

1

2 **Supporting information**

3 **Evolution of trade-offs across environments following**
4 **experimental evolution of the generalist *Drosophila suzukii***
5 **to different fruit media**

6

7

8

9 **Appendix S1: Performance of the F -test for local**
10 **adaptation (Blanquart et al., 2013) when applied to non-**
11 **normally distributed traits**

12
13 **Motivation**

14 The F -test proposed by Blanquart et al. (2013) to test for local adaptation assumes that the
15 fitness related trait under study is normally and independently distributed with equal
16 variances. We here evaluated the power and robustness of the test when applied to log-
17 transformed non-normally distributed count data simulated under a scenario mimicking our
18 experimental setup (i.e., similar number of populations, habitats and individuals) and with
19 different levels of local adaptation.

20
21 **Methods**

22 The count data Y_{ijk} for the trait (i.e., the number of eggs laid or number of emerged adults
23 that emerged) observed in population i on the host plant j for individual k was simulated as
24 follows:

25
26
$$Y_{ijk} \sim \text{Poisson}(\lambda_{ijk}) \quad (1),$$

27
28 where $\lambda_{ijk} = \exp(\mu + s \cdot I_{ij} + a_i + b_j + c_{ij} + \varepsilon_{ijk})$.

29 μ is a constant term corresponding to the overall mean counts measured in a log scale, $a_i \sim$
30 $N(0, \sigma_a^2)$ is the population i effect, $b_j \sim N(0, \sigma_b^2)$ is the habitat j effect, $c_{ij} \sim N(0, \sigma_c^2)$ is the
31 population by habitat interaction effect, and $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ is an error term that introduces
32 overdispersion among individuals sampled from population i sampled in habitat j . The binary
33 auxiliary variable I_{ij} indicates whether the population/habitat combination is allopatric ($I_{ij} = 0$

34 when $i \neq j$) or sympatric ($I_{ij} = 1$ when $i = j$). s is the magnitude of the fitness advantage of
35 being in sympatry (relative to allopatry).

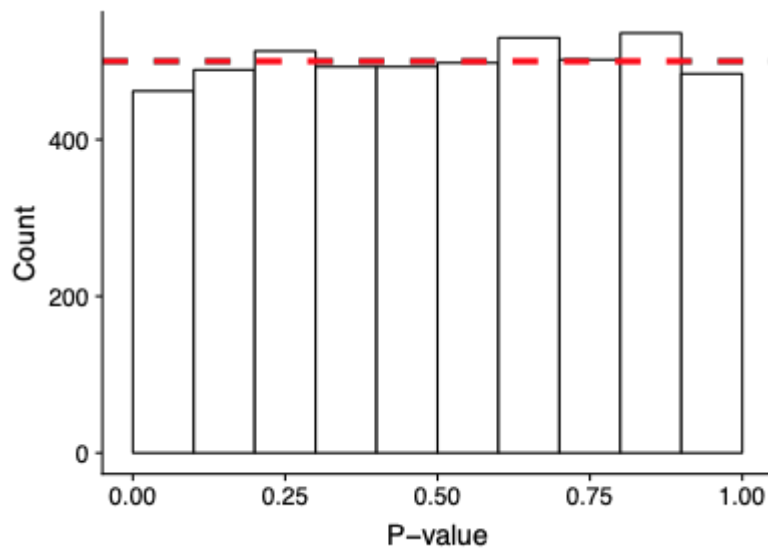
36 To mimic our experimental design, we simulated $I = 11$ populations and $J = 3$ habitats
37 and the per population/habitat combination sample size was set by default to $n=30$. Similarly,
38 in our experiment, we estimated $\mu = 3.053$, $\sigma_a = 0.165$, $\sigma_b = 0.0636$; $\sigma_c = 0.345$; $\sigma_e = 0.868$ and s
39 $= 0.295$ for the number of adults emerged during the final phenotyping. We thus considered
40 these estimations, as default values, for the corresponding simulation parameter values. We
41 also simulated a range of values for four parameters that were each modified one at a time
42 (i.e., other simulation parameter being set to their default values): (i) $s = 0, 0.1, 0.2, 0.3, 0.4$
43 or 0.5 to evaluate the power of the method as a function of the magnitude of local adaptation
44 with; (ii) $\sigma_e = 0, 1, 2$ or 3 to evaluate the impact of overdispersion (from absent to three times
45 as high as the one we observed); (iii) $\mu = 0, 1, 2, 3, 4$ or 5 to evaluate the effect of the overall
46 mean count (from $n = 1$ to ca. 150 on a natural scale); and (iv) the sample size per
47 population/habitat combination was set to $2, 10, 20$ or 30 to evaluate the effect of the number
48 of replicates per combination.

49 For each simulation scenario, $5,000$ data sets were generated and analyzed as
50 described in the main text. To compare the performance of the model for different parameter
51 values, we used the R package PRROC (Grau et al. 2015) to compute for various p-value
52 thresholds the (i) true positive rates (TPR) or power which corresponds to the proportion of
53 data sets with $s > 0$ among the ones declared significant for local adaptation); and (ii) false
54 positive rates (FPR) which corresponds to the proportion of data sets with $s = 0$ among the
55 ones declared non-significant for local adaptation. From these estimates, standard receiver
56 operating curves (ROC) plotting TPR against FPR could then be drawn and the area under
57 the ROC curve (AUC) computed. Note that $AUC = 1$ corresponds to an optimal classifier.

58

59 **Results**

60 Figure A1 shows that the distribution of p-values obtained after analyzing data sets simulated
61 with $s = 0$ (no local adaptation) is uniform. The Blanquart F test applied to log-transformed
62 count data is thus well calibrated under the null hypothesis of no local adaptation, at least
63 under the conditions of our experimental set-up.



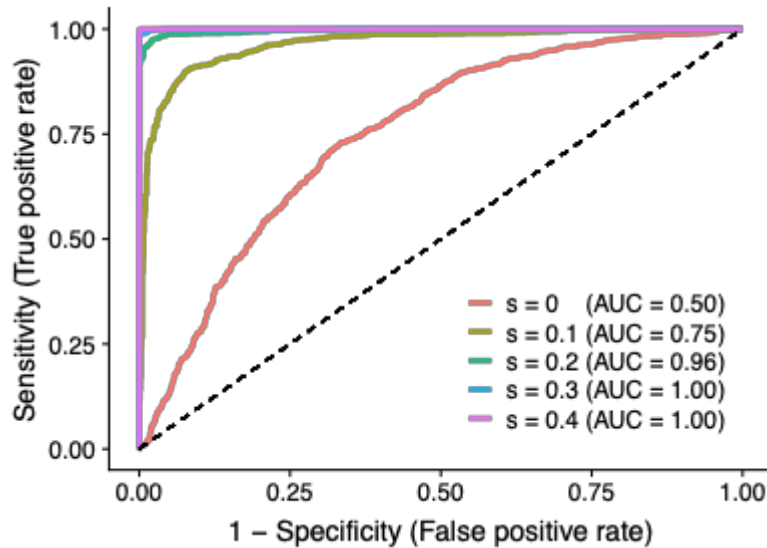
64

65 **Figure A1. Distribution of p-values calculated using Blanquart et al. (2013)'s F-test**
66 **processed after a log transformation of data simulated following our experimental setup**
67 **conditions and with a local adaptation s value equals to 0.**

68

69 The ROC curves obtained from the analysis of data sets simulated with varying
70 magnitude of local adaptation (from $s = 0.1$ to $s = 0.5$) are plotted on Figure A2. As expected,
71 the performance of the model improved with s , the ROC-AUC being above > 0.99 for $s \geq 0.3$,
72 the latter value being similar to the one we estimated on real data in our experiment.

73



74

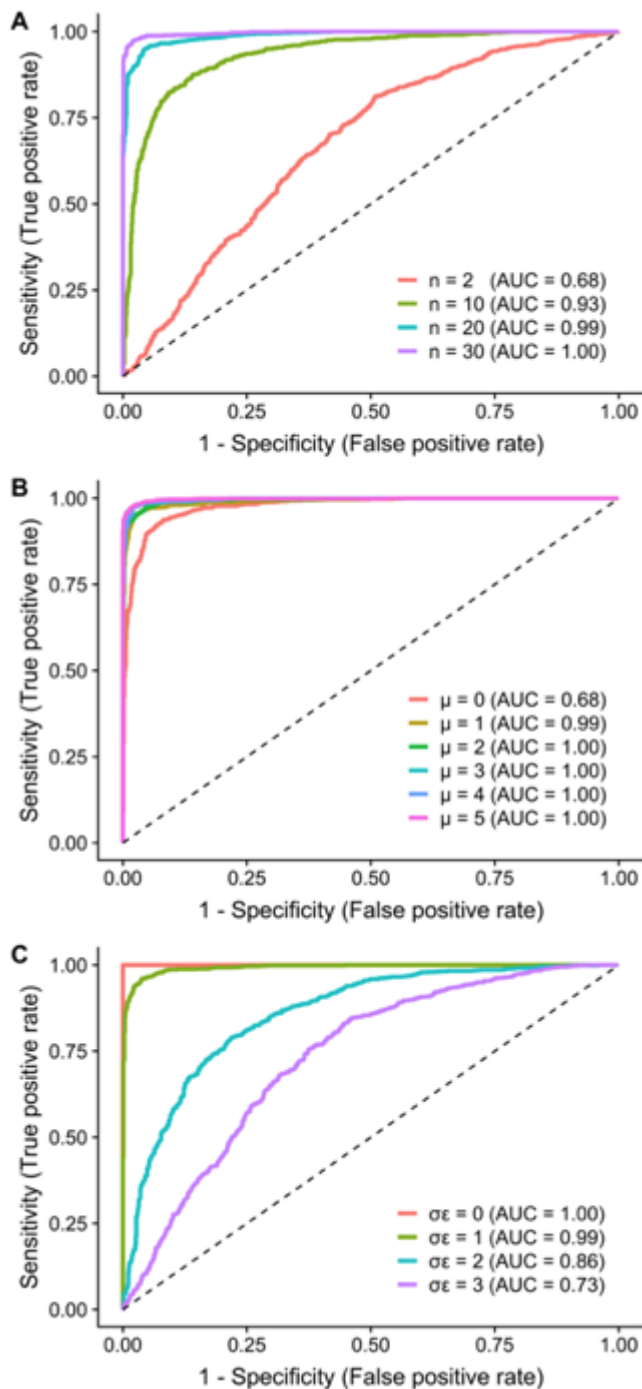
75 **Figure A2. Evaluation of the performance of the method for varying magnitude of local**
 76 **adaptation (measured by s).**

77 Count data sets were simulated with five different values of s ranging from $s = 0.1$ to 0.5
 78 (other simulation parameters being set to their default value). In each case, 5,000 data sets
 79 were then analyzed with the F -test by Blanquart et al. (2013) after log-transformation to
 80 estimate TPR and FPR (averaged over all the data sets). The corresponding ROC curves are
 81 plotted and ROC-AUC are given in parentheses in the figure legend.

82

83 Figure A3 gives the ROC curves obtained from the analyses of simulated data sets
 84 when varying i) the population by habitat sample size ($n = 2$ to $n = 30$); ii) overall mean count
 85 (from $\mu = 0$ to $\mu = 5$ in log-scale); and iii) overdispersion (from $\sigma_\epsilon = 0$ to $\sigma_\epsilon = 3$). The
 86 approach was mainly found sensitive to a smaller number of individuals ($n < 20$) and to
 87 higher overdispersion ($\sigma_\epsilon > 2$) than the ones corresponding to our experiment. Interestingly,
 88 the overall average mean count had only minor effect on the performance of the method.

89



90

91 **Figure A3. Evaluation of the performance of the method when varying (A) the**
 92 **population by habitat sample size (measured by n), (B) the overall mean count**
 93 **(measured by μ) and (C) the overdispersion of data (measured by σ_ε).**

94 Count data sets were simulated with four different values of n ranging from n = 2 to 30 (other
 95 simulation parameters being set to their default value), six different values of μ ranging from
 96 n = 0 to 5 (other simulation parameters being set to their default value), and four different
 97 values of σ_ε ranging from n = 0 to 3 (other simulation parameters being set to their default
 98 value). In each case, 5,000 data sets were then analyzed with the *F*-test of Blanquart et al.
 99 (2013) after log-transformation to estimate TPR and FPR (averaged over all the data sets).

100 The corresponding ROC curves are plotted and ROC-AUC are given in parentheses in the
101 figure legend.

102
103

104 **General conclusion**

105 From the power analysis results given above, we conclude that the F-test proposed by
106 Blanquart et al. (2013) to detect local adaptation can be applied to log-transformed count data
107 provided that the population by habitat sample size is high enough (> 20) and overdispersion
108 of the data remained limited, as previously discussed in O'Hara et Kotze (2010).

109

110 **References cited**

111 Blanquart, F., O. Kaltz, S. L. Nuismer, and S. Gandon. 2013. A practical guide to
112 measuring local adaptation. *Ecol. Lett.* 16:1195–1205.

113 Grau, J., I. Grosse, and J. Keilwagen. 2015. PRROC: computing and visualizing
114 precision-recall and receiver operating characteristic curves in R. *Bioinformatics* 31:2595–
115 2597.

116 O'Hara, R., and D. Kotze. 2010. Do not log-transform count data. *Methods Ecol.*
117 *Evol.* 1:118–122.

118

119

120

121

122

123 **Supplementary Table S1**

124 **Table S1. Composition for 18 factors of the three fruit media (cherry, cranberry and**
 125 **strawberry) used in the experimental evolution (data from Olazuaga et al. 2019).** The
 126 ash is the measure of residues obtained after heating so that water and organic materials such
 127 as fat and protein are removed. The measure of total sugars is the sum of the measures of
 128 fructose, glucose, saccharose, maltose and lactose. The measure of carbohydrates is the sum
 129 of measures of these sugars (total sugar) and all the other carbohydrates (e.g., amidon or
 130 polyols). Statistical analyses are based on the *F*-test proposed by Blanquart et al. (2013) using
 131 equation 3.

Variables	Cherry	Cran- berry	Straw- berry	<i>P</i> -values	Corrected <i>P</i> -values	Units
Moisture content	78.3	92.9	91.5	0.92	0.92	g/100g
Ash	0.6	0.1	0.32	0.77	0.92	g/100g
Protein	0.7	0	0.5	0.54	0.92	g/100g
Lipid	0	0	0	NA	NA	g/100g
Fiber	1.1	1.5	0.8	0.13	0.6	g/100g
Carbohydrate	19.3	5.5	6.9	0.92	0.92	g/100g
Energy	349	106	132	0.92	0.92	kJ/100g
Fructose	4.8	2.2	2.7	0.9	0.92	g/100g
Glucose	4.6	1.6	1.9	0.92	0.92	g/100g
Saccharose	0	0	0	NA	NA	g/100g
Total sugars	9.4	3.8	4.6	0.91	0.92	g/100g
Sodium: Na	11.8	3.7	10.5	0.43	0.92	mg/100g
Sodium chloride: NaCl	0.03	0	0.03	0.33	0.92	g/100g
Zinc: Zn	0.69	0.59	1.3	0.16	0.6	mg/100g
Calcium: Ca	230	57.4	90	0.9	0.92	mg/100g
Magnesium: Mg	139	24.9	91.9	0.65	0.92	mg/100g
Iron: Fe	3.6	3.1	7.3	0.17	0.6	mg/100g
Phosphore: P	236	0	128	0.69	0.92	mg/100g

132

133 **Supplementary Figures S1 to S4**



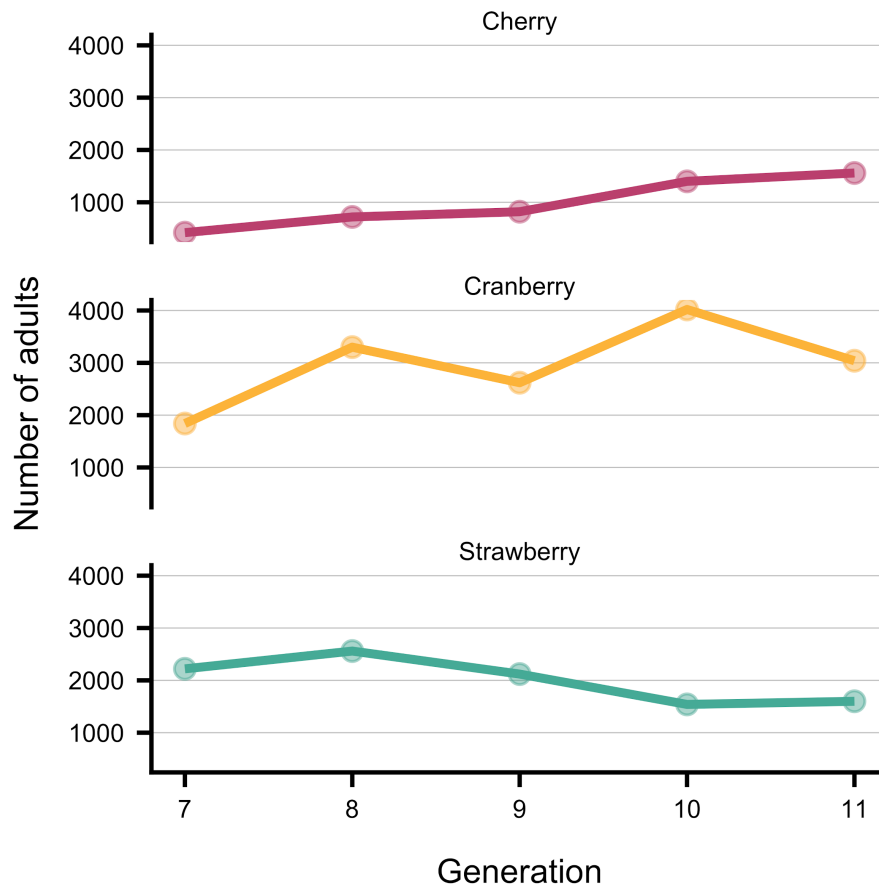
134

135

136 **Figure S1. Principal component analysis of the chemical composition of the eight fruit**
 137 **purees.** A principal component analysis (PCA) of fruit composition facilitated the selection
 138 of fruits that differed in carbohydrate, lipid, protein and mineral composition. The nutritional
 139 composition of the 12 purees of fruits was determined by a private company (Cereco
 140 Laboratoire Sud, France), providing estimates of energy, carbohydrate, protein, lipids, fibers,
 141 sodium, zinc, calcium, magnesium, iron and phosphorus, among others (see also Table S1).
 142 PCA axes 1 and 3 are shown in the panel A and PCA axes 2 and 3 are shown in panel B.

143

144



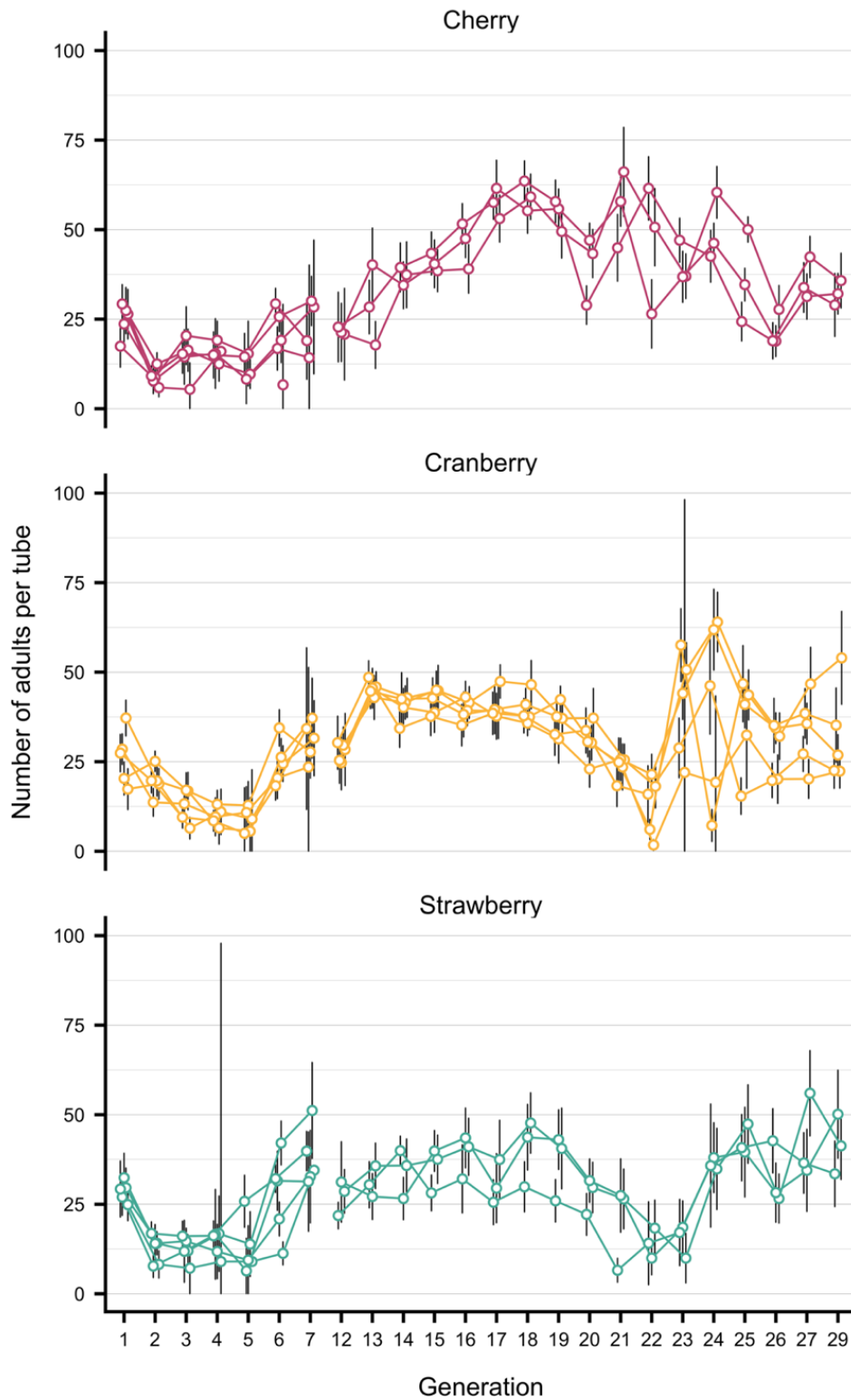
145

146 **Figure S2. Sizes of the three pooled populations in cherry, cranberry and strawberry**
 147 **media for generations 7 to 11.**

148

149

150



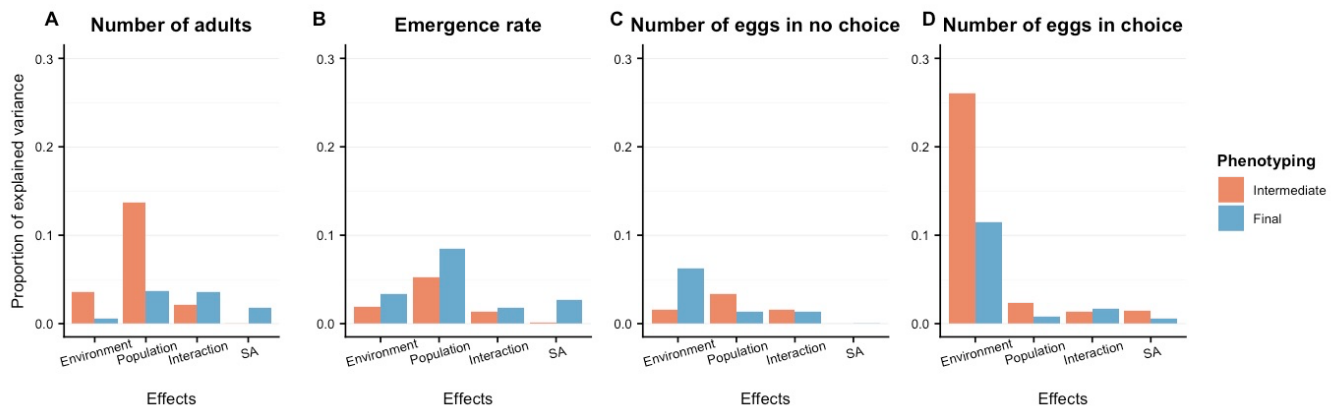
151

152 **Figure S3. Evolution of the number of adults produced per tube at each generation on**
 153 **the evolution fruit for the different populations evolved on cherry, cranberry and**
 154 **strawberry.**

155

156

157



158

159 **Figure S4. Proportion of variance explained by different effects on performance (A and**

160 **B), oviposition stimulation (C) and preference (D) during the intermediate and final**

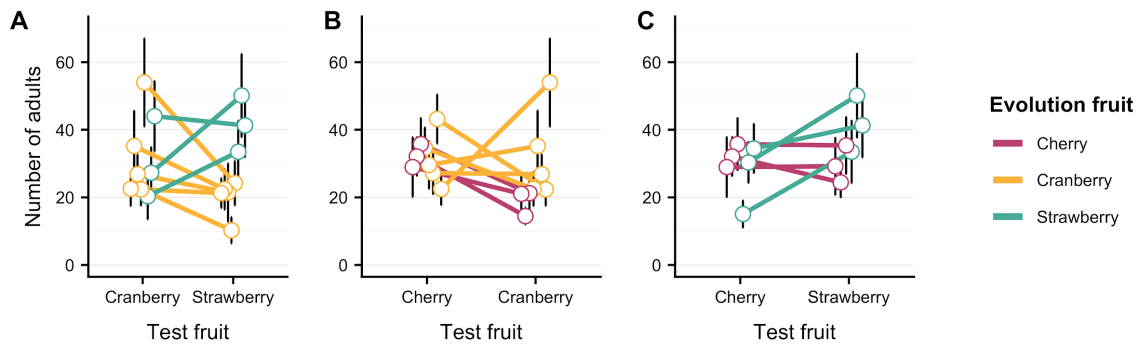
161 **phenotyping steps. “Interaction” refers to the interaction between test fruit and evolution**

162 **fruit (see eq. 3).**

163

164

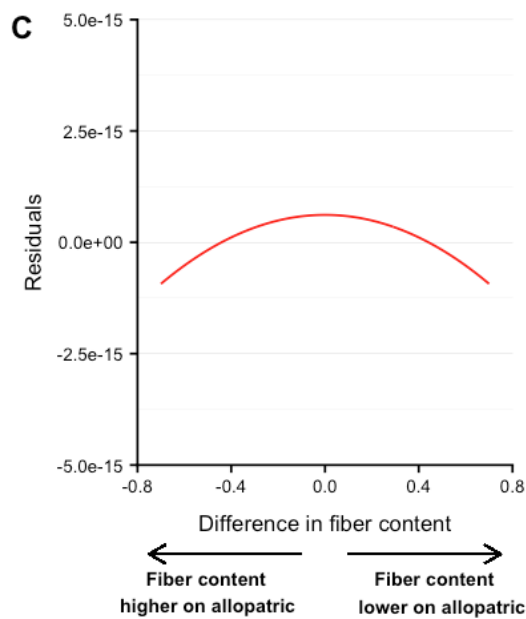
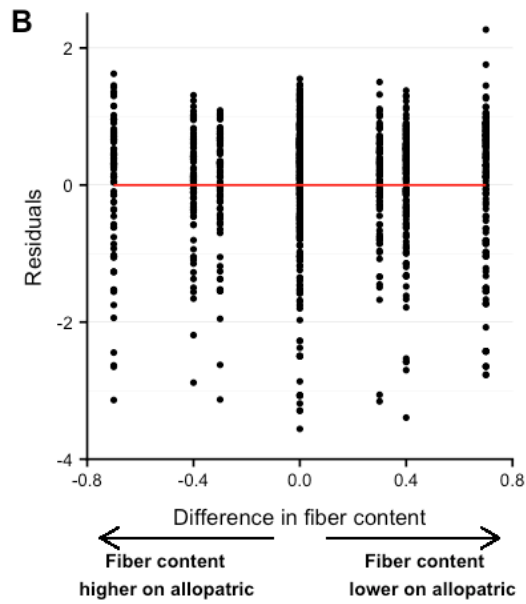
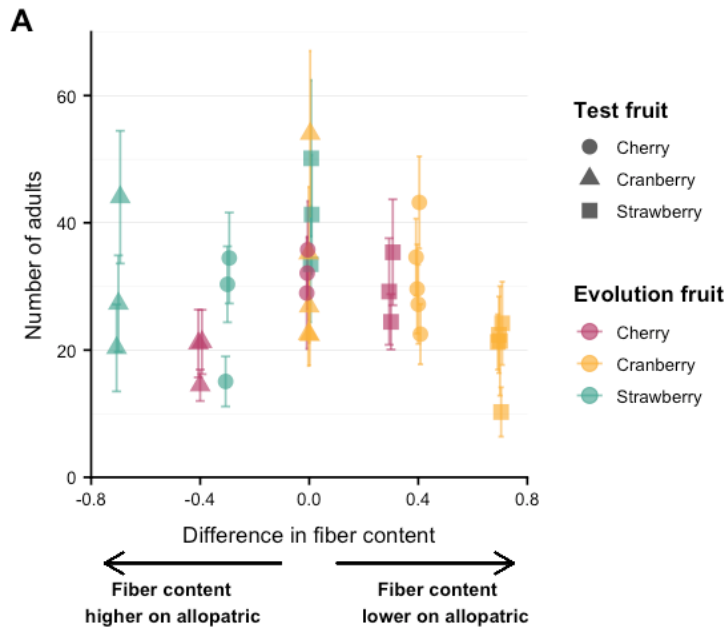
165



166

167

168 **Figure S5. Relationship between the performance of populations during the final**
169 **phenotyping step measured in one fruit and in another fruit, for the populations evolved**
170 **in one of the two fruits.** Performance is measured as the number of adults produced in a
171 tube. Cherry vs cranberry (A), cranberry vs strawberry (B) and strawberry vs cherry (C). The
172 dots represents the joint mean with 95% CI error-bars for each population.



174 **Figure S6. Relationship between the number of emerging adults and the squared**
175 **difference in fiber content between the evolution fruit and the test fruit during the final**
176 **phenotyping step.** Observed data are shown in panel A where each symbol represents the
177 joint mean of observed data of each population with 95% CI error-bars. Fitted data are shown
178 as a red line in panel B and C (two different y-axis scales). In panel B, circles represent the
179 residuals from a model that only fits population, environment and interaction effects to the
180 number of adults. The red line represents the fitted line based on a model that fits the squared
181 effect of the difference in fiber content, Δ_{ij} , to these residuals.

182
183
184
185
186
187
188
189
190
191