bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Title:**

2 Individual variation of natural *D. melanogaster* associated bacterial communities

3

4 **Running title:**

5 individual fly microbiota

6

7 **Authors:**

8 Yun Wang, Fabian Staubach

9

10 **Author affiliations:**

11 Yun Wang, Albert Ludwigs University Freiburg

12 Fabian Staubach, Albert Ludwigs University Freiburg, Stanford University

13

14 **Corresponding author:**

15 Fabian Staubach, fabian.staubach@biologie.uni-freiburg.de

16

17

18

19

20

21

22

23

24

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

25 **Abstract**

26 *D. melanogaster* has become an important model organism to study host-microbe interaction.
27 However, we still know little about the natural microbial communities that are associated with *D.*
28 *melanogaster*. Especially, information on inter-individual variation is still lacking because most
29 studies so far have used pooled material from several flies. Here, we collected bacterial 16S
30 rRNA gene community profiles from a set of 32 individuals from a single population and
31 compare the variation to that of samples collected from different substrates and locations. While
32 community differences were on average larger between samples collected from different
33 substrates, there was still a surprising amount of variation of microbial communities between
34 individual flies. The samples clustered into two groups suggesting that there are yet unknown
35 factors that affect the composition of natural fly associated microbial communities and need
36 research.

37

38 **Importance**

39 *D. melanogaster* is an important model organism in evolutionary biology and also for the study
40 of host-microbe interaction. In order to connect these to aspects of *D. melanogaster* biology, it is
41 crucial to better understand the natural *D. melanogaster* microbiota because only the natural
42 microbiota can affect the evolution of the host. We present, to our knowledge, the first data set
43 that captures inter-individual variation of *D. melanogaster* associated bacterial communities.
44 Clustering of communities into two larger groups suggests that there are important drivers of
45 these communities that we do not understand yet suggesting in return that more research on the
46 natural microbiota of *D. melanogaster* is needed.

47

48

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

49 **Introduction**

50 *Drosophila melanogaster* has become an important model for the investigation of host-microbe
51 interaction (1). Interactions with bacteria can affect *D. melanogaster* phenotype in many aspects.
52 Bacteria mediated *D. melanogaster* phenotypic effects range from pathogenic (2), over effects on
53 cold tolerance (3), and effects on *D. melanogaster* nutritional status (4, 5) to highly beneficial
54 effects. For example, certain acetic acid bacteria (Acetobacteraceae) and Lactobacillaceae
55 significantly promote *D. melanogaster* larval growth on amino acid poor diet (6, 7) and can
56 ensure longtime fertility and longevity under nutrient poor conditions (8). Because bacteria from
57 the same families that have these highly fitness relevant effects on *D. melanogaster* can also be
58 found in association with wild-caught flies (9–12), it seems reasonable to assume that they could
59 also play a role in fly evolution. However, in most studies bacterial strains and communities that
60 were isolated from lab-reared flies are investigated and we still know rather little about natural
61 *Drosophila* associated bacterial communities and the factors shaping them. Yet, only natural
62 microbial communities can have played a role in *Drosophila* evolutionary history. This is the
63 more important since it was shown by Chandler et al. (2011) (11) and Staubach et al. (2013) (12)
64 that bacterial communities associated with wild-caught flies are different from those associated
65 with lab-reared flies. Under controlled laboratory conditions, host genetic make up can influence
66 *D. melanogaster* associated bacterial communities (5, 13). Under natural conditions, the substrate
67 flies were collected from strongly correlates with bacterial community composition (12), while
68 species differences between *D. melanogaster* and *D. simulans* are detectable but have a smaller
69 effect.

70 These studies on natural microbial communities relied on pooling material from several
71 flies for bacterial community profiling. However, information from individual, wild-caught flies
72 is required to put the size of substrate related effects on communities into the perspective of

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

73 microbial community variation between individuals. A better understanding of inter-individual
74 variation within populations can also help us to assess if the current practice to pool material
75 from several individuals is appropriate to represent the microbial community of a local
76 population. Finally, bacterial community profiles of individuals collected from the same diet at
77 the same time could help to evaluate whether there are other factors beyond diet that drive
78 variation in natural fly associated microbial communities. Yet, contrary to humans, where
79 collecting individual microbial profiles has been in the focus (14), such data, to our knowledge,
80 does not exist for *D. melanogaster.*

81 In order to assess bacterial community variability between individuals under otherwise
82 constant conditions, we collected flies from the same substrate and location at the same time and
83 assessed the bacterial communities of individual male flies using 16S rRNA gene profiling. We
84 compared and contrasted these communities to communities from flies collected from different
85 populations and substrates to place variation between individuals into the context of variation
86 caused by known factors that influence microbial communities. Furthermore, we assess and
87 evaluate how the common practice of pooling material from several individuals influences
88 diversity measurements and composition of fly associated microbes.

89

90 **Results**

91 We assessed bacterial community composition and diversity by sequencing the V4 region of the
92 16S rRNA gene for 32 individual and a pool of 5 wild-caught *D. melanogaster* that were
93 collected from the same substrate (plums) at the same location and time. Additionally a pool of 5
94 male flies from the same location and substrate, but collected 1 year earlier was analyzed. In
95 order to assess between population variation 11 pools of flies collected from 7 different substrates
96 and locations were analyzed. All pools were based on 5 flies except Orange3, for which we were

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

97 able to obtain only 3 flies. Orange3 was no outlier in any of the analyses and was included with
98 the pools of 5 flies (see Table S1 for an overview of all samples). A total of ~ 2,240,000
99 sequences passed quality filtering. ~1,050,000 sequences matched the *Wolbachia* 16s rRNA gene
100 sequence and were removed. Individual fly sample #6 was removed because only 11 sequences
101 remained after *Wolbachia* removal. At least 1008 sequences per sample were collected for the
102 remaining samples.
103
104 **Bacterial diversity of pools and individual flies is similar**
105 We were interested in finding out whether individual flies carry reduced bacterial diversity or a
106 skew in bacterial abundance patterns, as might be expected from stronger stochastic effects in
107 smaller samples, when compared to pools of 5 flies. For comparing bacterial alpha diversity
108 between individual flies and pools of 5 flies, we grouped sequences into 97% identity OTUs
109 (Operational Taxonomic Units).
110 The mean number of OTUs observed when sampling 1008 sequences did not differ
111 between individual flies (44.4±17.8 OTUs) and the pool of 5 flies (40 OTUs) that was collected
112 from the same substrate at the same time ($P = 1$, Mann-Whitney test, Figure 1A, Table 1). This
113 indicates that bacterial OTU richness of the plum population is well represented by a pool of 5
114 flies. Surprisingly, OTU richness of pools of 5 flies collected across substrates and locations also
115 did not differ from the average richness found associated with individual flies collected from
116 plums (39.3±23.8 OTUs, $P = 0.23$, Mann-Whitney test), providing no evidence for a sample size
117 related reduction in diversity. The same holds true for Chao's richness estimate (15), Shannon's H
118 (16), and Simpsons D (17) (Table 1). Furthermore, there was also no difference in variance of
119 OTU richness between communities from individual flies and that of pools of 5 (Levene-test, P
120 $= 0.21$, Figure 1B) indicating that the variance in bacterial richness between individuals from the

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

121 same population and substrate can be similar to that between *D. melanogaster* populations from
122 different substrates and locations.
123 Analysis of evenness using Shannon's E and Simpson's E also revealed no significant
124 differences between individuals and pools of 5 flies. This indicates that there are no significant
125 skews introduced into the OTU distribution by sampling individuals when compared to pools of 5
126 flies.
127
128 **Bacterial communities are dominated by acetic acid bacteria and vary in composition**
129 **between individuals from the same population**
130 For a detailed view of the variability in bacterial community composition of individual and
131 pooled fly samples, we classified the 16S rRNA gene sequences taxonomically.
132 The bacterial communities are dominated by acetic acid bacteria (Acetobacteraceae
133 62.6%) representing the 4 of the 5 most common genera (*Saccharibacter* 24.1%, *Gluconobacter*
134 18.1%, *Acetobacter* 13.4%, and *Gluconacetobacter* 6.7% average relative abundance). The
135 relative abundance of the different taxa is highly variable between individuals from the same
136 population (*Saccharibacter* 2.0% - 93.9% relative abundance, *Gluconobacter* 1.7% - 64.9%,
137 *Acetobacter* 0.5% - 37.2%, and *Gluconacetobacter* 0.4% - 43.9%). Please note that bootstrap
138 support for the *Saccharibacter* classification is often relatively low (between 40% and 60%,
139 Table S2) indicating that there are several sequences in the SILVA database that match sequences
140 from this taxonomic group similarly well. Blast search for representative sequences from this
141 taxonomic group produced perfect matches to bacteria classified as *Commensalibacter* in
142 Chandler et al. (2011) (11) and *Acetobacter* in Corby-Harris et al. (2007) (9) (see File S1 for blast
143 search results). Acetic acid bacteria represent a major bacterial group associated with wild-caught
144 flies (9, 11, 12). Enterobacteriaceae are also common (16.8% average relative abundance) with

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

145 *Buttiauxella* (8.2% average relative abundance) and *Serratia* (6.8% average relative abundance)
146 being the most common. Members of the genus *Serratia* can be *Drosophila* pathogens and occur
147 at high relative abundance in individual samples. This pattern of low abundance or absence in
148 most samples and sharp increase in individual samples (sample orange2, 44.0% rel. abundance)
149 has been associated with *Drosophila* pathogens (12). A similar pattern is visible for *Enterococcus*
150 in sample orange1 (46.5% rel. abundance compared to 6.9% average relative abundance) and the
151 tomato sample (53.6% rel. abundance). *Enterococcus* can reach high titers in *Drosophila* and
152 cause mortality (10).
153
154 **Beta diversity between individual samples is smaller than between pools collected across**
155 **locations and substrates**
156 As described above, microbial community composition varies substantially even between
157 individual flies from the same population. In order to relate this variation between individuals to
158 variation between populations, we analyzed beta-diversity using Bray-Curtis community
159 distances (BCD) and included samples from different substrates and locations in the analysis.
160 The pairwise BCD between samples from individual flies is smaller than that of pools of 5
161 flies sampled across substrates and locations (Figure 3, $P = 2.3 \times 10^{-10}$, Mann-Whitney-test)
162 indicating that between population variation is larger than within population variation. However,
163 the difference is small (0.15) and the BCDs between populations fall almost completely within
164 the range of distances between individuals. These results hold true for Jaccard (Figure S1) as well
165 as weighted (Figure S2) and unweighted (Figure S3) Unifrac community distance.
166
167
168

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

169 **Individual fly samples from the same population fall into two groups**

170 In order to explore potential factors shaping fly associated bacterial communities, a Principal
171 Coordinate Analysis was carried out using BCD (Figure 4).

172 In this analysis most of the variation between individual fly samples falls onto the first
173 PCo, while the pools of 5 flies are distributed along the second PCo2. The two samples collected
174 from grapes that have a high relative abundance of *Buttiauxella* are apart from the other samples
175 at the top of the graph. Interestingly, Figure 4 suggests that the individual fly samples fall into
176 two groups. The pool of 5 flies from plums is at the center of one of these groups suggesting that
177 it represents only a part of the variation found in the plum population.

178 These two groups are supported by hierarchical clustering based on BCD (Figure 5),
179 Jaccard-Distances (Figure S4), and weighted Unifrac distances (Figure S5), but not by
180 unweighted Unifrac distances (Figure S6). In order to identify signature taxa that contribute to
181 differences between the two groups, we applied the SIMPER method combining the samples in
182 the two groups into two different habitats. This identifies OTU1 that was classified as
183 *Saccharibacter* as the largest contributor (19.8% of variation) to the dissimilarity between two
184 groups (a list of contributing OTUs can be found in Table S3).

185

186 **Discussion**

187 In this study we described bacterial community variation between individual wild-caught *D.*
188 *melanogaster*. By analyzing a sample of 32 individual flies from the same population, substrate,
189 and at the same time, we were able to address the questions (i) whether the common practice of
190 pooling material from several flies for bacterial community analysis leads to representative
191 community assessment and (ii) whether there is evidence for factors other than the substrate the
192 flies were collected from that shape natural bacterial communities, and (iii) compare within

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

193 population variation of bacterial communities to between population variation. As before (12), we
194 used entire flies for our study. This is important because fly pathogens can reach high titers in the
195 hemolymph (2, 10) and might be overlooked by focusing on the gut. Although bacteria on the fly
196 surface contribute only ~10% to the total bacterial load of flies (18), and hence their effect on
197 bacterial community composition should be minor, they could still play a role in inoculating fruit
198 and shaping the microbial environment of *D. melanogaster*.
199
200 **Alpha diversity of individual fly and pooled fly communities**
201 Bacterial diversity varied extensively between individuals from the same substrate and location
202 with the standard deviation of OTUs discovered at 40% of the mean. Pooling of 5 individuals
203 from the same population allowed for a good estimate of the mean OTU richness and evenness of
204 individual samples. Interestingly, the variation in richness and evenness between individuals from
205 the same population was as large as between populations that were collected from different
206 locations and substrates. This further indicates that variation in community diversity between
207 individuals of a population can be rather large.
208 Shannon's diversity (H = 1.94 +/- 0.62) was comparable to that found associated with
209 wild-caught flies in Staubach et al. 2013 (12) (H = 1.79 +/- 0.44) and hence to that from Chandler
210 et al. 2011 (11) and Cox and Gilmore (2007) (10) who also investigated bacterial communities of
211 wild-caught flies. Please see Staubach et al. (2013) (12) for extensive diversity comparisons
212 between studies.
213
214 **Composition of bacterial communities**
215 Concordant with many other studies acetic acid bacteria dominated the bacterial communities
216 (10–12, 19). Also concordant with earlier studies on wild-caught flies enterobacteria occur at

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

217 sometimes high relative abundance in some samples. This pattern has been connected to
218 pathogens before and could indicate systemic infections (12).
219 We were surprised to find sequences classified as *Halomonas* in our samples as this is a
220 halophile neither expected on rotting fruit nor flies. Blast search of a representative sequence
221 from the largest *Halomonas* OTU identified plant mitochondria as best hits (see File S2 for blast
222 search). The occurrence of plant mitochondrial sequence in our samples appears much more
223 likely than *Halomonas.*
224 A representative sequence of the unclassified gammaproteobacterium (grey in Figure 2)
225 that can be found in several samples matches perfectly with sequences from uncultured
226 enterobacteria isolated from social corbiculate bees and nectar feeding bats (file S3). The high
227 sugar content that nectar and rotting fruit have in common might favor the growth of similar
228 bacteria.
229 A full list of the 25 most common OTUs and representative sequences can be found in
230 Table S2.
231
232 **Beta diversity**
233 Bacterial community distances are on average higher between samples that were collected across
234 different substrates and locations than between individual flies from a single population that was
235 sampled at one point in time. This is not surprising since it has been shown before that substrate
236 or a correlating variable, for example season, is an important factor for bacterial community
237 composition associated with *D. melanogaster* and *D. simulans* (12). Nonetheless, the community
238 distances between individual flies can be high and the distributions of pairwise community
239 distances largely overlap between individuals and populations revealing surprising variability
240 within a population. Assuming that flies continuously exchange microbes with their environment

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

241 and that they replenish their gut microbiota through uptake of environmental bacteria (20) the
242 large variation of bacterial communities of individuals could result from heterogeneity of the
243 plum substrate the flies were collected from.

244 Surprisingly, our samples clustered into two groups. The clustering correlated with the
245 most abundant OTU (OTU1, supposedly *Saccharibacter*). We can only speculate here how the
246 difference in abundance of OTU1 or the clustering was generated. Following the argument above
247 that microbes are taken up from the environment, we can think of two scenarios that could
248 generate this pattern. In the first scenario, the plum substrate although heterogeneous, falls into
249 two different categories from a microbial composition perspective. These categories could be
250 states of decay or states that result from the dynamic interplay of microbial metabolites (21). In
251 the second scenario, we might have a cohort of flies that entered the population only recently and
252 brought microbes with them from the previous substrate. We can not disentangle these options at
253 the moment. Similarly, because D. melanogaster communities change with age (18, 22, 23),
254 distinct age cohorts in the population could cause distinct microbial communities. Finally,
255 because flies shape their associated microbial communities (24) and fly host genetic makeup
256 affect the community composition at least under laboratory conditions (4, 5, 13) genetic
257 differences in the host population could play a role in generating the observed pattern.

258 Because the sample of 5 flies from the plum population clustered with the smaller group
259 consisting of 10 individual flies, it did not represent beta diversity of the whole population well.

260

261 **Conclusion**

262 While average bacterial alpha diversity in a *D. melanogaster* population was well represented by
263 a sample of 5 flies, composition of microbial communities is highly variable between individuals
264 from the same population. Larger samples of individual flies could be used to better represent

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

265 beta diversity in a population. The two clusters we found in our sample of individuals collected
266 from the same substrate, at the same time and location suggests that there are important factors
267 that shape natural *D. melanogaster* microbial communities that we do not understand yet.
268 Because understanding natural communities is what matters for understanding *D. melanogaster*
269 evolution, more research is needed to better understand the factors that shape natural *D.*
270 *melanogaster* associated microbial communities.
271

272 **Material and Methods**

273 **Fly samples**

274 Flies were collected as described previously (12). In short, live flies were collected and brought
275 to the lab in empty vials within 5 hours of collection. Male *D. melanogaster* were identified
276 based on morphology and frozen at -80°C until DNA extraction. Samples from oranges, peaches,
277 and apple1 were the same as in (12). See Table S1 for a full list of sampling locations.

278

279 **DNA Extraction, PCR and sequencing**

280 DNA was extracted either from individual male *D. melanogaster* or pools of five males, with the
281 exception *D. melanogaster* orange sample 3 (orange3), for which we were able to retrieve three
282 males only. DNA extraction was performed using the Qiagen QIAamp DNA extraction kit
283 (Qiagen, Carlsbad, CA) using bead beating as described in (12) and running negative extraction
284 controls (without fly material) in parallel.

285 Barcoded broad range primers, 515F and 806R, as described in Caporaso et al. 2010 (25)
286 were used to amplify the V4 region of the bacterial 16S rRNA gene. DNA was amplified using
287 Phusion® Hot Start DNA Polymerase (Finnzymes, Espoo, Finland) and the following cycling
288 conditions: 30 sec at 98°C; 30 cycles of 9 sec at 98°C, 60 sec at 50°C, and 90 sec at 72°C; final

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

289 extension for 10 min at 72°C. In order to reduce PCR bias, amplification reactions were
290 performed in duplicate and pooled. PCR products were run on an agarose gel for quantification
291 and pooled in equimolar amounts. Extraction control PCRs were negative and excluded. The
292 resulting pool was gel extracted using the Qiaquick gel extraction kit (Qiagen, Carlsbad, CA) and
293 sequenced on an illumina MiSeq sequencer reading 2 x 250bp.
294
295 **Data analysis**
296 Sequencing data was analyzed using mothur (26)(v 1.36.0) following the MiSeq SOP on
297 mothur.org. Sequences were taxonomically classified using the SILVA reference database (27) as
298 implemented in mothur. The ecodist (28) R package was used to calculate Bray-Curtis-Distances.
299 The vegan R package (29) was used for Jaccard and Unifrac distances incorporating the
300 GUniFrac package (30). The pvclust package was used for cluster analysis (31). A detailed
301 analysis script with all mothur and R commands can be found in file S5.
302
303 **Data availability**
304 Raw data is available at ncbi SRA under the accession number XX
305
306 **Acknowledgements**
307 This work was supported by DFG Forschungsstipendium 1154/1-1 and DFG grant 1154/2-1. We
308 thank Dmitri A Petrov (Stanford University) in whose lab data was collected. We thank Alan O.
309 Bergland and Heather Machado for support in fly sampling. We thank Till Bayer for helpful
310 comments on the manuscript. The authors declare no conflict of interest.
311
312

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

313 **References**

1. Erkosar B, Storelli G, Defaye A, Leulier F. 2013. Host-intestinal microbiota mutualism:" learning on the fly". Cell Host Microbe 13:8.

2. Galac MR, Lazzaro BP. 2011. Comparative pathology of bacteria in the genus Providencia to a natural host, Drosophila melanogaster. Microbes Infect 13:673–683.

3. Linderman JA, Chambers MC, Gupta AS, Schneider DS. 2012. Infection-Related Declines in Chill Coma Recovery and Negative Geotaxis in Drosophila melanogaster. PLoS ONE 7:e41907.

4. Dobson AJ, Chaston JM, Newell PD, Donahue L, Hermann SL, Sannino DR, Westmiller S, Wong AC-N, Clark AG, Lazzaro BP, Douglas AE. 2015. Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of Drosophila melanogaster. Nat Commun 6:6312.

5. Chaston JM, Dobson AJ, Newell PD, Douglas AE. 2016. Host Genetic Control of the Microbiota Mediates the Drosophila Nutritional Phenotype. Appl Environ Microbiol 82:671–679.

6. Shin SC, Kim S-H, You H, Kim B, Kim AC, Lee K-A, Yoon J-H, Ryu J-H, Lee W-J. 2011. Drosophila Microbiome Modulates Host Developmental and Metabolic Homeostasis via Insulin Signaling. Science 334:670–674.

7. Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F. 2011. Lactobacillus plantarum Promotes Drosophila Systemic Growth by Modulating Hormonal Signals through TOR-Dependent Nutrient Sensing. Cell Metab 14:403–414.

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8. Téfit MA, Leulier F. 2017. Lactobacillus plantarum favors the early emergence of fit and fertile adult Drosophila upon chronic undernutrition. J Exp Biol jeb.151522.

9. Corby-Harris V, Pontaroli AC, Shimkets LJ, Bennetzen JL, Habel KE, Promislow DEL. 2007. Geographical Distribution and Diversity of Bacteria Associated with Natural Populations of Drosophila melanogaster. Appl Environ Microbiol 73:3470–3479.

10. Cox CR, Gilmore MS. 2007. Native Microbial Colonization of Drosophila melanogaster and Its Use as a Model of Enterococcus faecalis Pathogenesis. Infect Immun 75:1565–1576.

11. Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A. 2011. Bacterial Communities of Diverse Drosophila Species: Ecological Context of a Host–Microbe Model System. PLoS Genet 7:e1002272.

12. Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA. 2013. Host Species and Environmental Effects on Bacterial Communities Associated with Drosophila in the Laboratory and in the Natural Environment. PLoS ONE 8:e70749.

13. Unckless RL, Rottschaefer SM, Lazzaro BP. 2015. The Complex Contributions of Genetics and Nutrition to Immunity in Drosophila melanogaster. PLoS Genet 11:e1005030.

14. Consortium THMP. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.

15. Chao A. 1984. Nonparametric estimation of the number of classes in a population. Scand J Stat 265–270.

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

16. Shannon CE, Weaver W, Blahut RE, Hajek B. 1949. The mathematical theory of communication. Bell Syst Tech J 117.

17. Simpson EH. 1949. Measurement of Diversity. Nature 163:688–688.

18. Ren C, Webster P, Finkel SE, Tower J. 2007. Increased Internal and External Bacterial Load during Drosophila Aging without Life-Span Trade-Off. Cell Metab 6:144–152.

19. Wong AC-N, Chaston JM, Douglas AE. 2013. The inconstant gut microbiota of Drosophila species revealed by 16S rRNA gene analysis. ISME J 7:1922–1932.

20. Blum JE, Fischer CN, Miles J, Handelsman J. 2013. Frequent Replenishment Sustains the Beneficial Microbiome of Drosophila melanogaster. mBio 4:e00860-13.

21. Fischer CN, Trautman EP, Crawford JM, Stabb EV, Handelsman J, Broderick NA. 2017. Metabolite exchange between microbiome members produces compounds that influence Drosophila behavior. eLife 6:e18855.

22. Brummel T, Ching A, Seroude L, Simon AF, Benzer S. 2004. Drosophila lifespan enhancement by exogenous bacteria. Proc Natl Acad Sci U S A 101:12974–12979.

23. Clark RI, Salazar A, Yamada R, Fitz-Gibbon S, Morselli M, Alcaraz J, Rana A, Rera M, Pellegrini M, Ja WW, Walker DW. 2015. Distinct Shifts in Microbiota Composition during Drosophila Aging Impair Intestinal Function and Drive Mortality. Cell Rep 12:1656–1667.

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

24. Wong AC-N, Luo Y, Jing X, Franzenburg S, Bost A, Douglas AE. 2015. The Host as the Driver of the Microbiota in the Gut and External Environment of Drosophila melanogaster. Appl Environ Microbiol 81:6232–6240.

25. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2010. Colloquium Paper: Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci 108:4516–4522.

26. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl Environ Microbiol 75:7537–7541.

27. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196.

28. Goslee S, Urban D. 2007. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. J Stat Softw 22:1–19.

29. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2017. vegan: Community Ecology Package.

30. Chen J. 2012. GUniFrac: Generalized UniFrac distances. R package version 1.0. http://CRAN.R-project.org.

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

31. Suzuki R, Shimodaira H. 2015. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 2.0-0. http://CRAN.R-project.org.

314

315 **Figures**

316 Figure 1

317 Rarefaction curves of 97% identity OTUs (A) for individual flies and a pool of 5 flies from the
318 same substrate (plum). The orange line shows the mean number of OTUs discovered for 31
319 individual fly samples with the shaded area indicating the standard deviation. The green line
320 represents a pool of 5 flies sampled from the same substrate. (B) for 13 pools of five flies
321 collected across substrates and locations. The green line represents the mean and the shaded area
322 the standard deviation.

323

324 Figure 2

325 Relative abundance of bacterial taxa as assessed by 16s rRNA gene sequences. *Wolbachia*
326 sequences were excluded. Bacterial genera of abundance <3% have been removed for clarity.

327

328 Figure 3

329 Pairwise Bray-Curtis distance for individual and pooled samples. P-value was computed using
330 the Mann-Whitney-U-test.

331

332 Figure 4

333 Principal Coordinates Analysis of Bray-Curtis distances.

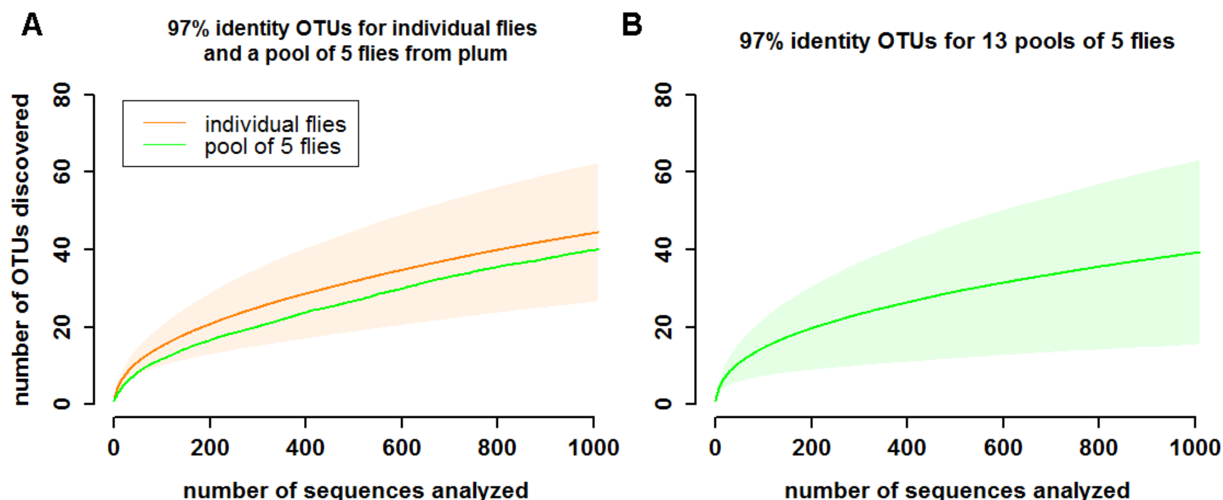
334

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

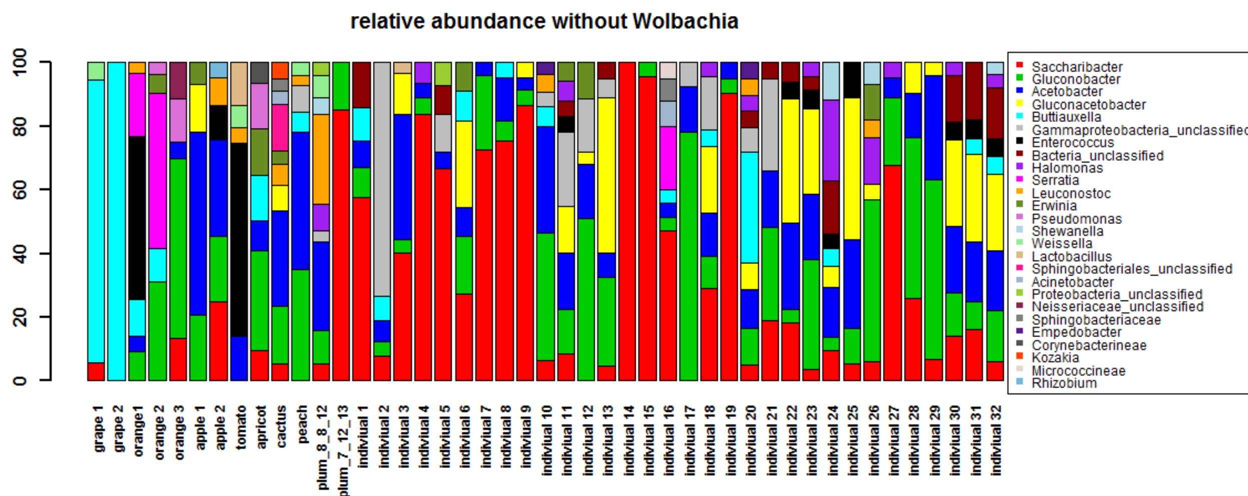
335 Figure 5. Hierarchical clustering of all samples based on Bray-Curtis dissimilarities. Values at
336 branches are AU (Approximately Unbiased) bootstrap support (31).

337

338 Figure 1



339 Figure 2

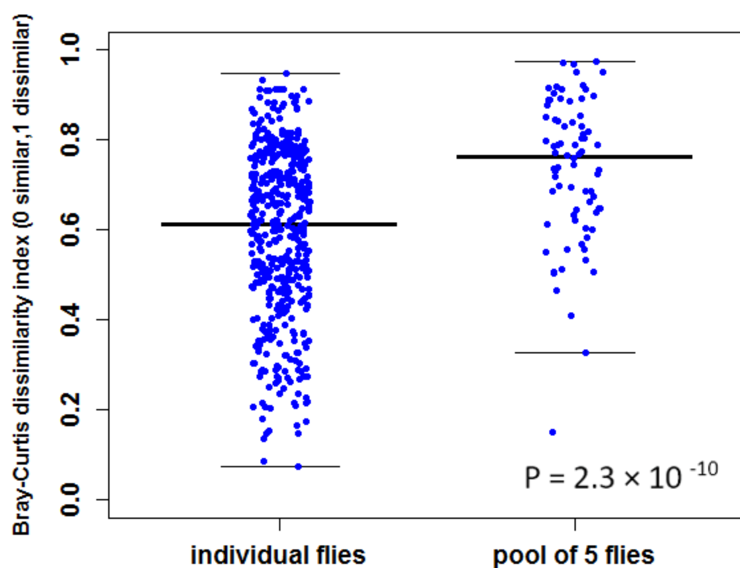


341

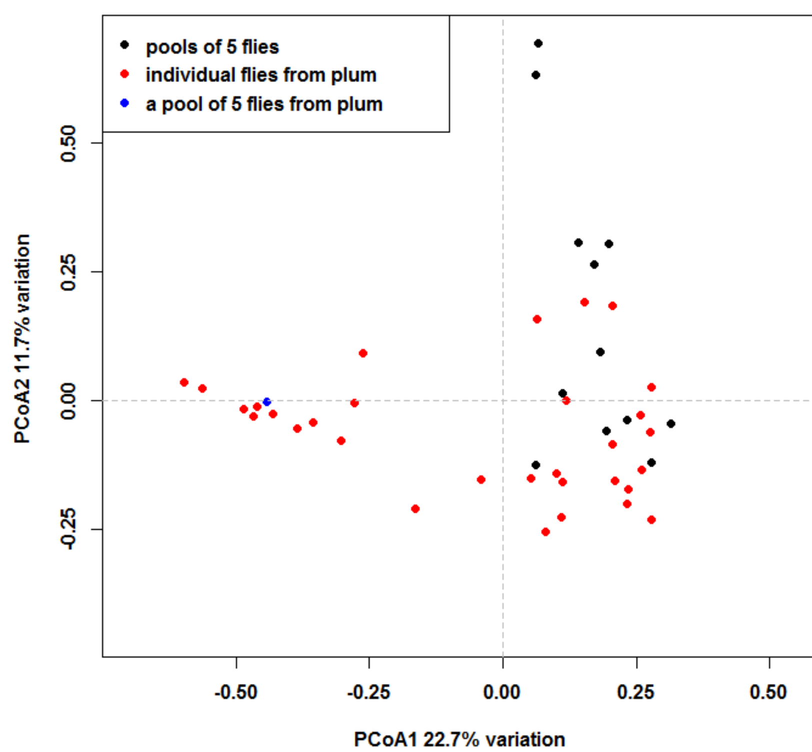
342

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

343 Figure 3

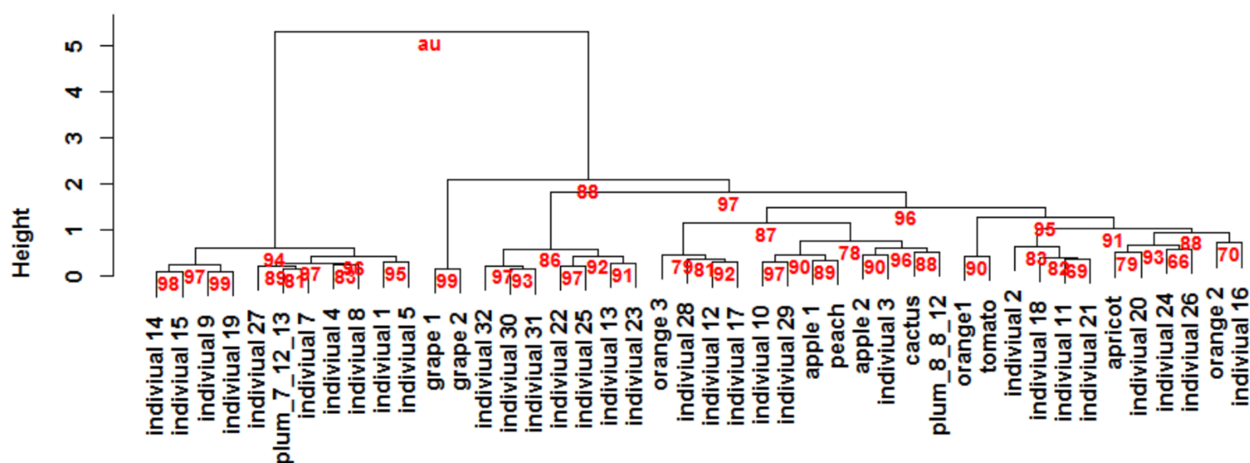


345 Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

347 Figure 5



349 **Tables**

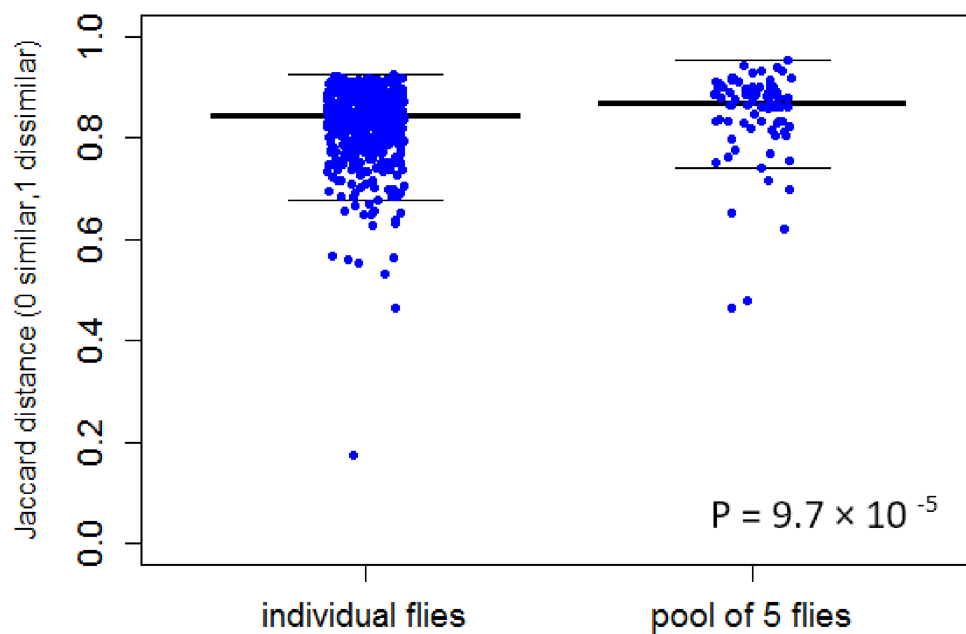
350 Table 1 Alpha diversity of individual and pooled samples

sample	#OTU	*P*	Chao	*P*	Shannon H	*P*	Simpson D	*P*	Shannon E	*P*	Simpson E	*P*
individual flies	44.4±17.8		75.9±32.0		1.94±0.62		0.29±0.20		0.51±0.13		0.11±0.04	
pools of 5 from plum	40	1	63.8	0.81	1.17	0.31	0.57	0.31	0.32	0.19	0.04	0.13
pools of 5	39.3±23.7	0.23	71.2±55.6	0.27	1.89±0.81	0.93	0.31±0.25	0.93	0.52±0.18	0.57	0.13±0.04	0.11

352
353
354
355
356
357
358
359
360
361

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

362 **Supplementary figures**

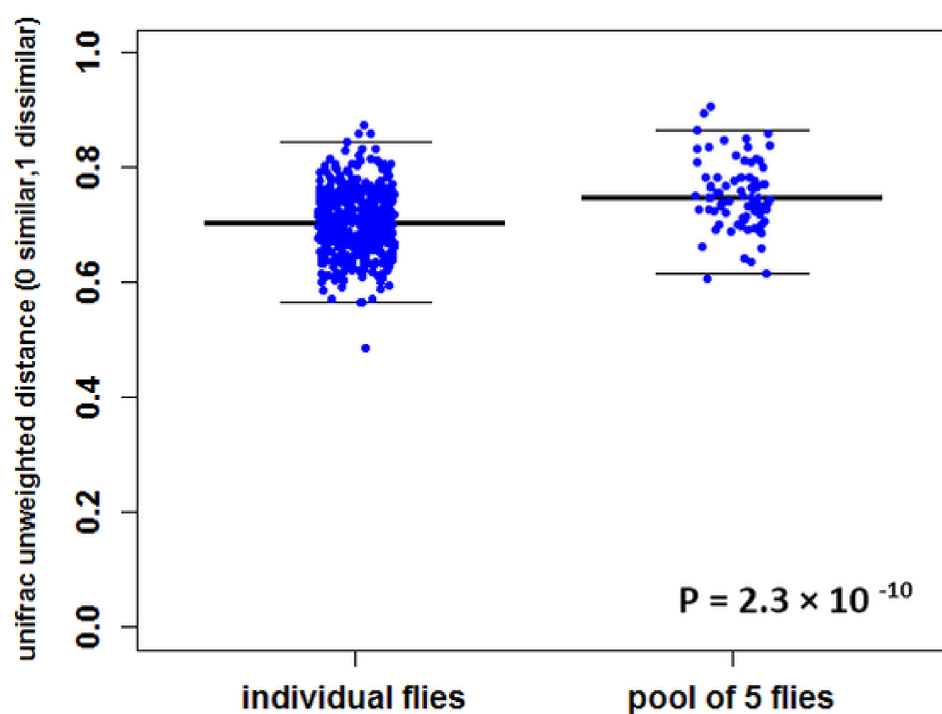


363 FigureS1

364 Pairwise Jaccard distances for individual and pooled samples . P-value was computed using

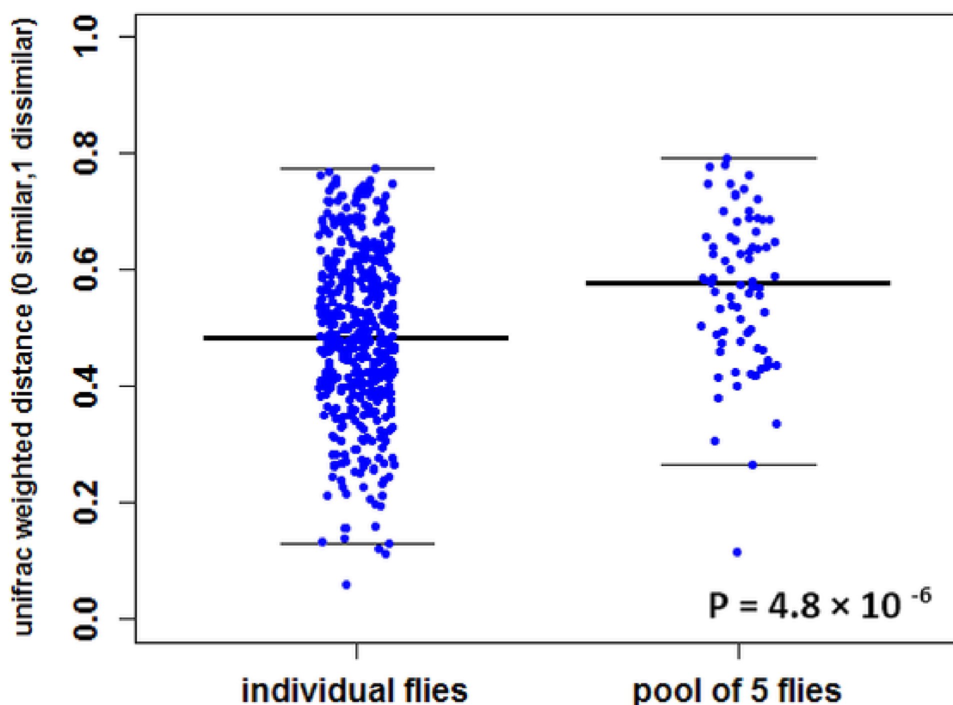
365 Mann-Whitney test.

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



366 Figure S2
367 Pairwise unweighted Unifrac distances for individual and pooled samples . P-value was
368 computed using Mann-Whitney test.

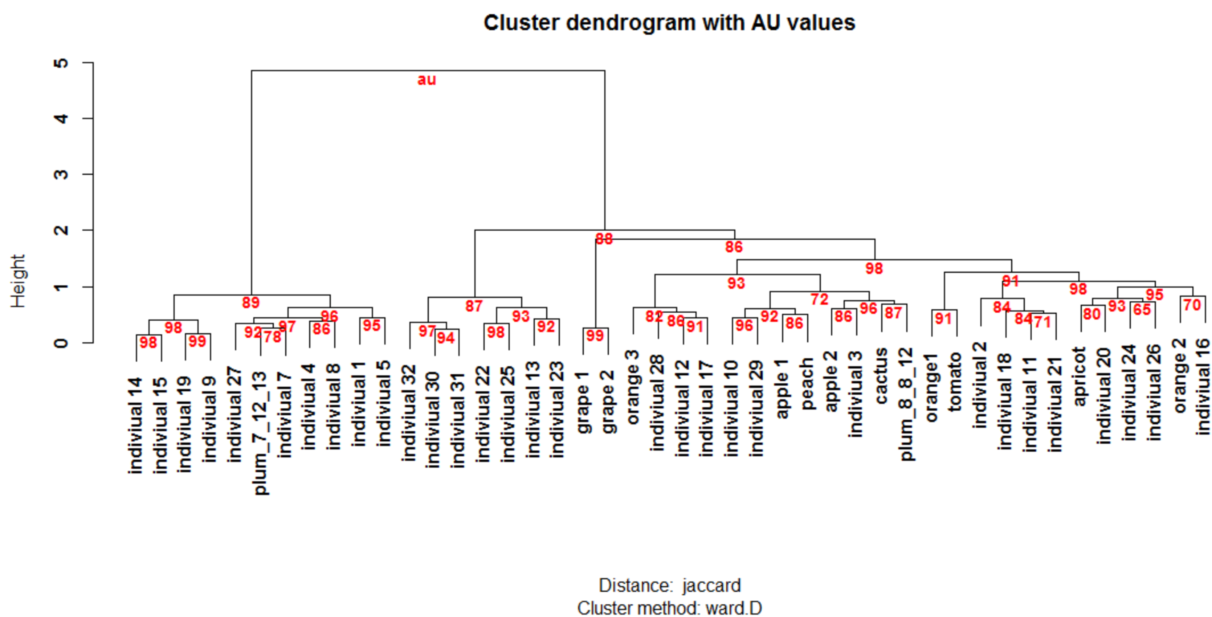
bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



369 Figure S3

370 Pairwise weighted Unifrac distances for individual and pooled samples. P-value was computed

371 using Mann-Whitney test.



373 Figure S4

374 Hierarchical clustering of all samples based on Jaccard distances. Values at branches are AU

375 (Approximately Unbiased) bootstrap support

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cluster dendrogram with AU values

Distance: Unifrac unweighted
Cluster method: Ward.D

376 Figure S5

377 Hierarchical clustering of all samples based on Unifrac unweighted distances. Values at branches

378 are AU (Approximately Unbiased) bootstrap support

Cluster dendrogram with AU values

Distance: Unifrac weighted
Cluster method: Ward.D

379 Figure S6

380 Hierarchical clustering of all samples based on Unifrac weighted distances. Values at branches

381 are AU (Approximately Unbiased) bootstrap support

382

383

384

385

386

387

388

bioRxiv preprint doi: https://doi.org/10.1101/126912; this version posted April 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

389 Table S1 list of all samples in the study
390
391
392

sample	date	location	coordinates
individual1_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual2_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual3_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual4_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual5_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual6_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual7_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual8_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual9_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual10_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual11_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual12_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual13_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual14_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual15_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual16_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual17_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual18_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual19_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual20_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual21_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual22_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual23_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual24_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual25_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual26_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual27_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual28_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual29_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual30_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual31_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
individual32_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
poolof5_plum_8_8_12	Aug 8 2012	Redwood City, CA	37.48256, -122.23427
poolof5_plum_7_12_13	July 12 2013	Redwood City, CA	37.48256, -122.23427
apple2	Nov 17 2012	Camino, CA	38.762123, -120.715358
apricot	July 27 2011	Guinda, CA	38.859975, -122.209494
grape1	Nov 15 2012	Acampo, CA	38.189685, -121.231378
cactus	Nov 15 2013	Stockton, CA	38.020093, -121.203840
grape2	Nov 15 2012	Lodi, CA	38.138774, -121.209475
tomatoes	July 27 2011	Capay, CA	38.708438, -122.075828
orange1	March 2010	Manteca, CA	37.79766, -121.13014
orange2	Feb 2010	Escalon, CA	37.79702, -120.99369
orange3	March 2010	Knightsen, CA	37.97451, -121.65161
apple1	August 2010	Johnston, RI	41.81636, -71.47705
peach	August 2010	Johnston, RI	41.81636, -71.47705