1 **Genomes of gut bacteria from *Nasonia* wasps shed light on phylosymbiosis and**
2 **microbe-assisted hybrid breakdown**
3
4 Karissa L. Cross[a,b], Brittany A. Leigh[a,b], E. Anne Hatmaker[a,b], Aram Mikaelyan[b,c], Asia K.
5 Miller[a,b], Seth R. Bordenstein[a,b,d,e,#]
6
7 [a] Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA
8 [b] Vanderbilt Microbiome Initiative, Vanderbilt University, Nashville, Tennessee, USA
9 [c] Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North
10 Carolina, USA
11 [d] Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical
12 Center, Nashville, Tennessee, USA
13 [e] Vanderbilt Institute for Infection, Immunology and Inflammation, Vanderbilt University Medical
14 Center, Nashville, Tennessee, USA
15
16 *Keywords: Proteus, Providencia, Nasonia, gut bacteria, microbiome, phage*
17
18
19 Running title: **Comparative genomics of bacteria from *Nasonia* wasps**
20
21
22 **ORCID iD:** Karissa L. Cross (http://orcid.org/0000-0001-5618-7642),
23 Brittany A. Leigh (https://orcid.org/0000-0001-5668-9925),
24 E. Anne Hatmaker (https://orcid.org/0000-0002-7821-8160)
25 Aram Mikaelyan (https://orcid.org/0000-0002-7276-0194),
26 Asia K. Miller (https://orcid.org/0000-0003-0378-232X),
27 Seth R. Bordenstein (https://orcid.org/0000-0001-7346-0954)
28
29 [#] Address correspondence to Seth R. Bordenstein, s.bordenstein@vanderbilt.edu
30
31 Abstract: 240 words
32 Text: 6,633 words
33

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 associates. *Proteus mirabilis* from *N. vitripennis, N. giraulti,* and their hybrid offspring are nearly identical and relatively distinct from human isolates based. 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.

56 **IMPORTANCE**

57 Animal and plant hosts often establish intimate relationships with their microbiomes. In varied

58 environments, closely-related host species share more similar microbiomes, a pattern termed

59 phylosymbiosis. When phylosymbiosis is functionally significant and beneficial, microbial

60 transplants between host species or host hybridization can have detrimental consequences on host

61 biology. In the *Nasonia* parasitoid wasp genus that contains a phylosymbiotic gut community, both

62 effects occur and provide evidence for selective pressures on the holobiont. Here, we show that

63 bacterial genomes in *Nasonia* differ from other environments and harbor genes with unique

64 functions that may regulate phylosymbiotic relationships. Furthermore, the bacteria in hybrids are

65 identical to parental species, thus supporting a hologenomic tenet that the same members of the

66 microbiome and the host genome impact phylosymbiosis, hybrid breakdown, and speciation.

67

68
INTRODUCTION
70
71   Microbiomes can play pivotal roles in host health and disease [1–6] and frequently

72 establish distinguishable associations with host lineages [7–14]. Often the evolutionary

73 relationships between closely-related host species mirror the ecological relationships of their

74 microbial communities in a pattern termed phylosymbiosis [15]. Across animal and plant clades,

75 closely-related host species harbor more similar microbial communities than divergent host species

76 [9, 15–20]. Breakdown of phylosymbiotic relationships can also occur when the host genome and

77 microbiome, or hologenome, is altered, such as in hybrid hosts [21–23] or parental hosts receiving

78 a microbiome transplantation from another species [9, 24]. It can lead to detrimental effects

79 including hybrid lethality and intestinal pathology [21, 22, 24, 25]. Resultantly, it is pertinent to

80 investigate how the bacterial identities present within parental species are or are not altered as a

81 result of hybridization.

82   A well-developed animal system for studying phylosymbiosis and hologenomic speciation

83 is the *Nasonia* parasitoid wasp genus. It is comprised of four species whose ancestor arose

84 approximately one million years ago (MYA) [26]. These species are interfertile in the absence of

85 *Wolbachia* endosymbionts and exhibit strong trends in phylosymbiosis even under identical

86 rearing conditions [7, 9, 27]. Crosses between *Nasonia* species long cured of their intracellular

87 *Wolbachia* endosymbionts readily produce fit F_1 hybrid females, but most of the F_2 hybrid males

88 die during larval development due to microbe-assisted hybrid lethality that occurs in association

89 with changes in the microbiome, hypermelanization, and transcriptome-wide upregulation of

90 immune gene expression [21, 28]. Moreover, transplants of microbiomes into larvae of each

91 *Nasonia* species elicits reductions in host developmental rates and survival, supporting the premise

92 that selective pressures drive phylosymbiosis in the system [24].

93   Early in development, *Nasonia* larvae possess a simple microbiome that changes

94 composition throughout development [27]. Larvae of the two most divergent species, *N.*

95 *vitripennis* and *N. giraulti,* possess gut microbiomes dominated by *Providencia* spp. bacteria (81-

96 96% of sequencing reads), whereas their F_2 larval hybrid offspring are dominated by *Proteus* spp.

97 bacteria (86% of sequencing reads) [21]. Furthermore, these F_2 interspecific hybrids exhibit ~90%

98 hybrid lethality between the third and fourth larval instars [21], and germ-free rearing can

99 remarkably rescue the hybrid lethality [21, 29]. The rescue of hybrid lethality via germ-free rearing

4

100   and recapitulation of death upon inoculation strongly supports the dependency of hybrid lethality

101   on gut bacteria [30]. Interspecific microbiome transplants between *Nasonia* species with heat-

102   killed bacterial communities also results in slowed larval growth and decreased pupation and adult

103   survival, demonstrating how host responses play an integral role in interacting with their

104   microbiomes [24]. The costs that result from gut microbiome transplants occur in an evolutionary-

105   informed manner, which further suggests that selective pressures can underpin phylosymbiosis

106   and holobiont composition [24].

107   The genomes of both *N. vitripennis* and *N. giraulti* are published [31, 32], and thus the

108   imperative turns to the bacterial genomes within these two species and their hybrid offspring to

109   understand the nature of the catastrophic events that result in gut bacteria-assisted hybrid lethality.

110   In particular, are the bacteria in *Nasonia* guts distinguishable from other environmental isolates

111   and thus specific to the species complex? And upon hybridization and phylosymbiotic breakdown,

112   are the gut bacteria in parental hosts identical or different to the bacteria in hybrid hosts? A key

113   aspect in this system is whether the same bacteria present in parental species contribute to

114   reproductive isolation in hybrids?

115   In *Nasonia*, an alteration in the abundances of *Proteus* and *Providencia* bacteria in hybrid

116   offspring associates with lethality [21]. *Proteus* and *Providencia* are well-characterized bacterial

117   genera due to their role as opportunistic pathogens in both humans and insects [33–36]. *Proteus*

118   spp. are present in low abundance in humans, but their overgrowth is often associated with urinary

119   tract infections [35, 37]. In *Drosophila melanogaster*, different strains of *Providencia* display

120   varying degrees of virulence towards the host whereby the host develops an immune response to

121   potentially combat bacterial infection [38–40]. Additionally, in *Caenorhabditis elegans*,

122   commensal *Providencia* bacteria in the gut produce a neurotransmitter that promotes fitness of

123   both the host and bacteria [41]. The evolution of varying host responses to different bacteria

124   demonstrates how intimate interactions between these bacteria and their hosts may mediate

125   pathogenic versus symbiotic relationships.

126   Since *Proteus* and *Providencia* bacteria are native members of the *Nasonia* gut microbiome

127   across multiple lines [27, 42], there are several questions as to whether or not changes in the strains

128   of gut bacteria mediate hybrid lethality and whether the strains in *Nasonia* are distinct from other

129   animals. As *Proteus* and *Providencia* bacteria are readily culturable in the laboratory, we isolated

130   and sequenced the genomes of representative species from *Nasonia* and their hybrid larvae to

5

131   investigate their genomic diversity of (i) *Proteus* and *Providencia* isolates between *Nasonia* and

132   publicly available genomes from insects and humans, (ii) *Proteus* and *Providencia* isolates

133   between parental *N. vitripennis* and *N. giraulti* species, and (iii) *Proteus* isolates between parental

134   and F_2 hybrid offspring to determine if they are the same. Annotation and evolutionary-guided

135   comparisons of the gene content may inform the nature of what drives phylosymbiosis in *Nasonia*

136   from the microbial side and what is the nature of bacterial-dependent hybrid lethality.

137
138
139
140   **RESULTS AND DISCUSSION**

141
142   **Biofilm formation and genomics of *Providencia* and *Proteus* bacterial isolates from *Nasonia***
143   **spp. and their hybrids**

144

145   We previously found that the dominant bacterial genus present in the microbiome of the

146   larvae of *N. giraulti* and *N. vitripennis* is the genus *Providencia*, whereas in F_2 recombinant male

147   hybrids of the same developmental stage, the bacterial genus *Proteus* becomes the dominant taxa

148   associated with severe hybrid mortality [21, 27]. Both of these bacteria are easily cultivable, well-

149   studied in arthropods, and opportunistic human pathogens [6, 33–35, 43, 44]. Therefore, we sought

150   to isolate and sequence the genomes of representative *Proteus* and *Providencia* species from

151   parental *N. vitripennis* or *N. giraulti* and F_2 hybrids derived from the cross of *N. vitripennis* males

152   x *N. giraulti* females. We concurrently set up intra- and interspecific crosses and collected F_2 third

153   instar larvae (L3) from virgin F_1 females. F_2 male larvae were surface sterilized with 70% ethanol

154   and sterile water and then homogenized. Homogenate was serially diluted on Tryptic Soy Agar

155   (TSA) plates, and distinct bacterial colonies were randomly chosen, sub-cultured, and sent for

156   whole genome sequencing **(Figure 1a)**.

157   Bacteria that successfully colonize insect guts often adhere to surfaces and rapidly form

158   biofilms [45]. To assess the ability for *Proteus* and *Providencia* to adhere and form biofilms on

159   solid surfaces, we plated representatives from each genus in single culture and co-culture, and we

160   measured their resulting biofilms. Over 24 hours, *Proteus mirabilis* formed a more robust biofilm

161   than *Providencia rettgeri*, and when both isolates were plated together in a 1:1 ratio, the biofilm

162   was significantly more abundant than *Providencia rettgeri* alone **(Figure 1b)**. To determine

163   whether the slight, insignificant biofilm increase in the co-culture occurred in an additive manner

164 with both species contributing equally to the biofilm or if the increase was principally due to one

165 of the two species, we developed a qPCR assay to quantify their abundance. *Proteus mirabilis*

166 significantly outcompeted *Providencia rettgeri* when grown together ($p<0.0001$) with *Providencia*

167 *rettgeri* only making up about 6.4% of the biofilm composition (**Figure 1c**). This suggests that

168 *Nasonia* host genetic factors may keep the *Proteus* bacterial abundances under control *in vivo* in

169 parental wasp species; that regulation is then compromised in hybrids, where *Proteus* dominates

170 the microbiome or when the two bacteria are grown in coculture outside of the host. *Proteus*

171 *mirabilis* is a well-characterized human pathogen [46] that has a distinctive swarming behavior

172 which serves to reduce competition for nutrients between bacterial strains [47]. When *Nasonia*

173 hybridization produces a F_2 genotypic recombinant that shuffles the genes of the two host species,

174 the swarming behavior of *Proteus mirabilis* coupled with a presumptive breakdown in the ability

175 of the host to regulate bacterial abundances may create an environment that permits bacterial

176 overgrowth and ensuing host lethality.

177   Following isolation of colonies and DNA extraction, whole genome sequencing was

178 performed using an Illumina MiSeq. Data on genome statistics including number of reads

179 generated, sequencing coverage, and genome size and completion are provided in Tables 1 and 2.

180 For strain nomenclature, Ngir or G, Nvit or V, and Hybrid or H denote *N. giraulti*, *N. vitripennis*,

181 and hybrid, respectively; "L3", if noted, represents the third larval instar stage used for bacterial

182 isolation and cultivation; numbers at the end of the name are unique designations assigned upon

183 isolation (#1-3). Shorthand notation of all genomes presented throughout this study are provided

184 in Supplemental Table 1. Isolate identity was confirmed after sequencing by comparison of the

185 16S rRNA sequences extracted from each genome to the RefSeq database.

186   We sequenced *Providencia rettgeri* isolates from *N. giraulti* and *N. vitripennis* parental

187 lines. The genome size for *Providencia* strain Wasp.NgirL3.3 is 4,254,678 bp with 163-fold

188 coverage; and Wasp.NvitL3.3 is 4,324,254 bp with 189-fold coverage. Both strains have a G+C

189 content of 40.2%. A single plasmid of 75,229 bp was identified in *Providencia* strain

190 Wasp.NgirL3.3 with 324-fold coverage, and a single plasmid of 75,094 bp was identified in

191 Wasp.NvitL3.3 with a 116-fold coverage. Both plasmids have 100% nucleotide identity between

192 them.

193   The average genome size for the seven *Proteus mirabilis* isolates is 4,046,931 bp ranging

194 between 3,823,446 to 4,215,774 bp with an average coverage of 385-fold. The G+C content of all

195 *Proteus mirabilis* isolates is 38.6%. No circular elements were identified in any isolate. Using

196 CheckM [48] for genome completion and contamination estimates, the *Providencia rettgeri* and

197 *Proteus mirabilis* genomes are predicted to be 99.91-100% complete with 0.00-1.89%

198 contamination (Tables 1 and 2).

199

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

201

202 To characterize and compare the genomic diversity of the sequenced *Providencia rettgeri*

203 and *Proteus mirabilis* isolates, we surveyed the National Center for Biotechnology Information

204 (NCBI) databases for representative genomes. *Proteus mirabilis* is a commensal, gastrointestinal

205 bacteria in host-associated environments [46, 49], including in humans [50], tree shrews [51, 52],

206 and insects [21, 27, 53]. It is well characterized due to its role as an opportunistic pathogen in

207 different clinical settings [33–36], especially urinary tract infections involving catheters [37] as

208 well as bacteremia and wound infections [46]. However, less is known regarding *Providencia*

209 *rettgeri*, despite also serving as an opportunistic pathogen in both humans and insects [5, 6, 43, 44,

210 54]. We characterized our *Nasonia* isolates based on single copy gene phylogenies of the

211 housekeeping genes *DNA gyrase B* (*gyrB,* topoisomerase) and *RNA Polymerase beta subunit*

212 (*rpoB*, RNA synthesis), a 70-71 multi-protein concatenated phylogeny **(Figure 2),** and whole

213 genome average nucleotide identity (ANI) for all-against-all pairwise comparisons with publicly

214 available whole genome sequences **(Supplemental Figure 1)**. We report two key results.

215 First, the *Providencia rettgeri* insect isolates are distinct from human-associated isolates.

216 In the maximum likelihood (ML) phylogeny for both *rpoB* and *gyrB*, the *Nasonia* isolates from

217 this study form a well-supported clade with previously-published, insect-associated *Providencia*

218 *rettgeri* strains Wasp.Nvit03 (QUAF01000007) from *Nasonia vitripennis* [42] and Fly.Dmel1

219 (NZ_CM001774) from *Drosophila melanogaster* [54] **(Figure 2a,c)**. To explore these

220 relationships further, we also built a concatenated protein tree using 70 core bacterial proteins to

221 garner a higher resolution view of the relatedness between *Providencia rettgeri* strains used in this

222 study **(Supplemental Table 1)**. A split in the phylogeny of these strains was apparent based on

223 host origin, whereby *Nasonia* isolates grouped together with *D. melanogaster* and apart from

224 human isolates, consistent with the ML phylogenies **(Figure 2e)**. Lastly, based on whole genome

225 pairwise average nucleotide identity (ANI) [55], we classified Fly.Dmel1, Wasp.Nvit03,

226 Wasp.NvitL3.3, and Wasp.NgirL3.3 as the same bacterial strain based on their >99% ANI. The

8

227 closest human reference isolate to the insect *Providencia rettgeri* isolates was strain Human.PR1

228 with an ANI of ~92% followed by strains Human.RB151 and Human.FDA330 at an ANI of ~84%

229 **(Supplemental Figure 1)**. ANI values >=95% are considered the same species [55, 56]; therefore,

230 the human isolates of *Providencia rettgeri* may represent different species.

231 The *Proteus mirabilis* isolates sequenced from *Nasonia* have identical *gyrB* and *rpoB*

232 sequences at the nucleotide level, and a representative sequence was used to build the ML

233 phylogenies **(Figure 2b,d)**. The *Proteus mirabilis* ML trees **(Figure 2b,d)** provide moderate

234 support for the placement of the *Nasonia* isolates, with most publicly available human-associated

235 *Proteus mirabilis* genomes forming a separate clade with low support. The low phylogenetic

236 support is due to the high similarity of these genes between close relatives. Therefore, we built a

237 concatenated amino acid protein tree using 71 core bacterial genes to provide more resolution to

238 the diversity within the *Proteus mirabilis* isolates **(Figure 2f, Supplemental Table 1)**. The

239 phylogenomic tree showed that *Nasonia*-associated *Proteus mirabilis* are nearly 100% identical

240 and form a distinct grouping from human-associated strains. Lastly, based on ANI, all *Proteus*

241 *mirabilis* isolates from *Nasonia* were identified as the same strain with >99% ANI. When

242 comparing human-associated whole genome sequences of *Proteus mirabilis* with our *Nasonia*

243 isolates, all isolates are classified as the same subspecies with an ANI >98.5% (**Figure 2b,d**;

244 **Supplemental Figure 1**). Collectively, there is little to no genetic diversity present amongst all

245 the *Proteus mirabilis* bacterial isolates from *Nasonia* species and their hybrids.

246
247
248 **Comparative genomics within and between bacteria from *Nasonia* and other animals**

249

250 ***Providencia rettgeri* pangenomics**

251 For comparative genomic analyses, we utilized five publicly available high-quality whole

252 genome sequences of *Providencia rettgeri* to investigate the unique genomic changes that arose in

253 *Nasonia*-associated and other animal-associated *Providencia rettgeri* isolates. The *Providencia*

254 *rettgeri* genomes are provided in Supplemental Table 1. The public genomes are from *Nasonia*

255 *vitripennis* (Wasp.Nvit03) [42], *Drosophila melanogaster* (Fly.Dmel1) [54], and three human

256 samples (reference strain Human.RB151 [44], Human.FDA330 [57], and Human.PR1). Combined

257 with our two *Nasonia*-associated isolates of *Providencia rettgeri*, Wasp.NvitL3.3 and

258 Wasp.NgirL3.3, pangenomic analyses in anvi'o [58, 59] produced a total of 6,789 gene clusters

9

259   (COGs). We grouped these gene clusters into seven bins based on their presence/absence across

260   the genomes: (i) *Providencia* spp. Core (2,949 gene clusters), (ii) Human Specific (58 gene

261   clusters), (iii) *Nasonia* Specific (55 gene clusters), (iv) Insect Specific (142 gene clusters), (v)

262   *Drosophila* Specific (197 gene clusters), (vi) *Nasonia* – Nashville (168 gene clusters), and (vii)

263   *Nasonia* – Cambridge (269 gene clusters) **(Figure 3)**. The city nomenclature between isolates

264   Wasp.NvitL3.3, Wasp.NgirL3.3, and Wasp.Nvit03 is used to denote potential genomic differences

265   between laboratories and geographies as these two lines were derived from the same *N. vitripennis*

266   line less than six years apart and reared apart since then. More genomes from these lines will be

267   required to further evaluate the evolution of genomic differences. The core genome represents

268   gene clusters present in all genomes analyzed (human-associated and insect-associated). Most

269   gene clusters (43.4%) fall within the *Providencia* spp. Core bin, as expected, where the majority

270   of COG functions relate to [C] Energy production and conversion, [E] Amino acid transport and

271   metabolism, and [J] Translation, ribosomal structure and biogenesis **(Supplemental Table 2, 3)**.

272   As aforementioned, we determined the phylogenetic relationships among these

273   *Providencia rettgeri* whole genome sequences and noted a split in the phylogeny of these strains

274   based on host origin, which is consistent with their ANI divergence **(Figure 2, Supplemental**

275   **Figures 1)**. As adaptations to specific host environments may drive strain divergence in

276   *Providencia rettgeri*, we sought to investigate host-associated genomic variation in *Providencia*

277   *rettgeri* between (i) humans and insects, (ii) *Nasonia* and human/*Drosophila*, and (iii) various

278   *Nasonia* isolates to further explore how bacterial genomic variation partitions across host variation.

279

280   *Functional differences between Providencia rettgeri from insects and humans*

281   The human and insect bins encompass accessory proteins, also referred to as gene clusters

282   (GCs), present in either all human- or insect-associated genomes (*Nasonia* and *Drosophila*

283   combined) **(Figure 3)**. We divide the discussion of these unique bins to those with shared functions

284   and then those without shared functions. Distinct gene clusters between human and insect isolates

285   based on the Markov Cluster Algorithm (MCL) [60, 61] that share homologous functional

286   assignments (defined as distinct proteins having the same COG assignment) span hemolysins,

287   large exoproteins involved in heme utilization or adhesin, and adhesin and type 1 pilus proteins

288   **(Supplemental Table 2)** These functions are typically involved in host cell lysis [62], biofilm

289   formation [63], and attachment to mucosal surfaces [63–65]. Notably, COG3539 for pilin (type 1

290   fimbria component protein) is the most abundant functional annotation in all *Providencia rettgeri*

291   genomes, averaging around 58 occurrences in each genome. Different proteins encoding similar

292   functions between human and insect-associated bacterial genomes could be due to slightly

293   different adaptations of these genes as the bacteria evolved within their specified niche.

294   Next, we investigated what functions are present in the human-associated *Providencia*

295   *rettgeri* genomes and missing or reduced in insect-associated isolates. In the human-associated

296   *Providencia rettgeri* genomes, there are twice as many proteins relating to COG functions for [U]

297   intracellular trafficking, secretion, and vesicular transport (7.05% vs. 2.35% of total COG

298   functions) **(Supplemental Table 2, 3)**. This difference seems to be primarily driven by the

299   increased occurrence of proteins relating to some of the same functions above, namely hemolysins

300   and large exoproteins involved in heme utilization or adhesin. Large exoproteins involved in heme

301   utilization or adhesin (COG3210) are secreted to outer bacterial membranes and serve as

302   attachment factors to host cells [65]. For example, *Providencia stuartii* expressing higher levels of

303   MR/K hemagglutinin adhered better to catheters and led to persistence in catheter-associated

304   bacteriuria [66]. Similarly, the filamentous hemagglutinin (FHA) protein secreted by *Bordetella*

305   *pertussis* is used for attachment to the host epithelium early in its pathogenesis [67]. Collectively,

306   the increased presence of these genes in human-associated *Providencia rettgeri* suggests an

307   adaptive lifestyle towards human colonization and virulence. Related, *Drosophila* species have

308   developed unique antimicrobial peptides that suppress and resist *Providencia rettgeri* by reducing

309   bacterial burden [39, 40].

310   Furthermore, CRISPR-Cas system-associated endonucleases Cas1 and Cas3 were strictly

311   in human-associated *Providencia rettgeri* strains Human.FDA330 and Human.RB151 and absent

312   in all other strains. The adaptive CRISPR-Cas systems are multigenic and include short repeated

313   palindromic regions alongside spacers of phage DNA that serve as recognition factors triggering

314   an "immune" response when phages encoding the spacer attempt to infect the microbe [68].

315   CRISPR-Cas systems are crucial for subversion of phage, and as such, human-associated isolates

316   with CRIPSR-Cas have fewer integrated prophages (five total; n=2) than the *Nasonia*-associated

317   ones (12 total; n=3) **(Table 1).** The human-associated isolate Human.PR1 without CRISPR-Cas

318   maintains five unique predicted prophages within its genome, similar to the number found in each

319   *Nasonia*-associated isolate; and it also maintains the highest ANI to the *Nasonia* isolates at ~92%

320   as compared to the ~84% of all other human isolates. The emphasis of these functions in human-

321   associated *Providencia rettgeri* isolates, as opposed to insect-associated isolates, suggests human

322   isolates are better poised to combat phage invasions.

323   To further tease apart the annotated functions specific in *Nasonia* and its close relative in

324   *D. melanogaster*, we performed a functional enrichment analysis to statistically identify COGs

325   only in the insect isolates **(Figure 3, Supplemental Table 4)**. The functional enrichment analysis

326   computed in anvi'o uses a Generalized Linear Model with the logit linkage function to compute

327   an enrichment score, p-value, and a False Detection Rate q-value between the different genome

328   datasets [58, 59]. This analysis identified six COG functions unique to insect-associated genomes:

329   L-rhamnose isomerase (COG4806), L-rhamnose mutarotase (COG3254), predicted metal-

330   dependent hydrolase (COG1735), uncharacterized iron-regulated membrane protein (COG3182),

331   putative heme iron utilization protein (COG3721), and Ca2+/H+ antiporter (COG0387)

332   **(Supplemental Table 4)**. L-rhamnose is a common sugar component of plant and bacterial cell

333   walls [69, 70], serves as a carbon source for some bacteria [71], and may confer protection to host

334   immune proteins [72]. For example, adherent-invasive *E. coli* when grown on bile salts (a

335   component of the gut) showed an upregulation of genes involved in sugar degradation, including

336   the rhamnose pathway [73], and host-associated *L. monocytogenes* survival increases by

337   glycosylating L-rhamnose and thus decreasing the cell wall permeability to antimicrobial peptides

338   [72]. Additionally, as iron is a key metal for biological functions, this poses the question as to why

339   these genes may be enriched in insect-associated environments as compared to those in human-

340   associated settings. Iron-regulated membrane proteins and heme iron utilization proteins permit

341   enteric bacteria to readily sense and respond to iron limiting environments and play a role in iron

342   acquisition [74, 75]. Similarly, as mentioned above, the human-associated isolates encoded unique

343   heme utilization proteins, emphasizing how these functions are important for specific host-

344   associated environments. Insects differ from mammals in that they secrete ferritin, a protein that

345   contains iron, into their hemolymph 1,000-fold higher than what is found in mammalian blood

346   [76] and has altered expression in the presence of *Wolbachia* endosymbionts [77]. Perhaps the

347   presence of these iron-regulating genes, particularly in insect-environments, helps to regulate iron

348   homeostasis specific to what is available in insects. Lastly, the Ca2+/H+ antiporter may play a role

349   in ion homeostasis to protect bacteria in altered pH environments and in maintaining internal pH

350   homeostasis [78, 79]. It is important to note that although the unadjusted p-value for all genes in

351   the enrichment analysis is significant (p=0.008), the adjusted q-value (q=0.897) that controls for

12

352 false discovery rate does not reach significance due to the sample sizes for these genomes (n=7)

353 **(Supplemental Table 4)**; however, this doesn't change the presence/absence of these gene clusters

354 across isolates.

355

356 *Functional differences between Providencia rettgeri from Nasonia isolates versus humans and*

357 *Drosophila*

358   We were next interested in disentangling functions specific in the *Nasonia*-associated

359 isolates. Therefore, we performed the same functional enrichment analysis to identify COGs that

360 occur only in the *Providencia rettgeri* strains from *Nasonia* **(Figure 3, Supplemental Table 4)**.

361 Only two COG categories occurred in the *Nasonia*-associated genomes: phospholipid N-

362 methyltransferase (COG3963) and CYTH domain, found in class IV adenylate cyclase and various

363 triphosphatases (COG2954) (p=0.03, q=1). Phospholipid N-methyltransferase synthesizes

364 phosphatidylcholine (PC), a membrane-forming phospholipid present in only about 15% of

365 bacteria [80, 81]. PC may help mediate symbiotic host-microbe interactions as PC is required for

366 virulence in some bacteria and to establish beneficial symbioses with hosts [82]. Furthermore,

367 adenylate cyclases are responsible for converting adenosine triphosphate (ATP) to cyclic

368 adenosine monophosphate (cAMP) and play key regulatory roles such as mediating signal

369 transduction in cells [83], of which the CYTH (CyaB, thiamine triphosphatase) domain of the type

370 IV adenylate cyclases binds organic phosphate [84]. In *Pseudomonas aeruginosa*, when the

371 adenylate cyclase gene CyaB was deleted, virulence was attenuated in a mouse model [85].

372 Collectively, these two COGs could play a role in mediating symbiotic relationships, beneficial or

373 pathogenic, in the *Nasonia* host. Lastly, the Wasp.NvitL3.3 and Wasp.NgirL3.3 isolates did not

374 share any mobile elements with any human or *Drosophila*-associated isolates. However,

375 Fly.Dmel1 from *Drosophila* does share 62% nucleotide similarity with one phage spanning 46

376 genes from the previously published Wasp.Nvit03 *Nasonia* isolate (Phage 1; **Supplemental**

377 **Figure 2**).

378

379 *Functional differences within Providencia rettgeri Nasonia isolates*

380   We further investigated the differences between *Providencia rettgeri* isolates within

381 *Nasonia* species. Strains Wasp.NvitL3.3 and Wasp.NgirL3.3 isolated from the lab in Nashville,

382 TN have a 99.99% ANI with each other and 99.3% ANI with Wasp.Nvit03 isolated from a lab

13

383   located in Cambridge, MA (**Supplemental Figure 1**); therefore, most of their genomic content is

384   similar. No discernable difference could be identified between the two Nashville isolates; however,

385   a small difference exists between these and the isolate from Cambridge derived from the same *N.*

386   *vitripennis* line, AsymCx, less than six years apart (**Figure 3; Supplemental Table 2**).

387   *Providencia rettgeri* from Nashville encodes over three times as many genes relating to the COG

388   category for [V] defense mechanisms (10.19% vs. 3.03% of total COG functions, **Supplemental**

389   **Table 3).** Although we found no evidence of CRISPR-Cas gene cassettes in *Nasonia*-associated

390   genomes, a major difference is the presence of Type I restriction modification (RM) systems

391   exclusively in the Nashville-associated *Providencia* genomes and absent from *N. vitripennis* strain

392   Wasp.Nvit03 from Cambridge. Type I RM systems cleave away from the recognition site and have

393   three components: a restriction enzyme (*hsdR*), a methyltransferase (*hsdM*), and a specificity

394   subunit (*hsdS*) [86]. *Providencia rettgeri* strains Wasp.NvitL3.3 and Wasp.NgirL3.3 are predicted

395   to encode two Type I RM systems, one with all three component genes and one missing the

396   restriction enzyme. The Wasp.Nvit03 strain without the Type I RM system harbors six putative

397   prophage regions within its genome, while both the newly sequenced Wasp.NvitL3.3 and

398   Wasp.NgirL3.3 strains here contain three prophage regions each. The three phages found within

399   Wasp.NvitL3.3 and Wasp.NgirL3.3 are 99.5-100% identical at the nucleotide level across the

400   entire phage regions (**Supplemental Figure 2**), whereas the remaining three phage regions in

401   Wasp.Nvit03 are novel or contain similarity to phage regions in *Providencia rettgeri* from *D.*

402   *melanogaster*. The absence of both CRISPR-Cas and Type I R-M systems in the Wasp.Nvit03

403   strain could be directly correlated to the number of prophages within its genome, but a larger

404   sample size of genomes may be necessary to make any firm predictions.

405   We also performed functional enrichment analysis to identify genes present only in

406   *Providencia rettgeri* strains isolated from the lab in Nashville and absent in all other strains. We

407   identified seven unique COGs: (i) CDP-glycerol glycerophosphotransferase (COG1887), (ii)

408   uncharacterized conserved protein YBBC (COG3876), (iii) argonaute homolog implicated in RNA

409   metabolism and viral defense (COG1431), (iv) predicted restriction endonuclease (COG3440), (v)

410   DNA replication protein DnaD (COG3935), (vi) replicative superfamily II helicase (COG1204),

411   and (vii) predicted ATPase archaeal AAA+ ATPase superfamily (COG1672) (p=0.03, q=1)

412   (**Supplemental Table 4**). Of particular interest is the Argonaute homolog and restriction

413   endonuclease that function as viral and mobile element defense mechanisms [87, 88]. Recently, a

414   bacterial Argonaute nuclease from *Clostridium butyricum* was shown to target multicopy genetic

415   elements and suppress the propagation of plasmids and infection by phages via DNA interference

416   [89]. In addition to the RM systems, this diverse suite of genes in *Nasonia* specifically from

417   Nashville suggests enhanced protection against viral infection.

418   Conversely, strain Wasp.Nvit03 from Cambridge, MA encodes genes relating to type III

419   secretion systems that are not present in strains from Nashville (COG4791, COG4790, COG4794)

420   (p=0.03, q=1). Specifically, type III secretion systems serve to deliver effector proteins across

421   bacterial and host membranes that can influence host cell biology [90]. For example, this

422   machinery provides efficient protein transfer into eukaryotic cells that could inhibit phagocytosis

423   or downregulate pro-inflammatory responses of the host [91]. Most notably, the biggest difference

424   between the *Nasonia* isolates from Nashville vs. Cambridge is the unique phage machinery.

425   Although two of the three prophages identified in Wasp.NvitL3.3 and Wasp.NgirL3.3 strains were

426   also present in Wasp.Nvit03 based on amino acid identity, extensive rearrangements were evident

427   **(Supplemental Figure 2).** Notably, one additional phage in Wasp.Nvit03 has 73% nucleotide

428   homology to a phage present in human isolate Human.RB151.

429

430   ***Proteus mirabilis* pangenomics**

431   We employed the same methodology to compare the *Proteus mirabilis Nasonia*-associated

432   genomes from this study (n=7) **(Table 2)** with high quality, publicly available genomes from close

433   relatives in humans (n=8) **(Supplemental Table 1)**. We were interested in the diversity of *Proteus*

434   *mirabilis* bacteria across related *Nasonia* species as they diverged one MYA, and their

435   microbiomes exhibit strong signs of phylosymbiosis [24, 27, 30]. Importantly, *Proteus* bacteria

436   exhibit low abundance in parental species, but F_2 hybrids exhibit significant breakdown and

437   lethality between the L3 and L4 larval stages; approximately 90% of the F_2 hybrids die in

438   conjunction with *Proteus* becoming the dominant bacteria that causes lethality [21]. Consequently,

439   whether the *Proteus mirabilis* bacteria in the hybrids are the same or different from the parents

440   remains a key question to understanding the nature of the bacteria-assisted hybrid breakdown. Our

441   collection of *Proteus mirabilis* isolates encompasses two-three isolates each from *N. vitripennis*,

442   *N. giraulti*, and F_2 hybrid males to control for potential sex effects. Lastly, we also investigated

443   how *Nasonia*-associated isolates compare to those isolated from human-associated environments.

444   All *Proteus mirabilis* genomes are listed in Supplemental Table 1.

15

445   With all genomes combined (n=15), pangenomic analyses in anvi'o [58, 59] produced a

446   total of 5,421 gene clusters. We grouped these gene clusters into three bins: (i) *Proteus* spp. Core

447   (3,043 gene clusters), (ii) Human Specific (101 gene clusters), and (iii) *Nasonia* Specific (189 gene

448   clusters) **(Figure 4a)**. Most gene clusters (56.1%) fall within the *Proteus* spp. Core bin where the

449   majority of COG functions relate to [C] Energy production and conversion, [E] Amino acid

450   transport and metabolism, and [J] Translation, ribosomal structure and biogenesis, the same three

451   categories as reported for the *Providencia* species **(Supplemental Table 3, 5)**. Of the gene clusters

452   not found in the core, 1.8% and 3.5% are shared just within human-associated and *Nasonia*-

453   associated isolates, respectively, and the 38.5% remaining gene clusters are found within limited

454   subsets of the genomes across isolates. Indeed, human-associated and *Nasonia*-associated *Proteus*

455   *mirabilis* genomes share >= 98.6% average nucleotide identity (ANI) at the whole-genome level.

456   Moreover, within *Nasonia*-associated isolates, the ANI increases to >= 99.9% identity

457   **(Supplemental Figure 1)**. Collectively, this evidence demonstrates that *Nasonia Proteus mirabilis*

458   are highly similar and slightly distinct from human *Proteus mirabilis*. Furthermore, using a set of

459   71 core bacterial proteins, we determined *Nasonia*-associated *Proteus mirabilis* are 99.9%

460   identical and phylogenetically split from the human-associated strains **(Figure 2)**. Therefore, we

461   next investigated what differences may exist between the genomes of (i) *Nasonia* species and

462   hybrids and (ii) *Nasonia* and humans.

463

464   *Functional differences between Proteus mirabilis Nasonia isolates*

465   We investigated the strain-level diversity within our *Nasonia*-associated *Proteus mirabilis*

466   isolates to distinguish whether or not the *Proteus* strains in hybrid offspring are the same as those

467   in parental *Nasonia*. Pangenomic analyses produced a total of 3,699 gene clusters **(Figure 4b)**.

468   Functional enrichment analyses found no significant difference between *Proteus mirabilis* isolated

469   from either *N. vitripennis, N. giraulti,* or F$_2$ hybrids. Therefore, the hybrid *Proteus mirabilis*

470   genomes are functionally identical to the parental *Proteus mirabilis* isolates. We identified two

471   regions where there was variability in the distribution of gene clusters that totaled <0.05% of total

472   gene clusters **(Figure 4b)** and grouped them into a bin labeled "Variable Regions" for further

473   investigation. Inspection of this bin identified 92 gene clusters that occurred in subsets of the

474   *Nasonia*-associated *Proteus mirabilis* genomes, of which 68 (74%) are lacking known functional

475   annotations. No consistent trend emerged regarding gene clusters found only within specific

476   *Nasonia*-associated genomes (e.g., gene clusters were not consistently unique in only *N.*
477   *vitripennis* and hybrid strains or *N. giraulti* and hybrid strains). Hierarchical clustering based on
478   gene cluster frequencies places the F$_2$ hybrid strains as more similar to paternal *N. vitripennis*
479   strains based on their gene-cluster distribution across genomes, but *N. giraulti* strains are still
480   closely associated within these relationships **(Figure 4b),** signifying that these small number of
481   differences are not enough to designate the specific parental origin of the F$_2$ hybrid strains.

482   For three phages (Phage 1-3), there is 100% nucleotide similarity and identical gene
483   synteny across all seven *Proteus mirabilis* isolates in *Nasonia,* however key differences including
484   gene deletions and truncations are apparent in proteins within the five other phages also present
485   **(Supplemental Figure 3)**. These five phages have a nucleotide similarity ranging from 93.6-
486   99.9% across the genomes, but the most significant dissimilarities are found within the phage tail
487   regions, the part of the phage that directly interacts with its bacterial host and often determines
488   host specificity. For example, in Phages 4 and 5 from Wasp.NgirL3.1, an early stop codon results
489   in a predicted truncation in the tail spike protein, whereas across all other isolates containing these
490   phages, the tail spike protein is intact. Relative to other isolates, Phage 6 found in Wasp.NvitL3.2
491   and Wasp.HybL3.2 both maintained the same two nucleotide deletions (4-mer and 8-mer) as well
492   as mutations surrounding these deletions in the encoded host specificity protein J gene. Altogether,
493   comparative genomics of the phage coding regions demonstrate that these isolates are extremely
494   similar, and some of their phages exhibit distinct differences in the tail regions known to evolve
495   fast, namely those that determine host specificity [92].

496

497   *Functional differences between Proteus mirabilis Nasonia and human isolates*

498   Next, comparisons between the *Nasonia* and Human bins **(Figure 4a)** revealed that the
499   *Nasonia*-associated *Proteus mirabilis* are particularly enriched for gene clusters with functions for
500   [M] cell wall/membrane/envelope biogenesis (14.58% vs 1.59%) **(Supplemental Table 3)**. This
501   includes gene functions for glycosyltransferases involved in LPS biosynthesis (COG3306),
502   peptidoglycan/LPS O-acetylase OafA/YrhL containing acyltransferase and SGNH-hydrolase
503   domains (COG1835), and UDP-galactopyranose mutase (COG0562). LPS is a major component
504   of Gram-negative bacterial cells walls that can be modified by glycosotransferases and
505   acyltransferases, and variations in the structure of LPS can provide selective advantages in
506   different environments such as inhibiting bacteriophage binding, temperature tolerance, and

17

507 antimicrobial resistance [93–96]. Functional enrichment analyses in anvi'o identified the UDP-

508 galactopyranose mutase COG function as only present in *Nasonia*-associated genomes (q=0.0173)

509 and is responsible for the biosynthesis of galactofuranose, a sugar structure found in bacterial cell

510 walls that is absent in mammals [97, 98] **(Supplemental Table 6)**. UDP-galactopyranose was

511 shown to be essential for growth of *Mycobacterium tuberculosis* [99], although it remains unclear

512 what function it may serve specifically in *Nasonia*-associated environments. Additionally, all

513 *Nasonia*-associated genomes and a single human-associated genome (HI4320) encode

514 muramidase (urinary lysozyme) (COG4678; q=0.0452), although a similar homologous COG

515 function (COG3772) exists in all human-associated genomes and in three *Nasonia*-associated

516 genomes as well. Murimidase activity has been directly connected to the ability for *Proteus* to

517 differentiate from vegetative to swarmer cells [100], a phenotype that is visible from bacteria

518 isolated here **(Figure 1b)**.

519 We also investigated putative adaptive and innate bacterial defense systems within the

520 isolates including CRISPR-Cas, restriction-modification (RM), and BacteRiophage Exclusion

521 (BREX) systems. Although we found no evidence of CRISPR-Cas gene cassettes in any of our

522 isolates, there were predicted Type I RM and BREX systems. The specificity subunit (hsdS) of the

523 complete Type I RM had an identical amino acid sequence to hsdS proteins (>95% nucleotide

524 similarity) found in two other *Proteus mirabilis* strains (PmSDC32 from red junglefowl;

525 CNR20130297 from human, as of September 2020). The putative Type I BREX system [101,

526 102] component genes were intact without frameshifts or premature stop codons. Although the

527 BREX mechanism of foreign DNA removal is not yet understood completely, unlike CRISPR-Cas

528 and RM systems, the defense system does not involve restriction of foreign DNA [101, 102].

529 However, both RM and BREX systems utilize methylation patterns to determine self/non-self-

530 recognition. The absence of CRISPR-Cas systems in *Proteus* has been documented previously

531 [103] although the absence has not been linked to expanded presence of mobile elements. The

532 presence of the newly-characterized BREX defense system in *Proteus mirabilis* could prevent

533 phage integration, but the evolutionary history behind either the prophage integration events

534 in *Proteus mirabilis* or the history of the BREX system are unknown. No publicly

535 available *Proteus mirabilis* strains appear to have the same Type I RM (as determined by the

536 conserved methyltransferase protein). *Proteus mirabilis* strains MPE5139 (CP053684.1) and

537 Pm15C1 (KX268685) contain a similar BREX cassette (86% query coverage; 97.75% similarity),

18

538 but strains placed phylogenetically close to the *Nasonia*-associated isolates such as PmSC1111

539 (CP034090) and BC11-12 (CP026571) do not. The acquisition of the BREX system after

540 integration of the numerous phage genomes found within *Proteus mirabilis* could explain the

541 apparent lack of phage inhibition, although we cannot say whether this is the case in the current

542 study.

543   We next investigated the human-associated *Proteus mirabilis* bin and discovered it encodes

544 more unique gene clusters for functions relating to [P] inorganic ion transport and metabolism

545 (9.56% vs 0% in *Nasonia* Specific bin) **(Supplemental Table 3)**. These gene clusters encompass

546 multiple COG functions relating to outer membrane receptors for iron (COG1629),

547 ferrienterochelin, and colicins (COG4771) as well as coper chaperone proteins (COG2608) and

548 oxidoreductases (NAD-binding) involved in siderophore biosynthesis (COG4693) **(Supplemental**

549 **Table 5)**. In addition to exploring the bins, we also performed a functional enrichment analysis in

550 anvi'o [58, 59] **(Figure 4a)**. This analysis identified that siderophore biosynthesis (COG4693),

551 previously noted in the human-associated *Proteus*, occurs in 7 out of 8 of the human-associated

552 genomes and is absent in all *Nasonia*-associated genomes (q=0.045) **(Supplemental Table 6)**.

553 Interestingly, the previously noted outer membrane receptor for ferrienterochelin and colicins

554 (COG4771) has upregulated expression in pathogenic *E. coli* strains as compared to commensals

555 [104]. Although homologous functions (different proteins with the same functional assignment)

556 for outer membrane receptor proteins (COG1629 and COG4771) are also found in *Nasonia*-

557 associated *Proteus mirabilis* genomes, the presence of oxidoreductase involved in siderophore

558 biosynthesis (COG4693) and increased occurrence and diversity of genes relating to inorganic ion

559 transport and metabolism in human-associated environments suggests that the more pathogenic

560 strains may maintain a variety of antigens that could result in persistent inflammation and

561 infection. When infection ensues, metal homeostasis can change dramatically, and therefore

562 pathogens must be able to compete for the limited metal availability [105]. Vertebrate hosts,

563 including humans, can produce proteins such as the iron-sequestering lipocalin 2 to keep trace

564 metals away from bacteria, which in turn creates a selective pressure for bacteria to evolve diverse

565 iron-binding siderophores and a higher affinity towards binding trace metals for thriving in

566 vertebrate environments [105–107]. For example, uropathogenic *E. coli* in humans use the

567 siderophore enterobactin (of which ferrienterochelin is the iron complex) to resist lipocalin 2 [108].

568   Therefore, the value of these proteins in human-associated environments can assist pathogens in

569   resisting nutritional immunity imposed by the host [105, 109, 110].

570   Human-associated *Proteus mirabilis* isolates also encode 5x as many transposases than

571   *Nasonia*-associated *Proteus* isolates. Transposons permit the movement of DNA in and around the

572   bacterial chromosome, which can help facilitate growth and adaptation of bacteria to their host

573   environment [111, 112]. Using the same functional enrichment analysis approach described above,

574   we identified just two gene functions that are found in all human-associated *Proteus mirabilis*

575   isolates (n=8) and absent in all *Nasonia*-associated isolates (q=0.017): (i) the type IV secretory

576   pathway VirB4 component, and (ii) a predicted nuclease of restriction endonuclease-like (RecB)

577   superfamily, DUF1016 family (**Supplemental Table 6**). Type IV secretion systems secrete

578   virulence factors in which the VirB4 subunit acts as an ATPase and is essential for some bacteria

579   to cause infection [113, 114]. Additionally, restriction endonucleases, such as RecB, cleave DNA

580   at specific sites and act as defense mechanisms for bacteria against foreign DNA [115]. Altogether,

581   these differences in genomic content in human-associated strains suggest they may be more

582   predisposed to succeed in environments where selective pressure may drive strategies for

583   scavenging host metal nutrients and genome changes related to competition with mobile genetic

584   elements.

585

586   **CONCLUSION**

587

588   Host-associated microbial communities often establish intimate and distinguishable

589   relationships that assist host metabolism [116], development [117], behavior [118], immune

590   maturation [119], among others. Moreover, cross-system trends occur such as phylosymbiosis

591   wherein microbial community ecological relationships recapitulate the host phylogenetic

592   relationships. [15]. The *Nasonia* parasitoid wasp genus is a model system with phylosymbiosis in

593   both the bacterial and viral compositions, hybrid maladies associated with the microbiome, and

594   bacterial cultivability that permits further investigation of the host-microbe interactions that

595   mediate these processes. Two bacterial genera dominate the larval microbiome of these wasps in

596   pure species and hybrids, *Proteus mirabilis* and *Providencia rettgeri*, and here we have

597   genomically characterized them relative to each other and to isolates from invertebrate and

598   vertebrate animal hosts. There are three important findings from this work. First, this study reports

599   the genome sequencing of bacteria from hybrid hosts. Recent sequencing of hybrid deep-sea

20

600　mussels showed that their symbionts were genetically indistinguishable from parental mussels

601　[120], and this premise remains to be evaluated in other hybrid systems. Second, the bacterial

602　species in insect-associated environments are not human contaminants as they differ from other

603　host-associated environments and have adapted unique functions for survival that may more tightly

604　regulate symbiotic relationships in insects. Third, the *Proteus mirabilis* genomic diversity is not

605　unique between *Nasonia* species and hybrids, thus supporting a tenet of hologenomic speciation

606　whereby the dominant bacterium in hybrids is a resident microbial taxon in parental species.

607　Therefore, just as the same alleles of nuclear genes in parental species underpin lethality in hybrids,

608　so do bacteria from parental species in the case of *Proteus mirabilis* and *Nasonia* hybrids. Whole

609　genome sequencing of both host and microbial constituents of this association now permit a deeper

610　understanding of the multi-omic interactions between resident members of the microbiome and

611　the host, which in turn underpin phylosymbiosis and hybrid breakdown in these wasps.

612

613

614

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

22

646   the North Carolina State University's Genomic Sciences Laboratory and sequenced using a single

647   flow cell of the Illumina MiSeq platform to produce 2 x 250 bp paired-end reads.

648

649   *Biofilm Growth*

650   Colonies from two isolates (NvitL3-1V and NvitL3-3V) from *N. vitripennis* were grown

651   in overnight cultures of LB broth at 37°C and 130 RPM. The next day, cultures were diluted to an

652   OD_{600} of 0.05 for each isolate, corresponding to a 5×10^5 colony forming units per milliliter. A

653   total of 1 milliliter was plated onto plastic cell culture 12-well plates in triplicate and incubated in

654   a humid chamber overnight at 37°C. To prepare the co-culture inoculate, *Proteus* and *Providencia*

655   were mixed 1:1 before adding a total of 1 milliliter to the wells. After 24 hours, the supernatant

656   was removed and the biofilm stained and measured using established protocols [122]. Briefly, the

657   culture was gently removed, and the plates were gently submerged into distilled water to removed

658   unadhered cells. The plates were allowed to dry within a biosafety hood. The dried plates were

659   then stained with 0.1% crystal violet for 10 minutes, washed three times in distilled water, and

660   allowed to dry again under the hood. The stain retained was then resuspended in 30% acetic acid,

661   and the OD_{600} was measured and reported. Values were graphed in GraphPad Prism 8, and

662   statistical significance was determined using a Kruskal Wallis test with Dunn's correction for

663   multiple comparisons.

664   To quantitatively determine the co-culture composition of the *Proteus mirabilis* and

665   *Providencia rettgeri* biofilm community we developed a qPCR assay using primers specific for

666   unique single-copy genes in each respective genome for *P. mirabilis* (5'-

667   GGTGAGATTTGTATTAATGG and 5'-ATCAGGAAGATGACGAG, annealing temperature

668   58°C) and *P. rettgeri* (5'-AACTCGGTCAGTTCCAAACG and 5'-

669   CTGCATTGTTCGCTTCTCAC, annealing temperature 66°C). *Proteus* primers were designed

670   for the ureR gene using a previously reported forward primer [123] and a new reverse primer.

671   *Providencia* primers were designed based on a phage gene only found within the *Providencia*

672   genomes. The biofilm experiment was repeated as described above except once the supernatant

673   was removed, the adherent cells in the biofilm were recovered from the 1:1 co-culture in 1 milliliter

674   of LB broth. The cells were then pelleted by spinning at 10,000g for 10 minutes and supernatant

675   removed. DNA from the resulting cell pellet was extracted using the Gentra Puregene Tissue Kit

676   (QIAGEN) according to manufacturer's protocol and diluted to ~10 nanograms/microliter each.

677   Amplification was with BioRad iTaq Universal SYBR green supermix in a CFX96 real-time

678   C1000 thermal cycler (BioRad, Hercules, CA) and each qPCR reaction was performed at the

679   following thermal profile: 95°C for 3 minutes, followed by 40 cycles of 95°C for 15 seconds and

680   the respective annealing temperature for each qPCR primer for 1 minute. Samples were calculated

681   using a standard curve generated from dilutions of larger gene products amplified of the same

682   genes for *P. mirabilis* (5'- GCGATTTTACACCGAGTTTC and 5'- ATCCCCATTCTGACATCCAA)

683   and *P. rettgeri* (5'- CCGTTGTGTGTTTGGTATCG and 5'- GTAAGCTGCGTGGATTGGTT). Primer

684   specificity was determined in-silico by BLASTing each primer sequence against each genome and

685   by testing each primer pair on DNA from each bacterial species and observing no amplification

686   either by PCR or qPCR. Isolates and materials are available upon request.

687

688   *Bacterial genome assemblies and annotations*

689   All reads were trimmed using Trimmomatic v0.32 [124] and quality checked with FastQC

690   [125]. Reads were further filtered using the Geneious v. 11.1.5 filter and trim workflow [126]. The

691   reads of each isolate were assembled using SPAdes 3.13.0 [127], and the quality of each assembly

692   was determined using QUAST [128], reads under 2kb were discarded. Genomes were annotated

693   using PROKKA [129] and visualized in Geneious v11.0.3. Genome completion and contamination

694   estimates were calculated using CheckM [48] and Average Nucleotide Identity (ANI) using

695   FastANI [55] online in Kbase (www.kbase.us) [130]. Prophages were identified using VirSorter

696   [131], and plasmids were identified through PROKKA annotations and increased numbers of reads

697   mapping through Bowtie2 [132]. All whole genome sequences were deposited in GenBank under

698   BioProject PRJNA660265. BioSample accessions and further metadata are provided in

699   Supplemental Table 1.

700

701   *Genomic comparisons*

702   For phylogenetic placement of *Proteus* and *Providencia* bacteria isolated in this study, we

703   aligned nucleotide sequences of the gyrase B subunit (*gyrB)* and RNA polymerase B subunit

704   (*rpoB*) separately within Geneious 11.1.5 [126] using the multiple align tool. Using these

705   alignments, we constructed separate unrooted maximum likelihood (ML) phylogenies with 1,000

706   replicates for bootstrapping, using the GTR GAMMA model within RAxML for *gyrB* from both

24

707   *P. rettgeri* and *P. mirabilis*. Evolutionary models were determined with jModelTest [133]. Trees

708   were visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

709   Concatenated amino acid sequences for phylogenomic analyses were computed in anvi'o

710   [58] using 'anvi-get-sequences-for-hmm-hits' with the flags --return-best-hit, --get-aa-sequences,

711   and –concatenate with the --hmm-source Bacteria_71: [type: singlecopy] ADK,

712   AICARFT_IMPCHAs, ATP-synt, ATP-synt_A, Adenylsucc_synt, Chorismate_synt, EF_TS,

713   Exonuc_VII_L, GrpE, Ham1p_like, IPPT, OSCP, PGK, Pept_tRNA_hydro, RBFA, RNA_pol_L,

714   RNA_pol_Rpb6, RRF, RecO_C, Ribonuclease_P, Ribosom_S12_S23, Ribosomal_L1,

715   Ribosomal_L13, Ribosomal_L14, Ribosomal_L16, Ribosomal_L17, Ribosomal_L18p,

716   Ribosomal_L19, Ribosomal_L2, Ribosomal_L20, Ribosomal_L21p, Ribosomal_L22,

717   Ribosomal_L23, Ribosomal_L27, Ribosomal_L27A, Ribosomal_L28, Ribosomal_L29,

718   Ribosomal_L3, Ribosomal_L32p, Ribosomal_L35p, Ribosomal_L4, Ribosomal_L5,

719   Ribosomal_L6, Ribosomal_L9_C, Ribosomal_S10, Ribosomal_S11, Ribosomal_S13,

720   Ribosomal_S15, Ribosomal_S16, Ribosomal_S17, Ribosomal_S19, Ribosomal_S2,

721   Ribosomal_S20p, Ribosomal_S3_C, Ribosomal_S6, Ribosomal_S7, Ribosomal_S8,

722   Ribosomal_S9, RsfS, RuvX, SecE, SecG, SecY, SmpB, TsaE, UPF0054, YajC, eIF-1a,

723   ribosomal_L24, tRNA-synt_1d, tRNA_m1G_MT). Note: *Providencia rettgeri* strain Dmel1 was

724   missing the gene for AICARFT_IMPCHAs so this was manually removed from the *Providencia*

725   alignment in Geneious resulting in a total of 70 genes. Concatenated alignments were imported

726   into Geneious (https://www.geneious.com) [126] v11.0.3 and protein trees built with PhyML using

727   an LG substitution model and 100 bootstraps. Evolutionary models were determined with

728   Modeltest-ng for amino acids [133].

729

730   We used anvi'o [58] v6.2 following the pangenomics workflow [59] to analyze pangenomes

731   of our *Proteus* and *Providencia* isolates with publicly available reference genomes from the

732   National Center for Biotechnology Information (NCBI) (**Supplemental Table 1**), respectively. If

733   genomes contained plasmids, plasmids were included within the genome file. Briefly, we used the

734   program 'anvi-script-FASTA-to-contigs-db' to convert genome fasta nucleotide files into contigs

735   databases for each genome which uses Prodigal [134] v2.6.2 for gene calling. We then annotated

736   each gene using 'anvi-run-ncbi-contigs'. A genomes storage file was created to collect each

737   genome database using 'anvi-gen-genomes-storage' and the pangenome computed using 'anvi-

25

738 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

739

740 amino acid sequence similarity search. Genomes were classified by host (human vs insect) using

741 'anvi-import-misc-data' and functional enrichment analyses were performed using 'anvi-get-

742 enriched-functions-per-pan-group' with --annotation-source COG_FUNCTION. We defined

743 "core" genes of each species pangenome as gene clusters that were present in every genome and

744 accessory genes as those present in only a subset of genomes (e.g. *Nasonia* specific gene clusters).

745 Figures were visualized in anvi'o interactive interface, Inkscape version 1.0 (available from

746 https://inkscape.org/), and finalized in Microsoft Office PowerPoint (v16.37).

747

Acknowledgments

756

Author Contributions

758 A.M. and S.R.B. conceptualized the study and designed experiments. A.M. performed microbial

759 isolation and genome sequencing. K.L.C., B.A.L., E.A.H., and A.K.M. carried out experiments

760 and performed data analysis. K.L.C. and B.A.L. analyzed the genomes and performed pangenomic

761 analyses. K.L.C. and S.R.B. wrote the manuscript with input from B.A.L. S.R.B. supervised the

762 project.

763

REFERENCES:

765

766 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

767

768 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

771   3. Jing T-Z, Qi F-H, Wang Z-Y (2020) Most dominant roles of insect gut bacteria: digestion,
772   detoxification, or essential nutrient provision? *Microbiome*, 8(1):38.
773   https://doi.org/10.1186/s40168-020-00823-y

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

777   5. Wang Y, Rozen DE (2018) Gut microbiota in the burying beetle, *Nicrophorus vespilloides* ,
778   provide colonization resistance against larval bacterial pathogens. *Ecology and Evolution*,
779   8(3):1646–1654. https://doi.org/10.1002/ece3.3589

780   6. Msaad Guerfali M, Djobbi W, Charaabi K, Hamden H, Fadhl S, Marzouki W, Dhaouedi F,
781   Chevrier C (2018) Evaluation of Providencia rettgeri pathogenicity against laboratory
782   Mediterranean fruit fly strain (Ceratitis capitata). *PLOS ONE*, 13(5):e0196343.
783   https://doi.org/10.1371/journal.pone.0196343

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

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

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

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

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

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

801   13. Song SJ, Sanders JG, Delsuc F, Metcalf J, Amato K, Taylor MW, Mazel F, Lutz HL, Winker
802   K, Graves GR, Humphrey G, Gilbert JA, Hackett SJ, White KP, Skeen HR, Kurtis SM, Withrow
803   J, Braile T, Miller M, McCracken KG, Maley JM, Ezenwa VO, Williams A, Blanton JM,
804   McKenzie VJ, Knight R (2020) Comparative Analyses of Vertebrate Gut Microbiomes Reveal
805   Convergence between Birds and Bats. *mBio*, 11(1):e02901-19, /mbio/11/1/mBio.02901-19.atom.
806   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

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

847 25. Wang J, Kalyan S, Steck N, Turner LM, Harr B, Künzel S, Vallier M, Häsler R, Franke A,
848 Oberg H-H, Ibrahim SM, Grassl GA, Kabelitz D, Baines JF (2015) Analysis of intestinal
849 microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome.
850 *Nature Communications*, 6(1):6440. https://doi.org/10.1038/ncomms7440

851 26. Campbell BC, Steffen-Campbell JD, Werren JH (1993) Phylogeny of the Nasonia species
852 complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and
853 28S rDNA sequences. *Insect Molecular Biology*, 2(4):225–237.

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

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

860 29. Shropshire JD, Opstal EJ van, Bordenstein SR (2016) An optimized approach to germ-free
861 rearing in the jewel wasp Nasonia. *PeerJ*, 4:e2316. https://doi.org/10.7717/peerj.2316

862 30. Brucker RM, Bordenstein SR (2012) Speciation by symbiosis. *Trends Ecol Evol*, 27(8):443–
863 51. https://doi.org/10.1016/j.tree.2012.03.011

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

885 Wyder S, Yamada T, Yi SV, Zecher CN, Zhang L, Gibbs RA (2010) Functional and
886 evolutionary insights from the genomes of three parasitoid Nasonia species. *Science*,
887 327(5963):343–8. https://doi.org/10.1126/science.1178028

888 32. Wang X, Kelkar YD, Xiong X, Martinson EO, Lynch J, Zhang C, Werren JH, Wang X
889 (2020) Genome Report: Whole Genome Sequence and Annotation of the Parasitoid Jewel Wasp
890 *Nasonia giraulti* Laboratory Strain RV2X[u]. *G3&#58; Genes|Genomes|Genetics*,
891 10(8):2565–2572. https://doi.org/10.1534/g3.120.401200

892 33. Bitar I, Mattioni Marchetti V, Mercato A, Nucleo E, Anesi A, Bracco S, Rognoni V, Hrabak
893 J, Migliavacca R (2020) Complete Genome and Plasmids Sequences of a Clinical Proteus
894 mirabilis Isolate Producing Plasmid Mediated NDM-1 From Italy. *Microorganisms*, 8(3):339.
895 https://doi.org/10.3390/microorganisms8030339

896 34. Di Pilato V, Chiarelli A, Boinett CJ, Riccobono E, Harris SR, D'Andrea MM, Thomson NR,
897 Rossolini GM, Giani T (2016) Complete Genome Sequence of the First KPC-Type
898 Carbapenemase-Positive *Proteus mirabilis* Strain from a Bloodstream Infection. *Genome
899 Announcements*, 4(3):e00607-16, /ga/4/3/e00607-16.atom.
900 https://doi.org/10.1128/genomeA.00607-16

901 35. Pearson MM, Sebaihia M, Churcher C, Quail MA, Seshasayee AS, Luscombe NM, Abdellah
902 Z, Arrosmith C, Atkin B, Chillingworth T, Hauser H, Jagels K, Moule S, Mungall K,
903 Norbertczak H, Rabbinowitsch E, Walker D, Whithead S, Thomson NR, Rather PN, Parkhill J,
904 Mobley HLT (2008) Complete Genome Sequence of Uropathogenic Proteus mirabilis, a Master
905 of both Adherence and Motility. *Journal of Bacteriology*, 190(11):4027–4037.
906 https://doi.org/10.1128/JB.01981-07

907 36. Octavia S, Xu W, Ng OT, Marimuthu K, Venkatachalam I, Cheng B, Lin RTP, Teo JWP
908 (2020) Identification of AbaR4 Acinetobacter baumannii resistance island in clinical isolates of
909 blaOXA-23-positive Proteus mirabilis. *Journal of Antimicrobial Chemotherapy*, 75(3):521–525.
910 https://doi.org/10.1093/jac/dkz472

911 37. Schaffer JN, Pearson MM (2015) Proteus mirabilis and Urinary Tract Infections.
912 *Microbiology Spectrum*, 3(5)https://doi.org/10.1128/microbiolspec.UTI-0017-2013

913 38. Galac MR, Lazzaro BP (2012) Comparative genomics of bacteria in the genus Providencia
914 isolated from wild Drosophila melanogaster. *BMC Genomics*, 13(1):612.
915 https://doi.org/10.1186/1471-2164-13-612

916 39. Unckless RL, Howick VM, Lazzaro BP (2016) Convergent Balancing Selection on an
917 Antimicrobial Peptide in Drosophila. *Current Biology*, 26(2):257–262.
918 https://doi.org/10.1016/j.cub.2015.11.063

919 40. Hanson MA, Dostálová A, Ceroni C, Poidevin M, Kondo S, Lemaitre B (2019) Synergy and
920 remarkable specificity of antimicrobial peptides in vivo using a systematic knockout approach.
921 *eLife*, 8:e44341. https://doi.org/10.7554/eLife.44341

922   41. O'Donnell MP, Fox BW, Chao P-H, Schroeder FC, Sengupta P (2020) A neurotransmitter
923   produced by gut bacteria modulates host sensory behaviour. *Nature*, 583:23.
924   https://doi.org/10.1038/s41586-020-2395-5

925   42. Wang G-H, Brucker RM (2019) Genome Sequence of Providencia rettgeri NVIT03, Isolated
926   from Nasonia vitripennis . *Microbiology Resource Announcements*,
927   https://doi.org/10.1128/mra.01157-18

928   43. Tchuinte PLS, Rabenandrasana MAN, Ramparany L, Ratsima E, Enouf V, Randrianirina F,
929   Collard J-M (2020) Genome-based insights into the resistomes and mobilomes of two
930   Providencia rettgeri strains isolated from wound infections in Madagascar. *Journal of Global*
931   *Antimicrobial Resistance*, 20:178–182. https://doi.org/10.1016/j.jgar.2019.07.013

932   44. Marquez-Ortiz RA, Haggerty L, Sim EM, Duarte C, Castro-Cardozo BE, Beltran M,
933   Saavedra S, Vanegas N, Escobar-Perez J, Petty NK (2017) First Complete Providencia rettgeri
934   Genome Sequence, the NDM-1-Producing Clinical Strain RB151. *Genome Announcements*,
935   5(3):e01472-16, e01472-16. https://doi.org/10.1128/genomeA.01472-16

936   45. Maltz MA, Weiss BL, O'Neill M, Wu Y, Aksoy S (2012) OmpA-Mediated Biofilm
937   Formation Is Essential for the Commensal Bacterium Sodalis glossinidius To Colonize the
938   Tsetse Fly Gut. *Applied and Environmental Microbiology*, 78(21):7760–7768.
939   https://doi.org/10.1128/AEM.01858-12

940   46. Drzewiecka D (2016) Significance and Roles of Proteus spp. Bacteria in Natural
941   Environments. *Microbial Ecology*, 72(4):741–758. https://doi.org/10.1007/s00248-015-0720-6

942   47. Budding AE, Ingham CJ, Bitter W, Vandenbroucke-Grauls CM, Schneeberger PM (2009)
943   The Dienes Phenomenon: Competition and Territoriality in Swarming Proteus mirabilis. *Journal*
944   *of Bacteriology*, 191(12):3892–3900. https://doi.org/10.1128/JB.00975-08

945   48. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing
946   the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
947   *Genome Research*, 25(7):1043–1055. https://doi.org/10.1101/gr.186072.114

948   49. O'Hara CM, Brenner FW, Miller JM (2000) Classification, Identification, and Clinical
949   Significance of Proteus, Providencia, and Morganella. *CLIN. MICROBIOL. REV.*, 13:13.

950   50. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris
951   M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J,
952   Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI (2012)
953   Human gut microbiome viewed across age and geography. *Nature*, 486(7402):222–227.
954   https://doi.org/10.1038/nature11053

955   51. Gu W, Wang W, Tong P, Liu C, Jia J, Lu C, Han Y, Sun X, Kuang D, Li N, Dai J (2020)
956   Comparative genomic analysis of Proteus spp. isolated from tree shrews indicated unexpectedly
957   high genetic diversity. *PLOS ONE*, 15(2):e0229125.
958   https://doi.org/10.1371/journal.pone.0229125

959 52. Gu W, Tong P, Liu C, Wang W, Lu C, Han Y, Sun X, Kuang DX, Li N, Dai J (2019) The
960 characteristics of gut microbiota and commensal Enterobacteriaceae isolates in tree shrew
961 (Tupaia belangeri). *BMC Microbiology*, 19(1):203. https://doi.org/10.1186/s12866-019-1581-9

962 53. Ma Q, Fonseca A, Liu W, Fields AT, Pimsler ML, Spindola AF, Tarone AM, Crippen TL,
963 Tomberlin JK, Wood TK (2012) Proteus mirabilis interkingdom swarming signals attract blow
964 flies. *The ISME Journal*, 6(7):1356–1366. https://doi.org/10.1038/ismej.2011.210

965 54. Galac MR, Lazzaro BP (2012) Comparative genomics of bacteria in the genus Providencia
966 isolated from wild Drosophila melanogaster. *BMC Genomics*, https://doi.org/10.1186/1471-
967 2164-13-612

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

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

974 57. Sichtig H, Minogue T, Yan Y, Stefan C, Hall A, Tallon L, Sadzewicz L, Nadendla S, Klimke
975 W, Hatcher E, Shumway M, Aldea DL, Allen J, Koehler J, Slezak T, Lovell S, Schoepp R,
976 Scherf U (2019) FDA-ARGOS is a database with public quality-controlled reference genomes
977 for diagnostic use and regulatory science. *Nature Communications*, 10(1):3313.
978 https://doi.org/10.1038/s41467-019-11306-6

979 58. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO (2015)
980 Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, 3:e1319.
981 https://doi.org/10.7717/peerj.1319

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

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

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

988 62. M. V. Murthy A, Phan M-D, Peters KM, Nhu NTK, Welch RA, Ulett GC, Schembri MA,
989 Sweet MJ (2018) Regulation of hemolysin in uropathogenic *Escherichia coli* fine-tunes killing of
990 human macrophages. *Virulence*, 9(1):967–980. https://doi.org/10.1080/21505594.2018.1465786

991 63. Spaulding CN, Schreiber HL, Zheng W, Dodson KW, Hazen JE, Conover MS, Wang F,
992 Svenmarker P, Luna-Rico A, Francetic O, Andersson M, Hultgren S, Egelman EH (2018)
993 Functional role of the type 1 pilus rod structure in mediating host-pathogen interactions. *eLife*,
994 7:e31662. https://doi.org/10.7554/eLife.31662

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

997   65. Jacob-Dubuisson F, Locht C, Antoine R (2001) Two-partner secretion in Gram-negative
998   bacteria: a thrifty, specific pathway for large virulence proteins. *Molecular Microbiology*,
999   40(2):306–313. https://doi.org/10.1046/j.1365-2958.2001.02278.x

1000   66. Mobley HLT, Chippendale GR, Tenney JH, Mayrer AR, Crisp LJ, Penner JL, Warren JW
1001   (1988) MR/K Hemagglutination of Providencia stuartii Correlates with Adherence to Catheters
1002   and with Persistence in Catheter-Associated Bacteriuria. *Journal of Infectious Diseases*,
1003   157(2):264–271. https://doi.org/10.1093/infdis/157.2.264

1004   67. Locht C, Geoffroy MC, Renauld G (1992) Common accessory genes for the Bordetella
1005   pertussis filamentous hemagglutinin and fimbriae share sequence similarities with the papC and
1006   papD gene families. *The EMBO Journal*, 11(9):3175–3183. https://doi.org/10.1002/j.1460-
1007   2075.1992.tb05394.x

1008   68. Barrangou R, Marraffini LA (2014) CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive
1009   Immunity. *Molecular Cell*, 54(2):234–244. https://doi.org/10.1016/j.molcel.2014.03.011

1010   69. Giraud M-F, Naismith JH (2000) The rhamnose pathway. *Current Opinion in Structural*
1011   *Biology*, 10(6):687–696. https://doi.org/10.1016/S0959-440X(00)00145-7

1012   70. Rodionova IA, Li X, Thiel V, Stolyar S, Stanton K, Fredrickson JK, Bryant DA, Osterman
1013   AL, Best AA, Rodionov DA (2013) Comparative genomics and functional analysis of rhamnose
1014   catabolic pathways and regulons in bacteria. *Frontiers in Microbiology*,
1015   4https://doi.org/10.3389/fmicb.2013.00407

1016   71. Boronat A, Aguilar J (1981) Metabolism of L-fucose and L-rhamnose in Escherichia coli:
1017   differences in induction of propanediol oxidoreductase. *Journal of Bacteriology*, 147(1):181–
1018   185. https://doi.org/10.1128/JB.147.1.181-185.1981

1019   72. Carvalho F, Atilano ML, Pombinho R, Covas G, Gallo RL, Filipe SR, Sousa S, Cabanes D
1020   (2015) L-Rhamnosylation of Listeria monocytogenes Wall Teichoic Acids Promotes Resistance
1021   to Antimicrobial Peptides by Delaying Interaction with the Membrane. *PLOS Pathogens*,
1022   11(5):e1004919. https://doi.org/10.1371/journal.ppat.1004919

1023   73. Delmas J, Gibold L, Faïs T, Batista S, Leremboure M, Sinel C, Vazeille E, Cattoir V,
1024   Buisson A, Barnich N, Dalmasso G, Bonnet R (2019) Metabolic adaptation of adherent-invasive
1025   Escherichia coli to exposure to bile salts. *Scientific Reports*, 9(1):2175.
1026   https://doi.org/10.1038/s41598-019-38628-1

1027   74. Otto BR, Sparrius M, Verweij-van Vught AM, MacLaren DM (1990) Iron-regulated outer
1028   membrane protein of Bacteroides fragilis involved in heme uptake. *Infection and Immunity*,
1029   58(12):3954–3958. https://doi.org/10.1128/IAI.58.12.3954-3958.1990

1030   75. Klebba PE, McIntosh MA, Neilands JB (1982) Kinetics of Biosynthesis of Iron-Regulated
1031   Membrane Proteins in Escherichia coli. *Journal of Bacteriology*, 149(3):880–888.
1032   https://doi.org/10.1128/JB.149.3.880-888.1982

1033   76. Tang X, Zhou B (2013) Iron homeostasis in insects: Insights from *Drosophila* studies: Iron
1034   Homeostasis in Insects. *IUBMB Life*, 65(10):863–872. https://doi.org/10.1002/iub.1211

1035   77. Kremer N, Voronin D, Charif D, Mavingui P, Mollereau B, Vavre F (2009) Wolbachia
1036   Interferes with Ferritin Expression and Iron Metabolism in Insects. *PLoS Pathogens*,
1037   5(10):e1000630. https://doi.org/10.1371/journal.ppat.1000630

1038   78. Resch CT, Winogrodzki JL, Patterson CT, Lind EJ, Quinn MJ, Dibrov P, Häse CC (2010)
1039   The Putative Na$^+$/H$^+$ Antiporter of *Vibrio cholerae* , Vc-NhaP2, Mediates the Specific K$^+$/H$^+$
1040   Exchange in Vivo. *Biochemistry*, 49(11):2520–2528. https://doi.org/10.1021/bi902173y

1041   79. Waditee R, Buaboocha T, Kato M, Hibino T, Suzuki S, Nakamura T, Takabe T (2006)
1042   Carboxyl-terminal hydrophilic tail of a NhaP type Na+/H+ antiporter from cyanobacteria is
1043   involved in the apparent affinity for Na+ and pH sensitivity. *Archives of Biochemistry and
1044   Biophysics*, 450(1):113–121. https://doi.org/10.1016/j.abb.2006.02.013

1045   80. Geiger O, López-Lara IM, Sohlenkamp C (2013) Phosphatidylcholine biosynthesis and
1046   function in bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of
1047   Lipids*, 1831(3):503–513. https://doi.org/10.1016/j.bbalip.2012.08.009

1048   81. Sohlenkamp C, López-Lara IM, Geiger O (2003) Biosynthesis of phosphatidylcholine in
1049   bacteria. *Progress in Lipid Research*, 42(2):115–162. https://doi.org/10.1016/S0163-
1050   7827(02)00050-4

1051   82. Aktas M, Wessel M, Hacker S, Klüsener S, Gleichenhagen J, Narberhaus F (2010)
1052   Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic
1053   cells. *European Journal of Cell Biology*, 89(12):888–894.
1054   https://doi.org/10.1016/j.ejcb.2010.06.013

1055   83. Zhang G, Liu Y, Ruoho AE, Hurley JH (1997) Structure of the adenylyl cyclase catalytic
1056   core. *Nature*, 386(6622):247–253. https://doi.org/10.1038/386247a0

1057   84. Iyer LM, Aravind L (2002) The catalytic domains of thiamine triphosphatase and CyaB-like
1058   adenylyl cyclase define a novel superfamily of domains that bind organic phosphates. *BMC
1059   Genomics*, 3(1):33. https://doi.org/10.1186/1471-2164-3-33

1060   85. Smith RS, Wolfgang MC, Lory S (2004) An Adenylate Cyclase-Controlled Signaling
1061   Network Regulates Pseudomonas aeruginosa Virulence in a Mouse Model of Acute Pneumonia.
1062   *Infection and Immunity*, 72(3):1677–1684. https://doi.org/10.1128/IAI.72.3.1677-1684.2004

1063   86. Murray NE (2003) Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy
1064   of Bertani and Weigle). *Microbiology and Molecular Biology Reviews*,
1065   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

1103   98. Soltero-Higgin M, Carlson EE, Gruber TD, Kiessling LL (2004) A unique catalytic
1104   mechanism for UDP-galactopyranose mutase. *Nature Structural & Molecular Biology*,
1105   11(6):539–543. https://doi.org/10.1038/nsmb772

1106   99. Pan F, Jackson M, Ma Y, Mcneil M (2001) Cell Wall Core Galactofuran Synthesis Is
1107   Essential for Growth of Mycobacteria. *J. BACTERIOL.*, 183:8.

1108   100. Strating H, Vandenende C, Clarke AJ (2012) Changes in peptidoglycan structure and
1109   metabolism during differentiation of Proteus mirabilis into swarmer cells. *Canadian journal of*
1110   *microbiology*, 58(10):1183–1194.

1111   101. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir
1112   G, Sorek R (2015) BREX is a novel phage resistance system widespread in microbial genomes.
1113   *The EMBO Journal*, 34(2):169–183. https://doi.org/10.15252/embj.201489455

1114   102. Chaudhary K (2018) BacteRiophage EXclusion (BREX): A novel anti-phage mechanism in
1115   the arsenal of bacterial defense system. *Journal of Cellular Physiology*, 233(2):771–773.
1116   https://doi.org/10.1002/jcp.25973

1117   103. Alam MdS, Chakraborty S, Rahman T, Hosen MdI, Paul A, Hasan AKMM, Hossain MA
1118   (2018) Investigation of the Potential Association between Clustered Regularly Interspersed Short
1119   Palindromic Repeats (CRISPR) and Antibiotic Resistance Pattern of Bacterial Strains Isolated
1120   from Medical Waste and Environmental Water. *Open Journal of Medical Microbiology*,
1121   08(02):13–25. https://doi.org/10.4236/ojmm.2018.82002

1122   104. Sommer U, Petersen J, Pfeiffer M, Schrotz-King P, Morsczeck C (2010) Comparison of
1123   surface proteomes of enterotoxigenic (ETEC) and commensal Escherichia coli strains. *Journal of*
1124   *Microbiological Methods*, 83(1):13–19. https://doi.org/10.1016/j.mimet.2010.07.011

1125   105. Palmer LD, Skaar EP (2016) Transition Metals and Virulence in Bacteria. *Annual Review of*
1126   *Genetics*, 50(1):67–91. https://doi.org/10.1146/annurev-genet-120215-035146

1127   106. Flo TH, Smith KD, Sato S, Rodriguez DJ, Holmes MA, Strong RK, Akira S, Aderem A
1128   (2004) Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating
1129   iron. *Nature*, 432(7019):917–921. https://doi.org/10.1038/nature03104

1130   107. Abergel RJ, Wilson MK, Arceneaux JEL, Hoette TM, Strong RK, Byers BR, Raymond KN
1131   (2006) Anthrax pathogen evades the mammalian immune system through stealth siderophore
1132   production. *Proceedings of the National Academy of Sciences*, 103(49):18499–18503.
1133   https://doi.org/10.1073/pnas.0607055103

1134   108. Shields-Cutler RR, Crowley JR, Hung CS, Stapleton AE, Aldrich CC, Marschall J,
1135   Henderson JP (2015) Human Urinary Composition Controls Antibacterial Activity of
1136   Siderocalin. *Journal of Biological Chemistry*, 290(26):15949–15960.
1137   https://doi.org/10.1074/jbc.M115.645812

1138   109. Perry WJ, Spraggins JM, Sheldon JR, Grunenwald CM, Heinrichs DE, Cassat JE, Skaar EP,
1139   Caprioli RM (2019) *Staphylococcus aureus* exhibits heterogeneous siderophore production

1140 within the vertebrate host. *Proceedings of the National Academy of Sciences*, 116(44):21980–
1141 21982. https://doi.org/10.1073/pnas.1913991116

1142 110. Tiwari V, Rajeswari MR, Tiwari M (2019) Proteomic analysis of iron-regulated membrane
1143 proteins identify FhuE receptor as a target to inhibit siderophore-mediated iron acquisition in
1144 Acinetobacter baumannii. *International Journal of Biological Macromolecules*, 125:1156–1167.
1145 https://doi.org/10.1016/j.ijbiomac.2018.12.173

1146 111. Fan C, Wu Y-H, Decker CM, Rohani R, Gesell Salazar M, Ye H, Cui Z, Schmidt F, Huang
1147 WE (2019) Defensive Function of Transposable Elements in Bacteria. *ACS Synthetic Biology*,
1148 8(9):2141–2151. https://doi.org/10.1021/acssynbio.9b00218

1149 112. Hartl DL, Dykhuizen DE, Miller RD, Green L, Framond J de (1983) Transposable element
1150 IS50 improves growth rate of E. coli cells without transposition. *Cell*, 35(2):503–510.
1151 https://doi.org/10.1016/0092-8674(83)90184-8

1152 113. Sgro GG, Oka GU, Souza DP, Cenens W, Bayer-Santos E, Matsuyama BY, Bueno NF,
1153 Santos TR dos, Alvarez-Martinez CE, Salinas RK, Farah CS (2019) Bacteria-Killing Type IV
1154 Secretion Systems. *Frontiers in Microbiology*, 10:1078.
1155 https://doi.org/10.3389/fmicb.2019.01078

1156 114. Schulein R, Dehio C (2002) The VirB/VirD4 type IV secretion system of Bartonella is
1157 essential for establishing intraerythrocytic infection: Role of type IV secretion in Bartonella.
1158 *Molecular Microbiology*, 46(4):1053–1067. https://doi.org/10.1046/j.1365-2958.2002.03208.x

1159 115. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG, Murray NE (2014) Highlights of the
1160 DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research*, 42(1):3–19.
1161 https://doi.org/10.1093/nar/gkt990

1162 116. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Pettersson S (2012)
1163 Host-Gut Microbiota Metabolic Interactions. 336:6.

1164 117. Sommer F, Bäckhed F (2013) The gut microbiota — masters of host development and
1165 physiology. *Nature Reviews Microbiology*, 11(4):227–238. https://doi.org/10.1038/nrmicro2974

1166 118. Heijtz RD, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, Hibberd ML, Forssberg
1167 H, Pettersson S Normal gut microbiota modulates brain development and behavior. :6.

1168 119. Round JL, Mazmanian SK (2009) The gut microbiota shapes intestinal immune responses
1169 during health and disease. *Nat Rev Immunol.*, 9:313–23. https://doi.org/doi:
1170 10.1038/nri2515\nnri2515

1171 120. Ücker M, Ansorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N (2020) Deep-sea
1172 mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically indistinguishable
1173 symbionts. https://doi.org/10.1101/2020.09.11.282681

1174   121. Niehuis O, Judson AK, Gadau J (2008) Cytonuclear Genic Incompatibilities Cause
1175   Increased Mortality in Male F 2 Hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics*,
1176   178(1):413–426. https://doi.org/10.1534/genetics.107.080523

1177   122. O'Toole GA (2011) Microtiter Dish Biofilm Formation Assay. *Journal of Visualized*
1178   *Experiments*, (47):2437. https://doi.org/10.3791/2437

1179   123. Zhang W, Niu Z, Yin K, Liu P, Chen L (2013) Quick identification and quantification of
1180   Proteus mirabilis by polymerase chain reaction (PCR) assays. *Annals of Microbiology*,
1181   63(2):683–689. https://doi.org/10.1007/s13213-012-0520-x

1182   124. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina
1183   sequence data. *Bioinformatics*, 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170

1184   125. Simon A (2010) FastQC: a quality control tool for high throughput sequence data.

1185   126. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
1186   Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic:
1187   An integrated and extendable desktop software platform for the organization and analysis of
1188   sequence data. *Bioinformatics*, https://doi.org/10.1093/bioinformatics/bts199

1189   127. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
1190   Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G,
1191   Alekseyev MA, Pevzner PA (2012) SPAdes: A New Genome Assembly Algorithm and Its
1192   Applications to Single-Cell Sequencing. *Journal of Computational Biology*,
1193   https://doi.org/10.1089/cmb.2012.0021

1194   128. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for
1195   genome assemblies. *Bioinformatics*, 29(8):1072–1075.
1196   https://doi.org/10.1093/bioinformatics/btt086

1197   129. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*,
1198   30(14):2068–2069. https://doi.org/10.1093/bioinformatics/btu153

1199   130. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware
1200   D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY,
1201   Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ,
1202   Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M,
1203   Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I,
1204   Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski
1205   J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F,
1206   Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS,
1207   Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M,
1208   Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D (2018)
1209   KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature*
1210   *Biotechnology*, 36(7):566–569. https://doi.org/10.1038/nbt.4163

1211   131. Roux S, Enault F, Hurwitz BL, Sullivan MB (2015) VirSorter: mining viral signal from
1212   microbial genomic data. *PeerJ*, 3:e985. https://doi.org/10.7717/peerj.985

1213   132. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature*
1214   *Methods*, 9(4):357–359. https://doi.org/10.1038/nmeth.1923

1215   133. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T (2020) ModelTest-NG:
1216   A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular*
1217   *Biology and Evolution*, 37(1):291–294. https://doi.org/10.1093/molbev/msz189

1218   134. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal:
1219   prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*,
1220   11(1):119. https://doi.org/10.1186/1471-2105-11-119

1221   135. AltschuP SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment
1222   Search Tool. *Journal of Molecular Biology*, 215:403–410.

1223

1   **FIGURES**

2

3
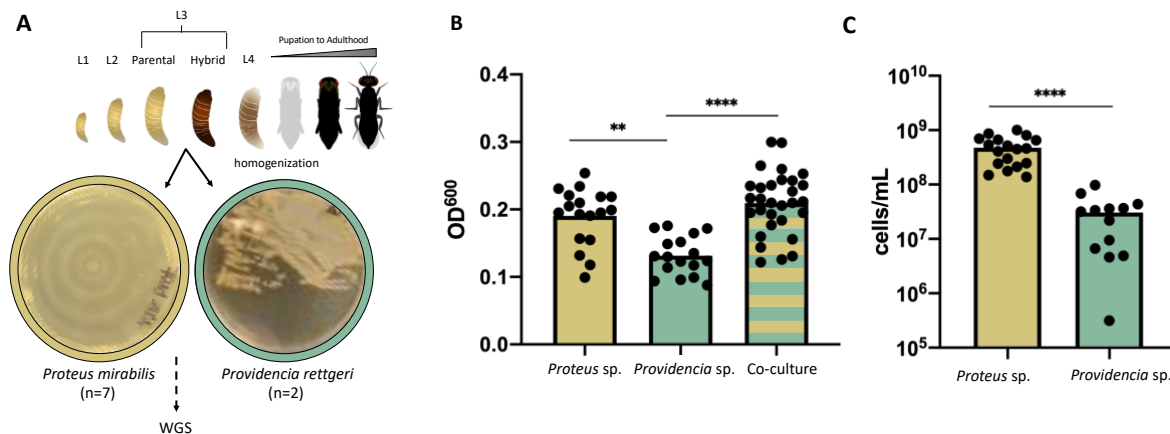
4   **Figure 1. Characteristics of *Proteus mirabilis* and *Providencia rettgeri* isolates. A)** Workflow

5   for sample collection and examples of unique colony morphologies of *Providencia* (n=2) and

6   *Proteus* (n=7) isolates. *Nasonia* development from larvae to adulthood is depicted on top

7   demonstrating the L3 larval stage in which samples were collected from parental and hybrid lines.

8   Following isolation, samples were prepared for whole genome sequencing (WGS). **B)** Biofilm

9   formation by representative *Proteus* and *Providencia* isolates show significantly different abilities

10   to form biofilms on solid surfaces. Bars denote significantly different values based on Kruskal-

11   Wallis test with Dunn's multiple comparison with sample size of n=17-30 of two biological

12   replicates (** *p*=0.0018; **** p=<0.0001). **C)** Composition of the biofilm co-culture shows

13   *Proteus mirabilis* significantly outcompetes *Providencia rettgeri* when grown together.

14   *Providencia rettgeri* makes up about 6.4% of the co-culture biofilm composition. Bars denote

15   significantly different values calculated by a Mann-Whitney U test with sample size of n = 13-18

16   (**** p=<0.0001).

40

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	F2Hybrid-1H	F2Hybrid-2H	F2Hybrid-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*)
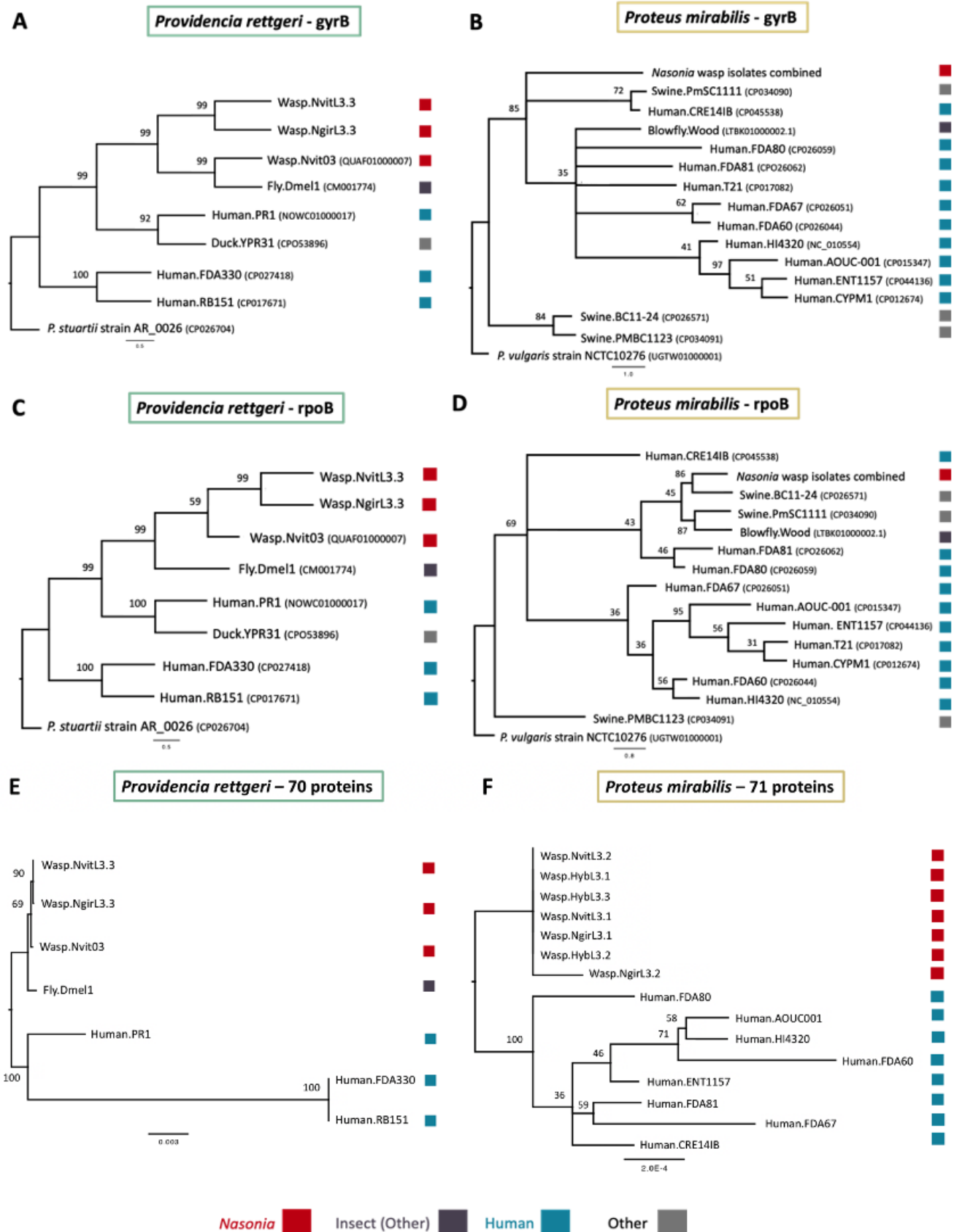
1   **Figure 2. Phylogenetic placement of *Nasonia* isolates relative to publicly available genomes. A-B)** Phylogeny

2   of *Providencia rettgeri* and *Proteus mirabilis* strains based on 2,415 nucleotides of the gyrase B subunit

3   reconstructed using RAxML. **C-D)** Phylogeny of *Providencia rettgeri* and *Proteus mirabilis* strains based on

4   4,029 nucleotides of the RNA Polymerase beta subunit reconstructed using RAxML. **E-F)** Phylogeny of

5   *Providencia rettgeri* and *Proteus mirabilis* strains used in this study for comparative genomics analyses

6   (Supplemental Table 1) based on concatenated amino acid alignments of 70-71 core bacterial proteins

7   reconstructed using PhyML.

8
9
10
11
12
13
14
15
16

17
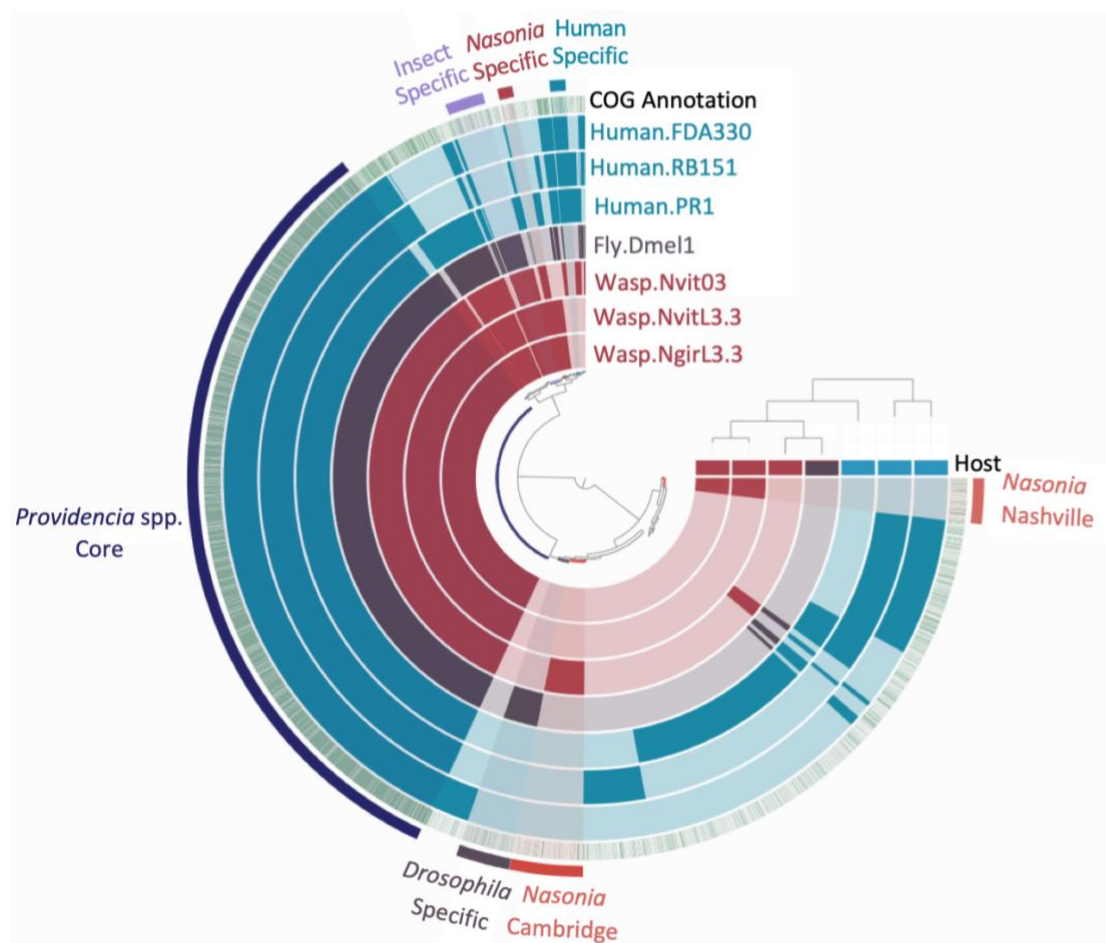
1
2 **Figure 3. Pangenome of *Providencia rettgeri* human and insect isolates.** The seven inner layers correspond to
3 the seven genomes, including the sequenced *Nasonia* isolates in maroon, previously published *Drosophila* isolate
4 in purple, and human associated genomes in blue. The bars in the inner circles show the presence of gene-clusters
5 (GCs) in a given genome with the outer green circle depicting known COGs (green) vs unknown (white)
6 assignments. The outermost layer of color-coded lines and text highlight groups of GCs that correspond to the
7 genome core or to group-specific GCs. Genomes are ordered based on their gene-cluster distribution across
8 genomes, which is shown at the top right corner (tree). Central dendrograms depict protein cluster hierarchy when
9 displayed as protein cluster frequency. The top horizontal layer underneath the tree represent host source (*Nasonia*
10 - maroon*, Drosophila* - purple*, or Human - blue).

11

12
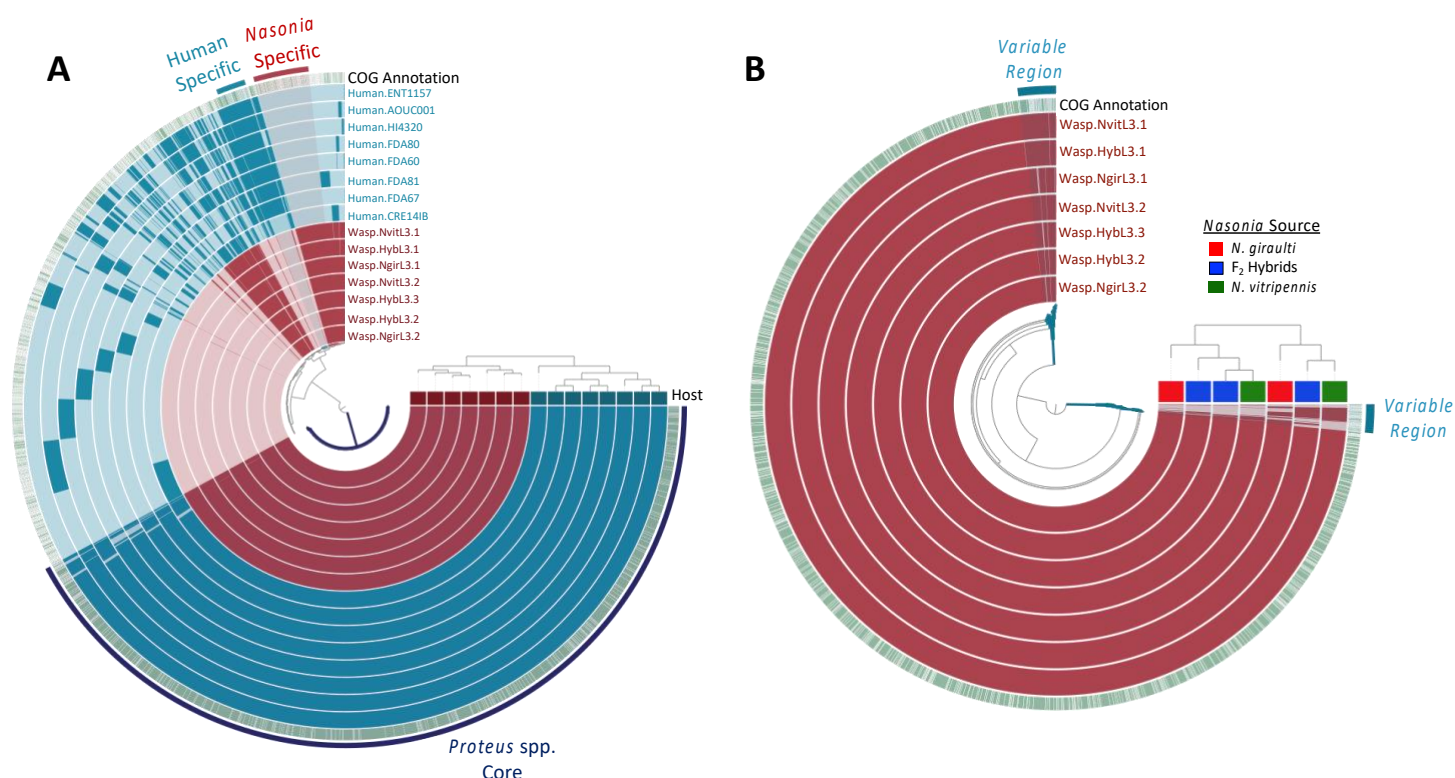
Figure 4. Pangenome of *Proteus mirabilis* human and *Nasonia* isolates. A) The fifteen inner layers correspond to all fifteen genomes, with *Nasonia* isolates shown in maroon and human isolates shown in blue. The top horizontal layer underneath the tree represent host source (*Nasonia* - maroon, Human - 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; F₂ 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

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

Columns (left to right): Human.FDA330, Human.RB151, Human.PR1, Fly.Dmel1, Wasp.Nvit03, Wasp.NvitL3.3, Wasp.NgirL3.3

ANI Estimate:
- >= 99.0%
- Insect (Other)
- *Nasonia*
- Human

Proteus mirabilis

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.NgirL3.2
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.HybL3.2
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.HybL3.3
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.NvitL3.2
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.NgirL3.1
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.HybL3.1
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	Wasp.NvitL3.1
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.CRE14IB
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.FDA67
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.FDA81
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.FDA60
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.FDA80
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.HI4320
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.AOUC001
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	Human.ENT1157

Columns (left to right): 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

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.

46

1

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.

1
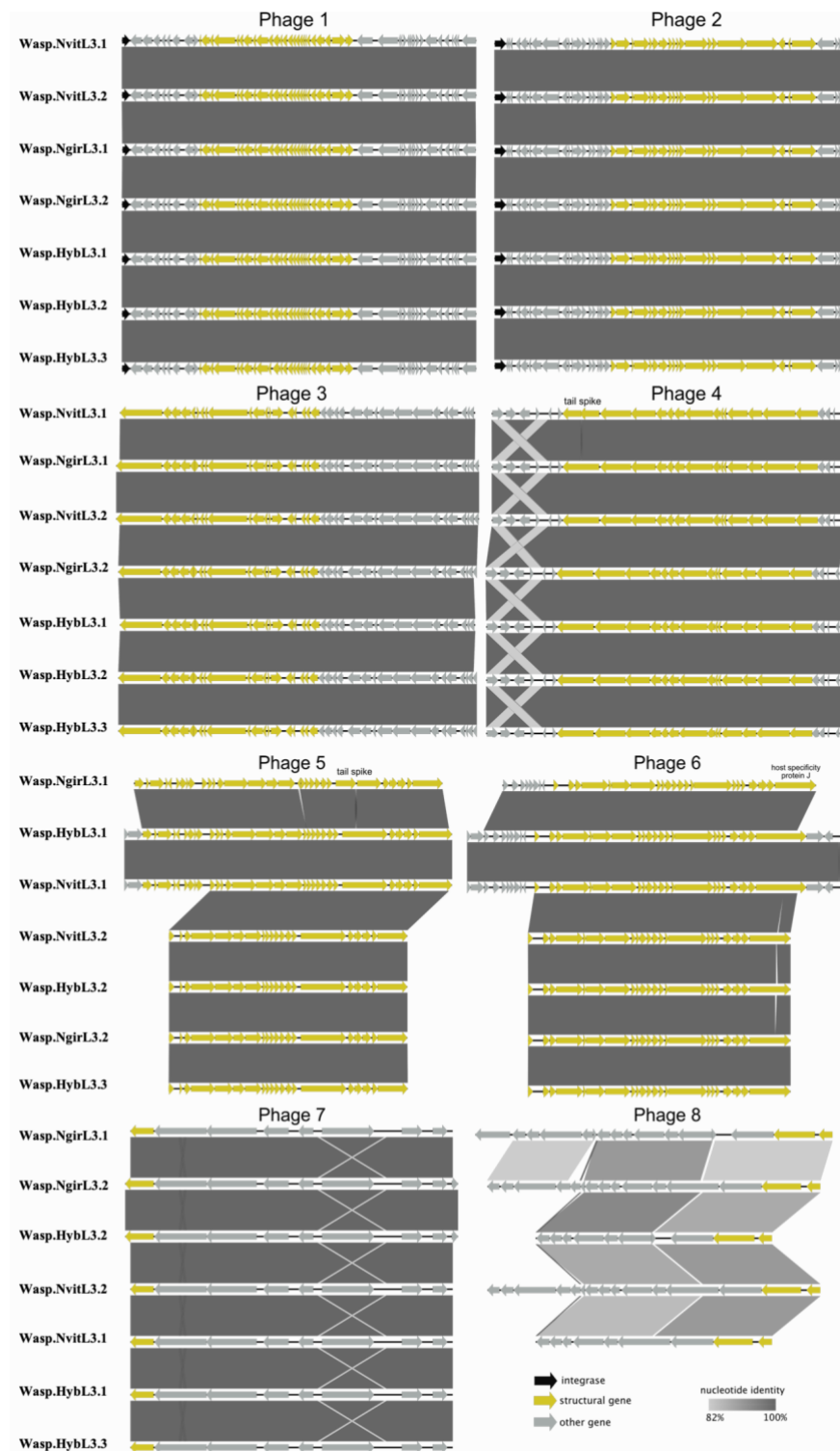
Supplemental Figure 3. Phages of *Proteus mirabilis Nasonia* isolates. Eight total phages were identified. Homologous phages were determined through whole genome alignments of putative prophages identified by VirSorter. Percent nucleotide identity as determined by blastn is indicated by shaded bars between each predicted coding gene. Disrupted tail proteins are labeled.

1   **Supplemental Table 1**. Whole genome sequences used for comparative genomics analyses.
2

3   **Supplemental Table 2.** Genome Bins associated with the *Providencia rettgeri* pangenomic analysis of the seven
4   *Providencia rettgeri* genomes.
5

6   **Supplemental Table 3.** COG analyses for comparison between the different *Providencia* anvi'o bins and the
7   different *Proteus* anvi'o bins.
8

9   **Supplemental Table 4.** *Providencia rettgeri* enrichment analysis.
10

11   **Supplemental Table 5.** Genome Bins associated with the *Proteus mirabilis* pangenomic analysis of the fifteen
12   *Proteus mirabilis* genomes.
13

14   **Supplemental Table 6.** *Proteus mirabilis* enrichment analysis.
15
16