Testing pleiotropy vs. separate QTL in multiparental populations

Frederick J. Boehm[*,1], **Elissa J. Chesler**[†], **Brian S. Yandell**[*,‡] **and Karl W. Broman**[§]

[*]Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin 53706, [†]The Jackson Laboratory, Bar Harbor, Maine 04609, [‡]Department of Horticulture, University of Wisconsin-Madison, Madison, Wisconsin 53706, [§]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 an opportunity 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 work of Jiang and Zeng (1995) to develop a likelihood ratio test, for a pair of traits, of the null hypothesis of pleiotropy against the alternative hypothesis of separate QTL, for the case of more than two alleles. We also incorporate polygenic random effects to account for variable relatedness among subjects. We use a parametric bootstrap to determine statistical significance. We apply our methods to a behavioral genetics data set from Diversity Outbred mice, where we find evidence for presence of two distinct loci in a 2.5 cM region. Our methods have been incorporated into the R package 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

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Saturday 19th January, 2019
[1]Department of Statistics, 1220 Medical Sciences Center, 1300 University Avenue, Madison, WI 53706. E-mail: frederick.boehm@gmail.com

results (Yang *et al.* 2014). Second, previous tests allowed for only two founder lines (Jiang and Zeng 1995). Finally, Jiang and Zeng (1995) assumed that the null distribution of the test statistic follows a chi-square distribution.

We developed a pleiotropy vs. separate QTL test for two traits in multiparental populations. Our test builds on research that Jiang and Zeng (1995), Knott and Haley (2000), Tian *et al.* (2016), and Zhou and Stephens (2014) initiated. Our innovations include the accommodation of k founder alleles per locus (compared to the traditional two founder alleles per locus) and the incorporation of multivariate polygenic random effects to account for relatedness. Furthermore, we implemented a parametric bootstrap to calibrate test statistic values (Efron 1979; Tian *et al.* 2016).

Below, we describe our likelihood ratio test for pleiotropy vs. separate QTL. In simulation studies, we find that it is slightly conservative, and that it has power to detect two separate loci when the univariate LOD peaks are strong. We further illustrate our approach with an application to data on a pair of behavior traits in a population of 261 DO mice (Logan *et al.* 2013; Recla *et al.* 2014). We find modest evidence for distinct QTL in a 2.5-cM region on mouse Chromosome 8.

Methods

Our strategy involves first identifying two traits that map to a common genomic region. We then perform a two-dimensional, two-QTL scan over the genomic region, with each trait affected by one QTL of varying position. We identify the QTL position that maximizes the likelihood under pleiotropy (that is, along the diagonal where the two QTL are at a common location), and the ordered pair of positions that maximizes the likelihood under the model where the two QTL are allowed to be distinct. The logarithm of the ratio of the two likelihoods is our test statistic. We calibrate this test statistic with a parametric bootstrap.

Data structures

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

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

Statistical Models

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

$$vec(Y) = Xvec(B) + vec(G) + vec(E) \tag{1}$$

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

We also impose distributional assumptions on G and E:

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

and

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

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

Parameter inference and log likelihood calculation

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

$$vec(Y) = X_0 vec(B) + vec(G) + vec(E) \tag{4}$$

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

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

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

where

$$\hat{\Sigma} = \hat{V}_g \otimes K + \hat{V}_e \otimes I_n \tag{6}$$

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

Pleiotropy vs. separate QTL hypothesis testing framework

Our test applies to two traits considered simultaneously. Below, λ_1 and λ_2 denote putative locus positions for traits one and two. We quantitatively state the competing hypotheses for our test as:

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

Our likelihood ratio test statistic is:

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

where L is the likelihood for fixed QTL positions, maximized over all other parameters.

Visualizing profile LOD traces

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

We define the LOD score for our test:

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

where ll_{10} denotes $\log_{10}$ likelihood.

We follow Zeng *et al.* (2000) and Tian *et al.* (2016) in defining profile LOD by the equation

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

We define profile $\text{LOD}_2(\lambda_2)$ analogously. The maximum value for the profile LOD_1 profile LOD_2 traces the same and are non-negative, and give the overall LOD test statistic.

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

$$LOD_p(\lambda) = ll_{10}(\lambda,\lambda) - \max ll_{10}(\lambda,\lambda) \tag{11}$$

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

Bootstrap for test statistic calibration

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

Data & Software Availability

Our methods have been implemented in an R package, `qtl2pleio`, available at GitHub:
https://github.com/fboehm/qtl2pleio
Custom R code for our analyses and simulations are at GitHub:
https://github.com/fboehm/qt2pleio-manuscript
The data from Recla *et al.* (2014) and Logan *et al.* (2013) are available at the Mouse Phenome Database:

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	Δ(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

https://phenome.jax.org/projects/Chesler4 and https://phenome.jax.org/projects/Recla1. They are also available in R/qtl2 format at https://github.com/rqtl/qtl2data.

Simulation studies

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

Type I error rate analysis

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

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

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

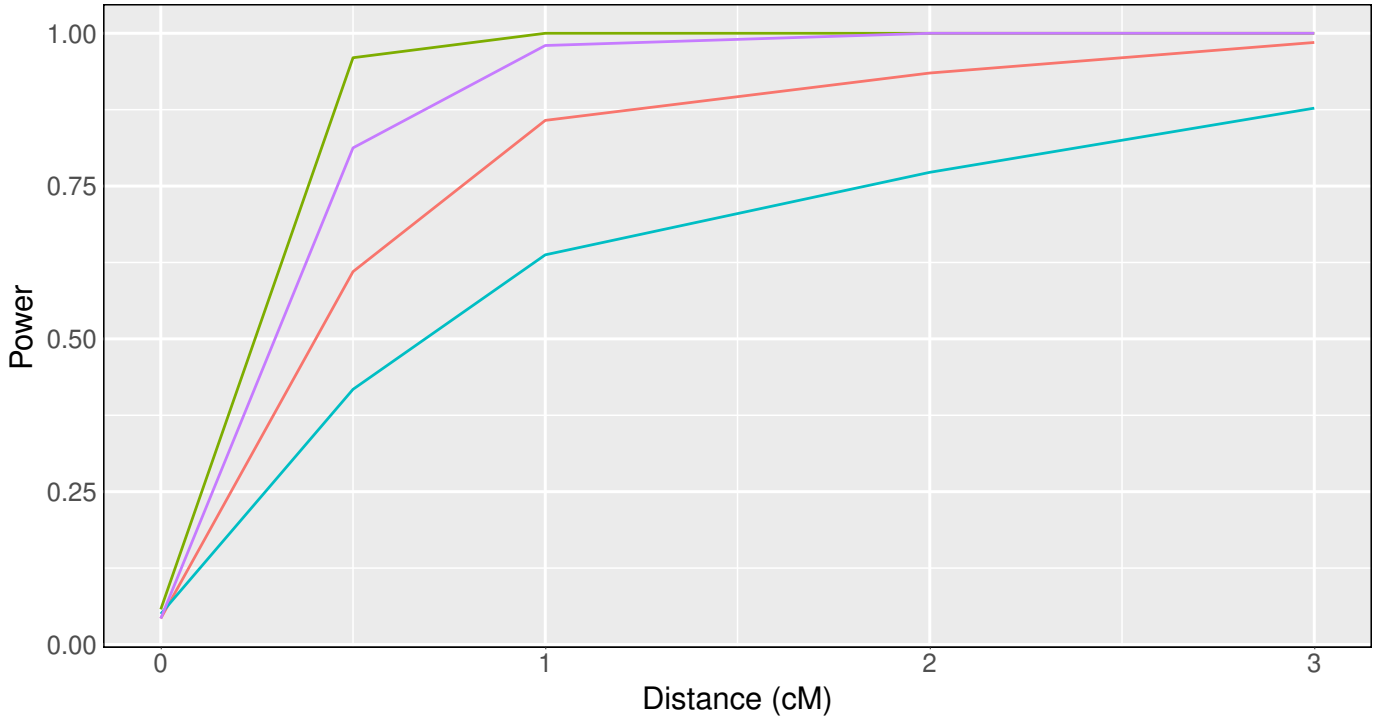

Figure 1 Pleiotropy vs. separate QTL power curves for each of four sets of parameter settings. Factors that differ among the four curves are allele effects difference and allele partitioning. Olive green denotes high allele effects difference and allele partitioning into two groups of four. Purple denotes high (2) allele effects difference with unequal allele partitioning. Red and blue both have the low allele effects difference with even and uneven allele partitioning, respectively.

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 calibrate the test statistics. 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 $400 * 400 = 160,000$ 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 didn't 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. The top two curves correspond to the case where the QTL effects are largest. For each value for the

QTL effect, power is greater when the QTL alleles are equally frequent, and smaller when a QTL allele is private to one strain. One can have high power to detect that the two traits have distinct QTL when they are separated by > 1 cM and when the QTL have large effect.

Application

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

QTL analysis with the LOCO method, and using sex as an additive covariate, showed multiple suggestive QTL for each phenotype (Figure **S2**; Table **S2**). For our investigation of pleiotropy, we focused on the interval 53–64 cM on Chromosome 8. The univariate QTL results for this region are shown in Figure 2.

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

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

The likelihood ratio test statistic for the test of pleiotropy was 1.2. Based on a parametric bootstrap with 1,000 bootstrap replicates, the estimated p-value was 0.11, indicating weak evidence for distinct QTL for the two traits.

Discussion

We developed a test of pleiotropy vs. separate QTL for multiparental populations, extending the work of Jiang and Zeng (1995) for multiple alleles and with a linear mixed models to account for population structure. Our simulation studies indicate that the test has power to detect presence of separate loci, especially when univariate trait associations are strong (Figure 1). Type I error rates indicate that our test is slightly conservative (Table 1).

In the application of our method to two behavioral phenotypes in a study of 261 Diversity Outbred mice (Recla *et al.* 2014; Logan *et al.* 2013), we obtained weak evidence (p=0.11) for the presence of two distinct QTL, with one QTL (which contained the *Hydin* gene) affecting only "hot plate latency" and a second QTL affecting "percent time in light" (Figure 4).

Founder allele effects plots provide further evidence for the presence of two distinct loci. As Macdonald and Long (2007) and King *et al.* (2012) have demonstrated in their analyses of multiparental Drosophila populations, a biallelic pleiotropic QTL would result in allele effects plots that have similar patterns. While we don't know that "percent time in light" and "hot plate latency" arise from biallelic QTL, the dramatic differences that we observe in allele effects patterns further support the argument for two distinct loci, with *Hydin* affecting only "hot plate latency".

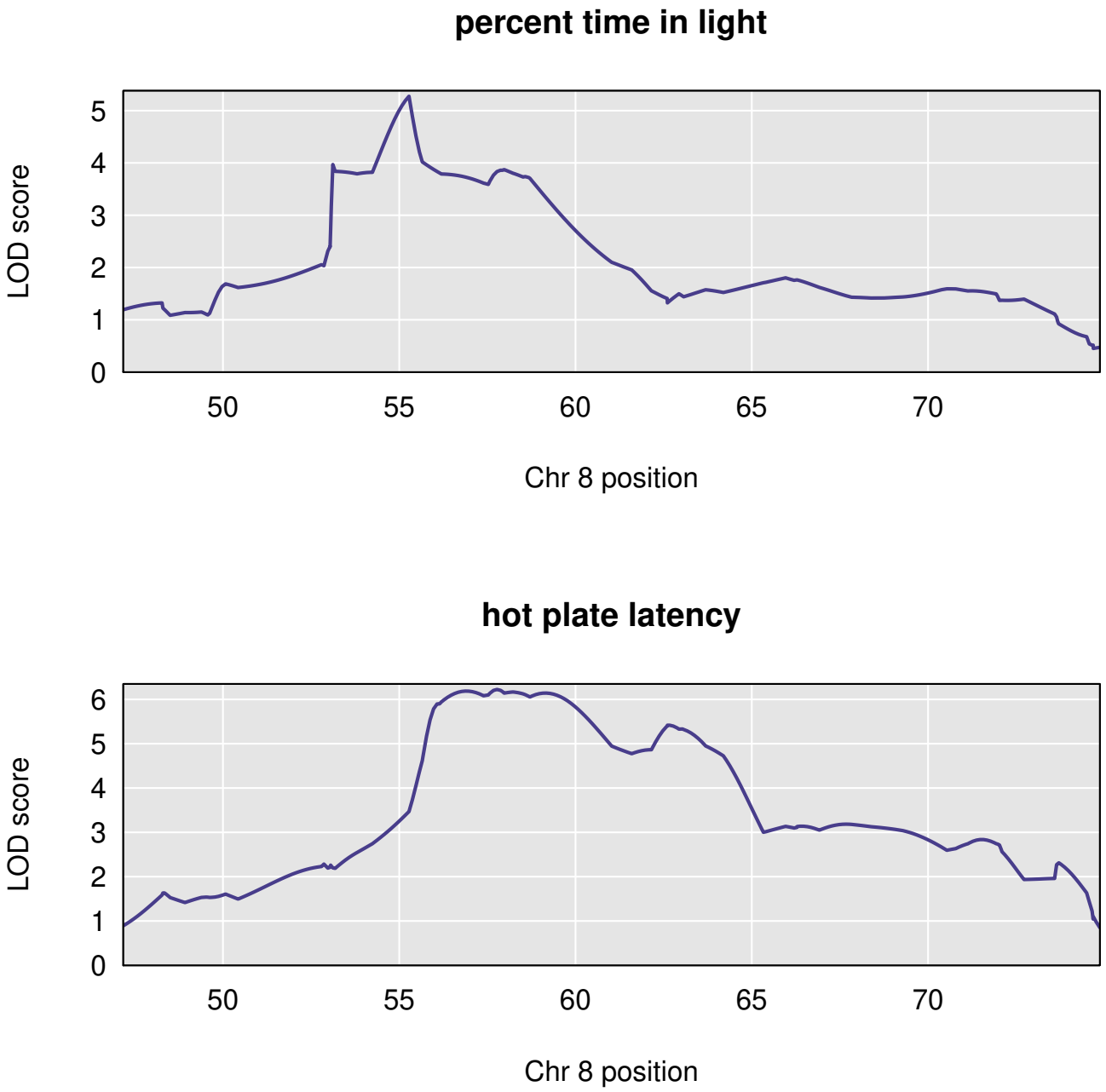

Figure 2 Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins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

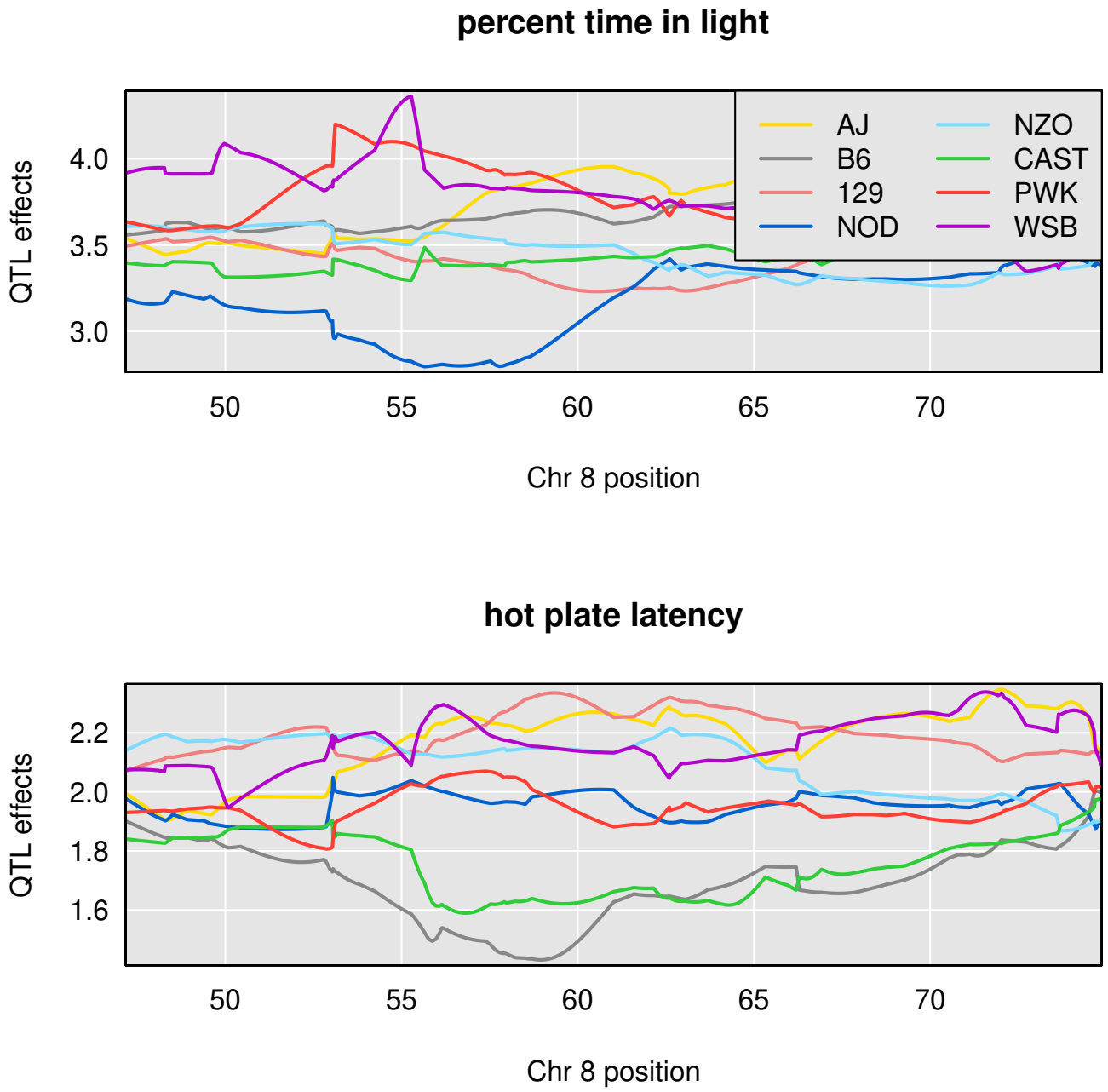

Figure 3 Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins near 56 cM and ends about 64 cM.

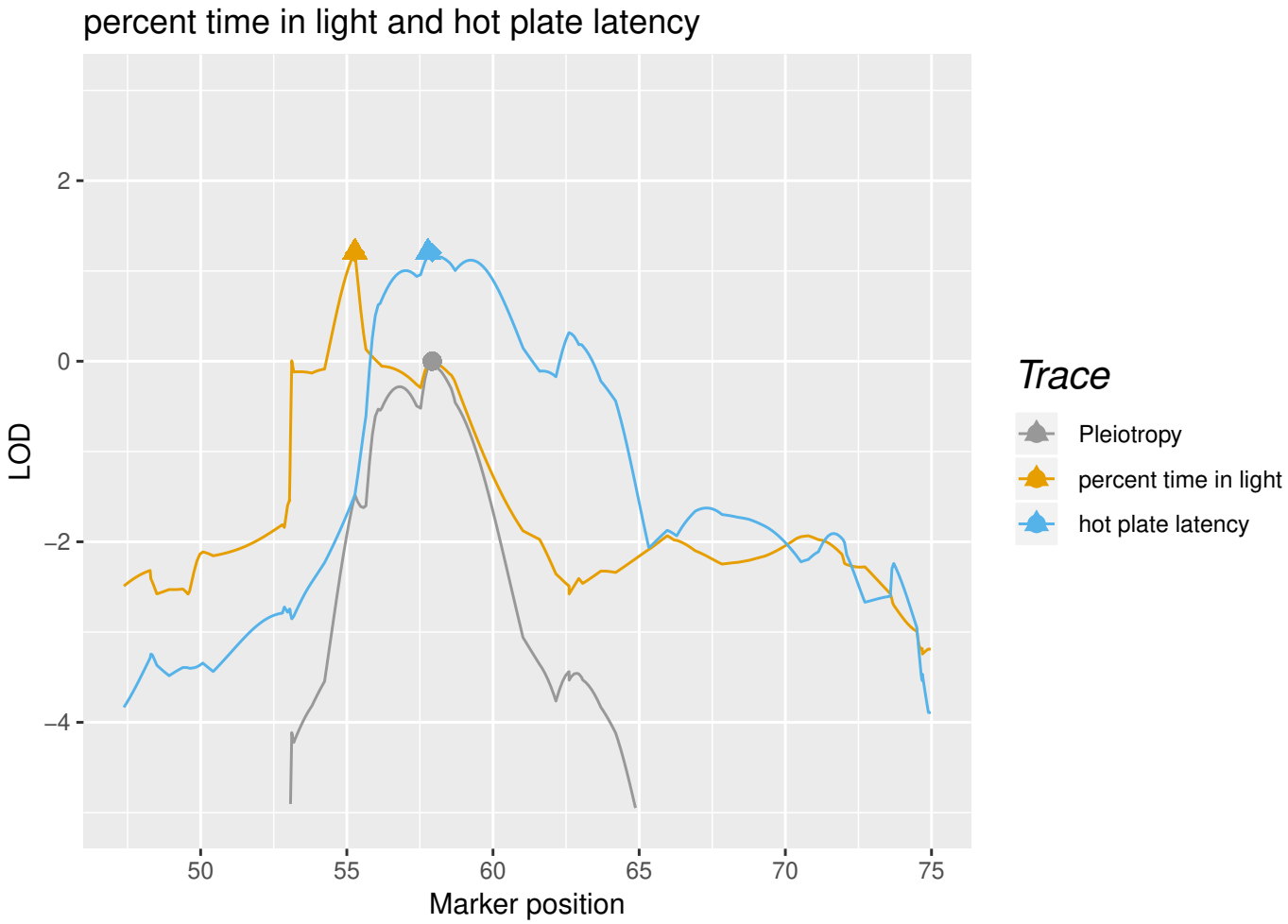

Figure 4 Profile LOD curves for the pleiotropy vs. separate QTL hypothesis test for "percent time in light" and "hot plate latency". Gray trace denotes pleiotropy LOD values. Triangles denote the univariate LOD maxima, while diamonds denote the profile LOD maxima. For "percent time in light", the brown triangle obscures the smaller brown diamond. Likelihood ratio test statistic value corresponds to the height of the blue and brown traces at their maxima.

We have implemented our methods in an R package `qtl2pleio`, but analyses can be computationally intensive and time consuming. `qtl2pleio` is written mostly in R, and so we could likely obtain improved computational speed by porting parts of the calculations to a compiled langugage such as C or C++. To accelerate our multi-dimensional QTL scans, we have integrated C++ code into `qtl2pleio`, using the Rcpp package (Eddelbuettel *et al.* 2011).

Another computational bottleneck is the estimation of the variance components V_g and V_e. In future research, we hope to extend our work for the joint analysis of more than two traits, and we will consider other strategies for variance component estimation, including that described by Meyer *et al.* (2018), who implement a bootstrap strategy to estimate variance components for lower-dimensional phenotypes before combining bootstrap variance component estimates into valid covariance matrices for the full multivariate phenotype.

We view tests of pleiotropy as complementary to the sorts of mediation tests and related methods that have become popular for inferring biomolecular causal relationships (Chick *et al.* 2016; Schadt *et al.* 2005; Baron and Kenny 1986). A mediation test proceeds by including a putative mediator as a covariate in the regression analysis of phenotype and QTL genotype; one looks for a substantial reduction in the association between genotype and phenotype. Mediation and pleiotropy tests have different goals: mediation analysis seeks to establish causal relationships, while pleiotropy tests examine whether two traits share a QTL.

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

Technological advances in mass spectrometry and RNA sequencing have enabled the acquisition of high-dimensional biomolecular phenotypes (Ozsolak and Milos 2011; Han *et al.* 2012). Multiparental populations in *Arabidopsis*, maize, wheat, oil palm, rice, *Drosophila*, yeast, and other organisms enable high-precision QTL mapping (Yu *et al.* 2008; Tisné *et al.* 2017; Stanley *et al.* 2017; Raghavan *et al.* 2017; Mackay *et al.* 2012; Kover *et al.* 2009; Cubillos *et al.* 2013). The need to analyze high-dimensional phenotypes in multiparental populations compels the scientific community to develop tools to study genotype-phenotype relationships and complex trait architecture. Our test, and its future extensions, will contribute to these ongoing efforts.

Acknowledgments

The authors thank Lindsay Traeger, Julia Kemis, and Rene Welch for valuable suggestions to improve the manuscript. This work was supported in part by National Institutes of Health grant R01GM070683 (to K.W.B.). The research made use of compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences at UW-Madison, which is supported by the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

Literature Cited

Akaike, H., 1974 A new look at the statistical model identification. IEEE transactions on automatic

control **19**: 716–723.
Baron, R. M. and D. A. Kenny, 1986 The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology **51**: 1173.
Boehm, F., 2018 *gemma2: Zhou & Stephens (2014) GEMMA multivariate linear mixed model*. R package version 0.0.1.
Broman, K. W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins, *et al.*, 2019 R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multi-parent populations. Genetics, to appear .
Chesler, E. J., D. M. Gatti, A. P. Morgan, M. Strobel, L. Trepanier, *et al.*, 2016 Diversity outbred mice at 21: Maintaining allelic variation in the face of selection. G3: Genes| Genomes| Genetics pp. g3–116.
Chick, J. M., S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, *et al.*, 2016 Defining the consequences of genetic variation on a proteome-wide scale. Nature **534**: 500.
Churchill, G. A., D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie, *et al.*, 2004 The collaborative cross, a community resource for the genetic analysis of complex traits. Nature genetics **36**: 1133–1137.
Churchill, G. A., D. M. Gatti, S. C. Munger, and K. L. Svenson, 2012 The diversity outbred mouse population. Mammalian genome **23**: 713–718.
Cubillos, F. A., L. Parts, F. Salinas, A. Bergström, E. Scovacricchi, *et al.*, 2013 High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. Genetics **195**: 1141–1155.
de Koning, D.-J. and L. M. McIntyre, 2014 Genetics and g3: Community-driven science, community-driven journals. Genetics **198**: 1–2.
Eddelbuettel, D., R. François, J. Allaire, J. Chambers, D. Bates, *et al.*, 2011 Rcpp: Seamless r and c++ integration. Journal of Statistical Software **40**: 1–18.
Efron, B., 1979 Bootstrap methods: another look at the jackknife. The Annals of Statistics **7**: 1–26.
Han, X., K. Yang, and R. W. Gross, 2012 Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. Mass spectrometry reviews **31**: 134–178.
Jiang, C. and Z.-B. Zeng, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics **140**: 1111–1127.
Keller, M. P., D. M. Gatti, K. L. Schueler, M. E. Rabaglia, D. S. Stapleton, *et al.*, 2018 Genetic drivers of pancreatic islet function. Genetics **209**: 335–356.
King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen, *et al.*, 2012 Genetic dissection of a model complex trait using the drosophila synthetic population resource. Genome research pp. gr–134031.
Knott, S. A. and C. S. Haley, 2000 Multitrait least squares for quantitative trait loci detection. Genetics **156**: 899–911.
Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich, *et al.*, 2009 A multiparent advanced generation inter-cross to fine-map quantitative traits in arabidopsis thaliana. PLoS genetics **5**: e1000551.
Logan, R. W., R. F. Robledo, J. M. Recla, V. M. Philip, J. A. Bubier, *et al.*, 2013 High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. Genes, Brain and Behavior **12**: 424–437.
Macdonald, S. J. and A. D. Long, 2007 Joint estimates of qtl effect and frequency using synthetic recombinant populations of drosophila melanogaster. Genetics .
Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, *et al.*, 2012 The drosophila melanogaster genetic reference panel. Nature **482**: 173.

Meyer, H. V., F. P. Casale, O. Stegle, and E. Birney, 2018 Limmbo: a simple, scalable approach for linear mixed models in high-dimensional genetic association studies. bioRxiv p. 255497.

Meyer, K., 1989 Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genetics Selection Evolution **21**: 317.

Meyer, K., 1991 Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. Genetics Selection Evolution **23**: 67.

Ozsolak, F. and P. M. Milos, 2011 Rna sequencing: advances, challenges and opportunities. Nature reviews genetics **12**: 87.

Raghavan, C., R. Mauleon, V. Lacorte, M. Jubay, H. Zaw, *et al.*, 2017 Approaches in characterizing genetic structure and mapping in a rice multiparental population. G3: Genes, Genomes, Genetics **7**: 1721–1730.

Recla, J. M., R. F. Robledo, D. M. Gatti, C. J. Bult, G. A. Churchill, *et al.*, 2014 Precise genetic mapping and integrative bioinformatics in diversity outbred mice reveals hydin as a novel pain gene. Mammalian genome **25**: 211–222.

Schadt, E. E., J. Lamb, X. Yang, J. Zhu, S. Edwards, *et al.*, 2005 An integrative genomics approach to infer causal associations between gene expression and disease. Nature genetics **37**: 710.

Stanley, P. D., E. Ng'oma, S. O'Day, and E. G. King, 2017 Genetic dissection of nutrition-induced plasticity in insulin/insulin-like growth factor signaling and median life span in a drosophila multiparent population. Genetics **206**: 587–602.

Tian, J., M. P. Keller, A. T. Broman, C. Kendziorski, B. S. Yandell, *et al.*, 2016 The dissection of expression quantitative trait locus hotspots. Genetics **202**: 1563–1574.

Tisné, S., V. Pomiès, V. Riou, I. Syahputra, B. Cochard, *et al.*, 2017 Identification of ganoderma disease resistance loci using natural field infection of an oil palm multiparental population. G3: Genes, Genomes, Genetics **7**: 1683–1692.

Yang, J., N. A. Zaitlen, M. E. Goddard, P. M. Visscher, and A. L. Price, 2014 Advantages and pitfalls in the application of mixed-model association methods. Nature genetics **46**: 100–106.

Yu, J., J. B. Holland, M. D. McMullen, and E. S. Buckler, 2008 Genetic design and statistical power of nested association mapping in maize. Genetics **178**: 539–551.

Zeng, Z.-B., J. Liu, L. F. Stam, C.-H. Kao, J. M. Mercer, *et al.*, 2000 Genetic architecture of a morphological shape difference between two drosophila species. Genetics **154**: 299–310.

Zhou, X. and M. Stephens, 2014 Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature 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

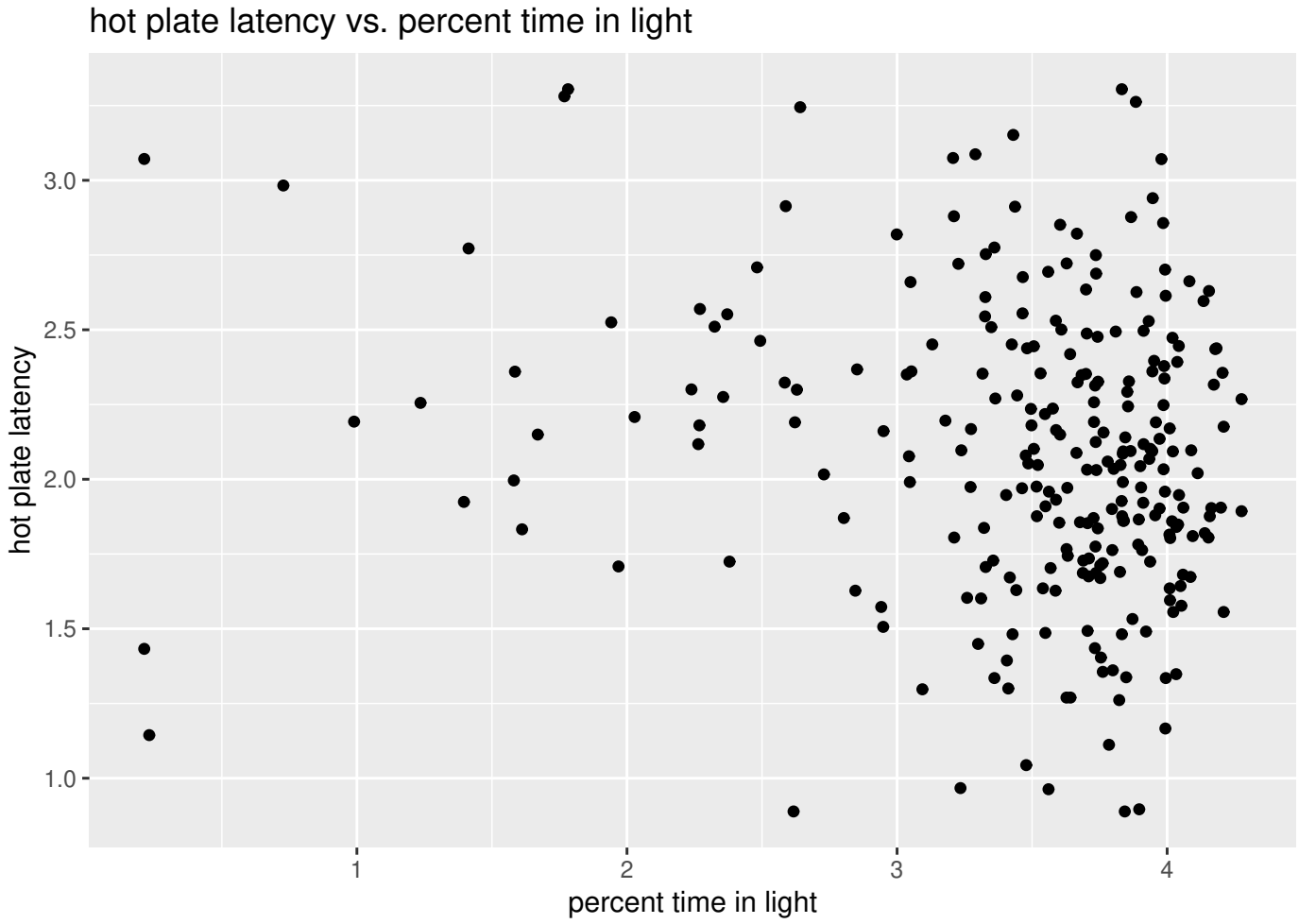

Figure S1 Scatter plot of "hot plate latency" against "percent time in light", after applying logarithm transformations and winsorizing both traits.

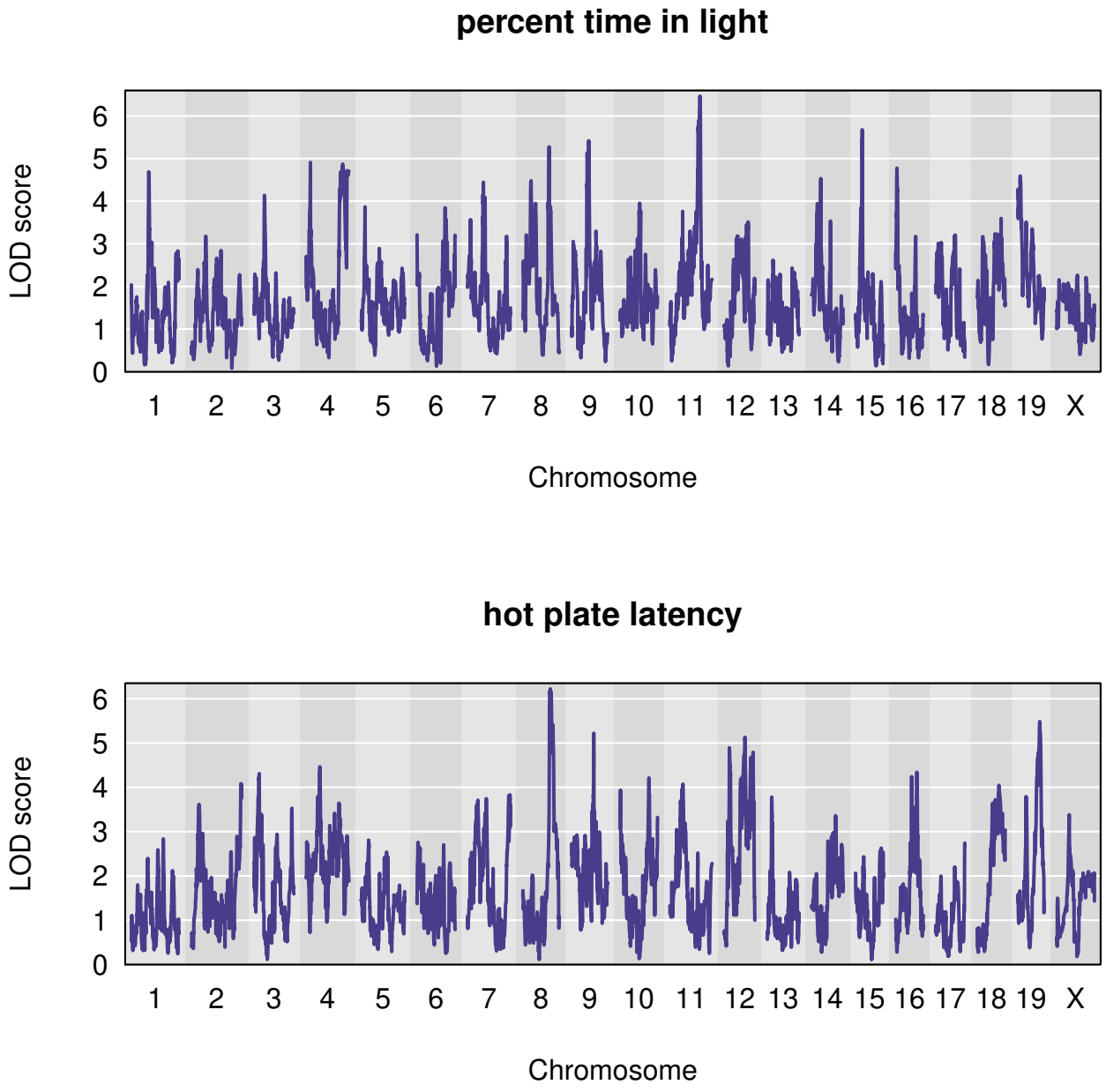

Figure S2 Genome-wide QTL scan for percent time in light reveals multiple QTL, including one on Chromosome 8.