1 **Characterisation of the symbionts in the Mediterranean fruitfly gut**

2

3 Darrington, M.[1]*, Leftwich, P.T.[1]*, Holmes, N.A.[1,4], Friend, L.A.[1], Clarke N.V.E.[1], Worsley, S.F.[1],

4 Margaritopolous, J.T.[2], Hogenhout, S.A.[3], Hutchings, M.I.[1,4] & Chapman, T^[1]

5

6 [1]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4

7 7TJ, UK.

8 [2]Department of Plant Protection, Institute of Industrial and Fodder Crops, Hellenic Agricultural

9 Organization–DEMETER, Volos, Greece

10 [3]Department of Crop Genetics, John Innes Centre, Norwich Research Park, NR4 7UH, Norwich,

11 UK

12 [4]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, NR4

13 7UH, UK.

14

15 *Joint first authors

16

17 ORCID: PTL 0000-0001-9500-6592

18 ORCID: TC 0000-0002-2401-8120

19 ORCID: NAH 0000-0002-4979-9680

20 ORCID: SFW 0000-0003-4736-0938

21 ORCID: SAH 0000-0003-1371-5606

22 ORCID: MH 0000-0001-6628-5940

23 ORCID: JM 0000-0002-5893-8400

24 ^Correspondence: tracey.chapman@uea.ac.uk

25

26

27 Running head: medfly gut microbiome

28

29 Keywords: *Klebsiella oxytoca*, symbiont, paratransgenesis, 16S rDNA sequencing.

30

31 **Abstract**

32 Symbioses between bacteria and their insect hosts can range from very loose associations through

33 to obligate interdependence. While fundamental evolutionary insights have been gained from the in-

34 depth study of obligate mutualisms, there is increasing interest in the evolutionary potential of flexible

35 symbiotic associations between hosts and their gut microbiomes. Understanding relationships

36 between microbes and hosts also offers the potential for exploitation for insect control. Here, we

37 investigate the gut microbiome of a global agricultural pest, the Mediterranean fruitfly (*Ceratitis*

38 *capitata*). We used 16S rRNA profiling to compare the gut microbiomes of laboratory and wild strains

39 raised on different diets and from flies collected from various natural plant hosts. The results showed

40 that medfly guts harbour a fairly simple microbiome, primarily determined by the larval diet in both

41 wild and laboratory flies. However, regardless of the laboratory diet or natural plant host on which

42 flies were raised, *Klebsiella* spp dominated the medfly microbiomes and resisted removal by

43 antibiotic treatment. We sequenced the genome of the dominant putative *Klebsiella* spp (designated

44 'Medkleb') isolated from the gut of the Toliman wild type fruitfly strain. Genome-wide ANI analysis

45 placed Medkleb within the *K. oxytoca / michiganensis* group. Molecular, sequence and phenotypic

46 analyses supported its identity as *K. oxytoca*. Medkleb has a genome size (5 825 435 bp) which is

47 1.6 standard deviations smaller than the mean genome size of free-living *Klebsiella* spp, and lacks

48 some genes involved in environmental sensing. Moreover, the Medkleb genome contains at least

49 two recently acquired unique genomic islands as well as genes that encode pectinolytic enzymes

50 capable of degrading plant cell walls. This may be advantageous given that the medfly diet includes

51 unripe fruits containing high proportions of pectin. These results suggest that the medfly harbours a

52 commensal gut bacterium that may have developed a mutualistic association with its host and

53 provide nutritional benefits.

54   **Introduction**

55   All eukaryotic organisms host bacteria (McFall-Ngai et al 2013) and some of the best-studied

56   associations are those that occur between insects and bacteria (Moran et al 2008; McCutcheon and

57   Moran 2011; Moran and Bennett 2014). The majority of microbe-insect interactions are temporary

58   associations. However, some bacteria and insects retain persistent associations over evolutionary

59   time and have evolved co-dependence (Moran et al 2008). In the most extreme examples, these

60   associations have persisted for millions of years and represent intimate co-evolutionary relationships

61   (Moran et al 2005; Moran et al 2008).

62

63   Heritable symbioses are one such example and are defined by the direct passage of bacteria from

64   insects to progeny, usually via maternal transmission (Moran et al 2008). These symbioses can be

65   facultative or obligate (Moran et al 2008). Over evolutionary time, facultative symbionts may lose

66   genes that facilitate life in varied environments (Moran and Bennett 2014). A transition to

67   evolutionary interdependence with hosts can often be identified by reduced genome size and

68   reduced GC content (Shigenobu et al 2000; van Ham et al 2003; Wu et al 2006; Hansen et al

69   2011; Bennett and Moran 2013; Husnik et al 2013). Novel loci can also be acquired via lateral

70   gene transfer or loci deleted through genome reduction (Ochman et al 2000; Ochman and Davalos

71   2006). Mutualistic benefits provided to insects by bacteria include synthesis of nutrients (Sabree et

72   al 2013; Storelli et al 2018; Sinotte et al 2018), carotenoids (Sloan and Moran 2012), and

73   antipredation molecules (Oliver et al 2003; Sinotte et al 2018). Heritable symbionts include

74   "reproductive parasites" that enhance their fitness by distorting the sex ratios of the hosts' offspring

75   (Hurst and Frost 2015). For example, male *Zyginidia pullulan* leafhopper embryos are feminised by

76   maternally-inherited *Wolbachia* bacteria (Negri et al 2009). Feminisation benefits *Wolbachia,*

77   because these bacteria preferentially reside within ovaries and therefore are transmitted across

78   generations at higher frequency (Negri et al 2009). The *Buchnera* symbionts of aphids have been

79   extensively studied and represent an example of an advanced obligate mutualism (Shigenobu et al

80   2000; Tamas et al 2002; van Ham et al 2003). *Buchnera spp* have been associated with aphids for

81   approximately 100M years (Von Dohlen and Moran 2000) and provide the host with essential

82   amino acids (Shigenobu et al 2000). There is an increasing body of research into the identities and

83   potential roles of diverse symbionts found across many different taxa (de Souza et al 2009; Weiss

84   and Kaltenpoth 2016; Holmes et al 2016; Whitten and Dyson 2017; Ballinger and Perlman 2017;

85   Anbutsu et al 2017; Heine et al 2018) including those that are only loosely associated with their

86   hosts, such as the gut symbionts of fruitflies that are the focus of this study.

87

88   There are many examples of studies in which the culturable and non-culturable members of gut

89   microbiome communities have been characterised via sequencing of specific regions of the

90   bacterial 16S rDNA. These works show that dipteran gut microbiomes are often relatively stable

91   and simple (Chandler et al 2011; Gould et al 2018; Deguenon et al 2019) and can in some species

92   be strongly influenced by host diet (Chandler et al 2011; Woruba et al 2019). For example, the gut

93   microbiome of *Drosophila melanogaster* is consistently formed by a core of five species:

94   *Lactobacillus plantarum, Lactobacillus brevis, Acetobacter pasteurianus, Acetobacter tropicalis and*

95   *Acetobacter orientalis* (Gould et al 2018). These simple microbiomes appear to contain bacteria

96   that reliably colonise the gut, suggesting the potential for hosts to actively regulate their gut

97   microbiomes (Lhocine et al 2008; Bosco-Drayon et al 2012; Lindberg et al 2018) and gain potential

98   benefits from doing so (Gould et al 2018). The five core members of the *D. melanogaster* gut

99   microbiome metabolise lactic acid and acetic acid, which may benefit larvae feeding on rotten fruit.

100   In contrast, the *Tephritidae* family of "true fruit-fly" pests hatch in unripe fruit and the culturable

101   species within their gut microbiomes are reported to contain pectinolytic bacteria, which could

102   assist the host in breaking down plant cell walls (Behar et al 2008b; Ben-Yosef et al 2014; Liu et al

103   2016).

104

105   Manipulation of the co-evolved intimacy of symbionts and hosts via paratransgenesis has the

106   potential to be used as a novel mode of pest insect control (Durvasula et al 1997; Whitten et al

107   2016). Gut symbionts have the potential to augment strategies for control as probiotics. For

108   example, they could boost fitness in insects such as Mediterranean fruitflies (medfly, *Ceratitis*

109   *capitata*) that are subjected to potentially damaging irradiation as part of sterile insect release

110   control programmes (Jurkevitch 2011). The potential utility of symbionts to either provide new

111   routes for pest control or to improve existing technologies (Leftwich et al 2016) has led to

112   increasing interest in investigating the symbiotic gut microbial communities of key global pests,

113   such as the medfly (Behar et al 2008a; Behar et al 2008b; Ben Ami et al 2010; Gavriel et al 2011),

114   which is the focus of investigation here.

115

116   Investigations of culturable gut bacteria in medfly have been predominantly performed using

117   amplified rDNA restriction analysis (ARDRA). These show that *Klebsiella* spp comprise at least 20-

118   30% of the larval, pupal and adult medfly gut microbiomes (Behar et al 2008b; Ben Ami et al 2010;

119   Aharon et al 2013). This has led to the hypothesis that *K. oxytoca* might benefit larval nutrition via

120   its reported pectinolytic activity against fruit sugars, or due to its ability to fix nitrogen (Behar et al

121   2008b). 16S rDNA analysis has also been used to show that irradiation can alter the microbiome,

122   and particularly diminishes the relative contribution of *Klebsiella* spp (Ben Ami et al 2010).

123   Subsequent reintroduction of *K. oxytoca* to irradiated flies significantly reduced mating latency in

124   comparison to males fed sterile diet (Ben Ami et al 2010), suggesting a potential host fitness benefit.

125

126   Although the culturable species within medfly microbiomes have been described, with *Klebsiella* spp

127   appearing to be a typical component, many details remain unknown. For example, we do not yet

128   know the contribution of non-culturable species to both laboratory and wild medfly, whether

129   microbiomes are stable, whether *Klebsiella* spp have the capability to confer a direct fitness benefit

130 to the medfly, and the extent to which *Klebsiella* is heritable. Behar et al. (Behar et al 2008a) suggest
131 that a gut symbiont identified as *K. oxytoca* is heritable and can be transmitted during oviposition.
132 However, *K. oxytoca* was detected in only 1 of 4 replicate samples (Behar et al 2008a) and the
133 potential transmission of *K. oxytoca* during oviposition could not be ruled out . Recovery of GFP
134 labelled *Klebsiella* bacteria in the guts of offspring of mothers into which those bacteria were
135 experimentally introduced provides stronger evidence of vertical transmission (Lauzon et al 2009),
136 though the relative importance of this mechanism is not yet clear. In terms of fitness benefits, Gavriel
137 et al. (Gavriel et al 2011) experimentally depleted the microbiome of male medflies by using
138 irradiation andfed males with a diet either containing *K. oxytoca* (*pro*) or a sterile diet (*ster*). *pro*
139 outcompeted *ster* flies for matings, and females mated to *pro* males were less inclined to re-mate.
140 These data suggest that *Klebsiella* spp could confer host benefits (Behar et al 2008b; Ben Ami et al
141 2010; Gavriel et al 2011), though it is not yet clear whether this occurs in the natural context.
142
143 Here we compared the culturable and non-culturable gut microbiomes of wild-collected adult
144 medflies from a range of different wild hosts, with those of the Toliman wild type laboratory strain
145 reared on a range of different larval diets in the presence and absence of antibiotics. We conducted
146 long-read genome sequencing and analysis of the genome of the dominant, putative *K. oxytoca*
147 (hereafter "Medkleb") spp extracted from the adult gut of Toliman wild type individuals. This was
148 done to confirm the phylogenetic placement of Medkleb and to interrogate the genome for features
149 characteristic of a nascent evolutionary interdependence with its medfly host. We assembled and
150 annotated the Medkleb genome sequence and tested for signatures of facultative transition, i.e. a
151 reduction in genome size or GC content (Moran et al 2008; McCutcheon and Moran 2011; Moran
152 and Bennett 2014). We conducted comparisons between Medkleb and other *Klebsiella* spp to reveal
153 phenotypes that might potentially facilitate a mutualistic relationship, or indicate restrictions to the
154 environments in which Medkleb might live. We investigated tests of phenotypic features by testing
155 whether Medkleb had the capacity to synthesise secondary metabolites, and by conducting direct
156 biochemical tests for pectinolytic activity.
157
158 **Materials and Methods**
159
160 **1. Toliman Wild type strain**
161 To directly test the effect of different dietary carbohydrates on the gut microbiome, we raised three
162 replicate samples each of individuals from the Toliman wild type strain on different larval diets. This
163 strain originated from Guatemala and has been reared in the laboratory since 1990. Our Toliman
164 colony has been maintained in non-overlapping generations in a controlled environment room
165 (humidity 50± 5%, temperature 25± 0.5°C) on a 12:12 light:dark cycle for over 30 generations
166 (Leftwich et al 2017). Under this regime, larvae are raised on a sugar-yeast-maize medium (1% agar,
167 7.4% sugar, 6.7% maize, 4.75% yeast, 2.5% Nipagin (10% in ethanol), 0.2% propionic acid) and

168 adults are given *ad libitum* access to sucrose-yeast food; 3:1 w/w sugar/ yeast hydrolysate, and
169 water.

2. Generation of laboratory and wild-derived gut microbiome samples

(i) Effect of the larval diet on the adult microbiome in the Toliman wild type

172 To generate the samples of wild type flies raised on different diets (± antibiotics) for subsequent gut
173 dissection and 16S rRNA amplicon sequencing of the gut microbiome, Toliman flies were cultured
174 from eggs collected over a 24h period placed in one of four larval diet treatments. Three of these
175 diets provided varying carbohydrate levels and sources, while maintaining the same yeast (~protein)
176 level: 1) Sucrose High Protein (SHP), 2) Glucose & 3) Starch. The fourth diet had a sucrose
177 carbohydrate base but only 60% of the yeast content: Sucrose Low Protein, SLP (Table S1). We
178 included Propionic acid as a food preservative (Leftwich et al 2018). All diet manipulations were done
179 in the presence and absence of antibiotics. For the antibiotic treatments, each larval diet contained
180 a final concentration of 100µg/ml kanamycin, 200 µg/ml ampicillin, 200 µg/ml streptomycin, 50 µg/ml
181 chloramphenicol, 100µg/ml apramycin, 100 µg/ml hygromycin and 200 µg/ml tetracycline.
182 Approximately 500 eggs were placed on 100 mL of the appropriate diet in a glass bottle. When third
183 instar larvae started to "jump" from the larval medium, the bottles were laid horizontally on sand and
184 pupae allowed to emerge for seven days. Pupae were then sieved from the sand and held in 9 mm
185 petri dishes until adult eclosion.

(ii) Effect of wild larval diets on the adult microbiome in wild flies under natural conditions

187 Wild flies were collected at adult eclosion from fallen argan fruit in Arzou, Ait Melloul, Morocco, in
188 July 2014, from Apricots, Oranges and Grapefruits in Chania, Crete, July-September 2014 and from
189 Peaches, Oranges and Tangerines in Ano Lechonia, Greece, July-September 2014. All samples
190 were preserved in 96% ethanol and sent to the UK before gut dissection and DNA extraction for the
191 16S rRNA amplicon sequencing described below.

3. 16S rRNA gene sequencing and bioinformatics analysis of the gut bacteria derived from laboratory- and wild- and derived adult medflies.

194 We analysed the composition of the gut microbiomes in the dissected guts of the laboratory and
195 wild-derived flies describe above, by using 16S rRNA amplicon sequencing. Each of the three
196 biological replicates was a pool of five adult flies (Supplementary Information). Batches of flies were
197 surface sterilized for 30 seconds in 0.5% sodium hypochlorite (bleach) (Sigma-Aldrich, Cat.
198 No.7681529) and washed for 30 seconds in sterile 1M PBS (pH 7.4) three times before being
199 homogenized. 100 µL of the third washes were used to check the surface sterilisation efficiency.
200 There was no microbial growth in any of these tests.

201 We used sterilised pestles to homogenise the surface-sterilised samples inside 2-mL microcentrifuge
202 tubes, with three freeze/thaw cycles in liquid nitrogen. DNA was extracted using the DNeasy Blood

203 and Tissue Kit (Qiagen) and quality checked using a NanoDrop (Thermoscientific Nanodrop 8000

204 Spectrophotometer). Approximately 100 ng of DNA per sample was used as the template for PCR

205 amplification with bacterial universal primers 515F (5′-GTG CCA GCM GCC GCG GTA A-3′) and

206 806R (5′-GGA CTA CHV GGG TWT CTA AT-3) against the 16S rRNA gene. Amplicon sequencing

207 was performed using paired-end 250 bp V2 chemistry (Illumina MiSeq platform, Earlham Institute

208 provider).

209 Demultiplexed sequences were obtained using mothur v38.2 (Schloss et al 2009), following their

210 standard MiSeq operating procedures. Sequence variants were assigned to operational taxonomic

211 units (OTUs) at a 97% similarity threshold. Taxonomy assignment of OTUs using the Silva database

212 (release 132). The minimum library sizes per sample were ~17K after passing quality control. All

213 statistical analyses of amplicon data were conducted in R v3.6.2 (R Core Team 2019) using the

214 phyloseq (McMurdie and Holmes 2013), vegan (Oksanen et al 2007) and tidyverse (Wickham 2017)

215 packages. Sequences were rarefied to normalise library sizes. Alpha diversity was estimated using

216 the Shannon species diversity index. We visualised differences in bacterial community structure

217 among samples (beta diversity) using non-metric multidimensional scaling (NMDS) plots of Bray-

218 Curtis distances and performed multivariate analysis of variance (PERMANOVA) with 999

219 permutations on Unifrac distance matrices.

220

221 **4. Genome sequencing of the dominant gut microbiome *Klebsiella* spp symbiont (Medkleb)**

222 **from the Toliman wild type**

223 We investigated the identity of the recurrent *Klebsiella* spp bacterial symbiont through isolation of

224 culturable *Klebsiella* spp. colonies from the Toliman strain.

225 *(i) Clonal isolation.* Clonal isolates of *Klebsiella* spp obtained from individuals of the Toliman wild

226 type strain were made by taking surface sterilized, homogenised samples from adults reared under

227 the standard conditions described above and plating them onto Simmon's Citrate LB Agar (with

228 bromothymol blue as a colour indicator). This is a substrate recommended for the isolation of

229 *Klebsiella oxytoca and K. pneumoniae* (Simmons 1926). Culture plates were made from 15 biological

230 replicates of medfly. Cultures were checked for morphological uniformity and their identity confirmed

231 with PCR amplification with universal bacterial primers 28F and 806R. Thirteen of the 15 isolates

232 had identical 16S rRNA gene sequences and were BLAST matched to *K. oxytoca* and our most

233 abundant OTU from 16S rRNA amplicon sequencing. Two isolates contained colonies which BLAST

234 matched to *Pantoea* spp. We chose a single *Klebsiella* spp colony at random for genomic

235 sequencing, as described below.

236 *(ii) DNA preparation.* The single clonal isolate selected for genome sequencing was streaked onto

237 LB media (15g/L Agar; 5g/L NaCl; 5g/L yeast extract; 1.5g/L glucose; 10g/L tryptone) and incubated

238 overnight at 25°C, then transferred into a 1.5ml microcentrifuge tube containing 1ml of 10% glycerol.

239 The sample was vortexed for 30 secs then centrifuged at 12000 RPM for 10min. Glycerol was then
240 removed, and the pelleted bacteria re-suspended in 2ml of SET buffer (65% v/v molecular grade
241 H_2O (ThermoFisher); 20% v/v Tris (pH8); 5% v/v 5M NaCl; 5% v/v 10% SDS; 5% v/v 0.5M EDTA)
242 before transfer to a 15ml falcon tube. 20µg of lysozyme (Sigma) and 0.4µg achromopeptidase
243 (Sigma) suspended in 40µl of molecular grade water (ThermoFisher) and 0.02µg of RNase
244 (Fermentas) were added. The sample was mixed gently then incubated at 37°C for 2hrs. 240µL of
245 10% sodium dodecyl sulphate and 56µL of proteinase K (20mg/ml) were added, before a second
246 incubation at 56°C for 2hrs, with manual mixing every 30 mins. 800µL of 5M NaCl and 2ml of
247 chloroform were added and the sample was mixed by hand for 10 mins, before centrifugation at
248 4000 RPM for 12 mins. The aqueous phase was then carefully transferred to a fresh tube. DNA was
249 precipitated in 0.6 volume isopropanol, then transferred to a 1.5ml microfuge tube by pipette. DNA
250 was washed once with 70% ethanol. Ethanol was removed, then 1ml of 70% ethanol was added,
251 and DNA was left to incubate overnight at 4°C. Ethanol was again removed, and DNA was re-
252 suspended in 200µL of molecular grade water (ThermoFisher).

253 *(iii) Single molecule real time (SMRT/PacBio) Medkleb genome sequencing.* DNA purity,
254 concentration, and average fragment size were analysed using Nanodrop (ThermoFisher), Qubit
255 v2.0 (Invitrogen) and Agilent Tapestation 4200 (Agilent) respectively. DNA was fragmented using a
256 G-tube (Covaris), and SMRTbell library construction was carried out using a Template Prep Kit 1.0
257 (PacBio). The library was then size selected to >7kb using the BluePippin system (Sage Science).
258 Sequencing was carried out on a Pacific Biosciences RSII instrument, using two RSII SMRTcells v3
259 and P6-C4 chemistry (PacBio, Earlham Institute provider). Each cell was sequenced using a 240-
260 minute movie, using the Magbead OCPW v1 protocol (PacBio).

261 *(iv) Medkleb genome assembly.* The Medkleb genome was assembled according to the Hierarchical
262 Genome-Assembly Process (HGAP.3) protocol (Chin et al. 2013) as follows. 1) Mapping – BLASR
263 (Chaisson et al., 2012) was used to map reads >500bp with a read quality >0.8 to seed reads
264 >6000bp. 2) Pre-assembly - the Directed Acyclic Graph Consensus (DAGCon) algorithm (Lee et al
265 2002) was used to produce a consensus sequence based on BLASR mapping. DAGCon then
266 trimmed the consensus, producing an error-corrected pre-assembled read. 3) *de novo* genome
267 assembly - the overlap-layout-consensus assembler Celera Assembler v8.1 was used to process
268 the pre-assembled read into a draft assembly. 4) Final consensus - the draft assembly was polished
269 using the Quiver multiread consensus algorithm (Chin et al 2013). 5) The final consensus sequence
270 was then manually trimmed to circularise the genome and place the stop codon (TGA) of the *dnaA*
271 gene at the 5' terminus.

272 *(v) Medkleb genome quality control.* An estimated Quiver quality value (QV) for the Medkleb final
273 consensus genome was provided (Earlham Institute). Genome completeness was estimated with
274 both benchmarking universal single copy orthologues (BUSCO) software 3.0.0 (Simão et al 2015),
275 and CheckM (Parks et al 2015). The *Enterobacteriales* order and *Enterobacteriaceae* family were

276 used as reference datasets for BUSCO and CheckM analyses respectively. Genome contamination

277 was estimated with CheckM, and mlplasmids (Arredondo-Alonso et al 2018) was used to classify

278 contigs as either chromosomal or plasmid DNA. The *Klebsiella pneumoniae* support-vector machine

279 (SVM) model was utilised for the analysis, with minimum posterior probability specified at 0.7 and

280 minimum contig length at 1,000nt.

281 *(vi) Annotation and genome mapping.* Coding sequences within the Medkleb chromosomal DNA and

282 plasmids, were called with the Prodigal algorithm (Hyatt et al 2010). Gene calls were then annotated

283 with Classic-RAST (Overbeek et al 2014). Ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs)

284 were called and annotated with Classic-RAST. Circular maps were created for the Medkleb

285 chromosomal DNA and plasmids using DNAplotter (Carver et al 2009).

286 **5. Phylogenetic analyses of the Medkleb *Klebsiella* gut symbiont genome sequence**

287 *(i) 16s rRNA gene sequence analyses.* The Medkleb genome was searched for regions homologous

288 to the 16S rRNA gene sequence of *K. oxytoca* strain ATCC 13182 (NR_118853.1) with BLASTn

289 (Altschul et al 1997). The region with greatest homology to NR_118853.1 (nts 341370-342821) was

290 then parsed with RNAmmer 1.2 (Lagesen et al 2007) which predicts ribosomal genes. Sixty-one 16S

291 rRNA gene nucleotide sequences representing 60 *Klebsiella* strains, *Pseudomonas aeruginosa*

292 strain JB2 and a putative Medkleb 16S sequence were used to create a phylogeny with the SILVA

293 ACT web app (Pruesse et al 2012). Where possible, non-redundant sequences were extracted from

294 the SILVA rRNA gene database (Quast et al 2013). All sequences were almost complete (>1400bp)

295 and met the standard operating procedure for phylogenetic inference (SOPPI) quality criteria set out

296 by Peplies et al. (Peplies et al 2008). The tree was computed with the FastTree2 maximum likelihood

297 programme (Price et al 2010) using the GTR evolutionary model and gamma distribution parameters

298 (Yang 1994). The resulting phylogeny was constructed with FigTree 1.4.3 (Rambaut and Drummond

299 2009).

300 *(ii) Average Nucleotide Identity (ANI) analyses.* 35 RefSeq whole genome entries extracted from the

301 NCBI database (Geer et al 2010), representing four *Klebsiella* species, were used to calculate a

302 hierarchical clustering based on Average Nucleotide Identity (Konstantinidis and Tiedje 2005). The

303 ANI Calculator (Figueras et al 2014) was used to compute the hierarchy using the BIONJ algorithm

304 (Gascuel 1997). The tree was constructed with FigTree 1.4.3 (Rambaut and Drummond 2009). To

305 predict recently acquired Medkleb sequences, the genome was aligned with three closely related

306 bacteria. Genomes were manually re-ordered to place the *dnaA* stop codon at the 5' terminus.

307 Synteny was then predicted with progressive Mauve (Darling et al 2010).

308

309 **6. Metabolic functions of the Medkleb *Klebsiella* gut symbiont**

310 *(i) Analysis of pectinolytic enzyme activity.* Medkleb's ability to degrade pectin was compared to two

311 bacterial species identified as positive and negative controls; *Erwinia carotovora* (+ve control) and

312   *Rhizobium leguminosarum* (-ve control) (Wegener 2002; Xie et al 2012). Medkleb bacteria were

313   cultured in LB broth (5g/L NaCl; 5g/L yeast extract; 1.5g/L glucose; 10g/L tryptone) which had been

314   stored in 50ml glass bottles and autoclaved prior to use. Each bottle was inoculated with a "loop" of

315   bacteria and incubated, shaking, in an orbital incubator (New Brunswick Scientific Innova 44) at

316   200RPM. Medkleb and *Erwinia carotovora* were incubated at 37°C and *Rhizobium leguminosarum*

317   was incubated at 28°C until optical density was greater than 1.0 at 600nm.

318

319   *(ii) Analysis of presence of pehX and 16S rRNA genes in medkleb.* Bacterial DNA was extracted

320   using a DNeasy blood and tissue kit (Qiagen) and microbe lysis buffer (MLB) (20 mg/ml of lysozyme

321   (Sigma) and 5mg/ml of achromopeptidase (Sigma) in 20 mM Tris-HCl, 2 mM EDTA, 1.2% Triton X

322   (pH 8.0)). 2ml of Medkleb, *Erwinia carotovora* and *Rhizobium leguminosarum* cultures were

323   centrifuged at 13K RPM for 5 mins, before the supernatants were removed. Pellets were then

324   homogenised with a clean pestle in liquid nitrogen. 180µl of MLB was added before the sample was

325   vortexed and incubated at 37°C for two hours. Samples were vortexed every 30 mins during the two-

326   hour incubation. Buffer AL and ethanol were mixed 1:1 (Buffer ALE) and warmed to 55°C. Each

327   sample had 400µl of warm Buffer ALE added before being immediately vortexed for 10-15 secs.

328   Samples were transferred to a spin column and centrifuged at 8000RPM for 60 secs. 500µl of Buffer

329   AW1 was added and the sample was centrifuged again at 8000RPM for 60 secs. 500µl of Buffer

330   AW2 was added and the samples were centrifuged at 13 000 RPM for 4 mins. 35µl of warm Buffer

331   AE (60°C) was added to the centre of the spin membrane and the samples were centrifuged at

332   6000RPM for 60 secs. DNA purity and concentration were measured using a Nanodrop

333   (ThermoFisher). The *K. oxytoca* polygalacturonase gene pehX (AY065648.1) was aligned to

334   genomes of *Klebsiella* bacteria using BLAST (Altschul et al 1997). The presence of

335   polygalacturonases in genomes of *Klebsiella* bacteria was assessed using the Carbohydrate-Active

336   enZymes database (CAZy) (Lombard et al 2010).

337

338   *(iii) Polygalacturonase enzyme assay.* Polygalacturonase production of Medkleb, *Erwinia carotovora*

339   and *Rhizobium leguminosarum* was measured using a DNS colorimetric method (Miller 1959) with

340   a protocol adapted from Sohail et al. (Sohail and Latif 2016) and Sigma Aldrich protocol EC 3.2.1.1.

341   Cultures were diluted with LB broth to an optical density of 1.0 at 600nm. Bacteria were then filtered

342   from culture media with 0.2µm PES syringe filters (ThermoFisher). Treatment reactions (which were

343   run in triplicate) were set up with 1ml of appropriate filtrate and 1ml of polysaccharide solution (PS)

344   (0.9% polygalacturonic acid (ThermoFisher) in 0.1M sodium acetate (ThermoFisher) (pH 4.5)), a

345   blank reaction was set up with 1ml of PS only. All reactions were incubated at 45°C for 30mins before

346   1ml of colour reagent solution (20% 5.3M potassium sodium tartrate, tetrahydrate in 2M sodium

347   hydroxide solution; 50% 96 mM 3,5-Dinitrosalicylic acid solution; 30% molecular water

348   (ThermoFisher)) was added. All reactions were incubated at 100°C for 15 mins, then placed on ice

349   to cool to room temperature. Once cooled, 12ml of molecular water (ThermoFisher) was added to

350 each reaction, followed by hand mixing. Absorbance at 530nm (ΔA530) was measured using a

351 spectrophotometer (Biochrom) which had been blanked for air, with the corrected ΔA530 for

352 treatment reactions being calculated as:

353   *ΔA530 (Treatment reaction) = ΔA530 (Treatment reaction) – ΔA530 (Blank reaction)*

354 Units of polygalacturonase in the filtrate were calculated via comparison to a standard curve of

355 galacturonic acid (Sigma). Standards were made with between 50μl and 1 ml of monosaccharide

356 solution (MS) (1.8% galacturonic acid (ThermoFisher) in 0.1M sodium acetate (ThermoFisher) (pH

357 4.5) and topped up to 2ml total volume with molecular water (ThermoFisher). A standard blank was

358 set up containing 2ml of molecular biology grade water only. 1ml of colour reagent solution was then

359 added before incubation at 100°C for 15 mins. Standards were placed on ice to cool to room

360 temperature before 12ml of molecular water was added and ΔA530 was measured using a

361 spectrophotometer (Biochrom) which had been blanked for air. The corrected ΔA530 for standards

362 was calculated as: *ΔA530 (Standard) = A530 (Standard) – A530 (Blank)*. The standard curve was

363 used to estimate mg of galactose released in treatment reactions with linear regression, and units of

364 polygalacturonase per ml of filtrate were then calculated as: *Units/ml enzyme = (mg of galactose*

365 *released)/ml of filtrate).*

366

367 *(iv) Prediction of higher order metabolic functions and secondary metabolites.* The higher order

368 metabolic functions of genes were predicted using the Kyoto Encyclopaedia of Genes and Genomes

369 (KEGG) (Kanehisa et al 2016). Secondary metabolites were predicted using antiSMASH 6.0 beta

370 (Blin et al 2019).

371

372

373   **Results**

374

375   **1. Characterisation of the medfly gut microbiome using amplicon sequencing**

376

377   *(i) Bacterial community diversity*. We found that OTU richness and diversity varied significantly

378   among host samples according to treatment and origin (Figure 1 A &B). Community structure (beta

379   diversity) was affected primarily by diet, in both wild and lab flies (PERMANOVA: $F_{8,48}$ = 2.74, P

380   <0.001, R^2 = 0.27, Figure 1A). Within lab flies, age ($F_{1,30}$ = 4.9, P<0.001, R^2 = 0.18), diet ($F_{3,30}$ = 2.03,

381   P =0.023, R^2 = 0.12), and antibiotic treatment ($F_{1,30}$ = 4.9, P =0.002, R^2 = 0.09), all affected

382   community structure, with decreasing levels of effect. Bacterial species diversity (alpha diversity) did

383   not vary significantly in the wild or laboratory population strains but was lower in antibiotic treated

384   flies. However, the Shannon index was not significantly different between any samples (Figure 1B).

385   Overall, we found no evidence of large-scale changes in diversity or composition of the microbiome

386   despite testing multiple wild food sources, effects of laboratory diets and antibiotic treatment (Figure

387   1C).

388

389   *(ii) Dominant bacterial taxa.* Four bacterial families representing two bacterial phyla made up over

390   90% of the sequences in our dataset. These were the Proteobacteria, Enterobacteriaceae (79%),

391   Moraxellaceae (2.3%) and Xanthomonadaceae (3.6%), and the Firmicutes, Enterococcus (7.3%)

392   (Figure 1C). Of these, a single bacterial genus *Klebsiella* spp emerged as a core member of the

393   bacterial microbiome. A putative *Klebsiella* spp was found in every medfly population sample and

394   comprised 73.6% of the entire dataset. This suggests that, although medflies are extremely

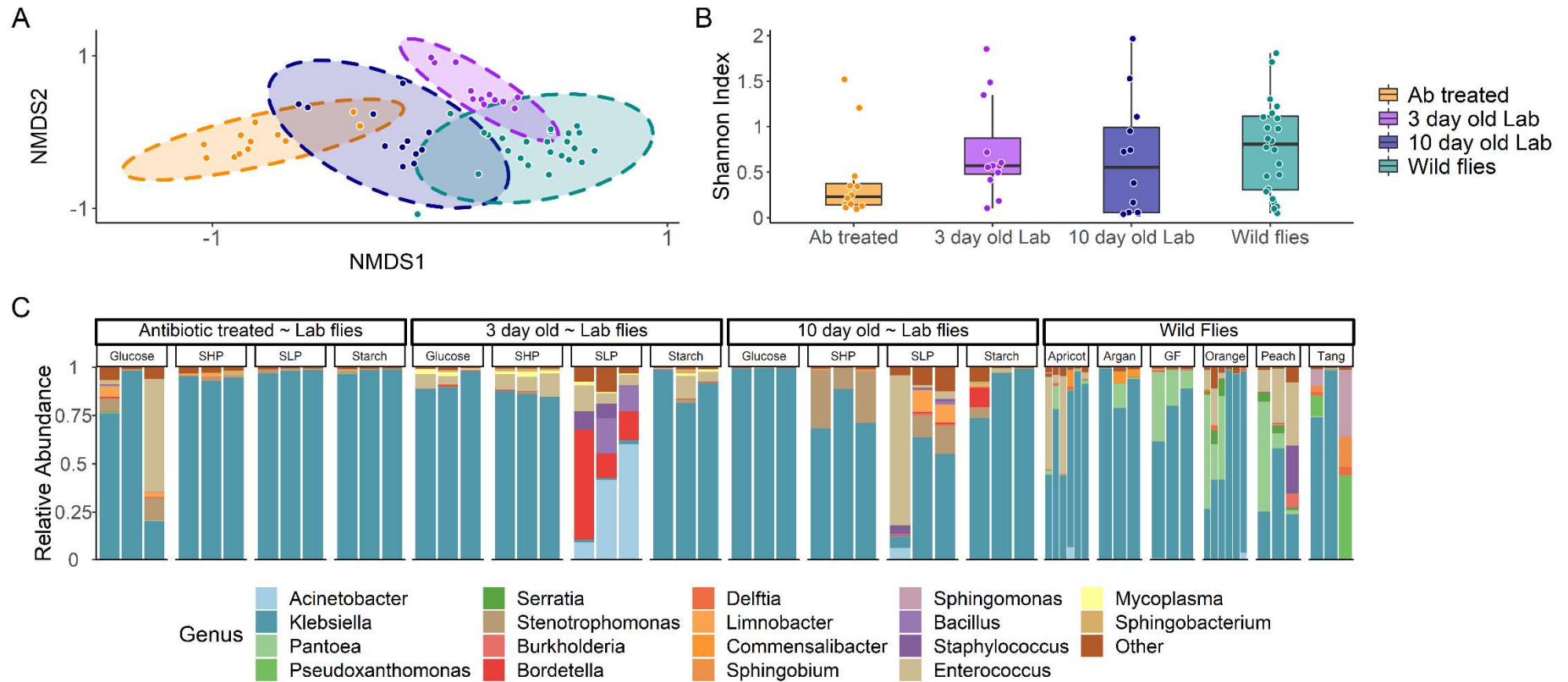395   polyphagous, they have a stable microbiome, containing a recurrent *Klebsiella* spp symbiont.

396

Figure 1. 16S rRNA profiles of wild and laboratory strains of medfly on varied larval diets and in the presence and absence of antibiotic treatment. Microbiome composition was measured as **(A)** community structure/beta diversity visualised as NMDS plots using a Bray-Curtis Dissimilarity Index with 95% confidence ellipses, **(B)** species richness/alpha diversity using the Shannon Index. Boxplot displays median, hinges are first and third quartiles, whiskers extend from hinge to 1.5* the interquartile range. **(C)** Bar plot of microbiome profiles. Ab = antibiotic treated (100μg/ml kanamycin, 200 μg/ml ampicillin, 200 μg/ml streptomycin, 50 μg/ml chloramphenicol, 100μg/ml apramycin, 100 μg/ml hygromycin and 200 μg/ml tetracycline), 3 day old Lab = 3 day old adults from the Toliman laboratory strain; 10 day old Lab = 10 day old adults from the Toliman laboratory strain, wild flies = flies from fallen argan fruit in Arzou, Ait Melloul, Morocco; Apricots, Oranges and Grapefruits in Chania, Crete; and from Peaches, Oranges and Tangerines in Ano Lechonia, Greece (n = 5 flies per replicate, 3 biological replicates).

2. Medkleb genome sequencing

(i) Classification of Medkleb sequencing contigs. Total Medkleb DNA was sequenced on a PacBio RSII module and assembled using the HGAP.3 algorithm and Quiver (Chin et al., 2013). This process detected one large contig that sequestered 94% of total gene space (Figure 2) and four small contigs (mkp1-4; Figure S1). This suggests that the total Medkleb DNA complement is formed of one chromosome and four plasmids. Consistent with this, the mlplasmids software (Arredondo-Alonso et al., 2018) classified the large Medkleb contig as chromosomal and the four smaller contigs as plasmids (Table S2). In addition, mkp2, mkp3 and mkp4 were sequenced with relatively high coverage depth (Figure S2) and showed evidence for high copy numbers, a common plasmid trait (Providenti et al., 2006). Finally, mkp3 and mkp5 were demonstrated to have low GC content relative to the putative chromosome (Table S2), which again is characteristic of plasmid DNA (Nishida 2012). mkp2 and mkp4 both exhibited a similar GC content to that of the chromosome, suggesting that they have been acquired more recently (Rocha and Danchin 2002).

(ii) Quality control. PacBio coverage depth >100X is considered sufficient for resolving nucleotide sequences (Rhoads et al., 2015) and this threshold was met by all contigs other than mkp4 (Figure S2). The QV sequence resolution was 48.9 (an average error rate of 1 base in every 80100) and hence the Medkleb chromosome was of high quality (Figure S2). The plasmid QV's were: mkp1) 48, mkp2) 47.3, mkp3) 45.3 and mkp4) 44.8 (Figure S2). Although the plasmid sequences had lower resolution, they were robust, with accuracy >99.994% in all cases. The completeness of the Medkleb genome was measured with BUSCO (Simão et al 2015), which was used to search the assembly for 781 marker genes associated with bacteria of the *Enterobacteriales* order. BUSCO estimated that the Medkleb chromosomal sequence was 99.5% complete (777 of 781 genes complete and single copy), far exceeding general quality thresholds (Bowers et al 2017). Furthermore, three genes that were predicted by BUSCO to be fragmented were likely heterozygous alleles that failed to collapse during the annotation. CheckM (Parks et al 2015) was used as a second method to assess the Medkleb chromosome for completeness and contamination. By using as reference 1162 marker genes associated with the family *Enterobacteriaceae*, the CheckM software estimated that the Medkleb genome was 99.7% complete and 0.212% contaminated, again far exceeding standard quality thresholds (Bowers et al 2017).

(iii) Medkleb genome. The general characteristics of the Medkleb genome were consistent with those published as *Klebsiella oxytoca* (Shin et al 2012; Bao et al 2013). The Medkleb genome was 5 825 435 nt in length, with 5 388 coding sequences and a GC content of 56.03% (Figure 2; Table S3). At 87.8%, overall coding sequence was within the expected range (Kuo et al 2009), and, as predicted by Reva et al. (2004), genes were distributed symmetrically between the two DNA strands. There were 2 541 coding sequences on the positive strand, which were predicted to code for 2 473 proteins, 50 tRNAs and 18 rRNAs. On the negative strand there were 2 847 coding sequences which coded for 2 805 proteins, 35 tRNAs and 7 rRNAs. KEGG (Kanehisa et al 2016) predicts that the Medkleb

442   genome encodes for genes with 2 842 distinct molecular functions. The genes for 16 (64%)

443   ribosomal RNAs (rRNA) clustered between nts 62 435 – 70 701 near the origin of replication (*oriC*).

444   Medkleb's GC content was 57.28% for protein coding genes, 53.83% for rRNA and 58.93% for tRNA.

445   Consistent with Lobry (1996) GC skew was asymmetric, with an overrepresentation of Gs on the

446   leading strand and Cs on the lagging strand, indicating that the genome is largely stable with few

447   recent recombination events. However, there is one obvious exception, in which GC skew was

448   inverted (>50% C's) between bases 654 265-677 695. This is indicative of a recent introgression

449   that has resulted in the acquisition of a new gene island (designated GI1) (Lawrence and Ochman

450   1997; Wixon 2001).

451   *(iv) Medkleb plasmids.* The lengths and GC contents of mkps 1-4 (Table S2) were all within

452   expected range for plasmids associated with *Klebsiella oxytoca*. mkps 1-4 are predicted by KEGG

453   (Kanehisa et al 2016) to code for 48 functional orthologues and devote 14-22% of gene space to

454   plasmid associated genes and mobile element coding sequences. This is substantial in comparison

455   to the main chromosome, which only allocated 0.02% to such features. mkps 1, 3 and 4 exhibited

456   asymmetric gene distribution between DNA strands (coding bias), which is common for plasmid

457   genomes (Reva and Tümmler 2004). The coding bias of mkp3 was particularly clear, with ~90% of

458   the total gene complement found on the positive strand. mkp4 was the only plasmid predicted by

459   antiSMASH 6.0 beta (Blin et al 2019) to code for secondary metabolites, which included cloacin (de

460   Graaf et al 1969) and colicin bacteriocins (Cascales et al 2007).
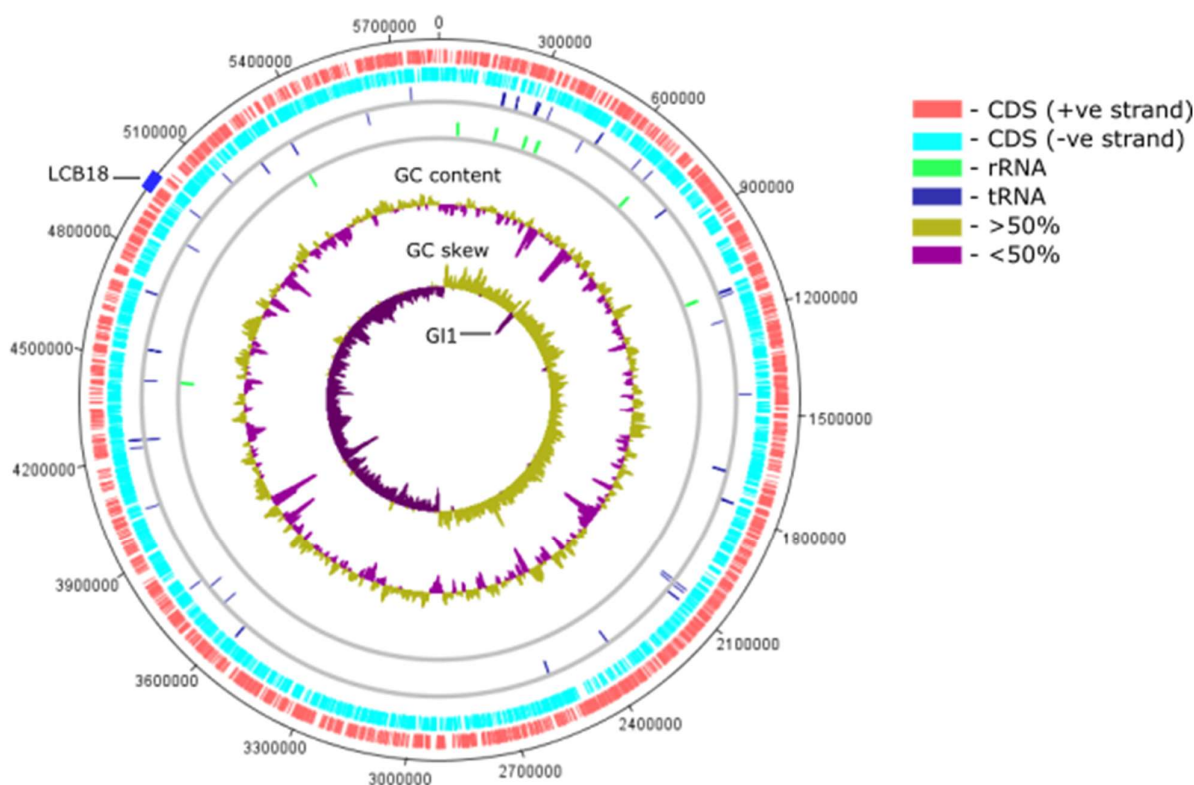
461

462

Figure 2. Circular summary map of the Medkleb chromosome. The Medkleb chromosome is represented with the stop codon (TGA) of the *dnaA* gene at position 0. Track one; 2 541 red ticks represent gene coding sequences on the positive strand. Track two; 2847 light blue ticks represent gene coding sequences on the negative strand. Track three; 50 dark blue ticks above the grey line represent tRNAs on the positive strand, and 35 ticks below the line represent tRNAs on the negative strand. Track four; 18 green ticks above the grey line represent rRNAs on the positive strand, and 7 ticks below the line represent rRNAs on the negative strand. 64% of rRNAs are found close to *oriC*. Track five – GC content; regions containing >50% GC content are mustard and regions containing <50% GC content are purple. Track six – GC skew; regions containing >50% Gs are mustard and regions containing <50% Gs are purple. GC skew was asymmetric between leading and lagging strands, with the purple spike between bases 654 265-677 695 (GI1) representing recent horizontal gene transfer (Lawrence and Ochman 1997). The LCB18 gene island (≈nts 4.97 x 10^6 ~ 5 x 10^6) is represented by a blue block on the outer ring.

475

476 **3. Taxonomic placement of the Medkleb *Klebsiella* gut symbiont**

477 *(i) Taxonomic identification of Medkleb –16S rDNA and ANI analysis.* Medkleb was predicted to be

478 a strain of *K. oxytoca,* which has previously been proposed as a major component of the medfly

479 microbiome (Behar et al 2008a)*.* The 16S rRNA gene sequence of *K. oxytoca* strain ATCC 13182

480 (NR_118853.1) was used to locate homologous sequences in the Medkleb genome using the

481 BLASTn algorithm (Altschul et al 1997). This revealed that the Medkleb genome contains eight

482 sequences >99% related to NR_118853.1, which were all predicted to code for 16S rRNAs by

483 RNAmmer 1.2 (Lagesen et al 2007). Nucleotides 341370-342821 (mk16S) which had the greatest

484 homology with NR_118853.1 were therefore used to represent Medkleb in subsequent taxonomic

485 analyses. These sequences were used in an initial taxonomic description (as described in the

486 Supplementary Information) which indicated that Medkleb was indeed likely to fall within a group of

487 *Klebsiella oxytoca / michiganensis* spp (Figure S3).

488 A more stringent Average Nucleotide Identity (ANI) (Konstantinidis et al., 2005) analysis was also

489 conducted. The Medkleb genome was positioned in an ANI matrix with 35 RefSeq *Klebsiella*

490 genomes that were extracted from the NCBI database (Geer et al 2010) (Figure 3). This analysis

491 placed Medkleb in a lineage with 13 strains classified as both *K. oxytoca* and *K. michiganensis*. ANI

492 scores >95% are required to classify bacteria as the same species (Richter and Rosselló-Móra 2009;

493 Kim et al 2014). This 95% similarity threshold was met by all thirteen members of Medkleb's ANI

494 clade (henceforth referred to as the Medkleb group) (Figure S4). Hence both 16S rRNA gene and

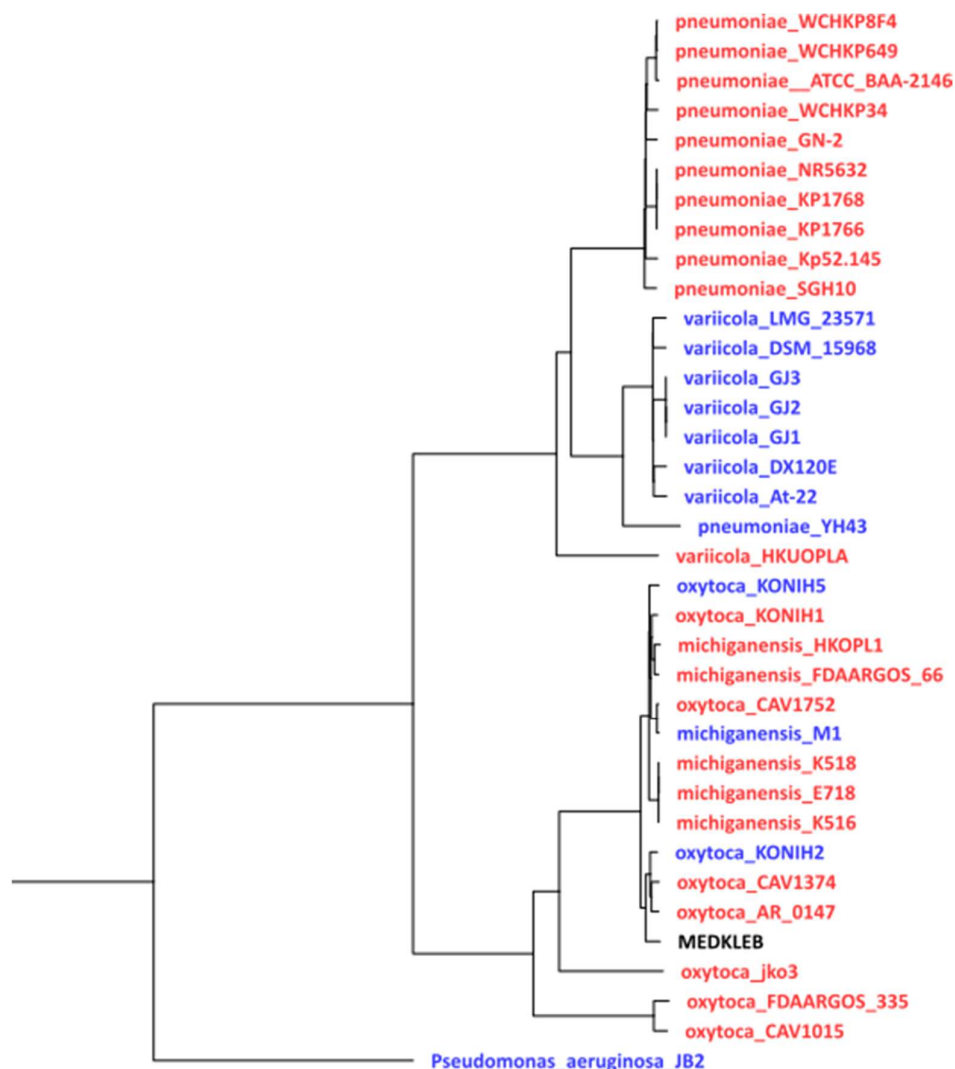495 ANI analyses did not differentiate between *K. oxytoca* and *K. michiganensis* as they are currently

496 named.

pneumoniae_WCHKP8F4
pneumoniae_WCHKP649
pneumoniae__ATCC_BAA-2146
pneumoniae_WCHKP34
pneumoniae_GN-2
pneumoniae_NR5632
pneumoniae_KP1768
pneumoniae_KP1766
pneumoniae_Kp52.145
pneumoniae_SGH10
variicola_LMG_23571
variicola_DSM_15968
variicola_GJ3
variicola_GJ2
variicola_GJ1
variicola_DX120E
variicola_At-22
pneumoniae_YH43
variicola_HKUOPLA
oxytoca_KONIH5
oxytoca_KONIH1
michiganensis_HKOPL1
michiganensis_FDAARGOS_66
oxytoca_CAV1752
michiganensis_M1
michiganensis_K518
michiganensis_E718
michiganensis_K516
oxytoca_KONIH2
oxytoca_CAV1374
oxytoca_AR_0147
MEDKLEB
oxytoca_jko3
oxytoca_FDAARGOS_335
oxytoca_CAV1015
Pseudomonas_aeruginosa_JB2

Figure 3. ANI hierarchical clustering showing the evolutionary relationship of environmentally-derived and host-derived *Klebsiella* bacteria. The tree was created using the ANI calculator (Figueras et al 2014), with *Pseudomonas aeruginosa* strain JB2 selected as the outgroup. Bacteria derived from animal hosts (red and black), and environmentally derived bacteria (blue), generally fell into three clades: 1) *K. pneumoniae*, 2) *K. variicola* and 3) *K. oxytoca/michiganensis*. With one exception in each group (YH43 and HKUOPLA), all *K. pneumoniae* are host derived and all *K. variicola* are environmentally derived. *K. oxytoca/K. michiganensis* have been isolated from both environmental and animal sources, but their sequences did not cluster according to source status. According to the ANI species threshold set by Kim et al. (2014), Medkleb is conspecific with twelve strains which have been classified as both *K. oxytoca* and *K. michiganensis*.

508 **4. Metabolic functions of the Medkleb *Klebsiella* gut symbiont**

509 *(i) Bioinformatic analyses of pectic lyases in Medkleb.* The ability to degrade pectin is a defining

510 phenotype of *K. oxytoca* and is proposed to be conferred by the polygalacturonase enzyme, pehX

511 (AY065648.1) (Kovtunovych et al 2003). We investigated whether *K. oxytoca* and *K. michiganensis*

512 *strains* of the Medkleb group could be distinguished via homology to pehX but our analysis detected

513 no significant differences between strains (analysed with BLAST; signed-rank linear model $F_{1,11}$ =

514 2.57, p = 0.137). Unexpectedly, *K. michiganensis* shared greater homology with AY065648.1 than

515 did *K. oxytoca* (Table 1), despite the fact that *K. michiganensis* does not degrade pectin (Saha et al

516 2013). In addition, the CAZy database (Lombard et al 2010) identified two pectate lyases (one from

517 family 2 and one from family 9) in all Medkleb group genomes available on the database (Table S5).

518 **Table 1 Genomic features of the 'Medkleb group' of bacteria.** *pehX* sequence identity (calculated with
519 BLAST (Altschul et al., 1997)) was not significantly different between species (analysed with BLAST; signed-
520 rank linear model $F_{1,11}$ = 2.57, p = 0.137). *K. michiganensis* genomes contained sequences more closely
521 related to AY065648.1 than was found for *K. oxytoca* genomes (mean relatedness = 88.4% vs 87.9%). All
522 genomes available on the CAZy database (Lombard et al 2010) are predicted to code for 2 pectate lyases.
523 Medkleb had the smallest genome in the group[1], and the second highest GC content[2].

Species	Strain	Genome size (nt)	GC (%)	*pehX* identity (%)	CAZy lyases
oxytoca	Medkleb	5825435[1]	56.03	87.43	NA
michiganensis	M1	5865090	56.13 [2]	88.47	2
michiganensis	HKOPL1	5914407	55.92	88.56	2
oxytoca	CAV1752	5992008	55.16	88.47	2
michiganensis	FDAARGOS_66	6071464	55.94	88.39	NA
michiganensis	E718	6097032	56.02	88.31	2
michiganensis	K518	6138996	55.95	88.39	2
michiganensis	K516	6139574	55.96	88.31	2
oxytoca	KONIH1	6152190	55.91	88.61	2
oxytoca	KONIH5	6179177	55.81	87.82	2
oxytoca	KONIH2	6190364	55.77	87.43	2
oxytoca	CAV1374	6257473	55.75	87.87	2
oxytoca	AR_0147	6350620	55.57	87.89	2

[1] smallest genome
[2] highest GC content

524

525

526 *(ii) Polygalacturonase enzyme assay.* Medkleb was pehX positive when analysed biochemically

527 (Figure S6), which suggested that it should be classified as *K. oxytoca* and can degrade pectin

528 (Kovtunovych et al 2003; Saha et al 2013). We quantified Medkleb's capacity to degrade

529 polygalacturonic acid- in comparison to positive (*E carotovora)* and negative (*R leguminosarum*)

530 control specimens, using a DNS colorimetric method (Miller 1959). A standard curve of galacturonic

531 acid incubated with DNS (Figure 4A), showed a positive relationship with colour change when the

532   OD was measured at 530nm (linear model, $F_{1,4}$ = 176.4, p <0.001, R^2 = 0.97). Filtrate of Medkleb
533   and *E. carotovora* culture media both contained pectinolytic enzymes, producing measurable colour
534   change when incubated with polygalacturonic acid and DNS (Figure 4B; (Sohail and Latif 2016)).
535   Medkleb filtrate reduced polygalacturonic acid with around 65% the efficiency of the *Erwinia*
536   *carotovora,* but *R leguminosarum* filtrate did not demonstrate any measurable pectinolytic activity
537   (Figure 4C). These data support the placement of Medkleb as *K. oxytoca*.

538   *(iii) Predicted metabolic functions of Medkleb.* KEGG (Kanehisa et al 2016) was used to produce a
539   list of gene functions for chromosomes and plasmids associated with all *K. oxytoca* bacteria in the
540   Medkleb group (Table S5). The dataset was filtered to produce three smaller lists: 1) Atypical - gene
541   functions unique to Medklebn, 2) Absent - gene functions present in all genomes analysed other
542   than Medkleb, 3) Duplicated – gene functions encoded in multiple copies by Medkleb but not by any
543   other genome analysed The Medkleb genome contained 21 atypical functional orthologues (Figure
544   S7), and of these, 16 were chromosomally derived and five located on plasmids. The atypical genes
545   included two putative transposases, enzymes involved in various modes of metabolism (e.g., *mtlA,*
546   *ACADSB, egsA, FAAH2*) and four transport protein genes (e.g. *natA* & *gatC*). When compared to
547   conspecifics, Medkleb has 35 absent gene functions, seven of which may indicate adaptation to the
548   relatively innoxious medfly gut (Figure S8). The largest cluster of absent genes was associated with
549   copper resistance (*copB, cusS, cusR, cusC, cusF)* and genes that degrade arsenic (*arsB*) and nitriles
550   (*nthA*) were also missing. The Medkleb genome codes for multiple copies of 11 functional
551   orthologues, that other *K. oxytoca* bacteria in the Medkleb group retain in single copy at most (Figure
552   S9). Medkleb's duplicated genes may have the potential to promote mutualistic phenotypes for the
553   medfly, as they syntheise amino acids (*proC, trpA*) and metabolise beneficial nutrients such as fatty
554   acids (*ACAT*) and citrate (*citE*).

555

556   *(iv) Secondary metabolites.* The Medkleb genome and plasmid sequences were interrogated using
557   antiSMASH 6.0 beta (Blin et al 2019) for the presence of novel secondary metabolites.
558   Chromosomes of all *K. oxytoca* members of the Medkleb group (other than KONIH2) coded for three
559   common secondary metabolite clusters: 1) non-ribosomal polypeptide synthetase (30% similarity to
560   turnerbactin), 2) thiopeptide antibiotic (14% similarity to O-antigen), 3) Ribosomally synthesized and
561   post-translationally modified peptides (RiPPs). The Medkleb chromosome contained two unique
562   secondary metabolite clusters that were not present in any other analysed *K. oxytoca* chromosomes
563   (i) a butyrolactone, a signalling molecule utilised by *Streptomyces* bacteria to to regulate antibiotic
564   production and cell cycle processes (Takano 2006; Kitani et al 2011), and (ii) an N-acyl amino acid
565   cluster, common to soil dwelling bacteria and involved in cell-to-cell communication (Craig et al 2011;
566   Battista et al 2019) (Table 2). Plasmid mkp4 was also predicted to code for cloacin and colicin
567   bacteriocins that were not coded by any other *K. oxytoca* strain in the Medkleb group. The cloacin
568   cluster encoded on mkp4 contains two mobile elements (MKleb_5887, MKleb_5890) and is generally
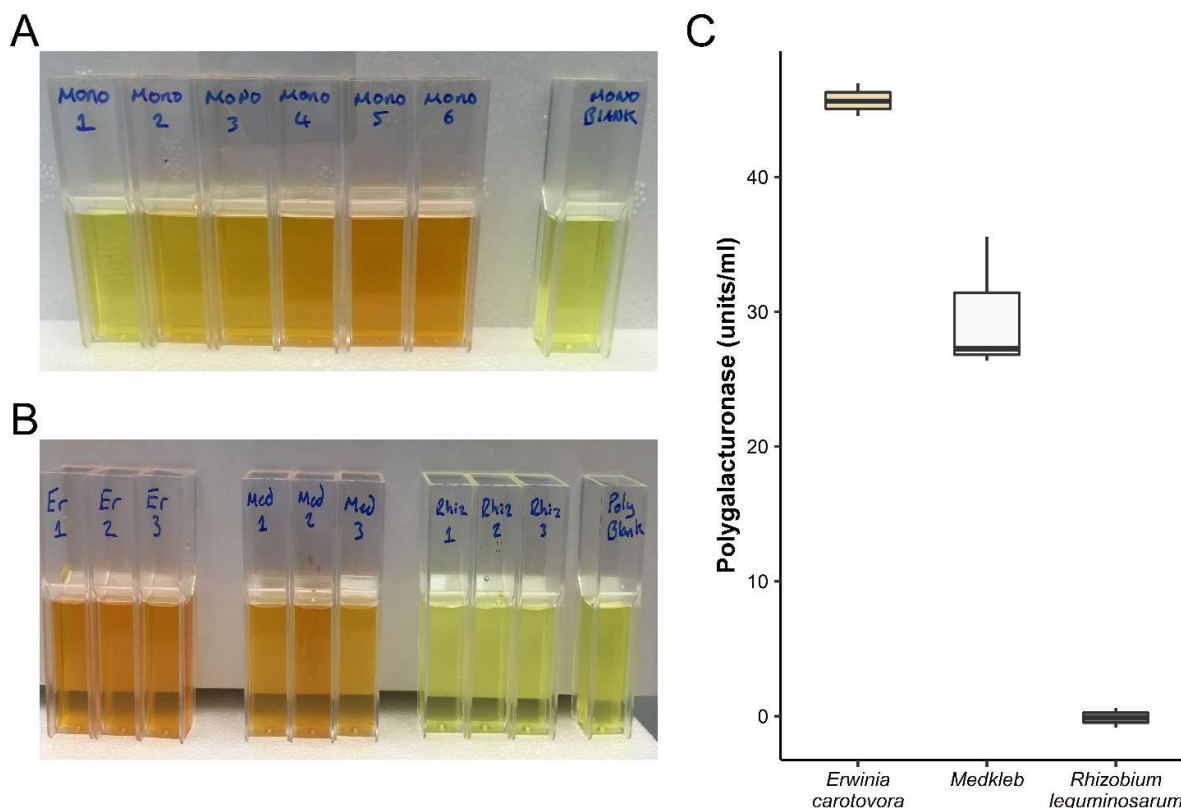569   considered to be toxic for *Klebsiella* bacteria (de Graaf et al 1969).

Figure 4. Polygalacturonase production by *Erwinia carotovora* (Er), Medkleb (Med) and *Rhizobium leguminosarum* (Rhiz). A) Monosaccharide colorimetric reaction mixtures, standard curve – The quantity of sugar in standards ranged from 0.9mg (mono 1) to 18mg (mono 6). Mono blank contained no sugar. When incubated with DNS, colour change (measured at ΔA530) occurred for all reactions relative to the blank. The relationship between sugar quantity and colour change was significant (linear model, $F_{1,4} = 176.4$, $p < 0.001$, $R^2 = 0.98$). B) Bacteria reaction mixtures – Polygalacturonase production of three bacteria, *Erwinia carotovora* (Er), Medkleb (Med) and *Rhizobium leguminosarum* (Rhiz) was measured via the quantity of reduced sugar in solution, following incubation of culture filtrate with polygalacturonic acid. Units of polygalacturonase were quantified in terms of colour change via extrapolation from the standard curve. Med and Er filtrate contained 30 and 46 units of polygalacturonase per ml respectively. Rhiz filtrate did not contain any polygalacturonase. C) Units of polygalacturonase contained in bacterial filtrate. Units of polygalacturonase per ml are represented on the y-axis. 1000ul aliquots of bacterial filtrate processed from three different species of bacteria are represented of the x-axis. Bacteria were assessed for the presence of polygalacturonases using a standard curve of galacturonic acid. The upper and lower hinges of boxplots represent the first and third quartiles of enzyme concentrations in filtrate, calculated from three technical replicates. *Erwinia carotovora* was the highest producer, with an average of 45.7 units/ml and Medkleb produced polygalacturonase with an average of 29.7 units/ml. *Rhizobium leguminosarum* did not produce any polygalacturonase.

589   **5. Searching for symbiotic signatures in the Medkleb genome**

590   *(i) Genome size and GC content.* At 5_825_435bp, the Medkleb genome was the smallest in the
591   'Medkleb group' and 1.6 standard deviations smaller than the mean genome size of the clade (n=13)
592   ($\bar{x}$ = 6.07 x 10^6 bp ± 1.49 x 10^5 s.d.). However, we conclude this is not diagnostic of symbiotic
593   transition as the genome of free-living *K. michiganensis* strain M1, was only 0.7% larger than
594   Medkleb (5.86 x 10^6 bp). Medkleb's GC content (56.03%) was also high, even for bacteria with a
595   free-living life history (Moran et al 2008) (Table 1). We conclude that there were no obvious
596   diagnostic signatures of a strong host-association.

597   *(ii) Local genomic rearrangement.* The Medkleb genome was aligned, using progressive Mauve
598   (Darling et al 2010) to three closely related strains of *K. oxytoca* (Figure S5). This analysis revealed
599   the presence of 21 local colinear blocks (LCBs) of conserved DNA in all four genomes (Figure S5).
600   These LCBs were not uniformly distributed between the genomes. *K. oxytoca* strains AR0147 and
601   CAV1374 both contained inversions between LCBs 11-16 and strain KONIH2 contained several
602   instances of translocation and inversion. However, despite inversion and translocation events,
603   nucleotide sequence in all LCBs other than LCB18 was very highly conserved. The annotation of
604   Medkleb LCB18 (≈nts 4.97 x 10^6 ~ 5 x 10^6) predicts 25 coding sequences in total, including 13
605   "hypothetical   proteins",   2   DNA   helicases   (Mkleb_4576/4577),   2   methyltransferases
606   (Mkleb_4579/4597) and an anti-restriction protein (Mkleb_4581). LCB18 may have been recently
607   acquired as it also contains genes associated with horizontal transfer including an integrase
608   (Mkleb_4574), a mobile element protein (Mkleb_4594) and a prophage protein (Mkleb_4585). In
609   addition, two hypothetical proteins encoded by LCB18 (MKleb_4591 & Mkleb_4592) may be
610   virulence factors, as they are predicted by TMHMM 2.0 {Krogh, 2001) to encode N-terminal peptides.
611   However, there is as yet no definitive evidence that LCB18 contains genes that might confer fitness
612   to the medfly (Table 2).

613

614   The Medkleb genome carries an inversion of GC skew on the leading strand (>50% C; region GI1)
615   indicating a putative horizontal gene transfer event. GI1 is thought to have been transferred from a
616   plasmid, as it codes for *TraY* which facilitates plasmid conjugal transfer (Nelson et al 1995). Though
617   plasmid integration into host chromosome is common (Dobrindt et al 2004; Bire and Rouleux-Bonnin
618   2012), GI1 does not appear to have integrated from any of the Medkleb plasmids (mkps 1-4) (Table
619   2). In total GI1 is predicted to code for 25 proteins including Bll0873, first sequenced in
620   *Bradyrhizobium diazoefficiens*, a bacterium that is a known nitrogen fixing symbiont of legumes
621   (Kaneko et al 2002). Interestingly, both GI1 (Mkleb_0586) and LCB18 (Mkleb_4596) code for genes
622   predicted to facilitate molybdopterin biosynthesis, which could have the potential to benefit the
623   medfly host, as this group of co-factors aid nitrate reduction (Moreno-Vivián et al 1999).

624

625 According to KEGG analysis, the Medkleb genome contains coding sequences for 21 atypical
626 functional orthologue genes that are not encoded in the genomes of any other closely related
627 *Klebsiella* bacteria in the Medkleb group (Table 2). Some of these are potentially mutualistic
628 functions such as sugar metabolism (*mtlA*) and nitrogen fixation (*nifZ*). Medkleb also has duplicates
629 of 11 functional orthologues that *K. oxytoca* generally retains in only single copy. These duplicated
630 functional orthologues do not cluster by genomic location. Several of these duplicated gene functions
631 have strong mutualistic potential such as the biosynthesis of amino acids (*proC*, *trpA*) and
632 breakdown of essential nutrients (*cite*, *pydC*). In contrast, Medkleb is also missing some clusters of
633 genes that are present in closely related bacteria, e.g. for specific enzymes related to copper
634 resistance (*copB, cusS, cusR, cusC, cusF*) and phosphonate transport (*phnC, phnD, phnE).*

635 **Table 2 Unique features of the Medkleb genome in comparison to conspecifics.**

Unique feature	Identified via	Possible mutualistic function
LCB18	progressiveMauve analysis of Medkleb and three conspecifics	Codes for genes related to molybdopterin biosynthesis and several proteins with unknown function. These genes may confer fitness to the medfly.
GI1	Inversion of GC skew	Codes for genes related to molybdopterin biosynthesis and several proteins with unknown function. These genes may confer fitness to the medfly.
Atypical metabolic genes	Comparison of KEGG gene functions between Medkleb group members.	*mtlA, ACADSB, egsA* and *FAAH2* expand the range of nutrients available to the medfly.
Absent sensory genes	Comparison of KEGG gene functions between Medkleb group members.	The medfly gut may be relatively non-toxic allowing Medkleb to survive without genes that detect and metabolise certain chemical threats in its environment (*copB, cusS, cusR, cusC, cusF, arsB, nthA*).
Duplicated genes	Comparison of KEGG gene functions between Medkleb group members	Duplication of 11 genes associated with biosynthesis of amino acids and breakdown of essential nutrients. These genes may confer fitness to the medfly by providing access to extra nutritional resources

Unique feature	Identified via	Possible mutualistic function
Butyrolactone biosynthesis	antiSMASH 6.0 beta	Possible regulation of antibiotic products (Takano 2006; Kitani et al 2011)
N-acyl amino acid biosynthesis	antiSMASH 6.0 beta	An important family of endogenous signalling molecules in which an amide bond covalently links an amino acid to the acyl moiety of a long-chain fatty acid. Primarily involved in cell-to-cell communication (Craig et al 2011; Battista et al 2019_

636

Discussion

638
639 **The medfly microbiome**

640 In this study, we report the use of laboratory and field-reared adult medflies to characterise key
641 features of the bacterial microbiome of this important agricultural pest. Using analyses of 16S rRNA
642 gene amplicon sequencing of culturable and non--culturable members of the gut microbiome, we
643 found that overall species richness was fairly stable between the laboratory-reared flies raised on
644 different diets and the wild medfly samples. Direct comparisons of beta diversity indicated that larval
645 diet, rather than exposure to antibiotics or wild vs lab rearing was the primary driver of microbial
646 diversity of gut microbiomes. Wild flies obtained from distinct geographical regions and different
647 hosts and flies reared on different substrates and exposed to antibiotic cocktails in the laboratory,
648 contained largely the same bacterial families. The data suggest that although medflies are highly
649 polyphagous, they have a stable microbiome that is dominated by the bacterial family
650 *Enterobacteriaceae,* including a putative symbiont *Klebsiella* spp. These findings are consistent with
651 previous analyses of medfly microbiomes made using culture-based methods (Behar et al 2005;
652 Behar et al 2008a; Behar et al 2008b; Behar et al 2009a). The picture may be more complex,
653 however, as one recent next-generation sequencing study did not isolate *Klebsiella* in medfly
654 microbiomes from wild populations (Malacrinò et al 2018), and another identified possible
655 geographic or host-genetic structure links with bacterial dominance (with samples from Greece being
656 dominated by *Klebsiella* spp) (Nikolouli et al 2020).

657

658 **Genome sequencing and analysis of the putative Medkleb gut symbiont**
659 We obtained a fully sequenced Medkleb genome and both 16S rRNA gene and ANI phylogenies
660 identified this as a *Klebsiella* species (Stackebrandt 2006; Kim et al 2014). Medkleb was *pehX*
661 positive when analysed with PCR, and a producer of pectinolytic enzymes, both of which support its
662 identification as *K. oxytoca* (Kovtunovych et al 2003; Saha et al 2013).

663

664 Genome size and GC content are generally reduced when a bacterium adopts a facultative lifestyle
665 (McCutcheon and Moran 2011). We found that, although the Medkleb genome was the smallest of
666 all *Klebsiella* bacteria in the 'Medkleb group', the free-living *K. michiganensis* strain M1 was only 40
667 kb (0.7%) larger. The Medkleb GC content was also the second highest in the 'Medkleb group', again
668 counter to what is expected for a facultative mutualist. A strongly symbiotic transition is generally
669 associated with an increased mutation rate, which causes facultative symbionts to occupy extended
670 branches of phylogenetic trees (McCutcheon and Moran 2011). Though Medkleb was slightly distinct
671 from the three bacterial species most closely related to it, it was not markedly so.

672

673 We also used high resolution comparative genetic techniques to test for signals of putative
674 mutualism. These analyses were designed to test: (i) If Medkleb possessed genetic loci/functions
675 that were absent from closely related *Klebsiella* bacteria and possessed an obvious mutualistic
676 capacity for the fly; (ii) If Medkleb was lacking genetic loci/functions found in all closely related free-
677 living *Klebsiella* spp., that would be considered necessary for life in varied environments. Set against
678 this is that gene loss may be unpredictable in the early stages of facultative mutualistic transitions
679 (Moran and Bennett 2014).

680

681 Comparative genomics highlighted that the Medkleb genome contained two discrete regions (GI1
682 and LCB18) which may have been acquired comparatively recently and are not typically found in *K.
683 oxytoca*. Both GI1 and LCB18 contain genes associated with horizontal transfer and encode several
684 unannotated hypothetical proteins with mutualistic potential. Medkleb was demonstrated to be
685 functionally pectinolytic and global gene function analyses highlighted several more atypical
686 functions were encoded exclusively by Medkleb in comparison to conspecifics. These atypical genes
687 are dispersed throughout the genome which suggests that they are ancestral and not recently
688 inherited. Some of these genes are predicted to encode metabolic enzymes that could allow the
689 medfly to utilise otherwise unattainable nutrients. In addition, Medkleb has several duplicated genes
690 that are randomly distributed throughout the genome which could provide the medfly with beneficial
691 nutrients and amino acids. Together, these genomic signatures suggest that Medkleb is a symbiont
692 that improves host fitness through increased access to vital nutrients. If substantiated, this could
693 help to partially explain the success of the medfly as a generalist pest. However, potentially
694 mutualistic traits are not evidence of mutualism itself and many signals that are commonly associated
695 with facultative mutualism are not evident in the Medkleb genome. Therefore there is insufficient
696 evidence, so far, to definitively classify Medkleb as a facultative mutualist of the medfly.

697

698 *Klebsiella* bacteria and the medfly appear to co-occur almost universally (Behar et al 2009b; Leftwich
699 2012). Therefore, Medkleb might provide a platform for paratransgenic control of the medfly. This
700 study is one of only a handful of comprehensive characterisations of the bacterial symbionts of the
701 tephritid fruit fly *C. capitata* using culture-independent methods, and the first to characterise the

genome of a putative symbiont. Future studies should sample all developmental stages of the medfly, and sample more intensively to establish the potential for phylosymbiosis. Complementing this should be experimental studies to establish localisation of *Klebsiella* within the medfly gut, to determine transmission mechanisms and assess metabolic communication and host benefit (Leftwich et al 2020).

Data Deposition:

Sequencing data has been submitted to the NCBI Sequence Read Archive and is available under BioProject ID PRJNA638617.

References

Aharon Y, Pasternak Z, Ben Yosef M, et al (2013) Phylogenetic, metabolic, and taxonomic diversities shape Mediterranean fruit fly microbiotas during ontogeny. Appl Environ Microbiol 79:303–313. doi: 10.1128/AEM.02761-12

Altschul SF, Madden TL, Schäffer AA, et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. doi: 10.1093/nar/25.17.3389

Anbutsu H, Moriyama M, Nikoh N, et al (2017) Small genome symbiont underlies cuticle hardness in beetles. Proc Natl Acad Sci USA 114:E8382–E8391. doi: 10.1073/pnas.1712857114

Arredondo-Alonso S, Rogers MRC, Braat JC, et al (2018) mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. Microb Genom. doi: 10.1099/mgen.0.000224

Ballinger MJ, Perlman SJ (2017) Generality of toxins in defensive symbiosis: Ribosome-inactivating proteins and defense against parasitic wasps in Drosophila. PLoS Pathog 13:e1006431. doi: 10.1371/journal.ppat.1006431

Bao G, Zhang Y, Du C, et al (2013) Genome Sequence of Klebsiella oxytoca M5al, a Promising Strain for Nitrogen Fixation and Chemical Production. Genome Announc. doi: 10.1128/genomeA.00074-12

Battista N, Bari M, Bisogno T (2019) N-Acyl Amino Acids: Metabolism, Molecular Targets, and Role in Biological Processes. Biomolecules. doi: 10.3390/biom9120822

Behar A, Ben-Yosef M, Lauzon CR, et al (2009a) Structure and function of the bacterial community associated with the Mediterranean fruit fly. Insect Symbiosis

Behar A, Ben-Yosef M, Lauzon CR, et al (2009b) Structure and function of the bacterial community associated with the Mediterranean fruit fly. 251–271.

Behar A, Jurkevitch E, Yuval B (2008a) Bringing back the fruit into fruit fly-bacteria interactions. Mol Ecol 17:1375–1386. doi: 10.1111/j.1365-294X.2008.03674.x

Behar A, Yuval B, Jurkevitch E (2008b) Gut bacterial communities in the Mediterranean fruit fly (Ceratitis capitata) and their impact on host longevity. J Insect Physiol 54:1377–1383. doi: 10.1016/j.jinsphys.2008.07.011

Behar A, Yuval B, Jurkevitch E (2005) Enterobacteria-mediated nitrogen fixation in natural populations of the fruit fly Ceratitis capitata. Mol Ecol 14:2637–2643. doi: 10.1111/j.1365-294X.2005.02615.x

Ben Ami E, Yuval B, Jurkevitch E (2010) Manipulation of the microbiota of mass-reared Mediterranean fruit flies *Ceratitis capitata* (Diptera: Tephritidae) improves sterile male sexual performance. ISME J 4:28–37. doi: 10.1038/ismej.2009.82

Ben-Yosef M, Pasternak Z, Jurkevitch E, Yuval B (2014) Symbiotic bacteria enable olive flies (*Bactrocera oleae*) to exploit intractable sources of nitrogen. J Evol Biol 27:2695–2705. doi: 10.1111/jeb.12527

Bennett GM, Moran NA (2013) Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. Genome Biol Evol 5:1675–1688. doi: 10.1093/gbe/evt118

Bire S, Rouleux-Bonnin F (2012) Transposable elements as tools for reshaping the genome: it is a huge world after all! Methods Mol Biol 859:1–28. doi: 10.1007/978-1-61779-603-6_1

Blin K, Shaw S, Steinke K, et al (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. doi: 10.1093/nar/gkz310

Bosco-Drayon V, Poidevin M, Boneca IG, et al (2012) Peptidoglycan sensing by the receptor PGRP-LE in the Drosophila gut induces immune responses to infectious bacteria and tolerance to microbiota. Cell Host Microbe 12:153–165. doi: 10.1016/j.chom.2012.06.002

Bowers RM, Kyrpides NC, Stepanauskas R, et al (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. doi: 10.1038/nbt.3893

Carver T, Thomson N, Bleasby A, et al (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25:119–120. doi: 10.1093/bioinformatics/btn578

Cascales E, Buchanan SK, Duché D, et al (2007) Colicin biology. Microbiol Mol Biol Rev 71:158–229. doi: 10.1128/MMBR.00036-06

773   Chandler JA, Lang JM, Bhatnagar S, et al (2011) Bacterial communities of diverse Drosophila
774   species: ecological context of a host-microbe model system. PLoS Genet 7:e1002272. doi:
775   10.1371/journal.pgen.1002272
776   Chin C-S, Alexander DH, Marks P, et al (2013) Nonhybrid, finished microbial genome assemblies
777   from long-read SMRT sequencing data. Nat Methods 10:563–569. doi:
778   10.1038/nmeth.2474
779   Craig JW, Cherry MA, Brady SF (2011) Long-chain N-acyl amino acid synthases are linked to the
780   putative PEP-CTERM/exosortase protein-sorting system in Gram-negative bacteria. J
781   Bacteriol 193:5707–5715. doi: 10.1128/JB.05426-11
782   Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene
783   gain, loss and rearrangement. PLoS One 5:e11147. doi: 10.1371/journal.pone.0011147
784   de Graaf FK, Spanjaerdt Speckman EA, Stouthamer AH (1969) Mode of action of a bacteriocin
785   produced by *Enterobacter cloacae* DF13. Antonie Van Leeuwenhoek 35:287–306.
786   de Souza DJ, Bézier A, Depoix D, et al (2009) *Blochmannia* endosymbionts improve colony growth
787   and immune defence in the ant *Camponotus fellah*. BMC Microbiol 9:29. doi:
788   10.1186/1471-2180-9-29
789   Deguenon JM, Travanty N, Zhu J, et al (2019) Exogenous and endogenous microbiomes of wild-
790   caught *Phormia regina* (Diptera: Calliphoridae) flies from a suburban farm by 16S rRNA
791   gene sequencing. Sci Rep 9:20365. doi: 10.1038/s41598-019-56733-z
792   Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and
793   environmental microorganisms. Nat Rev Microbiol 2:414–424. doi: 10.1038/nrmicro884
794   Durvasula RV, Gumbs A, Panackal A, et al (1997) Prevention of insect-borne disease: an
795   approach using transgenic symbiotic bacteria. Proc Natl Acad Sci USA 94:3274–3278. doi:
796   10.1073/pnas.94.7.3274
797   Figueras MJ, Beaz-Hidalgo R, Hossain MJ, Liles MR (2014) Taxonomic affiliation of new genomes
798   should be verified using average nucleotide identity and multilocus phylogenetic analysis.
799   Genome Announc. doi: 10.1128/genomeA.00927-14
800   Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of
801   sequence data. Mol Biol Evol 14:685–695. doi: 10.1093/oxfordjournals.molbev.a025808
802   Gavriel S, Jurkevitch E, Gazit Y, Yuval B (2011) Bacterially enriched diet improves sexual
803   performance of sterile male Mediterranean fruit flies. J Appl Entomology 135:564–573. doi:
804   10.1111/j.1439-0418.2010.01605.x
805   Geer LY, Marchler-Bauer A, Geer RC, et al (2010) The NCBI BioSystems database. Nucleic Acids
806   Res 38:D492–6. doi: 10.1093/nar/gkp858
807   Gould AL, Zhang V, Lamberti L, et al (2018) Microbiome interactions shape host fitness. Proc Natl
808   Acad Sci USA 115:E11951–E11960. doi: 10.1073/pnas.1809349115
809   Hansen TB, Wiklund ED, Bramsen JB, et al (2011) miRNA-dependent gene silencing involving
810   Ago2-mediated cleavage of a circular antisense RNA. EMBO J 30:4414–4422. doi:
811   10.1038/emboj.2011.359
812   Heine D, Holmes NA, Worsley SF, et al (2018) Chemical warfare between leafcutter ant symbionts
813   and a co-evolved pathogen. Nat Commun 9:2208. doi: 10.1038/s41467-018-04520-1
814   Holmes NA, Innocent TM, Heine D, et al (2016) Genome analysis of two
815   *Pseudonocardia* phylotypes associated with *Acromyrmex* leafcutter ants reveals their
816   biosynthetic potential. Front Microbiol 7:2073. doi: 10.3389/fmicb.2016.02073
817   Hurst GDD, Frost CL (2015) Reproductive parasitism: maternally inherited symbionts in a
818   biparental world. Cold Spring Harb Perspect Biol. doi: 10.1101/cshperspect.a017699
819   Husnik F, Nikoh N, Koga R, et al (2013) Horizontal gene transfer from diverse bacteria to an insect
820   genome enables a tripartite nested mealybug symbiosis. Cell 153:1567–1578. doi:
821   10.1016/j.cell.2013.05.040
822   Hyatt D, Chen G-L, Locascio PF, et al (2010) Prodigal: prokaryotic gene recognition and
823   translation initiation site identification. BMC Bioinformatics 11:119. doi: 10.1186/1471-2105-
824   11-119
825   Jurkevitch E (2011) Riding the Trojan horse: combating pest insects with their own symbionts.
826   Microb Biotechnol 4:620–627. doi: 10.1111/j.1751-7915.2011.00249.x
827   Kanehisa M, Sato Y, Kawashima M, et al (2016) KEGG as a reference resource for gene and
828   protein annotation. Nucleic Acids Res 44:D457–62. doi: 10.1093/nar/gkv1070

829 Kaneko T, Nakamura Y, Sato S, et al (2002) Complete genomic sequence of nitrogen-fixing
830   symbiotic bacterium *Bradyrhizobium japonicum* USDA110 (supplement). DNA Res 9:225–
831   256. doi: 10.1093/dnares/9.6.225
832 Kim M, Oh H-S, Park S-C, Chun J (2014) Towards a taxonomic coherence between average
833   nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of
834   prokaryotes. Int J Syst Evol Microbiol 64:346–351. doi: 10.1099/ijs.0.059774-0
835 Kitani S, Miyamoto KT, Takamatsu S, et al (2011) Avenolide, a Streptomyces hormone controlling
836   antibiotic production in *Streptomyces avermitilis*. Proc Natl Acad Sci USA 108:16410–
837   16415. doi: 10.1073/pnas.1113908108
838 Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for
839   prokaryotes. Proc Natl Acad Sci USA 102:2567–2572. doi: 10.1073/pnas.0409727102
840 Kovtunovych G, Lytvynenko T, Negrutska V, et al (2003) Identification of *Klebsiella oxytoca* using a
841   specific PCR assay targeting the polygalacturonase pehX gene. Res Microbiol 154:587–
842   592. doi: 10.1016/S0923-2508(03)00148-7
843 Kuo C-H, Moran NA, Ochman H (2009) The consequences of genetic drift for bacterial genome
844   complexity. Genome Res 19:1450–1454. doi: 10.1101/gr.091785.109
845 Lagesen K, Hallin P, Rødland EA, et al (2007) RNAmmer: consistent and rapid annotation of
846   ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. doi: 10.1093/nar/gkm160
847 Lauzon CR, McCombs SD, Potter SE, Peabody NC (2009) Establishment and vertical passage of
848   Enterobacter (Pantoea) agglomerans and Klebsiella pneumoniae through all life stages of
849   the Mediterranean fruit fly (Diptera: Tephritidae). Ann Entomol Soc Am 102:85–95.
850 Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: rates of change and
851   exchange. J Mol Evol 44:383–397. doi: 10.1007/pl00006158
852 Lee C, Grasso C, Sharlow MF (2002) Multiple sequence alignment using partial order graphs.
853   Bioinformatics 18:452–464. doi: 10.1093/bioinformatics/18.3.452
854 Leftwich PT (2012) Male reproductive success and population control in the Mediterranean fruit fly,
855   *Ceratitis capitata*. Doctoral Thesis, University of East Anglia
856 Leftwich PT, Bolton M, Chapman T (2016) Evolutionary biology and genetic techniques for insect
857   control. Evol Appl 9:212–230. doi: 10.1111/eva.12280
858 Leftwich PT, Clarke NVE, Hutchings MI, Chapman T (2018) Reply to Obadia et al.: Effect of methyl
859   paraben on host-microbiota interactions in *Drosophila melanogaster*. Proc Natl Acad Sci
860   USA 115:E4549–E4550. doi: 10.1073/pnas.1805499115
861 Leftwich PT, Edgington MP, Chapman T (2020) Transmission efficiency drives host-microbe
862   associations. Proc Biol Sci 287:20200820. doi: 10.1098/rspb.2020.0820
863 Leftwich PT, Nash WJ, Friend LA, Chapman T (2017) Adaptation to divergent larval diets in the
864   medfly, *Ceratitis capitata*. Evolution 71:289–303. doi: 10.1111/evo.13113
865 Lhocine N, Ribeiro PS, Buchon N, et al (2008) PIMS modulates immune tolerance by negatively
866   regulating Drosophila innate immune signaling. Cell Host Microbe 4:147–158. doi:
867   10.1016/j.chom.2008.07.004
868 Lindberg BG, Tang X, Dantoft W, et al (2018) Nubbin isoform antagonism governs Drosophila
869   intestinal immune homeostasis. PLoS Pathog 14:e1006936. doi:
870   10.1371/journal.ppat.1006936
871 Liu LJ, Martinez-Sañudo I, Mazzon L, et al (2016) Bacterial communities associated with invasive
872   populations of *Bactrocera dorsalis* (Diptera: Tephritidae) in China. Bull Entomol Res
873   106:718–728. doi: 10.1017/S0007485316000390
874 Lobry JR (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. Mol Biol Evol
875   13:660–665. doi: 10.1093/oxfordjournals.molbev.a025626
876 Lombard V, Bernard T, Rancurel C, et al (2010) A hierarchical classification of polysaccharide
877   lyases for glycogenomics. Biochem J 432:437–444. doi: 10.1042/BJ20101185
878 Malacrinò A, Campolo O, Medina RF, Palmeri V (2018) Instar- and host-associated differentiation
879   of bacterial communities in the Mediterranean fruit fly *Ceratitis capitata*. PLoS One
880   13:e0194131. doi: 10.1371/journal.pone.0194131
881 McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. Nat Rev
882   Microbiol 10:13–26. doi: 10.1038/nrmicro2670
883 McFall-Ngai M, Hadfield MG, Bosch TCG, et al (2013) Animals in a bacterial world, a new
884   imperative for the life sciences. Proc Natl Acad Sci USA 110:3229–3236. doi:
885   10.1073/pnas.1218525110

886 McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and
887   graphics of microbiome census data. PLoS One 8:e61217. doi:
888   10.1371/journal.pone.0061217

889 Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal
890   Chem 31:426–428. doi: 10.1021/ac60147a030

891 Moran NA, Bennett GM (2014) The tiniest tiny genomes. Annu Rev Microbiol 68:195–215. doi:
892   10.1146/annurev-micro-091213-112901

893 Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial
894   symbionts. Annu Rev Genet 42:165–190. doi: 10.1146/annurev.genet.41.110306.130119

895 Moran NA, Tran P, Gerardo NM (2005) Symbiosis and insect diversification: an ancient symbiont
896   of sap-feeding insects from the bacterial phylum Bacteroidetes. Appl Environ Microbiol
897   71:8802–8810. doi: 10.1128/AEM.71.12.8802-8810.2005

898 Moreno-Vivián C, Cabello P, Martínez-Luque M, et al (1999) Prokaryotic nitrate reduction:
899   molecular properties and functional distinction among bacterial nitrate reductases. J
900   Bacteriol 181:6573–6584. doi: 10.1128/JB.181.21.6573-6584.1999

901 Negri I, Franchini A, Gonella E, et al (2009) Unravelling the Wolbachia evolutionary role: the
902   reprogramming of the host genomic imprinting. Proc Biol Sci 276:2485–2491. doi:
903   10.1098/rspb.2009.0324

904 Nelson WC, Howard MT, Sherman JA, Matson SW (1995) The *traY* gene product and integration
905   host factor stimulate *Escherichia coli* DNA helicase I-catalyzed nicking at the F plasmid
906   oriT. J Biol Chem 270:28374–28380.

907 Nikolouli K, Augustinos AA, Stathopoulou P, et al (2020) Genetic structure and symbiotic profile of
908   worldwide natural populations of the Mediterranean fruit fly, *Ceratitis capitata*. BMC Genet
909   21:128. doi: 10.1186/s12863-020-00946-z

910 Nishida H (2012) Comparative analyses of base compositions, DNA sizes, and dinucleotide
911   frequency profiles in archaeal and bacterial chromosomes and plasmids. Int J Evol Biol
912   2012:342482. doi: 10.1155/2012/342482

913 Ochman H, Davalos LM (2006) The nature and dynamics of bacterial genomes. Science
914   311:1730–1733. doi: 10.1126/science.1119966

915 Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial
916   innovation. Nature 405:299–304. doi: 10.1038/35012500

917 Oksanen J, Kindt R, Legendre P, O'Hara B (2007) The vegan package. Community ecology …

918 Oliver KM, Russell JA, Moran NA, Hunter MS (2003) Facultative bacterial symbionts in aphids
919   confer resistance to parasitic wasps. Proc Natl Acad Sci USA 100:1803–1807. doi:
920   10.1073/pnas.0335320100

921 Overbeek R, Olson R, Pusch GD, et al (2014) The SEED and the Rapid Annotation of microbial
922   genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–14. doi:
923   10.1093/nar/gkt1226

924 Parks DH, Imelfort M, Skennerton CT, et al (2015) CheckM: assessing the quality of microbial
925   genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–
926   1055. doi: 10.1101/gr.186072.114

927 Peplies J, Kottmann R, Ludwig W, Glöckner FO (2008) A standard operating procedure for
928   phylogenetic inference (SOPPI) using (rRNA) marker genes. Syst Appl Microbiol 31:251–
929   257. doi: 10.1016/j.syapm.2008.08.003

930 Price MN, Dehal PS, Arkin AP (2010) FastTree 2 — approximately maximum-likelihood trees for
931   large alignments. PLoS One 5:e9490. doi: 10.1371/journal.pone.0009490

932 Pruesse E, Peplies J, Glöckner FO (2012) SINA: accurate high-throughput multiple sequence
933   alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829. doi:
934   10.1093/bioinformatics/bts252

935 Quast C, Pruesse E, Yilmaz P, et al (2013) The SILVA ribosomal RNA gene database project:
936   improved data processing and web-based tools. Nucleic Acids Res 41:D590–6. doi:
937   10.1093/nar/gks1219

938 R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for
939   Statistical Computing, Vienna, Austria

940 Rambaut A, Drummond AJ (2009) FigTree version 1.4.3.

Reva ON, Tümmler B (2004) Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns. BMC Bioinformatics 5:90. doi: 10.1186/1471-2105-5-90

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126–19131. doi: 10.1073/pnas.0906412106

Rocha EPC, Danchin A (2002) Base composition bias might result from competition for metabolic resources. Trends Genet 18:291–294. doi: 10.1016/S0168-9525(02)02690-2

Sabree ZL, Huang CY, Okusu A, et al (2013) The nutrient supplying capabilities of Uzinura, an endosymbiont of armoured scale insects. Environ Microbiol 15:1988–1999. doi: 10.1111/1462-2920.12058

Saha R, Farrance CE, Verghese B, et al (2013) *Klebsiella michiganensis* sp. nov., a new bacterium isolated from a tooth brush holder. Curr Microbiol 66:72–78. doi: 10.1007/s00284-012-0245-x

Schloss PD, Westcott SL, Ryabin T, et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. doi: 10.1128/AEM.01541-09

Shigenobu S, Watanabe H, Hattori M, et al (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–86. doi: 10.1038/35024074

Shin SH, Kim S, Kim JY, et al (2012) Complete genome sequence of *Klebsiella oxytoca* KCTC 1686, used in production of 2,3-butanediol. J Bacteriol 194:2371–2372. doi: 10.1128/JB.00026-12

Simão FA, Waterhouse RM, Ioannidis P, et al (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. doi: 10.1093/bioinformatics/btv351

Simmons JS (1926) A culture medium for differentiating organisms of typhoid-colon aerogenes groups and for isolation of certain fungi. J. Infect. Dis.

Sinotte VM, Freedman SN, Ugelvig LV, Seid MA (2018) *Camponotus floridanus* ants incur a trade-off between phenotypic development and pathogen susceptibility from their mutualistic endosymbiont *Blochmannia*. Insects. doi: 10.3390/insects9020058

Sloan DB, Moran NA (2012) Endosymbiotic bacteria as a source of carotenoids in whiteflies. Biol Lett 8:986–989. doi: 10.1098/rsbl.2012.0664

Sohail M, Latif Z (2016) Phylogenetic analysis of polygalacturonase-producing *Bacillus* and *Pseudomonas* isolated from plant waste material. Jundishapur J Microbiol 9:e28594. doi: 10.5812/jjm.28594

Stackebrandt E (2006) Taxonomic parameters revisited : tarnished gold standards. Microbiol Today 33:152–155.

Storelli G, Strigini M, Grenier T, et al (2018) Drosophila perpetuates nutritional mutualism by promoting the fitness of its intestinal symbiont *Lactobacillus plantarum*. Cell Metab 27:362–377.e8. doi: 10.1016/j.cmet.2017.11.011

Takano E (2006) Gamma-butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. Curr Opin Microbiol 9:287–294. doi: 10.1016/j.mib.2006.04.003

Tamas I, Klasson L, Canbäck B, et al (2002) 50 million years of genomic stasis in endosymbiotic bacteria. Science 296:2376–2379. doi: 10.1126/science.1071278

van Ham RCHJ, Kamerbeek J, Palacios C, et al (2003) Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci USA 100:581–586. doi: 10.1073/pnas.0235981100

Von Dohlen CD, Moran NA (2000) Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. Biological Journal of the Linnean Society 71:689–717. doi: 10.1006/bijl.2000.0470

Wegener CB (2002) Induction of defence responses against Erwinia soft rot by an endogenous pectate lyase in potatoes. Physiol Mol Plant Pathol 60:91–100. doi: 10.1006/pmpp.2002.0377

Weiss B, Kaltenpoth M (2016) Bacteriome-localized intracellular symbionts in pollen-feeding beetles of the genus Dasytes (Coleoptera, Dasytidae). Front Microbiol 7:1486. doi: 10.3389/fmicb.2016.01486

Whitten M, Dyson P (2017) Gene silencing in non-model insects: Overcoming hurdles using symbiotic bacteria for trauma-free sustainable delivery of RNA interference: Sustained RNA

998   interference in insects mediated by symbiotic bacteria: Applications as a genetic tool and
999   as a biocide. Bioessays. doi: 10.1002/bies.201600247

1000 Whitten MMA, Facey PD, Del Sol R, et al (2016) Symbiont-mediated RNA interference in insects.
1001   Proc Biol Sci 283:20160042. doi: 10.1098/rspb.2016.0042

1002 Wickham H (2017) Tidyverse: Easily Install and Load the' ' 'Tidyverse.

1003 Wixon J (2001) Featured organism: reductive evolution in bacteria: *Buchnera* sp., *Rickettsia*
1004   *prowazekii* and *Mycobacterium leprae*. Comp Funct Genomics 2:44–48. doi:
1005   10.1002/cfg.70

1006 Woruba DN, Morrow JL, Reynolds OL, et al (2019) Diet and irradiation effects on the bacterial
1007   community composition and structure in the gut of domesticated teneral and mature
1008   Queensland fruit fly, *Bactrocera tryoni* (Diptera: Tephritidae). BMC Microbiol 19:281. doi:
1009   10.1186/s12866-019-1649-6

1010 Wu D, Daugherty SC, Van Aken SE, et al (2006) Metabolic complementarity and genomics of the
1011   dual bacterial symbiosis of sharpshooters. PLoS Biol 4:e188. doi:
1012   10.1371/journal.pbio.0040188

1013 Xie F, Murray JD, Kim J, et al (2012) Legume pectate lyase required for root infection by rhizobia.
1014   Proc Natl Acad Sci USA 109:633–638. doi: 10.1073/pnas.1113992109

1015 Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable
1016   rates over sites: approximate methods. J Mol Evol 39:306–314. doi: 10.1007/BF00160154
1017