic analyses in anvi'o [58, 59] produced a total of 5,421 gene clusters. We grouped these gene clusters into three bins: (i) *Proteus* spp. Core (3,043 gene clusters), (ii) Human Specific (101 gene clusters), and (iii) *Nasonia* Specific (189 gene clusters) **(Figure 4a)**. Most gene clusters (56.1%) fall within the *Proteus* spp. 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 **(Supplemental Table 3, 5)**. 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 >= 98.6% average nucleotide identity (ANI) at the whole-genome level. Moreover, within *Nasonia*-associated isolates, the ANI increases to >= 99.9% identity **(Supplemental Figure 1)**. Collectively, this evidence demonstrates that *Nasonia Proteus mirabilis* are highly similar and slightly distinct from human *Proteus mirabilis*. Furthermore, using a set of 71 core bacterial proteins, we determined *Nasonia*-associated *Proteus mirabilis* are 99.9% identical and phylogenetically split from the human-associated strains **(Figure 2)**. Therefore, we next investigated what differences may exist between the genomes of (i) *Nasonia* species and hybrids and (ii) *Nasonia* and humans.

Functional differences between Proteus mirabilis Nasonia isolates

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 **(Figure 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 distribution of gene clusters that totaled <0.05% of total gene clusters **(Figure 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 **(Figure 4b),** signifying that these small number of differences are not enough to designate the specific parental origin of the F_2 hybrid strains.

For three phages (Phage 1-3), 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 **(Supplemental Figure 3)**. These five phages have a nucleotide similarity ranging from 93.6-99.9% across the genomes, but the most significant dissimilarities are found within the phage tail regions, 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 Wasp.HybL3.2 both maintained the same two nucleotide deletions (4-mer and 8-mer) as well as mutations surrounding these deletions in the encoded 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 [92].

Functional differences between Proteus mirabilis Nasonia and human isolates

Next, comparisons between the *Nasonia* and Human bins **(Figure 4a)** revealed that the *Nasonia*-associated *Proteus mirabilis* are particularly enriched for gene clusters with functions for [M] cell wall/membrane/envelope biogenesis (14.58% vs 1.59%) **(Supplemental Table 3)**. This includes gene functions for glycosyltransferases involved in 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 cells walls that can be modified by glycosotransferases 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 [93–96]. Functional enrichment analyses in anvi'o identified the UDP-galactopyranose mutase COG function as only present in *Nasonia*-associated genomes (q=0.0173) and is responsible for the biosynthesis of galactofuranose, a sugar structure found in bacterial cell walls that is absent in mammals [97, 98] **(Supplemental Table 6)**. UDP-galactopyranose was shown to be essential for growth of *Mycobacterium tuberculosis* [99], 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. Murimidase activity has been directly connected to the ability for *Proteus* to differentiate from vegetative to swarmer cells [100], a phenotype that is visible from bacteria isolated here **(Figure 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 had an identical amino acid sequence 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 [101, 102] 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 [101, 102]. However, both RM and BREX systems utilize methylation patterns to determine self/non-self-recognition. The absence of CRISPR-Cas systems in *Proteus* has been documented previously [103] 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 are unknown. No publicly available *Proteus mirabilis* strains appear to have the same Type I RM (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 BC11-12 (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 encodes more unique gene clusters for functions relating to [P] inorganic ion transport and metabolism (9.56% vs 0% in *Nasonia* Specific bin) **(Supplemental Table 3)**. These gene clusters encompass multiple COG functions relating to outer membrane receptors for iron (COG1629), ferrienterochelin, and colicins (COG4771) as well as coper chaperone proteins (COG2608) and oxidoreductases (NAD-binding) involved in siderophore biosynthesis (COG4693) **(Supplemental Table 5)**. In addition to exploring the bins, we also performed a functional enrichment analysis in anvi'o [58, 59] **(Figure 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) **(Supplemental Table 6)**. Interestingly, the previously noted outer membrane receptor for ferrienterochelin and colicins (COG4771) has upregulated expression in pathogenic *E. coli* strains as compared to commensals [104]. 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 suggests 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 [105]. 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 towards binding trace metals for thriving in vertebrate environments [105–107]. For example, uropathogenic *E. coli* in humans use the siderophore enterobactin (of which ferrienterochelin is the iron complex) to resist lipocalin 2 [108].

Therefore, the value of these proteins in human-associated environments can assist pathogens in resisting nutritional immunity imposed by the host [105, 109, 110].

Human-associated *Proteus mirabilis* isolates also encode 5x as many transposases than *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 [111, 112]. 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 restriction endonuclease-like (RecB) superfamily, DUF1016 family **(Supplemental Table 6)**. 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 [113, 114]. Additionally, restriction endonucleases, such as RecB, cleave DNA at specific sites and act as defense mechanisms for bacteria against foreign DNA [115]. 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 [116], development [117], behavior [118], immune maturation [119], 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 [120], 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 permit a deeper understanding of the multi-omic 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 remain interfertile, in the absence of incompatible *Wolbachia* infections, as a result of their recent evolutionary divergence [121], 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 F_0 crosses. We used the *Wolbachia* uninfected lines *N. vitripennis* AsymCx, *N. giraulti* RV2x(u), and F_2 hybrids from paternal AsymCx x maternal RV2x(u) crosses. *Nasonia* 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 48h 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 x maternal RV2x(u)), using Tryptic Soy Agar (TSA) plates (Difco) containing 1.5% Agar. For each wasp line, 10 larvae were collected, and surface sterilized with 70% ethanol for one minute, washed with 1X Phosphate-buffered saline (PBS), and resuspended in 20 uL of 1X PBS. The resuspended individuals were homogenized using sterile pestles and serial dilutions were plated on TSA plates. Colonies with distinct morphology were sub-cultured 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 rpms 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 (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 x 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 OD_{600} of 0.05 for each isolate, corresponding to a 5 x 10^5 colony forming units per milliliter. A total of 1 milliliter was plated onto plastic cell culture 12-well plates in triplicate and incubated in a humid chamber overnight at 37°C. To prepare the co-culture inoculate, *Proteus* and *Providencia* were mixed 1:1 before adding a total of 1 milliliter to the wells. After 24 hours, the supernatant was removed and the biofilm stained and measured using established protocols [122]. Briefly, the culture was gently removed, and the plates were gently submerged into distilled water to removed 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 minutes, washed three times in distilled water, and allowed to dry again under the hood. The stain retained was then resuspended in 30% acetic acid, and the 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 co-culture 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'-GGTGAGATTTGTATTAATGG and 5'-ATCAGGAAGATGACGAG, annealing temperature 58°C) and *P. rettgeri* (5'-AACTCGGTCAGTTCCAAACG and 5'-CTGCATTGTTCGCTTCTCAC, annealing temperature 66°C). *Proteus* primers were designed for the ureR gene using a previously reported forward primer [123] and a new reverse primer. *Providencia* primers were designed based on a phage gene only found within the *Providencia* genomes. The biofilm experiment was repeated as described above except once the supernatant was removed, the adherent cells in the biofilm were recovered from the 1:1 co-culture in 1 milliliter of LB broth. The cells were then pelleted by spinning at 10,000g for 10 minutes and supernatant removed. DNA from the resulting cell pellet was extracted using the Gentra Puregene Tissue Kit (QIAGEN) according to manufacturer's protocol and diluted to ~10 nanograms/microliter each.

Amplification was with BioRad iTaq Universal SYBR green supermix in a CFX96 real-time C1000 thermal cycler (BioRad, Hercules, CA) and each qPCR reaction was performed at the following thermal profile: 95°C for 3 minutes, followed by 40 cycles of 95°C for 15 seconds and the respective annealing temperature for each qPCR primer for 1 minute. Samples were calculated using a standard curve generated from dilutions of larger gene products amplified of the same genes for *P. mirabilis* (5'- GCGATTTTACACCGAGTTTC and 5'- ATCCCCATTCTGACATCCAA) 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 qPCR. Isolates and materials are available upon request.

Bacterial genome assemblies and annotations

All reads were trimmed using Trimmomatic v0.32 [124] and quality checked with FastQC [125]. Reads were further filtered using the Geneious v. 11.1.5 filter and trim workflow [126]. The reads of each isolate were assembled using SPAdes 3.13.0 [127], and the quality of each assembly was determined using QUAST [128], reads under 2kb were discarded. Genomes were annotated using PROKKA [129] and visualized in Geneious v11.0.3. Genome completion and contamination estimates were calculated using CheckM [48] and Average Nucleotide Identity (ANI) using FastANI [55] online in Kbase (www.kbase.us) [130]. Prophages were identified using VirSorter [131], and plasmids were identified through PROKKA annotations and increased numbers of reads mapping through Bowtie2 [132]. All whole genome sequences were deposited in GenBank under BioProject PRJNA660265. BioSample accessions and further metadata are provided in Supplemental Table 1.

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 [126] 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 [133]. 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 [58] 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: *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 (https://www.geneious.com) [126] v11.0.3 and protein trees built with PhyML using an LG substitution model and 100 bootstraps. Evolutionary models were determined with Modeltest-ng for amino acids [133].

We used anvi'o [58] v6.2 following the pangenomics workflow [59] to analyze pangenomes of our *Proteus* and *Providencia* isolates with publicly available reference genomes from the National Center for Biotechnology Information (NCBI) (**Supplemental Table 1**), 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 contigs databases for each genome which uses Prodigal [134] v2.6.2 for gene calling. We then annotated each gene using 'anvi-run-ncbi-contigs'. A genomes storage file was created to collect each genome database using 'anvi-gen-genomes-storage' and the pangenome computed using 'anvi-

pan-genome' with flags --mcl-inflation 10 and --use-ncbi-blast which uses the MCL algorithm [60, 61] to identify clusters in amino acid sequence similarity search results and blastp [135] for the amino acid sequence similarity search. Genomes were classified by host (human vs 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).

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. Mikaelyan, 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 authors(s) and do not necessarily reflect the views of the National Institutes of Health, the National Science Foundation, or Vanderbilt University.

Author Contributions

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:

1. Shropshire JD, Bordenstein SR (2016) Speciation by Symbiosis: the Microbiome and Behavior. *MBio*, 7(2):e01785. https://doi.org/10.1128/mBio.01785-15

2. 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(13):4303–12. https://doi.org/10.1128/AEM.01886-10

3. 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(1):38. https://doi.org/10.1186/s40168-020-00823-y

4. Duplouy A, Minard G, Saastamoinen M (2020) The gut bacterial community affects immunity but not metabolism in a specialist herbivorous butterfly. *Ecology and Evolution*, :15. https://doi.org/DOI: 10.1002/ece3.6573

5. Wang Y, Rozen DE (2018) Gut microbiota in the burying beetle, *Nicrophorus vespilloides* , provide colonization resistance against larval bacterial pathogens. *Ecology and Evolution*, 8(3):1646–1654. https://doi.org/10.1002/ece3.3589

6. 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(5):e0196343. https://doi.org/10.1371/journal.pone.0196343

7. Leigh BA, Bordenstein SR, Brooks AW, Mikaelyan A, Bordenstein S (2018) Finer-scale phylosymbiosis: insights from insect viromes. *mSystems*, 3(6)

8. Kohl KD, Dearing MD, Bordenstein SR (2018) Microbial communities exhibit host species distinguishability and phylosymbiosis along the length of the gastrointestinal tract. *Molecular Ecology*, 27(8):1874–1883. https://doi.org/10.1111/mec.14460

9. Brooks AW, Kohl KD, Brucker RM, Opstal EJ van, Bordenstein SR (2016) Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biol*, 14(11):e2000225. https://doi.org/10.1371/journal.pbio.2000225

10. Bordenstein SR, Theis KR (2015) Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biol*, 13(8):e1002226. https://doi.org/10.1371/journal.pbio.1002226

11. Díaz-Sánchez S, Estrada-Peña A, Cabezas-Cruz A, Fuente J de la (2019) Evolutionary Insights into the Tick Hologenome. *Trends in Parasitology*, 35(9):725–737. https://doi.org/10.1016/j.pt.2019.06.014

12. Dunaj SJ, Bettencourt BR, Garb JE, Brucker RM (2020) Spider phylosymbiosis: divergence of widow spider species and their tissues' microbiomes. *BMC Evolutionary Biology*, 20(1):104. https://doi.org/10.1186/s12862-020-01664-x

13. 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(1):e02901-19, /mbio/11/1/mBio.02901-19.atom. https://doi.org/10.1128/mBio.02901-19

14. Pollock FJ, McMinds R, Smith S, Bourne DG, Willis BL, Medina M, Thurber RV, Zaneveld JR (2018) Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nature Communications*, 9(1):4921. https://doi.org/10.1038/s41467-018-07275-x

15. Lim SJ, Bordenstein SR (2020) An introduction to phylosymbiosis. *Proceedings of the Royal Society B: Biological Sciences*, 287(1922):20192900. https://doi.org/10.1098/rspb.2019.2900

16. Dietrich C, Köhler T, Brune A (2014) The Cockroach Origin of the Termite Gut Microbiota: Patterns in Bacterial Community Structure Reflect Major Evolutionary Events. *Applied and Environmental Microbiology*, 80(7):2261–2269. https://doi.org/10.1128/AEM.04206-13

17. Amato KR, G. Sanders J, Song SJ, Nute M, Metcalf JL, Thompson LR, Morton JT, Amir A, J. McKenzie V, Humphrey G, Gogul G, Gaffney J, L. Baden A, A.O. Britton G, P. Cuozzo F, Di Fiore A, J. Dominy N, L. Goldberg T, Gomez A, Kowalewski MM, J. Lewis R, Link A, L. Sauther M, Tecot S, A. White B, E. Nelson K, M. Stumpf R, Knight R, R. Leigh S (2019) Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *The ISME Journal*, 13(3):576–587. https://doi.org/10.1038/s41396-018-0175-0

18. Chiarello M, Auguet J-C, Bettarel Y, Bouvier C, Claverie T, Graham NAJ, Rieuvilleneuve F, Sucré E, Bouvier T, Villéger S (2018) Skin microbiome of coral reef fish is highly variable and driven by host phylogeny and diet. *Microbiome*, 6(1):147. https://doi.org/10.1186/s40168-018-0530-4

19. 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(1):93. https://doi.org/10.1186/s40168-020-00840-x

20. Bouffaud M-L, Poirier M-A, Muller D, Moënne-Loccoz Y (2014) Root microbiome relates to plant host evolution in maize and other Poaceae: Poaceae evolution and root bacteria. *Environmental Microbiology*, 16(9):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(6146):667–9. 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: Changes in rumen microbiome and metabolites. *Environmental Microbiology Reports*, 8(6):1016–1023. https://doi.org/10.1111/1758-2229.12482

23. Edwards JE, Schennink A, Burden F, Long S, Doorn DA van, Pellikaan WF, Dijkstra J, Saccenti E, Smidt H (2020) Domesticated equine species and their derived hybrids differ in their fecal microbiota. *Animal Microbiome*, 2(1):8. https://doi.org/10.1186/s42523-020-00027-7

24. Opstal EJ van, Bordenstein SR (2019) Phylosymbiosis Impacts Adaptive Traits in Nasonia Wasps. *MBio*, 10(4)https://doi.org/10.1128/mBio.00887-19

25. Wang J, Kalyan S, Steck N, Turner LM, Harr B, Künzel S, Vallier M, Häsler 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. *Nature Communications*, 6(1):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 Molecular Biology*, 2(4):225–237.

27. Brucker RM, Bordenstein SR (2012) The Roles of Host Evolutionary Relationships (Genus: Nasonia) and Development in Structuring Microbial Communities. *Evolution*, 66(2):349–362. https://doi.org/10.1111/j.1558-5646.2011.01454.x

28. Beukeboom LW, Koevoets T, Morales HE, Ferber S, Zande L van de (2015) Hybrid incompatibilities are affected by dominance and dosage in the haplodiploid wasp Nasonia. *Frontiers in Genetics*, 6https://doi.org/10.3389/fgene.2015.00140

29. Shropshire JD, Opstal EJ van, 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(8):443–51. 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 G, Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJ, Kitts P, Lynch JA, Murphy T, Oliveira DC, Smith CD, Zande L van de, Worley KC, Zdobnov EM, Aerts M, Albert S, Anaya VH, Anzola JM, Barchuk AR, Behura SK, Bera AN, Berenbaum MR, Bertossa RC, Bitondi MM, Bordenstein SR, Bork P, Bornberg-Bauer E, Brunain M, Cazzamali G, Chaboub L, Chacko J, Chavez D, Childers CP, Choi JH, Clark ME, Claudianos C, Clinton RA, Cree AG, Cristino AS, Dang PM, Darby AC, Graaf DC de, Devreese B, Dinh HH, Edwards R, Elango N, Elhaik E, Ermolaeva O, Evans JD, Foret S, Fowler GR, Gerlach D, Gibson JD, Gilbert DG, Graur D, Grunder S, Hagen DE, Han Y, Hauser F, Hultmark D, Hunter HC th, Hurst GD, Jhangian SN, Jiang H, Johnson RM, Jones AK, Junier T, Kadowaki T, Kamping A, Kapustin Y, Kechavarzi B, Kim J, Kim J, Kiryutin B, Koevoets T, Kovar CL, Kriventseva EV, Kucharski R, Lee H, Lee SL, Lees K, Lewis LR, Loehlin DW, Logsdon JM Jr, Lopez JA, Lozado RJ, Maglott D, Maleszka R, Mayampurath A, Mazur DJ, McClure MA, Moore AD, Morgan MB, Muller J, Munoz-Torres MC, Muzny DM, Nazareth LV, Neupert S, Nguyen NB, Nunes FM, Oakeshott JG, Okwuonu GO, Pannebakker BA, Pejaver VR, Peng Z, Pratt SC, Predel R, Pu LL, Ranson H, Raychoudhury R, Rechtsteiner A, Reese JT, Reid JG, Riddle M, Robertson HM, Romero-Severson J, Rosenberg M, Sackton TB, Sattelle DB, Schluns H, Schmitt T, Schneider M, Schuler A, Schurko AM, Shuker DM, Simoes ZL, Sinha S, Smith Z, Solovyev V, Souvorov A, Springauf A, Stafflinger E, Stage DE, Stanke M, Tanaka Y, Telschow A, Trent C, Vattathil S, Verhulst EC, Viljakainen L, Wanner KW, Waterhouse RM, Whitfield JB, Wilkes TE, Williamson M, Willis JH, Wolschin F,

Wyder S, Yamada T, Yi SV, Zecher CN, Zhang L, Gibbs RA (2010) Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science*, 327(5963):343–8. 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: Genes|Genomes|Genetics*, 10(8):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 Plasmids Sequences of a Clinical Proteus mirabilis Isolate Producing Plasmid Mediated NDM-1 From Italy. *Microorganisms*, 8(3):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 GM, Giani T (2016) Complete Genome Sequence of the First KPC-Type Carbapenemase-Positive *Proteus mirabilis* Strain from a Bloodstream Infection. *Genome Announcements*, 4(3):e00607-16, /ga/4/3/e00607-16.atom. https://doi.org/10.1128/genomeA.00607-16

35. Pearson MM, Sebaihia 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, Rabbinowitsch 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. *Journal of Bacteriology*, 190(11):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. *Journal of Antimicrobial Chemotherapy*, 75(3):521–525. https://doi.org/10.1093/jac/dkz472

37. Schaffer JN, Pearson MM (2015) Proteus mirabilis and Urinary Tract Infections. *Microbiology Spectrum*, 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(1):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. *Current Biology*, 26(2):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*, 583:23. 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 . *Microbiology Resource Announcements*, https://doi.org/10.1128/mra.01157-18

43. Tchuinte 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. *Journal of Global Antimicrobial Resistance*, 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 Announcements*, 5(3):e01472-16, 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. *Applied and Environmental Microbiology*, 78(21):7760–7768. https://doi.org/10.1128/AEM.01858-12

46. Drzewiecka D (2016) Significance and Roles of Proteus spp. Bacteria in Natural Environments. *Microbial Ecology*, 72(4):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. *Journal of Bacteriology*, 191(12):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 Research*, 25(7):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.*, 13:13.

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(7402):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(2):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 Microbiology*, 19(1):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. *The ISME Journal*, 6(7):1356–1366. https://doi.org/10.1038/ismej.2011.210

54. Galac MR, Lazzaro BP (2012) Comparative genomics of bacteria in the genus Providencia isolated from wild Drosophila melanogaster. *BMC Genomics*, https://doi.org/10.1186/1471-2164-13-612

55. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, 9(1):5114. https://doi.org/10.1038/s41467-018-07641-9

56. Van Rossum T, Ferretti P, Maistrenko OM, Bork P (2020) Diversity within species: interpreting strains in microbiomes. *Nature Reviews Microbiology*, https://doi.org/10.1038/s41579-020-0368-1

57. 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. *Nature Communications*, 10(1):3313. https://doi.org/10.1038/s41467-019-11306-6

58. 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

59. Delmont TO, Eren AM (2018) Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ*, 6:e4320. https://doi.org/10.7717/peerj.4320

60. Enright AJ (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, 30(7):1575–1584. https://doi.org/10.1093/nar/30.7.1575

61. Van Dongen S (2008) Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and Applications*, 30(1):121–141. https://doi.org/10.1137/040608635

62. M. V. Murthy A, 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(1):967–980. https://doi.org/10.1080/21505594.2018.1465786

63. 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

64. Martinez JJ (2000) Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *The EMBO Journal*, 19(12):2803–2812. https://doi.org/10.1093/emboj/19.12.2803

65. Jacob-Dubuisson F, Locht C, Antoine R (2001) Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. *Molecular Microbiology*, 40(2):306–313. https://doi.org/10.1046/j.1365-2958.2001.02278.x

66. Mobley HLT, Chippendale GR, Tenney JH, Mayrer 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. *Journal of Infectious Diseases*, 157(2):264–271. https://doi.org/10.1093/infdis/157.2.264

67. Locht C, Geoffroy MC, Renauld G (1992) Common accessory genes for the Bordetella pertussis filamentous hemagglutinin and fimbriae share sequence similarities with the papC and papD gene families. *The EMBO Journal*, 11(9):3175–3183. https://doi.org/10.1002/j.1460-2075.1992.tb05394.x

68. Barrangou R, Marraffini LA (2014) CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Molecular Cell*, 54(2):234–244. https://doi.org/10.1016/j.molcel.2014.03.011

69. Giraud M-F, Naismith JH (2000) The rhamnose pathway. *Current Opinion in Structural Biology*, 10(6):687–696. https://doi.org/10.1016/S0959-440X(00)00145-7

70. 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. *Frontiers in Microbiology*, 4https://doi.org/10.3389/fmicb.2013.00407

71. Boronat A, Aguilar J (1981) Metabolism of L-fucose and L-rhamnose in Escherichia coli: differences in induction of propanediol oxidoreductase. *Journal of Bacteriology*, 147(1):181–185. https://doi.org/10.1128/JB.147.1.181-185.1981

72. 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 Pathogens*, 11(5):e1004919. https://doi.org/10.1371/journal.ppat.1004919

73. Delmas J, Gibold L, Faïs T, Batista S, Leremboure M, Sinel C, Vazeille E, Cattoir V, Buisson A, Barnich N, Dalmasso G, Bonnet R (2019) Metabolic adaptation of adherent-invasive Escherichia coli to exposure to bile salts. *Scientific Reports*, 9(1):2175. https://doi.org/10.1038/s41598-019-38628-1

74. Otto BR, Sparrius M, Verweij-van Vught AM, MacLaren DM (1990) Iron-regulated outer membrane protein of Bacteroides fragilis involved in heme uptake. *Infection and Immunity*, 58(12):3954–3958. https://doi.org/10.1128/IAI.58.12.3954-3958.1990

75. Klebba PE, McIntosh MA, Neilands JB (1982) Kinetics of Biosynthesis of Iron-Regulated Membrane Proteins in Escherichia coli. *Journal of Bacteriology*, 149(3):880–888. https://doi.org/10.1128/JB.149.3.880-888.1982

76. Tang X, Zhou B (2013) Iron homeostasis in insects: Insights from *Drosophila* studies: Iron Homeostasis in Insects. *IUBMB Life*, 65(10):863–872. https://doi.org/10.1002/iub.1211

77. Kremer N, Voronin D, Charif D, Mavingui P, Mollereau B, Vavre F (2009) Wolbachia Interferes with Ferritin Expression and Iron Metabolism in Insects. *PLoS Pathogens*, 5(10):e1000630. https://doi.org/10.1371/journal.ppat.1000630

78. Resch CT, Winogrodzki JL, Patterson CT, Lind EJ, Quinn MJ, Dibrov P, Häse CC (2010) The Putative Na^+/H^+ Antiporter of *Vibrio cholerae* , Vc-NhaP2, Mediates the Specific K^+/H^+ Exchange in Vivo. *Biochemistry*, 49(11):2520–2528. https://doi.org/10.1021/bi902173y

79. Waditee R, Buaboocha T, Kato M, Hibino T, Suzuki S, Nakamura T, Takabe T (2006) Carboxyl-terminal hydrophilic tail of a NhaP type Na+/H+ antiporter from cyanobacteria is involved in the apparent affinity for Na+ and pH sensitivity. *Archives of Biochemistry and Biophysics*, 450(1):113–121. https://doi.org/10.1016/j.abb.2006.02.013

80. Geiger O, López-Lara IM, Sohlenkamp C (2013) Phosphatidylcholine biosynthesis and function in bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1831(3):503–513. https://doi.org/10.1016/j.bbalip.2012.08.009

81. Sohlenkamp C, López-Lara IM, Geiger O (2003) Biosynthesis of phosphatidylcholine in bacteria. *Progress in Lipid Research*, 42(2):115–162. https://doi.org/10.1016/S0163-7827(02)00050-4

82. Aktas M, Wessel M, Hacker S, Klüsener S, Gleichenhagen J, Narberhaus F (2010) Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic cells. *European Journal of Cell Biology*, 89(12):888–894. https://doi.org/10.1016/j.ejcb.2010.06.013

83. Zhang G, Liu Y, Ruoho AE, Hurley JH (1997) Structure of the adenylyl cyclase catalytic core. *Nature*, 386(6622):247–253. https://doi.org/10.1038/386247a0

84. 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(1):33. https://doi.org/10.1186/1471-2164-3-33

85. Smith RS, Wolfgang MC, Lory S (2004) An Adenylate Cyclase-Controlled Signaling Network Regulates Pseudomonas aeruginosa Virulence in a Mouse Model of Acute Pneumonia. *Infection and Immunity*, 72(3):1677–1684. https://doi.org/10.1128/IAI.72.3.1677-1684.2004

86. Murray NE (2003) Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews*, https://doi.org/10.1128/mmbr.64.2.412-434.2000

87. Makarova KS, Wolf YI, Oost J van der, 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. *Biology Direct*, 4(1):29. https://doi.org/10.1186/1745-6150-4-29

88. Hegge JW, Swarts DC, Oost J van der (2018) Prokaryotic Argonaute proteins: novel genome-editing tools? *Nature Reviews Microbiology*, 16(1):5–11. https://doi.org/10.1038/nrmicro.2017.73

89. 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*, https://doi.org/10.1038/s41586-020-2605-1

90. Tseng T-T, Tyler BM, Setubal JC (2009) Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiology*, 9(Suppl 1):S2. https://doi.org/10.1186/1471-2180-9-S1-S2

91. Mota LJ, Cornelis GR (2005) The bacterial injection kit: Type III secretion systems. *Annals of Medicine*, 37(4):234–249. https://doi.org/10.1080/07853890510037329

92. 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. *Scientific Reports*, 10(1):307. https://doi.org/10.1038/s41598-019-57221-0

93. Li Y, Powell DA, Shaffer SA, Rasko DA, Pelletier MR, Leszyk JD, Scott AJ, Masoudi A, Goodlett DR, Wang X, Raetz CRH, Ernst RK (2012) LPS remodeling is an evolved survival strategy for bacteria. *Proceedings of the National Academy of Sciences*, 109(22):8716–8721. https://doi.org/10.1073/pnas.1202908109

94. Maldonado RF, Sá-Correia I, Valvano MA (2016) Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS Microbiology Reviews*, 40(4):480–493. https://doi.org/10.1093/femsre/fuw007

95. Knirel YA, Prokhorov NS, Shashkov AS, Ovchinnikova OG, Zdorovenko EL, Liu B, Kostryukova ES, Larin AK, Golomidova AK, Letarov AV (2015) Variations in O-Antigen Biosynthesis and O-Acetylation Associated with Altered Phage Sensitivity in Escherichia coli 4s. *Journal of Bacteriology*, 197(5):905–912. https://doi.org/10.1128/JB.02398-14

96. 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(3):310–316. https://doi.org/10.1134/S0026261717030080

97. Tanner JJ, Boechi L, Andrew McCammon J, Sobrado P (2014) Structure, mechanism, and dynamics of UDP-galactopyranose mutase. *Archives of Biochemistry and Biophysics*, 544:128–141. https://doi.org/10.1016/j.abb.2013.09.017

98. Soltero-Higgin M, Carlson EE, Gruber TD, Kiessling LL (2004) A unique catalytic mechanism for UDP-galactopyranose mutase. *Nature Structural & Molecular Biology*, 11(6):539–543. https://doi.org/10.1038/nsmb772

99. Pan F, Jackson M, Ma Y, Mcneil M (2001) Cell Wall Core Galactofuran Synthesis Is Essential for Growth of Mycobacteria. *J. BACTERIOL.*, 183:8.

100. Strating H, Vandenende C, Clarke AJ (2012) Changes in peptidoglycan structure and metabolism during differentiation of Proteus mirabilis into swarmer cells. *Canadian journal of microbiology*, 58(10):1183–1194.

101. 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. *The EMBO Journal*, 34(2):169–183. https://doi.org/10.15252/embj.201489455

102. Chaudhary K (2018) BacteRiophage EXclusion (BREX): A novel anti-phage mechanism in the arsenal of bacterial defense system. *Journal of Cellular Physiology*, 233(2):771–773. https://doi.org/10.1002/jcp.25973

103. Alam MdS, Chakraborty S, Rahman T, Hosen MdI, 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 Journal of Medical Microbiology*, 08(02):13–25. https://doi.org/10.4236/ojmm.2018.82002

104. Sommer U, Petersen J, Pfeiffer M, Schrotz-King P, Morsczeck C (2010) Comparison of surface proteomes of enterotoxigenic (ETEC) and commensal Escherichia coli strains. *Journal of Microbiological Methods*, 83(1):13–19. https://doi.org/10.1016/j.mimet.2010.07.011

105. Palmer LD, Skaar EP (2016) Transition Metals and Virulence in Bacteria. *Annual Review of Genetics*, 50(1):67–91. https://doi.org/10.1146/annurev-genet-120215-035146

106. 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 sequestrating iron. *Nature*, 432(7019):917–921. https://doi.org/10.1038/nature03104

107. 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. *Proceedings of the National Academy of Sciences*, 103(49):18499–18503. https://doi.org/10.1073/pnas.0607055103

108. Shields-Cutler RR, Crowley JR, Hung CS, Stapleton AE, Aldrich CC, Marschall J, Henderson JP (2015) Human Urinary Composition Controls Antibacterial Activity of Siderocalin. *Journal of Biological Chemistry*, 290(26):15949–15960. https://doi.org/10.1074/jbc.M115.645812

109. 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. *Proceedings of the National Academy of Sciences*, 116(44):21980–21982. https://doi.org/10.1073/pnas.1913991116

110. 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. *International Journal of Biological Macromolecules*, 125:1156–1167. https://doi.org/10.1016/j.ijbiomac.2018.12.173

111. 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 Synthetic Biology*, 8(9):2141–2151. https://doi.org/10.1021/acssynbio.9b00218

112. Hartl DL, Dykhuizen DE, Miller RD, Green L, Framond J de (1983) Transposable element IS50 improves growth rate of E. coli cells without transposition. *Cell*, 35(2):503–510. https://doi.org/10.1016/0092-8674(83)90184-8

113. Sgro GG, Oka GU, Souza DP, Cenens W, Bayer-Santos E, Matsuyama BY, Bueno NF, Santos TR dos, Alvarez-Martinez CE, Salinas RK, Farah CS (2019) Bacteria-Killing Type IV Secretion Systems. *Frontiers in Microbiology*, 10:1078. https://doi.org/10.3389/fmicb.2019.01078

114. Schulein R, Dehio C (2002) The VirB/VirD4 type IV secretion system of Bartonella is essential for establishing intraerythrocytic infection: Role of type IV secretion in Bartonella. *Molecular Microbiology*, 46(4):1053–1067. https://doi.org/10.1046/j.1365-2958.2002.03208.x

115. 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 Research*, 42(1):3–19. https://doi.org/10.1093/nar/gkt990

116. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Pettersson S (2012) Host-Gut Microbiota Metabolic Interactions. 336:6.

117. Sommer F, Bäckhed F (2013) The gut microbiota — masters of host development and physiology. *Nature Reviews Microbiology*, 11(4):227–238. https://doi.org/10.1038/nrmicro2974

118. Heijtz RD, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg H, Pettersson S Normal gut microbiota modulates brain development and behavior. :6.

119. Round JL, Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol.*, 9:313–23. https://doi.org/doi: 10.1038/nri2515\nnri2515

120. Ücker M, Ansorge 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. https://doi.org/10.1101/2020.09.11.282681

121. 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(1):413–426. https://doi.org/10.1534/genetics.107.080523

122. O'Toole GA (2011) Microtiter Dish Biofilm Formation Assay. *Journal of Visualized Experiments*, (47):2437. https://doi.org/10.3791/2437

123. Zhang W, Niu Z, Yin K, Liu P, Chen L (2013) Quick identification and quantification of Proteus mirabilis by polymerase chain reaction (PCR) assays. *Annals of Microbiology*, 63(2):683–689. https://doi.org/10.1007/s13213-012-0520-x

124. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

125. Simon A (2010) FastQC: a quality control tool for high throughput sequence data.

126. 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*, https://doi.org/10.1093/bioinformatics/bts199

127. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, https://doi.org/10.1089/cmb.2012.0021

128. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075. https://doi.org/10.1093/bioinformatics/btt086

129. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069. https://doi.org/10.1093/bioinformatics/btu153

130. 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, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D (2018) KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, 36(7):566–569. https://doi.org/10.1038/nbt.4163

131. 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

132. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359. https://doi.org/10.1038/nmeth.1923

133. 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. *Molecular Biology and Evolution*, 37(1):291–294. https://doi.org/10.1093/molbev/msz189

134. 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(1):119. https://doi.org/10.1186/1471-2105-11-119

135. AltschuP SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215:403–410.

FIGURES

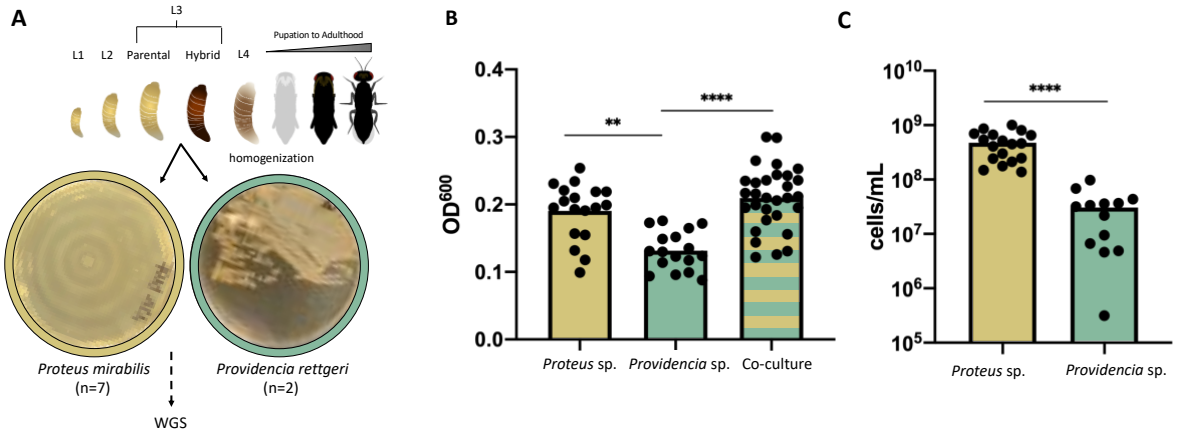


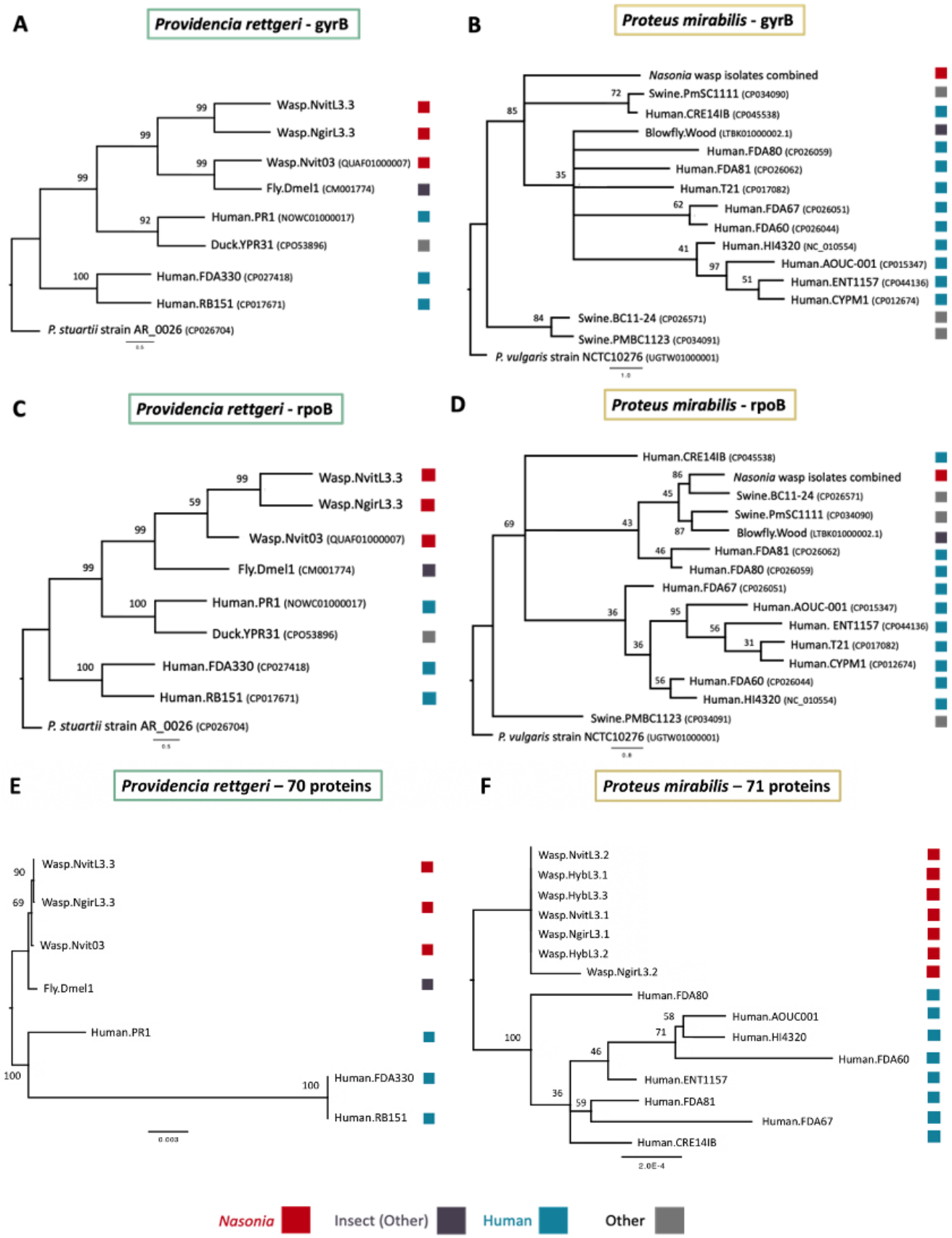
Figure 1. Characteristics of *Proteus mirabilis* and *Providencia rettgeri* isolates. A) Workflow for sample collection and examples of unique colony morphologies of *Providencia* (n=2) and *Proteus* (n=7) isolates. *Nasonia* development from larvae to adulthood is depicted on top demonstrating the L3 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 show 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 size of n=17-30 of two biological replicates (** *p*=0.0018; **** p=<0.0001). **C)** Composition of the biofilm co-culture shows *Proteus mirabilis* significantly outcompetes *Providencia rettgeri* when grown together. *Providencia rettgeri* makes up about 6.4% of the co-culture biofilm composition. Bars denote significantly different values calculated by a Mann-Whitney U test with sample size of n = 13-18 (**** p=<0.0001).

Strain Reference	Wasp.NgirL3.3	Wasp.NvitL3.3	Wasp.Nvit03	Fly.Dmel1	Human.RB151
NCBI Strain Name	NgirL3-3G	NvitL3-3V	NVIT03†	Dmel1†	RB151*
Host Source	*N. giraulti* L3 Larvae	*N. vitripennis* L3 Larvae	*N. vitripennis* gut	*D. melanogaster* hemolymph	Human (urine UTI)
Reads	3,774,124	1,321,295	191,241	680,000	167,518
Contigs	16	20	73	9	1
Average Coverage	163	189	164	30	581
Genome Size	4,254,678 bp	4,314,756 bp	4,397,852 bp	4,251,508 bp	4,780,676 bp
GC content	40.2	40.2	40.3	40.2	40.5
CDS (coding)	3,898	3,897	4,076	3,848	4,484
Plasmids	1	1	0	1	1
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 GH, et al. 2019	Galac MR, et al. 2012	Marquez-Ortiz RA, et al. 2017

Table 1. *Providencia rettgeri* sequencing and genome statistics.
† Previously published insect isolates (*Accession: NVIT03 QUAF01000007; Dmel1 CM001774*)
* Human isolate reference strain *(Accession: RB151 PRJNA342046*)

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*
NCBI Strain Name	NgirL3-1G	NgirL3-2G	F_2Hybrid-1H	F_2Hybrid-2H	F_2Hybrid-3H	NvitL3-1V	NvitL3-2V	HI4320*
Host Source	*N. giraulti* L3 Larvae	*N. giraulti* L3 Larvae	*Nasonia* L3 Hybrid	*Nasonia* L3 Hybrid	*Nasonia* L3 Hybrid	*N. vitripennis* L3 Larvae	*N. vitripennis* L3 Larvae	Human (urine)
Reads	4,180,770	4,215,774	3,928,914	3,823,446	3,926,886	4,193,134	4,059,594	60,288
Contigs	57	61	46	62	63	42	62	1
Average coverage	399	348	443	323	331	471	377	8
Genome Size	4,024,338 bp	4,023,051 bp	4,076,851 bp	4,023,528 bp	4,024,027 bp	4,070,909 bp	4,026,776 bp	4,063,606 bp
GC content	38.6	38.6	38.6	38.6	38.6	38.6	38.6	38.8
CDS (coding)	3,639	3,632	3,688	3,621	3,628	3,683	3,628	3,611
Plasmids	0	0	0	0	0	0	0	1
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 MM, et al. 2008

Table 2. *Proteus mirabilis* sequencing and genome statistics.
* Human isolate reference strain *(Accession: HI4320 PRJNA342046*)


A
Providencia rettgeri - gyrB
Wasp.NvitL3.3
Wasp.NgirL3.3
Wasp.Nvit03 (QUAF01000007)
Fly.Dmel1 (CM001774)
Human.PR1 (NOWC01000017)
Duck.YPR31 (CPO53896)
Human.FDA330 (CP027418)
Human.RB151 (CP017671)
P. stuartii strain AR_0026 (CP026704)
B
Proteus mirabilis - gyrB
Nasonia wasp isolates combined
Swine.PmSC1111 (CP034090)
Human.CRE14IB (CP045538)
Blowfly.Wood (LTBK01000002.1)
Human.FDA80 (CP026059)
Human.FDA81 (CPO26062)
Human.T21 (CP017082)
Human.FDA67 (CP026051)
Human.FDA60 (CP026044)
Human.HI4320 (NC_010554)
Human.AOUC-001 (CP015347)
Human.ENT1157 (CP044136)
Human.CYPM1 (CP012674)
Swine.BC11-24 (CP026571)
Swine.PMBC1123 (CP034091)
P. vulgaris strain NCTC10276 (UGTW01000001)
C
Providencia rettgeri - rpoB
Wasp.NvitL3.3
Wasp.NgirL3.3
Wasp.Nvit03 (QUAF01000007)
Fly.Dmel1 (CM001774)
Human.PR1 (NOWC01000017)
Duck.YPR31 (CPO53896)
Human.FDA330 (CP027418)
Human.RB151 (CP017671)
P. stuartii strain AR_0026 (CP026704)
D
Proteus mirabilis - rpoB
Human.CRE14IB (CP045538)
Nasonia wasp isolates combined
Swine.BC11-24 (CP026571)
Swine.PmSC1111 (CP034090)
Blowfly.Wood (LTBK01000002.1)
Human.FDA81 (CPO26062)
Human.FDA80 (CP026059)
Human.FDA67 (CP026051)
Human.AOUC-001 (CP015347)
Human. ENT1157 (CP044136)
Human.T21 (CP017082)
Human.CYPM1 (CP012674)
Human.FDA60 (CP026044)
Human.HI4320 (NC_010554)
Swine.PMBC1123 (CP034091)
P. vulgaris strain NCTC10276 (UGTW01000001)
E
Providencia rettgeri – 70 proteins
Wasp.NvitL3.3
Wasp.NgirL3.3
Wasp.Nvit03
Fly.Dmel1
Human.PR1
Human.FDA330
Human.RB151
F
Proteus mirabilis – 71 proteins
Wasp.NvitL3.2
Wasp.HybL3.1
Wasp.HybL3.3
Wasp.NvitL3.1
Wasp.NgirL3.1
Wasp.HybL3.2
Wasp.NgirL3.2
Human.FDA80
Human.AOUC001
Human.HI4320
Human.FDA60
Human.ENT1157
Human.FDA81
Human.FDA67
Human.CRE14IB
Nasonia
Insect (Other)
Human
Other

1

Figure 2. Phylogenetic placement of *Nasonia* isolates relative to publicly available genomes. A-B) Phylogeny of *Providencia rettgeri* and *Proteus mirabilis* strains based on 2,415 nucleotides of the gyrase B subunit reconstructed using RAxML. **C-D)** Phylogeny of *Providencia rettgeri* and *Proteus mirabilis* strains based on 4,029 nucleotides of the RNA Polymerase beta subunit reconstructed using RAxML. **E-F)** Phylogeny of *Providencia rettgeri* and *Proteus mirabilis* strains used in this study for comparative genomics analyses (Supplemental Table 1) based on concatenated amino acid alignments of 70-71 core bacterial proteins reconstructed using PhyML.

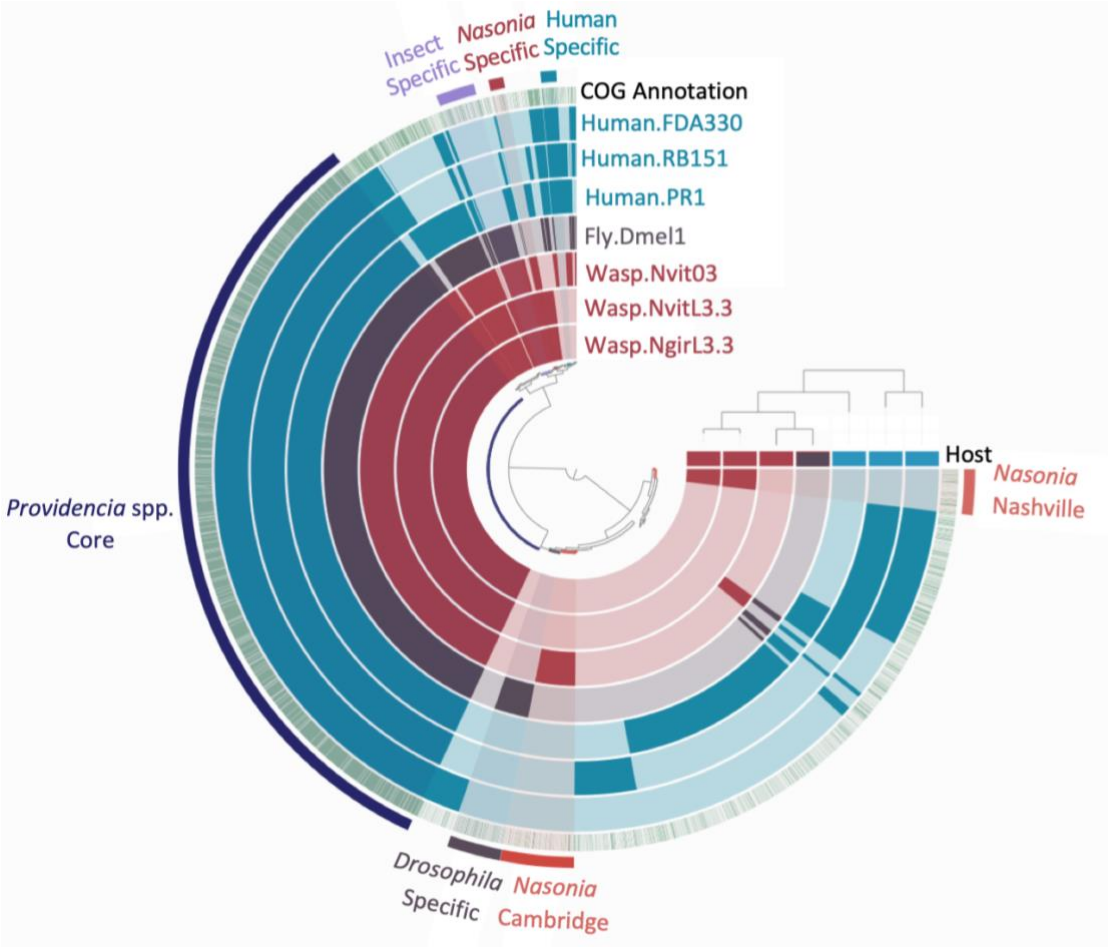


Figure 3. Pangenome of *Providencia rettgeri* human and insect isolates. The seven inner layers correspond to the seven genomes, including the sequenced *Nasonia* isolates in maroon, previously published *Drosophila* isolate in purple, and human associated genomes in blue. The bars in the inner circles show the presence of gene-clusters (GCs) in a given genome with the outer green circle depicting known COGs (green) vs unknown (white) assignments. The outermost layer of color-coded lines 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. The top horizontal layer underneath the tree represent host source (*Nasonia* - maroon, *Drosophila* - purple, or Human - blue).

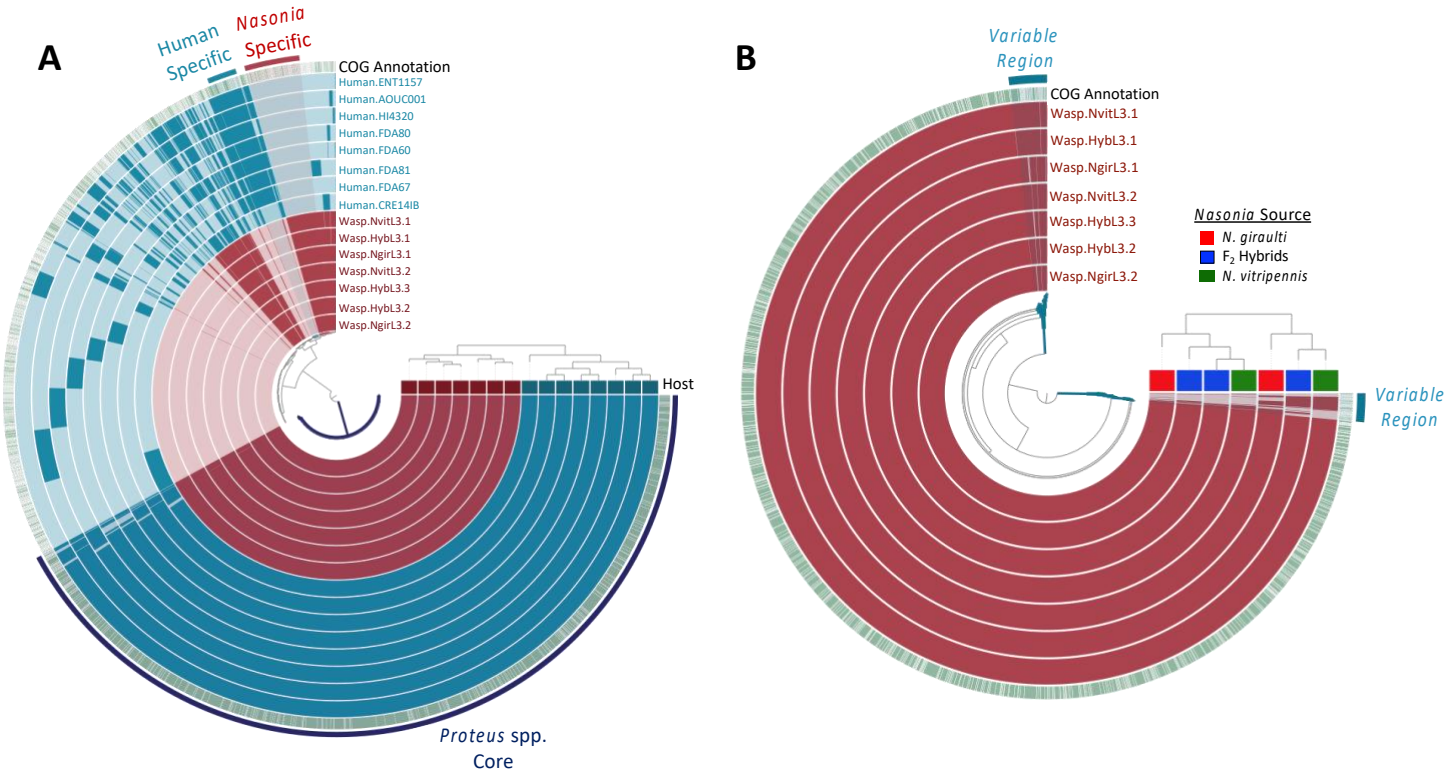


Figure 4. Pangenome of *Proteus mirabilis* human and *Nasonia* isolates. A) The fifteen inner layers correspond to all fifteen genomes, with *Nasonia* isolates shown in maroon and human isolates shown in blue. The top horizontal layer underneath the tree represent host source (*Nasonia* - maroon, Human - 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 **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) vs unknowns (white) assignments. The outermost layer of solid bars 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.

Providencia rettgeri

	Human.FDA330	Human.RB151	Human.PR1	Fly.Dmel1	Wasp.Nvit03	Wasp.NvitL3.3	Wasp.NgirL3.3
Human.FDA330	100	99.9	84.0	84.3	83.9	84.0	84.1
Human.RB151	99.9	100	84.7	84.4	84.1	84.0	84.1
Human.PR1	84.0	84.7	100	91.9	92.0	91.9	91.9
Fly.Dmel1	84.0	83.9	91.9	99.1	99.0	99.0	99.0
Wasp.Nvit03	83.9	84.0	91.9	99.3	100	99.3	99.3
Wasp.NvitL3.3	83.9	83.9	91.7	99.3	99.2	100	99.99
Wasp.NgirL3.3	84.0	83.9	91.7	99.3	99.2	99.99	100

Proteus mirabilis

	Wasp.NgirL3.2	Wasp.HybL3.2	Wasp.HybL3.3	Wasp.NvitL3.2	Wasp.NgirL3.1	Wasp.HybL3.1	Wasp.NvitL3.1	Human.CRE14IB	Human.FDA67	Human.FDA81	Human.FDA60	Human.FDA80	Human.HI4320	Human.AOUC001	Human.ENT1157
Wasp.NgirL3.2	100	99.9	99.9	99.9	99.9	99.9	99.9	98.8	98.8	98.6	98.7	98.6	98.7	98.7	98.7
Wasp.HybL3.2	99.9	100	99.9	99.9	99.9	99.9	99.9	98.8	98.8	98.7	98.7	98.6	98.7	98.7	98.8
Wasp.HybL3.3	99.9	99.9	100	99.9	99.9	99.9	99.9	98.8	98.7	98.6	98.6	98.6	98.7	98.7	98.8
Wasp.NvitL3.2	99.9	99.9	99.9	100	99.9	99.9	99.9	98.8	98.8	98.7	98.6	98.6	98.7	98.7	98.7
Wasp.NgirL3.1	99.9	99.9	99.9	99.9	100	99.9	99.9	98.9	98.7	98.7	98.6	98.6	98.7	98.7	98.7
Wasp.HybL3.1	99.9	99.9	99.9	99.9	99.9	100	99.9	98.9	98.7	98.7	98.7	98.6	98.7	98.6	98.7
Wasp.NvitL3.1	99.9	99.9	99.9	99.9	99.9	99.9	100	98.9	98.7	98.7	98.7	98.7	98.7	98.6	98.7
Human.CRE14IB	98.8	98.8	98.9	98.9	98.8	98.9	98.9	100	98.7	98.6	98.6	98.6	98.6	98.6	98.6
Human.FDA67	98.8	98.8	98.8	98.8	98.7	98.7	98.7	98.7	100	99.1	99.0	99.0	98.9	99.0	98.9
Human.FDA81	98.7	98.7	98.7	98.7	98.7	98.7	98.7	98.7	99.1	100	99.0	99.0	98.9	99.0	98.9
Human.FDA60	98.6	98.7	98.7	98.6	98.6	98.6	98.6	98.6	99.0	98.9	100	99.1	99.0	99.0	98.9
Human.FDA80	98.6	98.6	98.6	98.6	98.7	98.7	98.6	98.6	99.0	98.9	99.2	100	99.0	99.0	98.9
Human.HI4320	98.7	98.8	98.7	98.8	98.8	98.7	98.7	98.7	99.0	99.0	99.0	99.0	100	99.2	99.1
Human.AOUC001	98.7	98.7	98.7	98.7	98.6	98.7	98.7	98.6	99.0	98.9	99.0	99.0	99.1	100	99.6
Human.ENT1157	98.8	98.7	98.8	98.8	98.8	98.7	98.7	98.6	99.0	98.9	99.0	99.0	99.1	99.6	100

ANI Estimate:

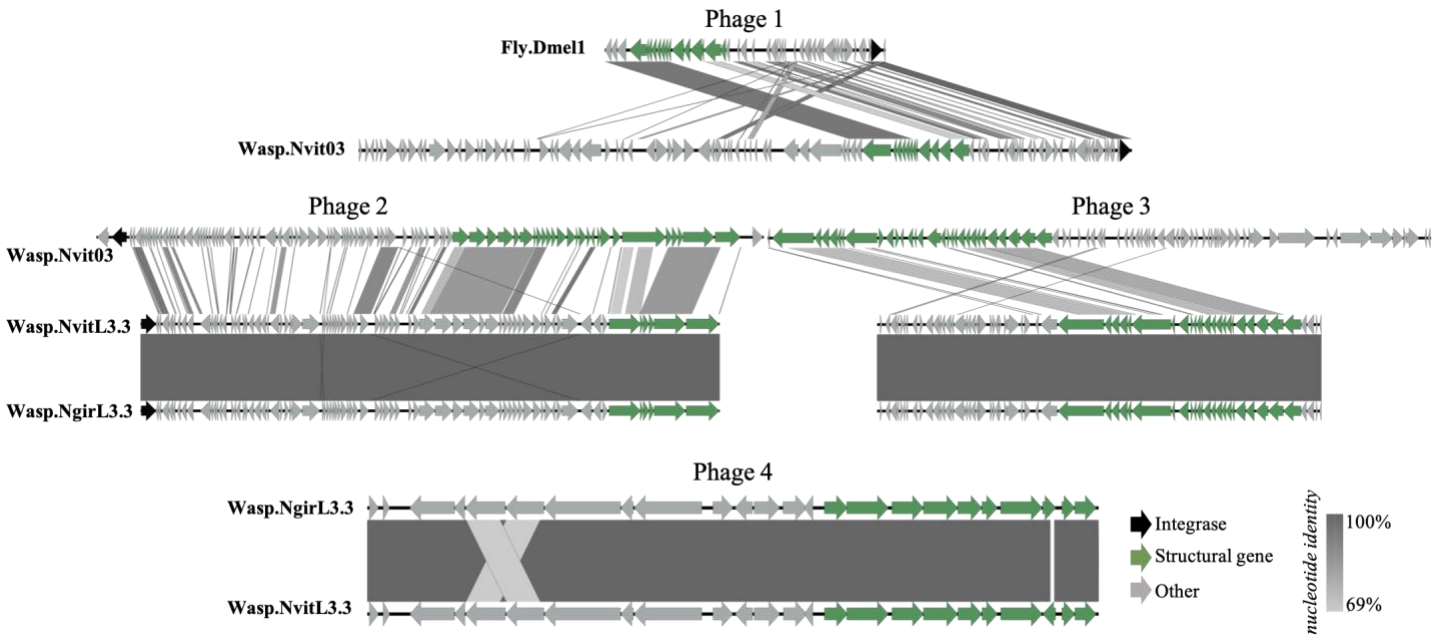
>= 99.0%

Insect (Other)

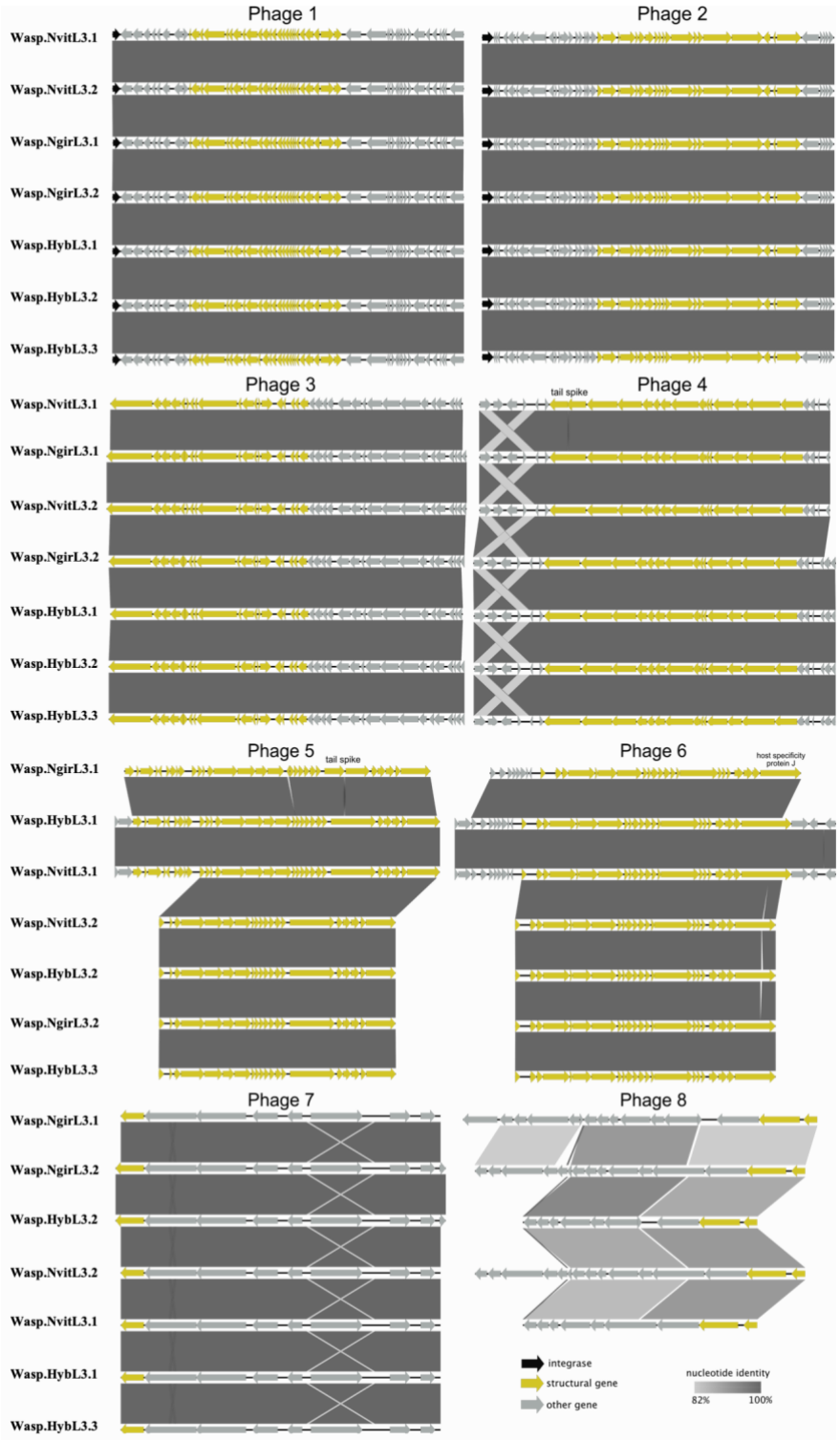
Nasonia

Human

Supplemental Figure 1. Whole genome average nucleotide identity comparisons between *Nasonia* isolates from this study and the publicly available genomes used for comparative genomic analyses.



Supplemental Figure 2. Phages of *Providencia rettgeri* insect isolates. Four total phages were identified among the insect isolates. Homologous phages were determined through whole genome alignments of putative prophages identified by VirSorter. Percent nucleotide identity as determined by blastn is indicated by shaded bars between each predicted coding gene.



Supplemental Figure 3. Phages of *Proteus mirabilis Nasonia* isolates. Eight total phages were identified. Homologous phages were determined through whole genome alignments of putative prophages identified by VirSorter. Percent nucleotide identity as determined by blastn is indicated by shaded bars between each predicted coding gene. Disrupted tail proteins are labeled.

Supplemental Table 1. Whole genome sequences used for comparative genomics analyses.

Supplemental Table 2. Genome Bins associated with the *Providencia rettgeri* pangenomic analysis of the seven *Providencia rettgeri* genomes.

Supplemental Table 3. COG analyses for comparison between the different *Providencia* anvi'o bins and the different *Proteus* anvi’o bins.

Supplemental Table 4. *Providencia rettgeri* enrichment analysis.

Supplemental Table 5. Genome Bins associated with the *Proteus mirabilis* pangenomic analysis of the fifteen *Proteus mirabilis* genomes.

Supplemental Table 6. *Proteus mirabilis* enrichment analysis.