

# Testing pleiotropy vs. separate QTL in multiparental populations

Frederick J. Boehm<sup>\*</sup>, Elissa J. Chesler<sup>†</sup>, Brian S. Yandell<sup>\*,‡</sup> and Karl W. Broman<sup>§,1</sup>

<sup>\*</sup>Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin 53706, <sup>†</sup>The Jackson Laboratory, Bar Harbor, Maine 04609, <sup>‡</sup>Department of Horticulture, University of Wisconsin-Madison, Madison, Wisconsin 53706, <sup>§</sup>Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin 53706

## ABSTRACT

The high mapping resolution of multiparental populations, combined with technology to measure tens of thousands of phenotypes, presents a need for quantitative methods to enhance understanding of the genetic architecture of complex traits. When multiple traits map to a common genomic region, knowledge of the number of distinct loci provides important insight into the underlying mechanism and can assist planning for subsequent experiments. We extend the method of [Jiang and Zeng \(1995\)](#), for testing pleiotropy with a pair of traits, to the case of more than two alleles. We also incorporate polygenic random effects to account for population structure. We use a parametric bootstrap to determine statistical significance. We apply our methods to a behavioral genetics data set from Diversity Outbred mice. Our methods have been incorporated into the R package `qtl2pleio`.

**KEYWORDS** Quantitative trait locus; pleiotropy; multivariate analysis; linear mixed effects models; systems genetics

Complex trait studies in multiparental populations present new challenges in statistical methods and data analysis. Among these is the development of strategies for multivariate trait analysis. The joint analysis of two or more traits allows one to address additional questions, such as whether two traits share a single pleiotropic locus.

Previous research addressed the question of pleiotropy vs. separate QTL in two-parent crosses. [Jiang and Zeng \(1995\)](#) developed a likelihood ratio test for pleiotropy vs. separate QTL for a pair of traits. Their approach assumed that each trait was affected by a single QTL. Under the null hypothesis, the two traits were affected by a common QTL, and under the alternative hypothesis the two traits were affected by distinct QTL. [Knott and Haley \(2000\)](#) used linear regression to develop a fast approximation to the test of [Jiang and Zeng \(1995\)](#), while [Tian \*et al.\* \(2016\)](#) used the methods from [Knott and Haley \(2000\)](#) to dissect QTL hotspots in a  $F_2$  population.

Multiparental populations, such as the Diversity Outbred (DO) mouse population ([Churchill \*et al.\* 2012](#)), enable high-precision mapping of complex traits ([de Koning and McIntyre 2014](#)). The DO mouse population began with progenitors of the Collaborative Cross (CC) mice ([Churchill \*et al.\* 2004](#)) Each DO mouse is a highly heterozygous genetic mosaic of alleles from the eight CC founder lines. Random matings among non-siblings have maintained the DO population for more than 23 generations ([Chesler \*et al.\* 2016](#)).

Several limitations of previous pleiotropy vs. separate QTL tests prevent their direct application in multiparental populations. First, multiparental populations can have complex patterns of relatedness among subjects, and failure to account for these patterns of relatedness may lead to spurious results ([Yang \*et al.\* 2014](#)). Second, previous tests allowed for only two founder lines ([Jiang and Zeng](#)

<sup>1</sup>Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, 2126 Genetics-Biotechnology Center, 425 Henry Mall, Madison, WI 53706. E-mail: [broman@wisc.edu](mailto:broman@wisc.edu)

42 1995). Finally, [Jiang and Zeng \(1995\)](#) assumed that the null distribution of the test statistic follows a  
43 chi-square distribution.

44 We developed a pleiotropy vs. separate QTL test for two traits in multiparental populations. Our  
45 test builds on research that [Jiang and Zeng \(1995\)](#), [Knott and Haley \(2000\)](#), [Tian et al. \(2016\)](#), and  
46 [Zhou and Stephens \(2014\)](#) initiated. Our innovations include the accommodation of  $k$  founder alle-  
47 les per locus (compared to the traditional two founder alleles per locus) and the incorporation of  
48 multivariate polygenic random effects to account for relatedness. Furthermore, we implemented  
49 a parametric bootstrap test to assess statistical significance ([Efron 1979](#); [Tian et al. 2016](#)). We focus  
50 on the case that two traits are measured in the same set of subjects (Design I in the notation of  
51 [Jiang and Zeng \(1995\)](#)).

52 Below, we describe our likelihood ratio test for pleiotropy vs. separate QTL. In simulation studies,  
53 we find that it is slightly conservative, and that it has power to detect two separate loci when the  
54 univariate LOD peaks are strong. We further illustrate our approach with an application to data on  
55 a pair of behavior traits in a population of 261 DO mice ([Logan et al. 2013](#); [Recla et al. 2014](#)).

## 56 **Methods**

57 Our strategy involves first identifying two traits that map to a common genomic region. We then per-  
58 form a two-dimensional, two-QTL scan over the genomic region, with each trait affected by one QTL  
59 of varying position. We identify the QTL position that maximizes the likelihood under pleiotropy  
60 (that is, along the diagonal where the two QTL are at a common location), and the ordered pair of  
61 positions that maximizes the likelihood under the model where the two QTL are allowed to be dis-  
62 tinct. The logarithm of the ratio of the two likelihoods is our test statistic. We determine statistical  
63 significance with a parametric bootstrap.

### 64 **Data structures**

65 The data consist of three objects. The first is an  $n$  by  $k$  by  $m$  array of allele probabilities for  $n$  subjects  
66 with  $k$  alleles and  $m$  marker positions on a single chromosome [derived from the observed SNP  
67 genotype data by a hidden Markov model; see [Broman et al. \(2019\)](#)]. The second object is an  $n$  by 2  
68 matrix of phenotype values. Each column is a phenotype and each row is a subject. The third object  
69 is an  $n$  by  $c$  matrix of covariates, where each row is a subject and each column is a covariate.

70 One additional object is the genotype-derived kinship matrix, which is used in the linear mixed  
71 model to account for population structure. We are focusing on a defined genomic interval, and  
72 we prefer to use a kinship matrix derived by the “leave one chromosome out” (LOCO) method  
73 ([Yang et al. 2014](#)), in which the kinship matrix is derived from the genotypes for all chromosomes  
74 except the chromosome under test.

### 75 **Statistical Models**

76 Focusing on a pair of traits and a particular genomic region of interest, the next step is a two-  
77 dimensional, two-QTL scan ([Jiang and Zeng 1995](#)). We consider two QTL with each affecting a  
78 different trait, and consider all possible pairs of locations for the two QTL. For each pair of posi-  
79 tions, we fit the multivariate linear mixed effects model defined in Equation 1. Note that we have  
80 assumed an additive genetic model throughout our analyses, but extensions to design matrices that  
81 include dominance are straightforward.

$$vec(Y) = Xvec(B) + vec(G) + vec(E) \quad (1)$$

82 where  $Y$  is the  $n$  by 2 matrix of phenotypes values;  $X$  is a  $2n$  by  $2(k + c)$  matrix that contains the  $k$   
 83 allele probabilities for the two QTL positions and the  $c$  covariates in diagonal blocks;  $B$  is a  $(k + c)$   
 84 by 2 matrix of allele effects and covariate effects;  $G$  is a  $n$  by 2 matrix of random effects; and  $E$  is a  
 85  $n$  by 2 matrix of random errors.  $n$  is the number of mice. The ‘vec’ operator stacks columns from  
 86 a matrix into a single vector. For example, a 2 by 2 matrix inputted to ‘vec’ results in a vector with  
 87 length 4. Its first two entries are the matrix’s first column, while the third and fourth entries are the  
 88 matrix’s second column.

89 We also impose distributional assumptions on  $G$  and  $E$ :

$$G \sim MN_{nx2}(0, K, V_g) \quad (2)$$

90 and

$$E \sim MN_{nx2}(0, I, V_e) \quad (3)$$

91 where  $MN_{nx2}(0, V_r, V_c)$  denotes the matrix-variate ( $n$  by 2) normal distribution with mean being the  
 92  $n$  by 2 matrix with all zero entries and row covariance  $V_r$  and column covariance  $V_c$ . We assume  
 93 that  $G$  and  $E$  are independent.

#### 94 **Parameter inference and log likelihood calculation**

95 Inference for parameters in multivariate linear mixed effects models is notoriously difficult and  
 96 can be computationally intense (Meyer 1989, 1991). Thus, we estimate  $V_g$  and  $V_e$  under the null  
 97 hypothesis of no QTL, and then take them as fixed and known in our two-dimensional, two-QTL  
 98 genome scan. We use restricted maximum likelihood methods to fit the model:

$$vec(Y) = X_0 vec(B) + vec(G) + vec(E) \quad (4)$$

99 where  $X_0$  is a  $2n$  by  $2(c + 1)$  matrix whose first column of each diagonal block in  $X_0$  has all entries  
 100 equal to one (for an intercept); the remaining columns are the covariates.

101 We draw on our R implementation (Boehm 2018) of the GEMMA algorithm for fitting a multivari-  
 102 ate linear mixed effects model with expectation-maximization (Zhou and Stephens 2014). We use  
 103 restricted maximum likelihood fits for the variance components  $V_g$  and  $V_e$  in subsequent calcula-  
 104 tions of the generalized least squares solution  $\hat{B}$ .

$$\hat{B} = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} vec(Y) \quad (5)$$

105 where

$$\hat{\Sigma} = \hat{V}_g \otimes K + \hat{V}_e \otimes I_n \quad (6)$$

106 where  $\otimes$  denotes the Kronecker product,  $K$  is the kinship matrix, and  $I_n$  is a  $n$  by  $n$  identity matrix.  
 107 We then calculate the log likelihood for a normal distribution with mean  $X vec(\hat{B})$  and covariance  $\hat{\Sigma}$   
 108 that depends on our estimates of  $V_g$  and  $V_e$  (Equation 6).

#### 109 **Pleiotropy vs. separate QTL hypothesis testing framework**

110 Our test applies to two traits considered simultaneously. Below,  $\lambda_1$  and  $\lambda_2$  denote putative locus  
 111 positions for traits one and two. We quantitatively state the competing hypotheses for our test as:

$$\begin{aligned} H_0 : \lambda_1 &= \lambda_2 \\ H_A : \lambda_1 &\neq \lambda_2 \end{aligned} \quad (7)$$

112 Our likelihood ratio test statistic is:

$$\text{LOD} = \log_{10} \left[ \frac{\max_{\lambda_1, \lambda_2} L(B, \Sigma, \lambda_1, \lambda_2)}{\max_{\lambda} L(B, \Sigma, \lambda, \lambda)} \right] \quad (8)$$

113 where  $L$  is the likelihood for fixed QTL positions, maximized over all other parameters. The denom-  
114 inator concerns the likelihood for the null hypothesis of pleiotropy, where  $\lambda = \lambda_1 = \lambda_2$ .

### 115 **Visualizing profile LOD traces**

116 The output of the above analysis is a two-dimensional  $\log_{10}$  likelihood surface. To visualize these  
117 results, we followed an innovation of [Zeng et al. \(2000\)](#) and [Tian et al. \(2016\)](#), and plot three traces:  
118 the results along the diagonal (corresponding to the null hypothesis of pleiotropy), and then the  
119 profiles derived by fixing one QTL's position and maximizing over the other QTL's position.

120 We define the LOD score for our test:

$$\text{LOD}(\lambda_1, \lambda_2) = l_{10}(\lambda_1, \lambda_2) - \max l_{10}(\lambda, \lambda) \quad (9)$$

121 where  $l_{10}$  denotes  $\log_{10}$  likelihood.

122 We follow [Zeng et al. \(2000\)](#) and [Tian et al. \(2016\)](#) in defining profile LOD by the equation

$$\text{profile LOD}_1(\lambda_1) = \max_{\lambda_2} \text{LOD}(\lambda_1, \lambda_2) \quad (10)$$

123 We define profile  $\text{LOD}_2(\lambda_2)$  analogously. The profile  $\text{LOD}_1$  and profile  $\text{LOD}_2$  traces have the same  
124 maximum value, which is non-negative and gives the overall LOD test statistic.

125 We construct the pleiotropy trace by calculating the log-likelihoods for the pleiotropic models at  
126 every position.

$$\text{LOD}_p(\lambda) = l_{10}(\lambda, \lambda) - \max l_{10}(\lambda, \lambda) \quad (11)$$

127 By definition, the maximum value for this pleiotropy trace is zero.

### 128 **Bootstrap for test statistic calibration**

129 We use a parametric bootstrap to determine statistical significance ([Efron 1979](#)). While [Jiang and Zeng](#)  
130 [\(1995\)](#) used quantiles of a chi-squared distribution to determine p-values, this does not account  
131 for the two-dimensional search over QTL positions. We follow the approach of [Tian et al. \(2016\)](#),  
132 and identify the maximum likelihood estimate of the QTL position under the null hypothesis of  
133 pleiotropy. We then use the inferred model parameters under that model and with the QTL at that  
134 position to simulate bootstrap data sets according to the model in equations 1–3. For each of  $b$  boot-  
135 strap data sets, we perform a two-dimensional QTL scan (over the genomic region of interest) and  
136 derive the test statistic value. We treat these  $b$  test statistics as the empirical null distribution, and  
137 calculate a p-value as the proportion of the  $b$  bootstrap test statistics that equal or exceed the ob-  
138 served one, with the original data,  $p = \#\{i : \text{LOD}_i^* \geq \text{LOD}\} / b$  where  $\text{LOD}_i^*$  denotes the LOD score  
139 for the  $i$ th bootstrap replicate and LOD is the observed test statistic.

### 140 **Data & Software Availability**

141 Our methods have been implemented in an R package, `qt12pleio`, available at GitHub:

142 <https://github.com/fboehm/qt12pleio>

143 Custom R code for our analyses and simulations are at GitHub:

144 <https://github.com/fboehm/qt2pleio-manuscript-clean>

**Table 1** Type I error rates for all runs in our  $2^3$  experimental design. We set (marginal) genetic variances (*i.e.*, diagonal elements of  $V_g$ ) to 1 in all runs.  $V_e$  was set to the 2 by 2 identity matrix in all runs. We used allele probabilities at a single genetic marker to simulate traits for all eight sets of parameter inputs. In the column “Allele effects partitioning”, “ABCD:EFGH” means that lines A–D carry one QTL allele while lines E–H carry the other allele. “F:ABCDEGH” means the QTL has a private allele in strain F.

Run	$\Delta$ (Allele effects)	Allele effects partitioning	Genetic correlation	Type I error rate
1	6	ABCD:EFGH	0	0.032
2	6	ABCD:EFGH	0.6	0.035
3	6	F:ABCDEGH	0	0.040
4	6	F:ABCDEGH	0.6	0.045
5	12	ABCD:EFGH	0	0.038
6	12	ABCD:EFGH	0.6	0.042
7	12	F:ABCDEGH	0	0.025
8	12	F:ABCDEGH	0.6	0.025

145 The data from [Recla \*et al.\* \(2014\)](#) and [Logan \*et al.\* \(2013\)](#) are available at the Mouse Phenome Database:  
146 <https://phenome.jax.org/projects/Chesler4> and <https://phenome.jax.org/projects/Recla1>.  
147 They are also available in R/qlt2 format at <https://github.com/rqtl/qlt2data>.

## 148 Simulation studies

149 We performed two types of simulation studies, one for type I error rate assessment and one to  
150 characterize the power to detect separate QTL. To simulate traits, we specified  $X$ ,  $B$ ,  $V_g$ ,  $K$ , and  $V_e$   
151 matrices (Equations 1–3). For both we used the allele probabilities from a single genomic region  
152 derived empirically from data for a set of 479 Diversity Outbred mice from [Keller \*et al.\* \(2018\)](#).

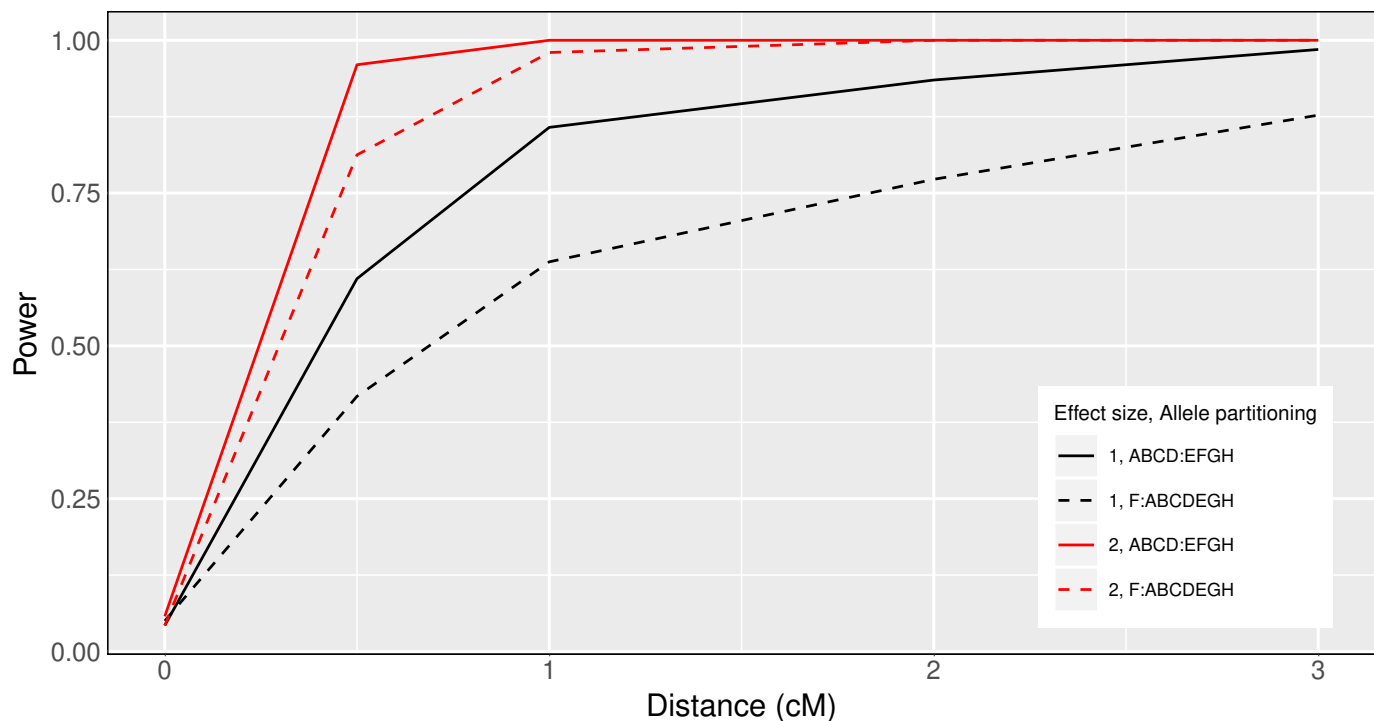
### 153 Type I error rate analysis

154 To quantify type I error rate (*i.e.*, false positive rate), we simulated 400 pairs of traits for each of eight  
155 sets of parameter inputs (Table 1). We used a  $2^3$  factorial experimental design with three factors:  
156 allele effects difference, allele effects partitioning, and genetic correlation, *i.e.*, the off-diagonal entry  
157 in the 2 by 2 matrix  $V_g$ .

158 We chose two strong allele effects difference values, 6 and 12. These ensured that the univariate  
159 phenotypes mapped with high LOD scores to the region of interest. For the allele partitioning factor,  
160 we used either equally frequent QTL alleles, or a private allele in the CAST strain (F). For the residual  
161 genetic correlation (the off-diagonal entry in  $V_g$ ), we considered the values 0 and 0.6. The marginal  
162 genetic variances (*i.e.*, the diagonal entries in  $V_g$ ) for each trait were always set to one.

163 We performed 400 simulation replicates per set of parameter inputs, and each used  $b = 400$   
164 bootstrap samples. For each bootstrap sample, we calculated the test statistic (Equation 8). We  
165 then compared the test statistic from the simulated trait against the empirical distribution of its 400  
166 bootstrap test statistics. When the simulated trait’s test statistic exceeded the 0.95 quantile of the





**Figure 1** Pleiotropy vs. separate QTL power curves for each of four sets of parameter settings. Factors that differ among the four curves are allele effects difference and allele partitioning. Red denotes high allele effects difference, while black is the low allele effects difference. Solid line denotes the even allele partitioning (ABCD:EFGH), while dashed line denotes the uneven allele partitioning (F:ABCDEGH).

empirical distribution of bootstrap test statistics, we rejected the null hypothesis. We observed that the test is slightly conservative over our range of parameter selections (Table 1), with estimated type I error rates  $< 0.05$ .

### Power analysis

We also investigated the power to detect the presence of two distinct QTL. We used a  $2 \times 2 \times 5$  experimental design, where our three factors were allele effects difference, allele effects partitioning, and inter-locus distance. The two levels of allele effects difference were 1 and 2. The two levels of allele effects partitioning were as in the type I error rate studies, ABCD:EFGH and F:ABCDEGH (Table S1). The five levels of interlocus distance were 0, 0.5, 1, 2, and 3 cM.  $V_g$  and  $V_e$  were both set to the 2 by 2 identity matrix in all power study simulations.

We simulated 400 pairs of traits per set of parameter inputs. For each simulation replicate, we calculated the likelihood ratio test statistic. We then applied our parametric bootstrap to determine the statistical significance of the results. For each simulation replicate, we used  $b = 400$  bootstrap samples. Because the bootstrap test statistics within a single set of parameter inputs followed approximately the same distribution, we pooled the  $400 \times 400 = 160,000$  bootstrap samples per set of parameter inputs and compared each test statistic to the empirical distribution derived from the 160,000 bootstrap samples. However, for parameter inputs with interlocus distance equal to zero, we did not pool the 160,000 bootstrap samples; instead, we proceeded by calculating power (*i.e.*, type I error rate, in this case), as we did in the type I error rate study above.

We present our power study results in Figure 1. Power increases as interlocus distance increases.

187 The top two curves correspond to the case where the QTL effects are largest. For each value for the  
188 QTL effect, power is greater when the QTL alleles are equally frequent, and smaller when a QTL  
189 allele is private to one strain. One can have high power to detect that the two traits have distinct  
190 QTL when they are separated by  $> 1$  cM and when the QTL have large effect. We provide example  
191 profile LOD plots from the power analysis in Figure S3.

## 192 Application

193 To illustrate our methods, we applied our test to data from Logan *et al.* (2013) and Recla *et al.* (2014),  
194 on 261 DO mice measured for a set of behavioral phenotypes. Recla *et al.* (2014) identified *Hydin* as  
195 the gene that underlies a QTL on Chromosome 8 at 57 cM for the “hot plate latency” phenotype (a  
196 measure of pain tolerance). The phenotype “percent time in light” in a light-dark box (a measure  
197 of anxiety) was measured on the same set of mice (Logan *et al.* 2013) and also shows a QTL near  
198 this location, which led us to ask whether the same locus affects both traits. The two traits show a  
199 correlation of  $-0.15$  (Figure S1).

200 QTL analysis with the LOCO method, and using sex as an additive covariate, showed multiple  
201 suggestive QTL for each phenotype (Figure S2; Table S2). For our investigation of pleiotropy, we  
202 focused on the interval 53–64 cM on Chromosome 8. Univariate analyses showed a QTL in this  
203 region for both traits (Figure 2).

204 The estimated QTL allele effects for the two traits are quite different (Figure 3). With the QTL  
205 placed at 55 cM, for “percent time in light”, the WSB and PWK alleles are associated with large  
206 phenotypes and NOD with low phenotypes. For “hot plate latency”, on the other hand, CAST and  
207 NZO show low phenotypes and NOD and PWK are near the center.

208 In applying our test for pleiotropy, we performed a two-dimensional, two-QTL scan for the pair  
209 of phenotypes. With these results, we created a profile LOD plot (Figure 4). The profile LOD for  
210 “percent time in light” (in brown) peaks near 55 cM, as was seen in the univariate analysis. The  
211 profile LOD for “hot plate latency” (in blue) peaks near 57 cM, also similar to the univariate analysis.  
212 The pleiotropy trace (in gray) peaks near 55 cM.

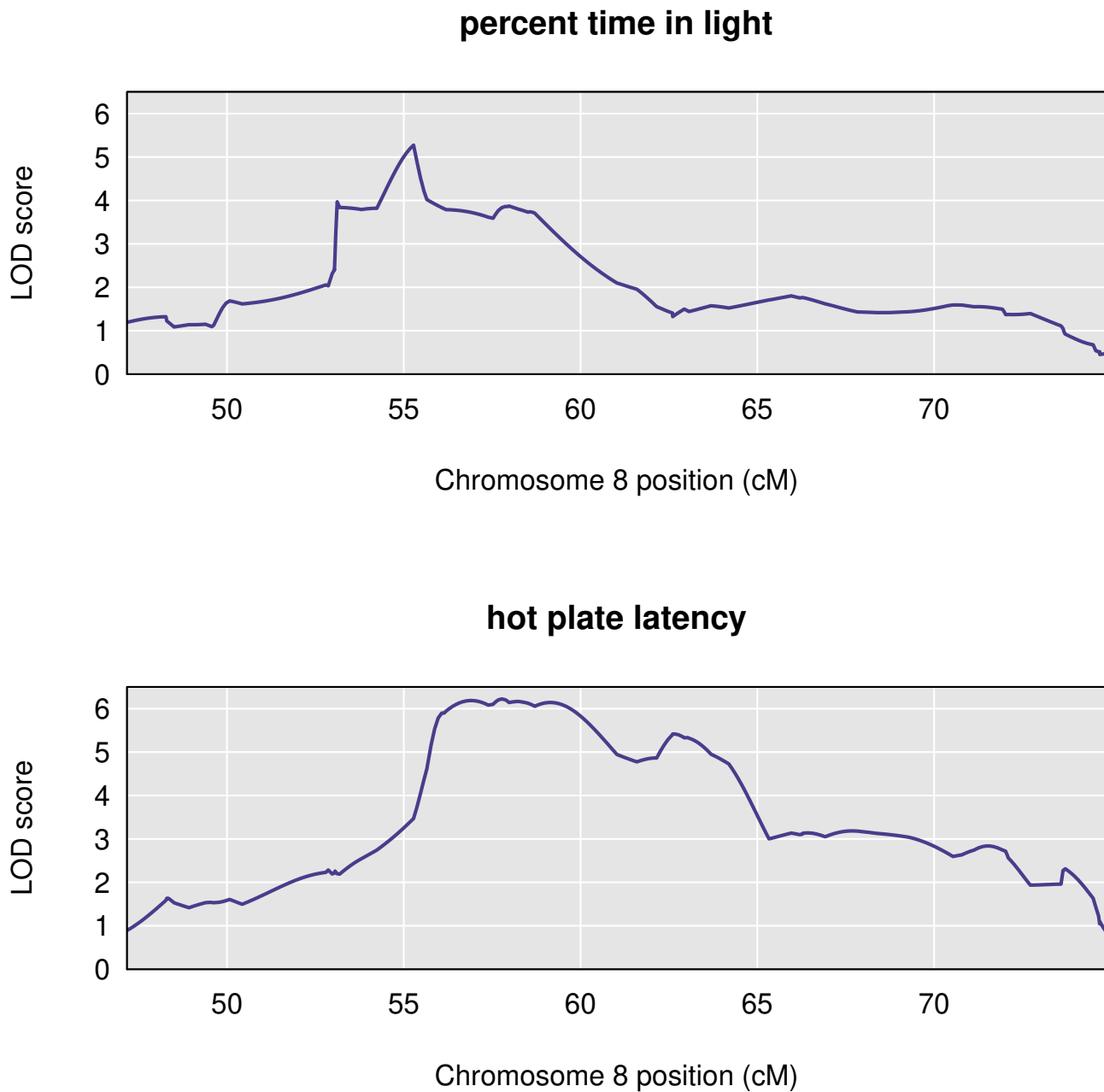
213 The likelihood ratio test statistic for the test of pleiotropy was 1.2. Based on a parametric bootstrap  
214 with 1,000 bootstrap replicates, the estimated p-value was 0.11. Thus, by our approach, the evidence  
215 for the two traits having distinct QTL is weak.

## 216 Discussion

217 We developed a test of pleiotropy vs. separate QTL for multiparental populations, extending the  
218 work of Jiang and Zeng (1995) for multiple alleles and with a linear mixed model to account for  
219 population structure (Kang *et al.* 2010; Yang *et al.* 2014). Our simulation studies indicate that our  
220 test is slightly conservative, with type I error rates below their nominal values (Table 1). Power is  
221 affected by many factors (including sample size, effect size, and allele frequencies). We studied the  
222 effects of interlocus distance and QTL effect on power, and we showed that our test has power to  
223 detect presence of separate loci, especially when univariate trait associations are strong (Figure 1).  
224 Type I error rates indicate that our test is slightly conservative (Table 1).

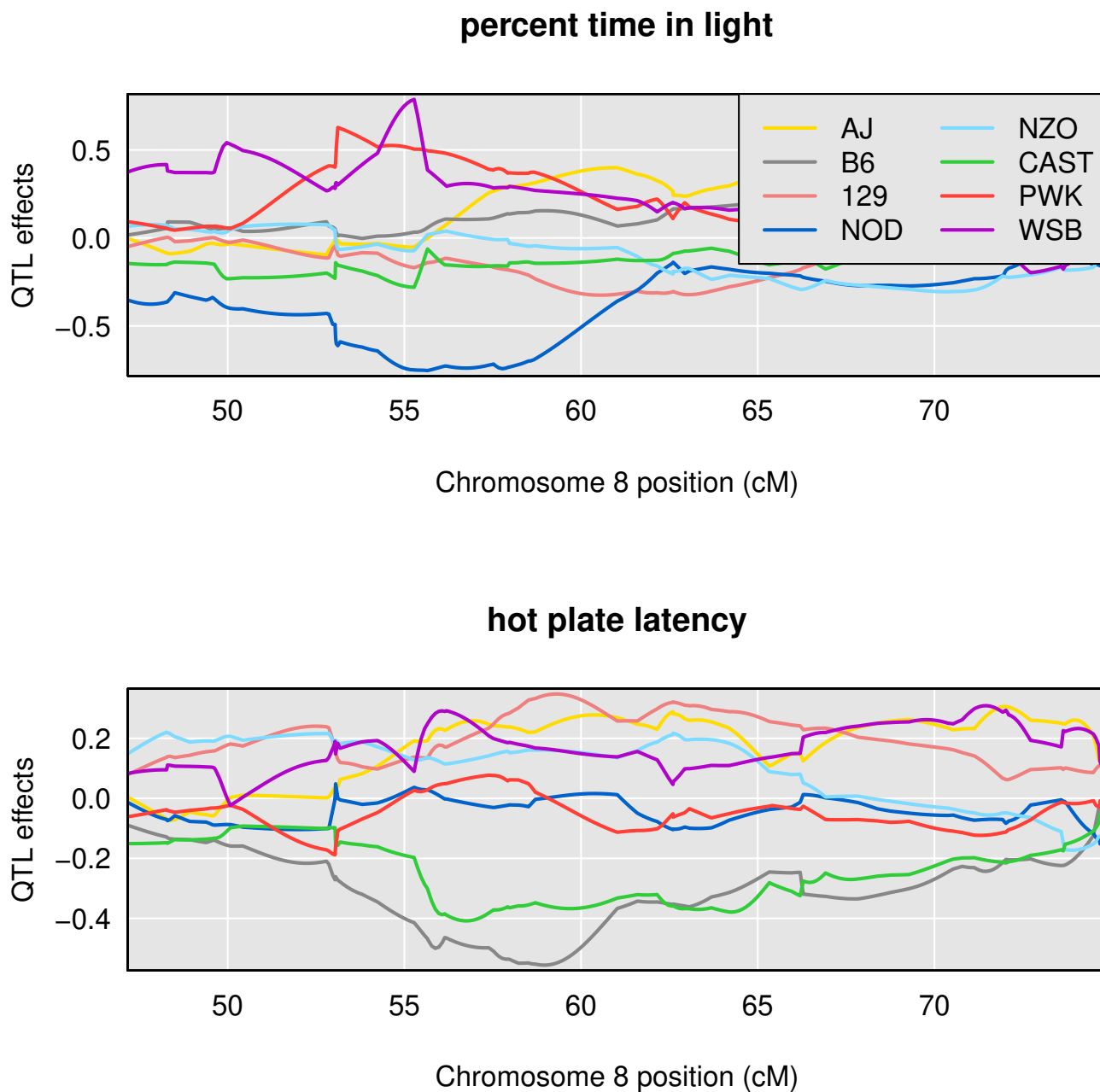
225 In the application of our method to two behavioral phenotypes in a study of 261 Diversity Out-  
226 bred mice (Recla *et al.* 2014; Logan *et al.* 2013), the evidence for the presence of two distinct QTL,  
227 with one QTL (which contains the *Hydin* gene) affecting only “hot plate latency” and a second QTL  
228 affecting “percent time in light” was weak ( $p = 0.11$ , Figure 4).

229 Founder allele effects plots provide further evidence for the presence of two distinct loci. As  
230 Macdonald and Long (2007) and King *et al.* (2012) have demonstrated in their analyses of multi-

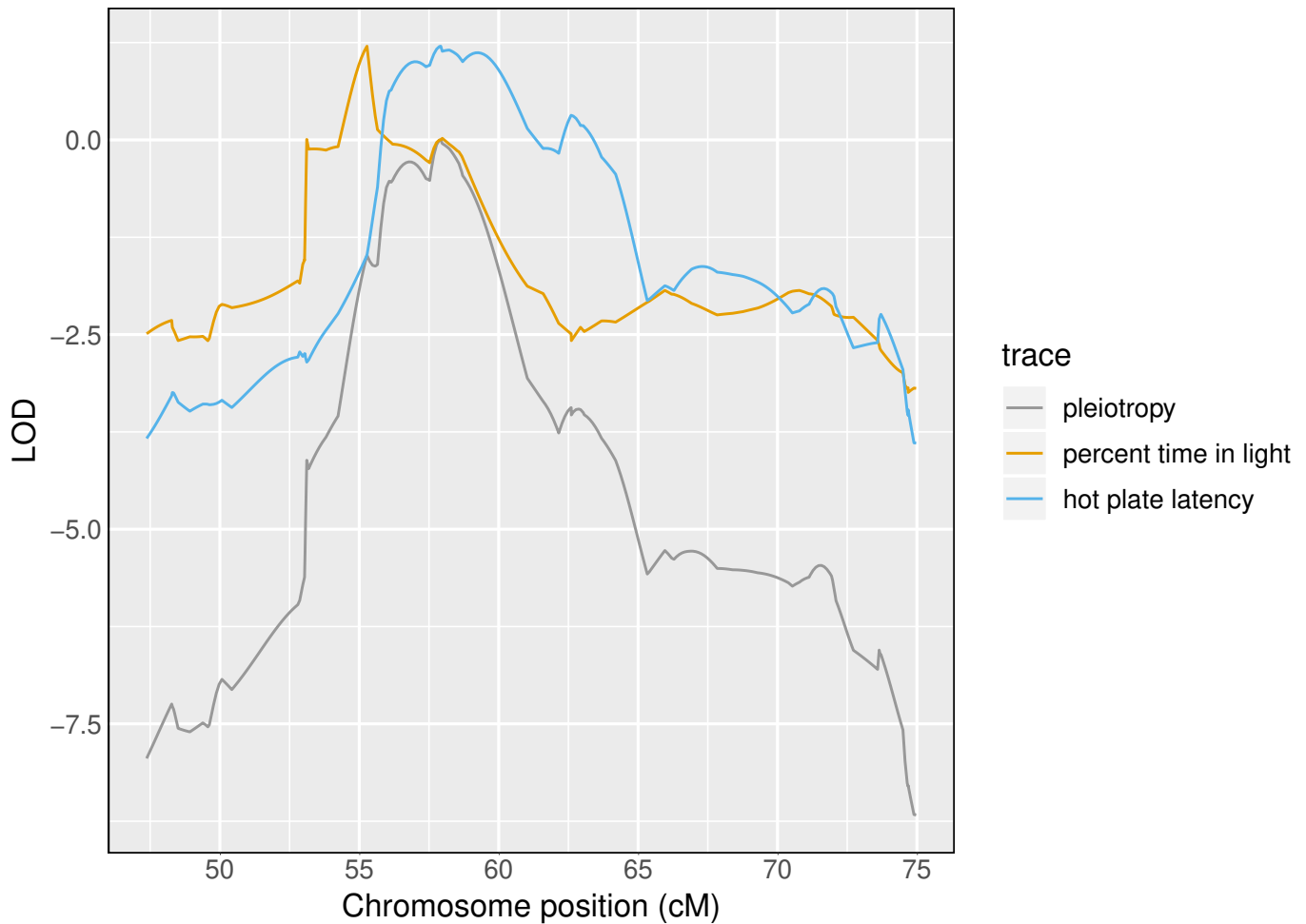


**Figure 2** Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins near 56 cM and ends about 64 cM.





**Figure 3** Chromosome 8 founder allele effects for percent time in light and hot plate latency demonstrate distinct allele patterns between 53cM and 64 cM.



**Figure 4** Profile LOD curves for the pleiotropy vs. separate QTL hypothesis test for “percent time in light” and “hot plate latency”. Gray trace denotes pleiotropy LOD values. Likelihood ratio test statistic value corresponds to the height of the blue and gold traces at their maxima.

231 parental *Drosophila* populations, a biallelic pleiotropic QTL would result in allele effects plots that  
232 have similar patterns. While we do not know that “percent time in light” and “hot plate latency”  
233 arise from biallelic QTL, the dramatic differences that we observe in allele effects patterns further  
234 support the argument for two distinct loci.

235 We have implemented our methods in an R package `qt12pleio`, but analyses can be computa-  
236 tionally intensive and time consuming. `qt12pleio` is written mostly in R, and so we could likely  
237 obtain improved computational speed by porting parts of the calculations to a compiled language  
238 such as C or C++. To accelerate our multi-dimensional QTL scans, we have integrated C++ code  
239 into `qt12pleio`, using the `Rcpp` package ([Eddelbuettel et al. 2011](#)).

240 Another computational bottleneck is the estimation of the variance components  $V_g$  and  $V_e$ . To  
241 accelerate this procedure, especially for the joint analysis of more than two traits, we will consider  
242 other strategies for variance component estimation, including that described by [Meyer et al. \(2018\)](#).  
243 [Meyer et al. \(2018\)](#), in joint analysis of dozens of traits, implement a bootstrap strategy to estimate  
244 variance components for lower-dimensional phenotypes before combining bootstrap estimates into  
245 valid covariance matrices for the full multivariate phenotype. Such an approach may ease some of  
246 the computational burdens that we encountered.

247 We view tests of pleiotropy as complementary to mediation tests and related methods that have  
248 become popular for inferring biomolecular causal relationships ([Chick et al. 2016](#); [Schadt et al. 2005](#);  
249 [Baron and Kenny 1986](#)). A mediation test proceeds by including a putative mediator as a covariate  
250 in the regression analysis of phenotype and QTL genotype; a substantial reduction in the association  
251 between genotype and phenotype corresponds to evidence of mediation.

252 Mediation analyses and our pleiotropy test ask distinct, but related, questions. Mediation analysis  
253 seeks to establish causal relationships among traits, including molecular traits, or dependent biolog-  
254 ical and behavioral processes. Pleiotropy tests examine whether two traits share a single source of  
255 genetic variation, which may act in parallel or in a causal network. Pleiotropy is required for causal  
256 relations among traits. In many cases, the pleiotropy hypothesis is the only reasonable one.

257 [Schadt et al. \(2005\)](#) argued that both pleiotropy tests and causal inference methods may contribute  
258 to gene network reconstruction. They developed a model selection strategy, based on the Akaike  
259 Information Criterion ([Akaike 1974](#)), to determine which causal model is most compatible with the  
260 observed data. [Schadt et al. \(2005\)](#) extended the methods of [Jiang and Zeng \(1995\)](#) to consider more  
261 complicated alternative hypotheses, such as the possibility of two QTL, one of which associates with  
262 both traits, and one of which associates with only one trait. As envisioned by [Schadt et al. \(2005\)](#), we  
263 foresee complementary roles emerging for our pleiotropy test and mediation tests in the dissection  
264 of complex trait genetic architecture.

265 Two related approaches for identifying and exploiting pleiotropy deserve mention. First, CAPE  
266 (Combined Analysis of Pleiotropy and Epistasis) is a strategy for identifying higher-order relation-  
267 ships among traits and marker genotypes ([Tyler et al. 2013, 2016](#)) and has recently been extended  
268 for use with multiparental populations, including DO mice ([Tyler et al. 2017](#)). CAPE exploits the  
269 pleiotropic relationship among traits in order to characterize the underlying network of QTLs, and  
270 it can suggest possible pleiotropic effects, but it does not provide an explicit test of pleiotropy. Sec-  
271 ond, [Schaid et al. \(2016\)](#) described a test for pleiotropy in the context of human genome-wide asso-  
272 ciation studies (GWAS). Their approach is fundamentally different from ours, in that rather than  
273 ask whether traits are affected by a common locus or distinct loci, they ask whether the traits are all  
274 affected by a particular SNP or only some are. The difference in these approaches may be attributed  
275 to the difference in mapping resolution between human GWAS and experimental populations.

276 Technological advances in mass spectrometry and RNA sequencing have enabled the acquisi-  
277 tion of high-dimensional biomolecular phenotypes ([Ozsolak and Milos 2011](#); [Han et al. 2012](#)). Mul-

278 tiparental populations in *Arabidopsis*, maize, wheat, oil palm, rice, *Drosophila*, yeast, and other  
279 organisms enable high-precision QTL mapping (Yu *et al.* 2008; Tisné *et al.* 2017; Stanley *et al.* 2017;  
280 Raghavan *et al.* 2017; Mackay *et al.* 2012; Kover *et al.* 2009; Cubillos *et al.* 2013). The need to analyze  
281 high-dimensional phenotypes in multiparental populations compels the scientific community to de-  
282 velop tools to study genotype-phenotype relationships and complex trait architecture. Our test, and  
283 its future extensions, will contribute to these ongoing efforts.

## 284 **Acknowledgments**

285 The authors thank Lindsay Traeger, Julia Kemis, Qiongshi Lu, Rene Welch, and two anonymous  
286 referees for valuable suggestions to improve the manuscript. This work was supported in part  
287 by National Institutes of Health grants R01GM070683 (to K.W.B.) and P50DA039841 (to E.J.C.).  
288 The research made use of compute resources and assistance of the UW-Madison Center For High  
289 Throughput Computing (CHTC) in the Department of Computer Sciences at UW-Madison, which  
290 is supported by the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation,  
291 the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active mem-  
292 ber of the Open Science Grid, which is supported by the National Science Foundation and the U.S.  
293 Department of Energy's Office of Science.

## 294 **Literature Cited**

- 295 Akaike, H., 1974 A new look at the statistical model identification. *IEEE T. Automat. Contr.* **19**: 716–  
296 723.
- 297 Baron, R. M. and D. A. Kenny, 1986 The moderator–mediator variable distinction in social psy-  
298 chological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51**:  
299 1173–1182.
- 300 Boehm, F., 2018 *gemma2: GEMMA multivariate linear mixed model*. R package version 0.0.1.
- 301 Broman, K. W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, *et al.*, 2019 R/qt12: Software for map-  
302 ping quantitative trait loci with high-dimensional data and multi-parent populations. *Genetics*  
303 **211**: 495–502.
- 304 Chesler, E. J., D. M. Gatti, A. P. Morgan, M. Strobel, L. Trepanier, *et al.*, 2016 Diversity outbred mice  
305 at 21: Maintaining allelic variation in the face of selection. *G3* **6**: 3893–3902.
- 306 Chick, J. M., S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, *et al.*, 2016 Defining the consequences  
307 of genetic variation on a proteome-wide scale. *Nature* **534**: 500–505.
- 308 Churchill, G. A., D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, *et al.*, 2004 The Collaborative Cross,  
309 a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36**: 1133–1137.
- 310 Churchill, G. A., D. M. Gatti, S. C. Munger, and K. L. Svenson, 2012 The diversity outbred mouse  
311 population. *Mamm. Genome* **23**: 713–718.
- 312 Cubillos, F. A., L. Parts, F. Salinas, A. Bergström, E. Scovacricchi, *et al.*, 2013 High-resolution map-  
313 ping of complex traits with a four-parent advanced intercross yeast population. *Genetics* **195**:  
314 1141–1155.
- 315 de Koning, D.-J. and L. M. McIntyre, 2014 Genetics and G3: Community-driven science, community-  
316 driven journals. *Genetics* **198**: 1–2.
- 317 Eddelbuettel, D., R. François, J. Allaire, J. Chambers, D. Bates, *et al.*, 2011 Rcpp: Seamless R and C++  
318 integration. *J. Stat. Softw.* **40**: 1–18.
- 319 Efron, B., 1979 Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**: 1–26.
- 320 Han, X., K. Yang, and R. W. Gross, 2012 Multi-dimensional mass spectrometry-based shotgun  
321 lipidomics and novel strategies for lipidomic analyses. *Mass Spectrom. Rev.* **31**: 134–178.

- 322 Jiang, C. and Z.-B. Zeng, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci.  
323 *Genetics* **140**: 1111–1127.
- 324 Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, *et al.*, 2010 Variance component model  
325 to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**: 348–354.
- 326 Keller, M. P., D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, *et al.*, 2018 Genetic drivers  
327 of pancreatic islet function. *Genetics* **209**: 335–356.
- 328 King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hooper, S. Sen, *et al.*, 2012 Genetic dissection of a model  
329 complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.* **22**: 1558–1566.
- 330 Knott, S. A. and C. S. Haley, 2000 Multitrait least squares for quantitative trait loci detection. *Genet-*  
331 *ics* **156**: 899–911.
- 332 Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich, *et al.*, 2009 A multiparent ad-  
333 vanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5**:  
334 e1000551.
- 335 Logan, R. W., R. F. Robledo, J. M. Recla, V. M. Philip, J. A. Bubier, *et al.*, 2013 High-precision genetic  
336 mapping of behavioral traits in the diversity outbred mouse population. *Genes Brain Behav.* **12**:  
337 424–437.
- 338 Macdonald, S. J. and A. D. Long, 2007 Joint estimates of QTL effect and frequency using synthetic  
339 recombinant populations of *Drosophila melanogaster*. *Genetics* **176**: 1261–1281.
- 340 Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, *et al.*, 2012 The *Drosophila*  
341 *melanogaster* genetic reference panel. *Nature* **482**: 173–178.
- 342 Meyer, H. V., F. P. Casale, O. Stegle, and E. Birney, 2018 LiMMBo: a simple, scalable ap-  
343 proach for linear mixed models in high-dimensional genetic association studies. bioRxiv. DOI:  
344 <https://doi.org/10.1101/255497>.
- 345 Meyer, K., 1989 Restricted maximum likelihood to estimate variance components for animal models  
346 with several random effects using a derivative-free algorithm. *Genet. Sel. Evol.* **21**: 317–340.
- 347 Meyer, K., 1991 Estimating variances and covariances for multivariate animal models by restricted  
348 maximum likelihood. *Genet. Sel. Evol.* **23**: 67–83.
- 349 Ozsolak, F. and P. M. Milos, 2011 RNA sequencing: advances, challenges and opportunities. *Nat.*  
350 *Rev. Genet.* **12**: 87–98.
- 351 Raghavan, C., R. Mauleon, V. Lacorte, M. Jubay, H. Zaw, *et al.*, 2017 Approaches in characterizing  
352 genetic structure and mapping in a rice multiparental population. *G3* **7**: 1721–1730.
- 353 Recla, J. M., R. F. Robledo, D. M. Gatti, C. J. Bult, G. A. Churchill, *et al.*, 2014 Precise genetic map-  
354 ping and integrative bioinformatics in Diversity Outbred mice reveals *Hydin* as a novel pain gene.  
355 *Mamm. Genome* **25**: 211–222.
- 356 Schadt, E. E., J. Lamb, X. Yang, J. Zhu, S. Edwards, *et al.*, 2005 An integrative genomics approach to  
357 infer causal associations between gene expression and disease. *Nat. Genet.* **37**: 710–717.
- 358 Schaid, D. J., X. Tong, B. Larrabee, R. B. Kennedy, G. A. Poland, *et al.*, 2016 Statistical methods for  
359 testing genetic pleiotropy. *Genetics* **204**: 483–497.
- 360 Stanley, P. D., E. Ng'oma, S. O'Day, and E. G. King, 2017 Genetic dissection of nutrition-induced  
361 plasticity in insulin/insulin-like growth factor signaling and median life span in a *Drosophila*  
362 multiparent population. *Genetics* **206**: 587–602.
- 363 Tian, J., M. P. Keller, A. T. Broman, C. Kendziorski, B. S. Yandell, *et al.*, 2016 The dissection of expres-  
364 sion quantitative trait locus hotspots. *Genetics* **202**: 1563–1574.
- 365 Tisné, S., V. Pomiès, V. Riou, I. Syahputra, B. Cochard, *et al.*, 2017 Identification of ganoderma disease  
366 resistance loci using natural field infection of an oil palm multiparental population. *G3* **7**: 1683–  
367 1692.
- 368 Tyler, A. L., L. R. Donahue, G. A. Churchill, and G. W. Carter, 2016 Weak epistasis generally stabi-

- 369 lizes phenotypes in a mouse intercross. *PLoS Genet.* **12**: e1005805.
- 370 Tyler, A. L., B. Ji, D. M. Gatti, S. C. Munger, G. A. Churchill, *et al.*, 2017 Epistatic networks jointly  
371 influence phenotypes related to metabolic disease and gene expression in diversity outbred mice.  
372 *Genetics* **206**: 621–639.
- 373 Tyler, A. L., W. Lu, J. J. Hendrick, V. M. Philip, and G. W. Carter, 2013 CAPE: an R package for  
374 combined analysis of pleiotropy and epistasis. *PLoS Comput. Biol.* **9**: e1003270.
- 375 Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price, 2014 Advantages and pitfalls  
376 in the application of mixed-model association methods. *Nat. Genet.* **46**: 100–106.
- 377 Yu, J., J. B. Holland, M. D. McMullen, and E. S. Buckler, 2008 Genetic design and statistical power of  
378 nested association mapping in maize. *Genetics* **178**: 539–551.
- 379 Zeng, Z.-B., J. Liu, L. F. Stam, C.-H. Kao, J. M. Mercer, *et al.*, 2000 Genetic architecture of a morpho-  
380 logical shape difference between two *Drosophila* species. *Genetics* **154**: 299–310.
- 381 Zhou, X. and M. Stephens, 2014 Efficient multivariate linear mixed model algorithms for genome-  
382 wide association studies. *Nat. Methods* **11**: 407–409.



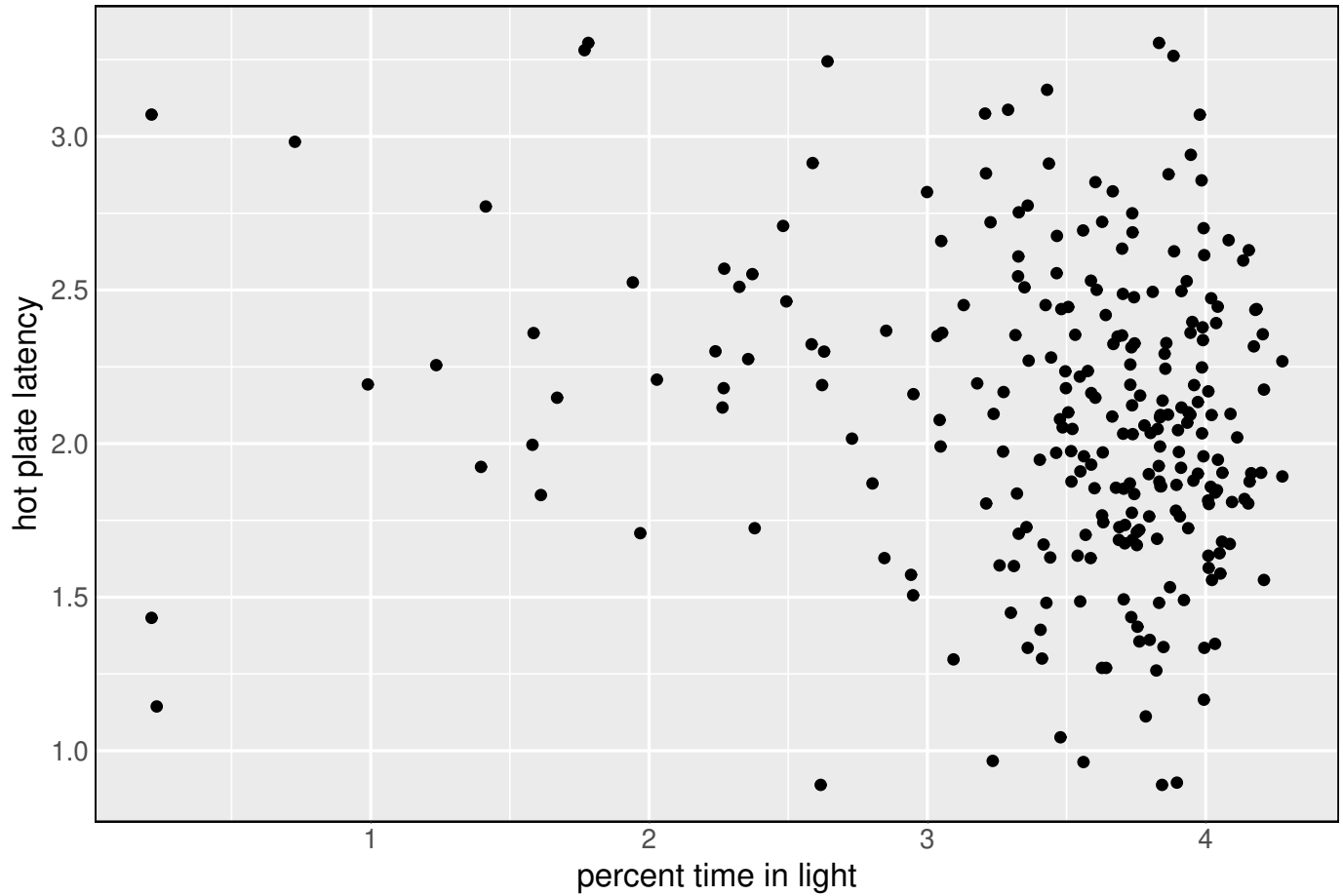
**Table S1** Eight founder lines and their one-letter abbreviations.

Founder allele	One-letter abbreviation
A/J	A
C57BL/6J	B
129S1/SvImJ	C
NOD/ShiLtJ	D
NZO/H1LTJ	E
Cast/EiJ	F
PWK/PhJ	G
WSB/EiJ	H

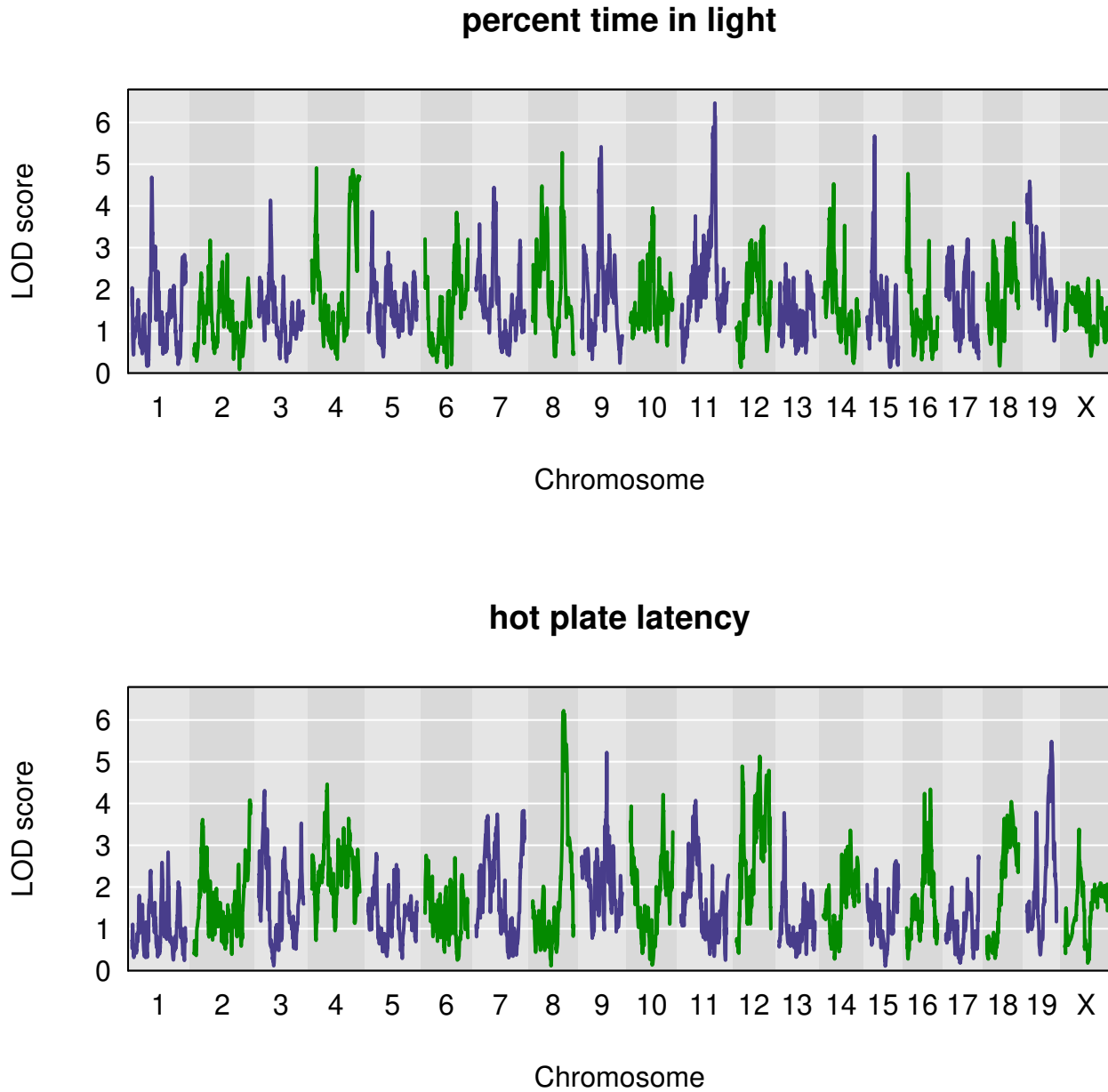
**Table S2** Both “hot plate latency” and “percent time in light” demonstrate multiple QTL peaks with LOD scores above 5.

phenotype	chr	pos	LOD score
percent time in light	8	55.28	5.27
hot plate latency	8	57.77	6.22
percent time in light	9	36.70	5.42
hot plate latency	9	46.85	5.22
percent time in light	11	63.39	6.46
hot plate latency	12	43.52	5.13
percent time in light	15	15.24	5.67
hot plate latency	19	47.80	5.48

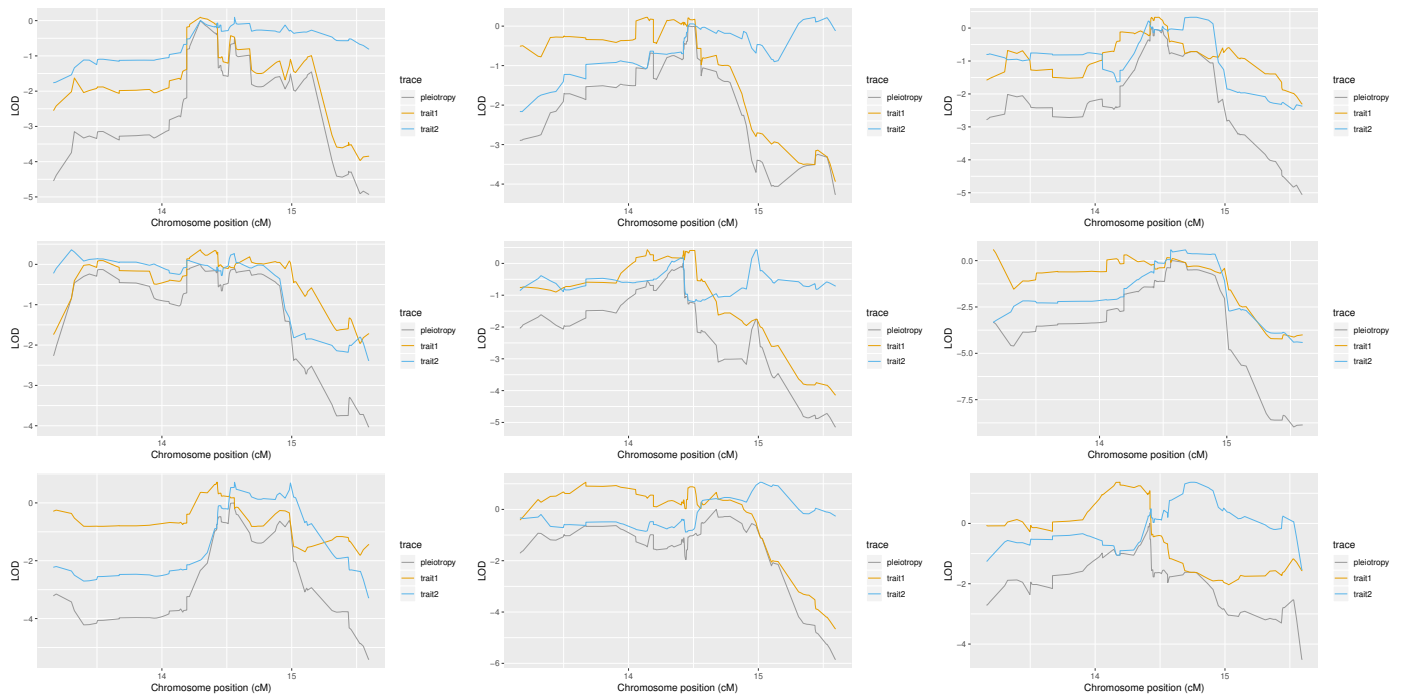
### hot plate latency vs. percent time in light



**Figure S1** Scatter plot of “hot plate latency” against “percent time in light”, after applying logarithm transformations and winsorizing both traits.



**Figure S2** Genome-wide QTL scan for percent time in light reveals multiple QTL, including one on Chromosome 8.



**Figure S3** Randomly sampled profile LOD plots for trait pairs simulated with interlocus distance equal to 0.5 cM, effect size 1, and allele partitioning F:ABCDEFGH. The test statistics, by row, starting at the top and proceeding left to right, are: 0.10, 0.22, 0.33, 0.36, 0.43, 0.59, 0.71, 1.06, and 1.37. The critical value (for  $\alpha = 0.05$ ) is 0.77.