Testing pleiotropy vs. separate QTL in multiparental populations

Frederick J. Boehm[*,1], **Brian S. Yandell**[*,†], **Karl W. Broman**[‡], **Author Four**[§] **and Author Five**[**]

[*]Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin 53706, [†]Department of Horticulture, University of Wisconsin-Madison, Madison, Wisconsin 53706, [‡]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin 53706, [§]Author four affiliation, [**]Author five affiliation

ABSTRACT Simultaneous developments of model organism multiparental populations with high mapping resolution and of technologies to measure tens of thousands of phenotypes present an opportunity for quantitative methods to enhance understanding of complex trait architecture. In multivariate quantitative trait locus mapping studies, one often observes multiple traits that map to a single genomic region. In this setting, knowing the number of distinct loci informs understanding of complex trait architecture and planning for subsequent experiments. We extend the work of Jiang and Zeng (1995) to develop a likelihood ratio test for the alternative hypothesis of separate QTL of two loci against the null hypothesis of pleiotropy for a pair of traits that map to a single genomic region. Unlike previous tests of these competing hypotheses, our test incorporates polygenic random effects to account for complex patterns of relatedness among subjects. Additionally, our test accommodates more than two founder alleles. We use a parametric bootstrap to determine statistical significance. Finally, 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-centiMorgan region on Chromosome 8, where one locus affects "percent time in light" and a second, distinct locus affects "hot plate latency". We share our software in an accessory R package, `qtl2pleio` on Github.

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. While examining one trait at a time has become standard practice in mapping quantitative trait loci (QTL), we argue below that simultaneous analysis of two or more traits in multiparental populations addresses additional questions, including the ability to ask whether two traits share a single pleiotropic locus. We present below a statistical hypothesis test to address this question. We characterize our test's statistical properties and apply it to publicly available data from Diversity Outbred mice to learn that two behavioral traits that map to a 2.5-cM region arise from two separate QTL.

Previous research addressed the question of pleiotropy vs. separate QTL in two-parent crosses. To gain insight into genetic architecture of complex traits, Jiang and Zeng (1995) developed a likelihood ratio test for pleiotropy vs. separate QTL. For two traits that map to a single genomic region, they fit null models in which a single QTL affects two traits and alternative models in which two distinct QTL affect one trait each. Jiang and Zeng (1995) argued that distinguishing between these two hypotheses - pleiotropy and separate QTL - informs understanding of genetic architecture of complex traits. 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. In analyzing genome-wide expression traits, Tian *et al.* (2016) found evidence for the presence of at least two QTL in five of six expression trait hotspots.

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Sunday 16th December, 2018
[1]Department of Statistics, 1220 Medical Sciences Center, 1300 University Avenue, Madison, WI 53706. E-mail: frederick.boehm@gmail.com

Newly developed, community-supported multiparental populations, such as the Diversity Outbred mouse population, have enabled high-precision mapping of complex traits (de Koning and McIntyre 2014). The Diversity Outbred mouse population began with progenitors of the Collaborative Cross mice (Churchill *et al.* 2004, 2012). Each Diversity Outbred mouse is a highly heterozygous genetic mosaic of alleles from the eight Collaborative Cross founder lines. Random matings among non-siblings have maintained the Diversity Outbred 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 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 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 then apply our test to two behavioral phenotypes. We find modest evidence for two distinct QTL in a region on Chromosome 8 (Logan *et al.* 2013; Recla *et al.* 2014).

Methods

Our strategy involves first identifying two traits that map to one genomic region. We then perform a two-dimensional QTL scan over the genomic region. We identify the marker that maximizes the likelihood under pleiotropy and the ordered pair of markers that maximizes the likelihood under the two separate QTL model. The logarithm of the ratio of the two likelihoods is our test statistic. We then calibrate our test statistic with a parametric bootstrap.

Data structures

We arrange the data into three objects. The first is a n by k by m array of allele probabilities for n subjects with k alleles per marker and m markers on a single chromosome. The second object is a n by d matrix of phenotype values, where $d = 2$ for our analyses. Each column is a single phenotype and each row is a single subject. The third object is a n by c matrix of covariates, where each row is one subject and each column is one covariate. In analyses below, we use four indicator variables to specify each subject's sex and membership in one of four waves of subjects.

One additional object is the genotype-derived collection of "leave one chromosome out" kinship matrices. The "leave one chromosome out" approach calculates a collection of kinship matrices, one for each chromosome. Each kinship matrix results from kinship estimation performed while omitting the genetic data from a single chromosome. For example, the chromosome 1 "leave one chromosome out" kinship matrix results from omitting the genetic data of chromosome 1 and using all other genetic data. Our motivation for using "leave one chromosome out" kinship calculations is to avoid using any marker's genetic data in both the kinship matrix and a design matrix for linear mixed effects models (Yang *et al.* 2014).

Statistical Models

The next step in a pleiotropy vs. separate QTL test, after identifying two traits and a genomic region of interest, is a two-dimensional QTL scan (Jiang and Zeng 1995). in a two-dimensional QTL scan, we fit the multivariate linear mixed effects model defined in Equation 1 once for every ordered pair of markers. Note that we assumed an additive genetic model throughout our analyses, but extensions to design matrices for dominance models are straightforward.

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

where Y is a n by 2 matrix of phenotypes values, where each row is one mouse and each column is one trait; X is a $2n$ by $2(k+c)$ matrix that contains two markers' k allele probabilities and 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 computationally intense (Meyer 1989, 1991). Thus, we make the simplifying assumption that V_g and V_e don't depend on allele probabilities. We use restricted maximum likelihood methods to fit the model:

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

X_0 is a $2n$ by $2(c+1)$ matrix. The first column of each diagonal block X_0 has all entries equal to one, and the remaining columns are covariates.

We draw on our implementation (Boehm 2018) (in `R`) 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 a ("leave one chromosome out") 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, ω_1 and ω_2 denote putative locus positions for traits one and two. We quantitatively state the competing hypotheses for our test as:

$$\begin{aligned} H_0 &: \omega_1 = \omega_2 \\ H_A &: \omega_1 \neq \omega_2 \end{aligned} \tag{7}$$

Our likelihood ratio test statistic is:

$$-\log_{10} \frac{\max L_0(B, \Sigma, \omega_1, \omega_2)}{\max L_A(B, \Sigma, \omega_1, \omega_2)} \tag{8}$$

where L_0 is the likelihood under the null hypothesis (pleiotropy) and L_A is the likelihood under the alternative hypothesis (separate QTL). We maximize each term in the fraction over its respective parameter space.

Visualizing profile LOD traces

Following an innovation of Zeng *et al.* (2000) and Tian *et al.* (2016), we plotted three traces: profile LOD_1 trace, profile LOD_2 trace, and the pleiotropy trace to aid in visualizing the two-dimensional likelihood surface. By definition (Equation 11), the maximum height of the pleiotropy trace is zero. Additionally, the maxima for profile LOD_1 trace and profile LOD_2 trace are equal and non-negative (Equations 9 and 10).

We define the LOD for our test:

$$LOD(\omega_1, \omega_2) = ll_{10}(\omega_1, \omega_2) - \max ll_{10}(\omega, \omega) \tag{9}$$

where ll_{10} denotes log-likelihood.

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

$$\text{profile LOD}_1(\omega_1) = \max_{\omega_2} \text{LOD}_{pvl}(\omega_1, \omega_2) \tag{10}$$

We define profile $\text{LOD}_2(\omega_2)$ analogously.

We constructed the pleiotropy trace by calculating the log-likelihoods for the pleiotropic models at every marker.

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

Bootstrap for test statistic calibration

We used 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, apparent deviations of the test statistic distribution from a chi-squared concerned us. Thus, we followed the approach of Tian *et al.* (2016) by identifying the marker that maximizes the likelihood under pleiotropy (the null hypothesis). We then used the inferred model parameters (under pleiotropy) at this marker to create bootstrap data sets, according to equations 1, 2, and 3. Specifically, we created b bootstrap data sets

for each analysis. We performed a two-dimensional QTL scan (over the genomic region of interest) with each of the b bootstrap samples to get b test statistic values. We treated these b test statistics as the empirical distribution with which we compared the test statistic for the observed data. We calculated the p-value as the proportion of the b test statistics (Λ_i) that equaled or exceeded the test statistic value for the true data, Λ:

$$p = \frac{\#\{i : \Lambda_i \geq \Lambda\}}{b} \tag{12}$$

where Λ_i denotes the likelihood ratio test statistic for the i^{th} bootstrap data set and Λ denotes the true likelihood ratio test statistic, *i.e.*, that for the observed data.

Data & Software Availability

We share an accessory R package that contains our functions at:
https://github.com/fboehm/qtl2pleio.
We also share our analysis and simulations R code at this Github repository:
https://github.com/fboehm/qt2pleio-manuscript (R Core Team 2018).
Data from Recla *et al.* (2014) and Logan *et al.* (2013) are in the Mouse Phenome Database at:
https://phenome.jax.org/projects/Chesler4 and https://phenome.jax.org/projects/Recla1.
We share these data in `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 statistical power to detect separate QTL. To simulate traits, we specified X, B, V_g, K, and V_e matrices (Equations 1, 2, 3). For all type I error rate studies, we used the allele probabilities from a single locus for 479 Diversity Outbred mice. For studies to assess power to detect separate QTL, we used the same 479 Diversity Outbred mice and allele probabilities of markers that are centered at the marker fro the type I error rate studies.

Type I error rate analysis

To quantify type I error rate (*i.e.*, false positive rate), we simulated 400 bivariate 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 explain below the reasoning behind our choice of levels for the three factors. 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 one level of the allele partitioning factor, we isolated the CAST allele (F). Three considerations motivated this choice. First, we observed in other analyses that CAST allele effects often differed from those of the other seven founder lines. Second, the phylogenetic distance between CAST and the other lines is greater than other inter-line distances (Didion and de Villena 2013). Third, we wanted to examine the effect of sample sizes (approximately equal vs unequal) in our analyses. We chose to evenly split the founder alleles into two groups of four for the other level of the allele partitioning factor (Table 1). We considered two values, 0 and 0.6, of the genetic correlation, *i.e.*, the off-diagonal entry in V_g. The marginal genetic variances (*i.e.*, the diagonal entries in V_g) for each trait were always set to one.

Run	Δ(Allele effects)	Allele effects partitioning	Genetic correlation	Type I error rate
1	6	ABCD:EFGH	0	0.0325
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	0.0375
6	12	ABCD:EFGH	0.6	0.0425
7	12	F:ABCDEGH	0	0.025
8	12	F:ABCDEGH	0.6	0.025

Table 1 Type I error rates for all eight 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. Founder allele one-letter abbreviations are listed in Table 2. In the column "Allele effects partitioning", "ABCD:EFGH" means that lines A, B, C, and D form one group while the other four lines are the second group. "F:ABCDEGH" means that the F allele is one group, while the seven other lines are a second group. Thus, for run 1, we assigned lines A, B, C, and D to have founder allele effects of 3, while E, F, G, and H had founder allele effects of -3.

We performed 400 tests per set of parameter inputs. Each test 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 the test is slightly conservative over our range of parameter selections (Table 1).

Power analysis

We also investigated the test's power to detect the presence of two distinct QTL. We used a 2 by 2 by 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 2). 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 bivariate traits per set of parameter inputs. For each simulated bivariate trait, we calculated a likelihood ratio test statistic. We then applied our parametric bootstrap to calibrate the test statistics. For each bivariate trait, 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, as we did in the type I error rate study above.

We present our power study results in Figure 1. Each curve tends upward monotonically as

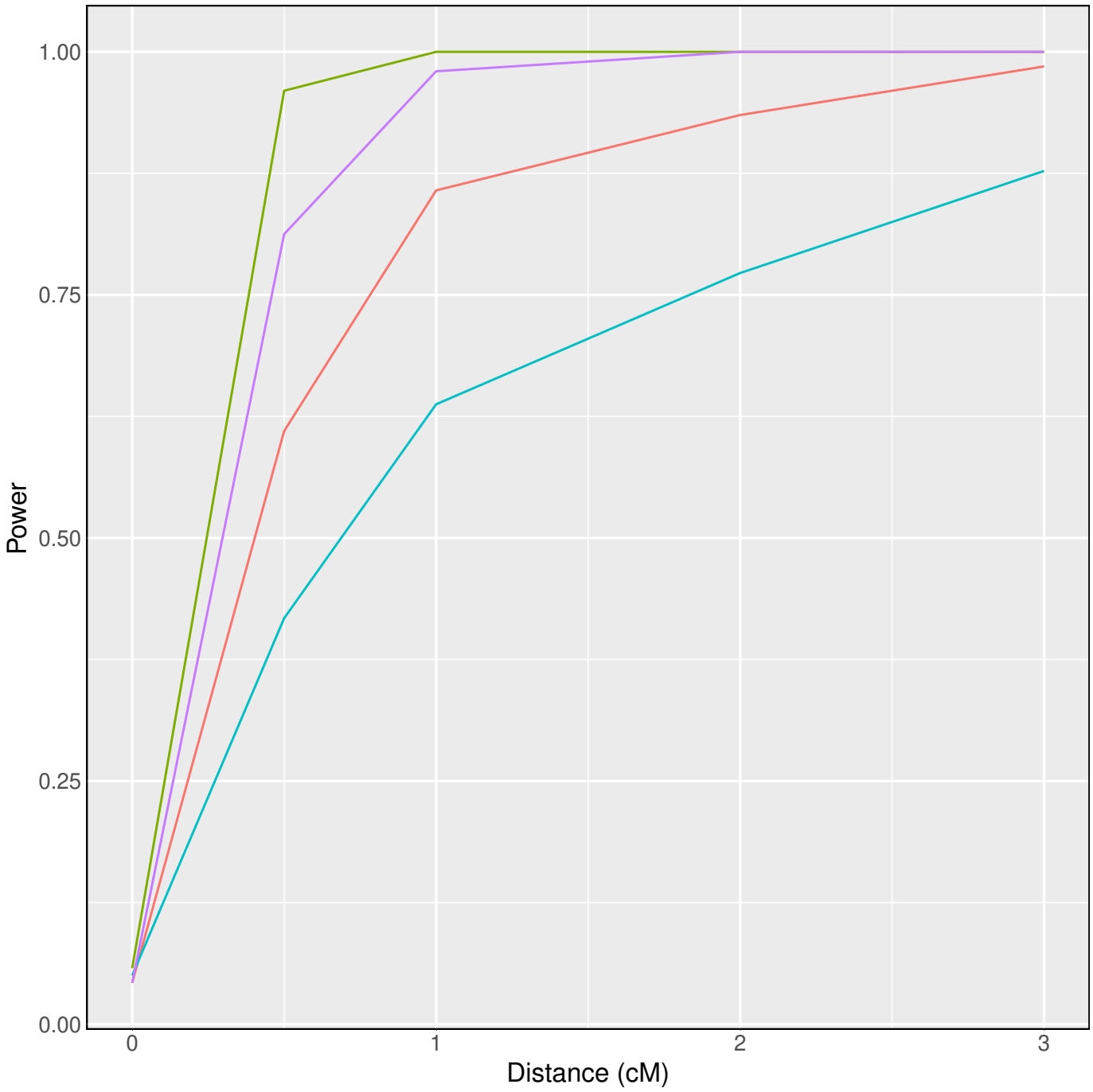

Figure 1 Pleiotropy vs. separate QTL power curves for each of four sets of parameter settings. Factors that differ among the four curves are allele effects difference and allele partitioning. Olive green denotes high (2) allele effects difference and allele partitioning into two groups of four. Purple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

interlocus distance increases. The curve with both the high value, 2, of allele effects difference and even allele partitioning into two groups of four founder alleles uniformly has the greatest power (olive green in Figure 1). The purple curve denotes the parameter input sets with uneven allele partitioning (F:ABCDEGH) and the high value (2) of allele effects difference. The two curves - red and blue - with the low value, 1, of allele effects difference are uniformly less powerful than those with the high value of allele effects difference. Within the two curves with allele effects difference equal to one, that with the even partitioning (ABCD:EFGH) of founder alleles has uniformly greater power.

Application

To illustrate our methods, we applied our test to data from Logan *et al.* (2013) and Recla *et al.* (2014). Logan *et al.* (2013) identified two phenotypes - "percent time in light" and "distance traveled in light" - that mapped to 55 cM on chromosome 8. This genomic region intrigued us because Recla *et al.* (2014) identified *Hydin* as the gene that underlies the mapping of "hot plate latency" to 57 cM (Supplementary figures). This led us to ask whether *Hydin* affects "percent time in light". We ignored "distance traveled in light" because of its high correlation (Pearson correlation 0.89) with "percent time in light".

We downloaded and formatted the data from the Mouse Phenome Database (Bogue *et al.* 2015). We inserted pseudomarkers at 0.1 cM intervals to bring the number of genome-wide markers to 20,183. We used three data structures: one for genotype data, one for phenotype data, and a third for covariates data. For each mouse and each marker, we have eight founder allele probabilities. We structured the allele probabilities for each chromosome as a 261 by 8 by m_i three-dimensional array, where m_i denotes the number of markers on chromosome i. We arranged phenotype data into a 261 by 3 matrix, where each column was a single phenotype and each row corresponded to a single mouse. We used only the sex covariate in our analyses.

We briefly describe the phenotypes that Logan *et al.* (2013) and Recla *et al.* (2014) measured. Both "percent time in light" and "distance traveled in light" were measured by placing mice in a light-dark box, consisting of an insert dividing an open-field apparatus into a light component and a dark component. Logan *et al.* (2013) then measured the percentage of 20 minutes time spent in the light component of the light-dark box. "Hot plate latency" reflects pain reflexes in response to a thermal stimulus, as recorded by a hot plate analgesia meter. Each mouse remained on the plate until it performed either of two behaviors regarded as indicative of nociception: hindpaw lick or hindpaw shake/flutter.

To determine whether *Hydin* affects "percent time in light", we examined a phenotype scatter plot, performed a test of pleiotropy vs. separate QTL, and created founder allele effects plots. A scatter plot revealed that "percent time in light" and "hot plate latency" are not highly correlated, with a Pearson correlation coefficient of -0.15 (Figure 2).

In examining the founder allele effects plot for "percent time in light", we see that the purple (WSB) and red (PWK) alleles have high effects near 55 cM, while the navy blue (NOD) has a dramatically lower value (Figure 4). The allele effects plot for "hot plate latency", on the other hand, features a noticeably distinct pattern of allele effects near 55 to 60 cM (Figure 4). Specifically, it has grey (B6) and green (CAST) alleles with much lower effects, while the pink (129), yellow (AJ), purple (WSB), and light blue (NZO) all show higher allele effects in this region.

We then performed a two-dimensional QTL scan for the pair of phenotypes. With these results, we created a profile LOD plot (Figure 5). The profile LOD plot for the analysis of "percent time in light" and "hot plate latency" displays a peak in the "percent time in light" profile LOD trace near 55 cM; in fact, it overlaps perfectly with the univariate LOD peak for "percent time in light". "Hot

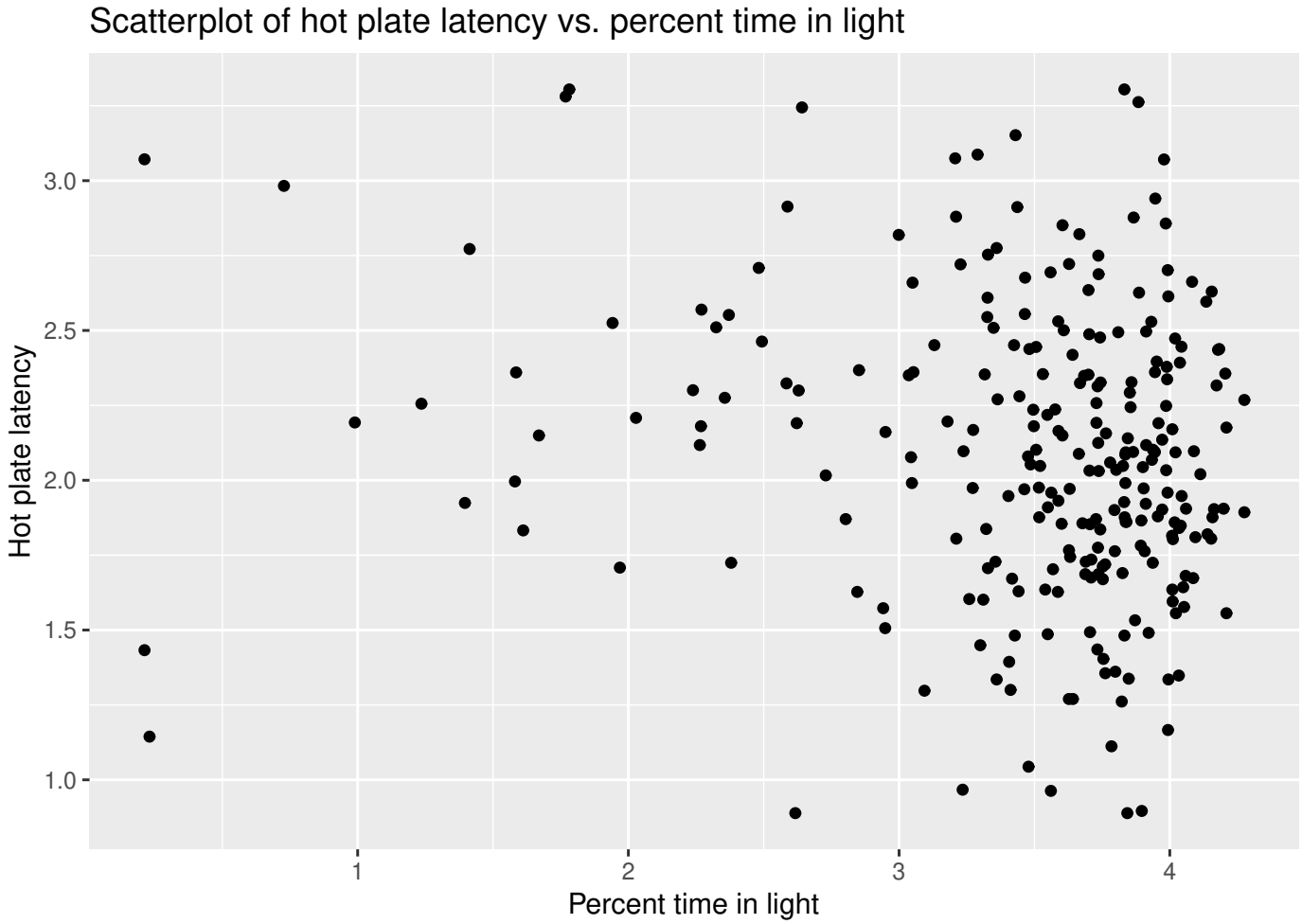

Figure 2 Scatter plot of "hot plate latency" against "percent time in light", after applying logarithm transformations and winsorizing both traits.

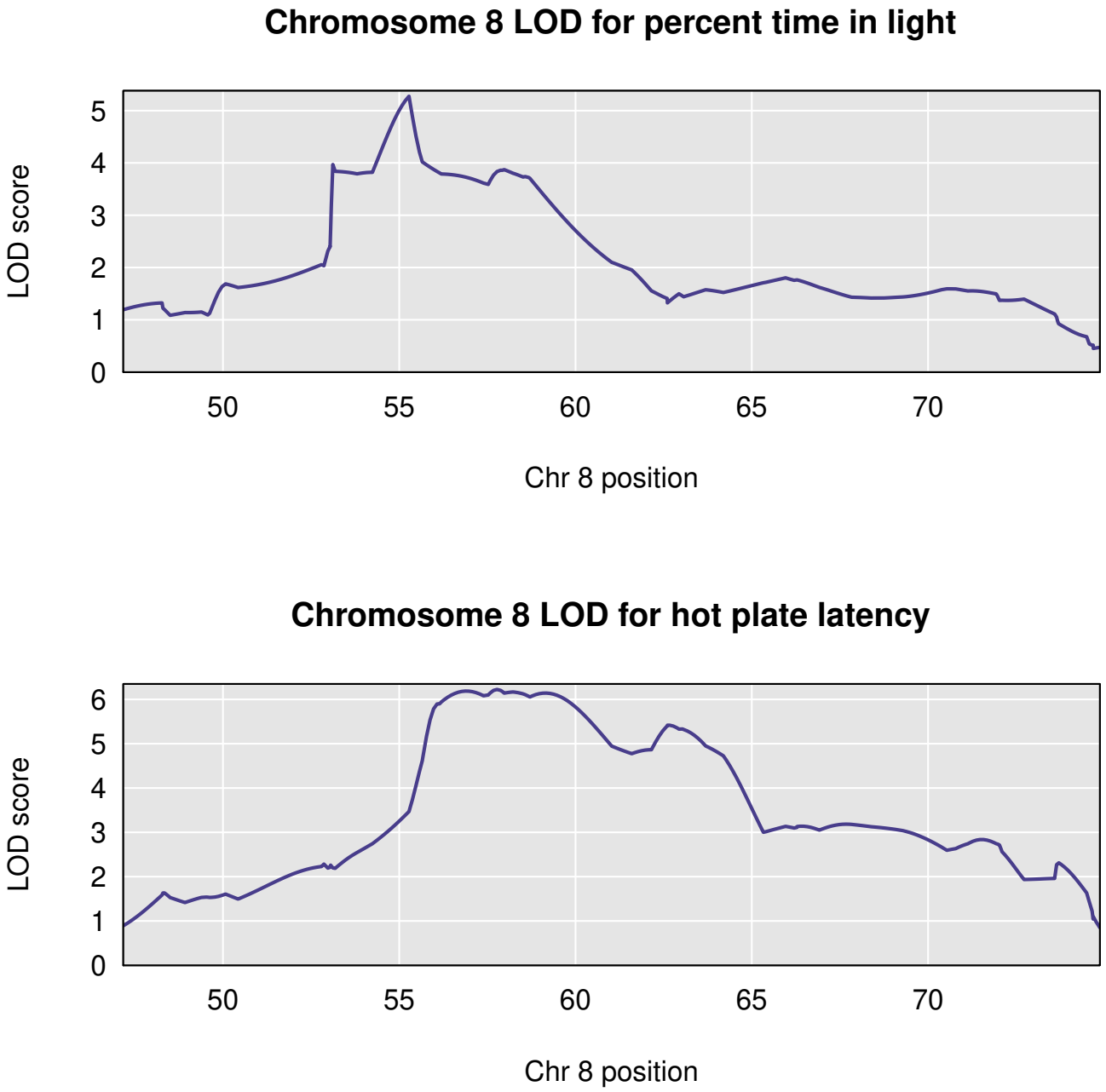

Figure 3 Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins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

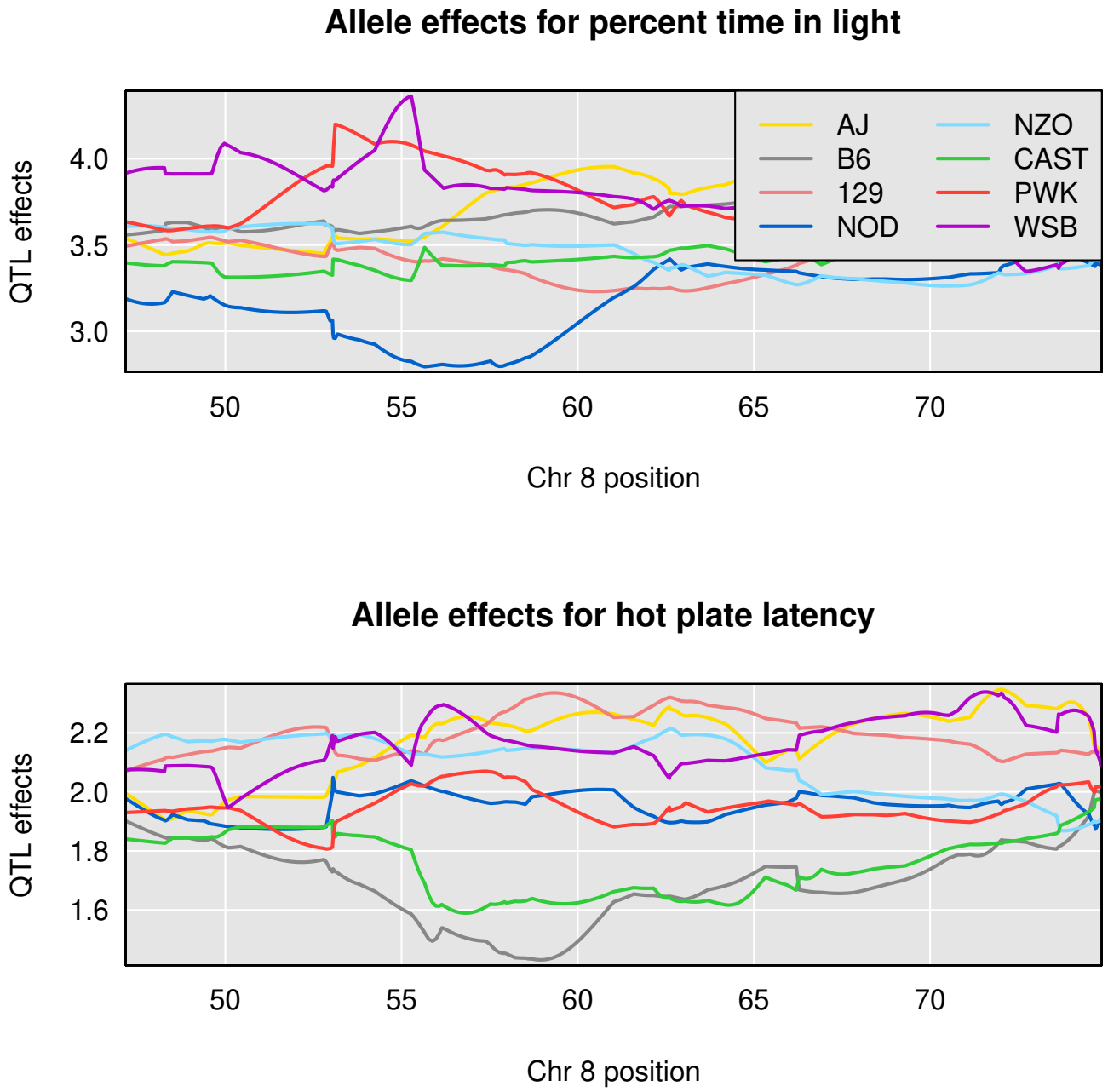

Figure 4 Chromosome 8 univariate LOD scores for percent time in light and hot plate latency reveal broad, overlapping peaks between 53 cM and 64 cM. The peak for percent time in light spans the region from approximately 53 cM to 60 cM, with a maximum near 55 cM. The peak for hot plate latency begins near 56 cM and ends about 64 cM.

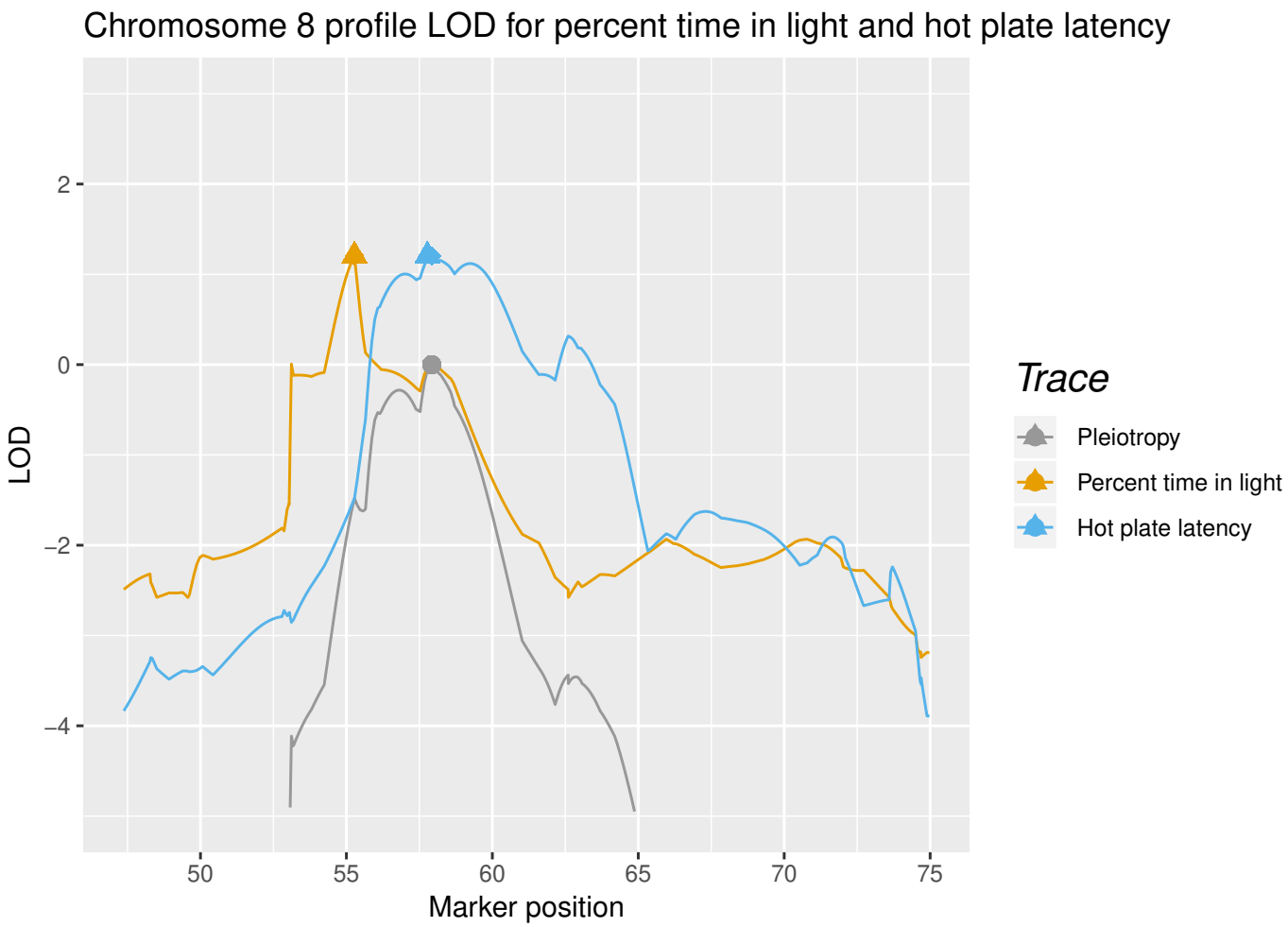

Figure 5 Profile LOD curves for the pleiotropy vs. separate QTL hypothesis test for "percent time in light" and "hot plate latency". Gray trace denotes pleiotropy LOD values. Triangles denote the univariate LOD maxima, while diamonds denote the profile LOD maxima. For "percent time in light", the yellow triangle obscures the smaller yellow diamond. Likelihood ratio test statistic value corresponds to the height of the blue and yellow traces at their maxima.

plate latency", on the other hand, demonstrates a broad peak in the profile LOD trace, with the maximum achieved near 57 cM (Figure 5A). The pleiotropy trace maximum occurs, too, near 57 cM. The maximal heights of the two profile LOD traces are near 1.2, which is the likelihood ratio test statistic value.

Using the two-dimensional QTL scan results, we calculated a likelihood ratio test statistic value: 1.20. We also calculated a bootstrap p-value with $b = 1000$. The p-value was 0.109. Having set type I error rate $\alpha = 0.05$, this p-values led us to fail to reject the null hypothesis of pleiotropy.

Discussion

We developed a test of pleiotropy vs. separate QTL for multiparental populations, characterized it with simulation studies, and applied it to answer a scientific question in behavioral genetics. 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).

We applied our test to two behavioral phenotypes in a study of 261 Diversity Outbred mice (Recla *et al.* 2014; Logan *et al.* 2013). Our results suggest the presence of two distinct QTL, with one QTL (that containing *Hydin*) affecting only "hot plate latency" and a second QTL affecting "percent time in light" (Figure 5). While our p-value of 0.109 led us to fail to reject the pleiotropy hypothesis, the modest p-value suggests the possibility of two distinct loci.

Founder allele effects plots provide further evidence for the presence of two distinct loci. As Macdonald and Long (2007) and King *et al.* (2012) argue 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".

Before discussing future directions for our research, we consider limitations of our current software and methods. Analyses with `qtl2pleio` are slow. One factor contributing to this is the fact that we wrote `qtl2pleio` mostly in R. R, due to its design, is slower than many compiled languages, such as C++ (Wickham 2014). To accelerate our multi-dimensional QTL scans, we integrated C++ code into `qtl2pleio` (Eddelbuettel *et al.* 2011).

Another computational bottleneck is the estimation of variance components V_g and V_e. In future research, where we may develop tests for simultaneous analysis of more than two traits, 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'll add this variance component estimation method to `qtl2pleio` by using the `reticulate` R package to expose objects from the python module `limmbo` (Allaire *et al.* 2018).

We now contrast pleiotropy vs. separate QTL tests with mediation tests in efforts to illustrate their complementary roles. Many systems genetics investigations apply mediation tests and related methods to infer 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. If a putative mediator sufficiently diminishes the LOD score, then the putative mediator is declared a mediator of the phenotype-QTL relationship.

An important distinction lies in the different goals for mediation tests and pleiotropy vs. separate QTL tests. In a mediation test, we seek to identify mediators of a phenotype-QTL relationship from a collection of putative mediators. In a pleiotropy vs. separate QTL test, one examines two (or more) phenotypes and inquires about the number of distinct QTL. A pleiotropy vs. separate QTL test provides information about the number of underlying QTL, while a mediation test offers conclusions about the relative ordering of biological intermediates within a molecular pathway.

Despite their distinct goals, Schadt *et al.* (2005) argue that both a pleiotropy vs. separate QTL test and causal inference methods, such as mediation tests, may contribute to gene network reconstruction. They develop a model selection strategy, based on Akaike information criterion (Akaike 1974), to determine which causal model is most compatible with the observed multivariate phenotype data (Schadt *et al.* 2005). In this context, Schadt *et al.* (2005) extend 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 vs. separate QTL test and mediation tests in the dissection of complex trait genetic architecture.

Two trends make our scientific contributions especially timely: the accelerating ability of biologists to measure tens of thousands of traits in model organisms and the proliferation of model

organism multiparental populations in diverse species. Technological advances in mass spectrometry and RNA sequencing enable 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

This research was performed using the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, 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.

We thank Lindsay Traeger, Julia Kemis, and Rene Welch for critiquing early versions of the manuscript.

This work was supported in part by National Institutes of Health grant R01GM070683 (to K.W.B.).

Literature Cited

Akaike, H., 1974 A new look at the statistical model identification. IEEE transactions on automatic control **19**: 716–723.

Allaire, J., K. Ushey, and Y. Tang, 2018 *reticulate: Interface to 'Python'*. R package version 1.6.

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.

Bogue, M. A., G. A. Churchill, and E. J. Chesler, 2015 Collaborative cross and diversity outbred data resources in the mouse phenome database. Mammalian Genome **26**: 511–520.

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.

Didion, J. P. and F. P.-M. de Villena, 2013 Deconstructing mus gemischus: advances in understanding ancestry, structure, and variation in the genome of the laboratory mouse. Mammalian genome **24**: 1–20.

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.

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.

R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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.

Wickham, H., 2014 *Advanced R*. CRC Press.

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.

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 2 Eight founder lines and their one-letter abbreviations.

Supplementary tables

lodcolumn	chr	pos	lod
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

Table 3 Both "hot plate latency" and "percent time in light" demonstrate multiple QTL peaks with LOD scores above 5.

Supplementary figures

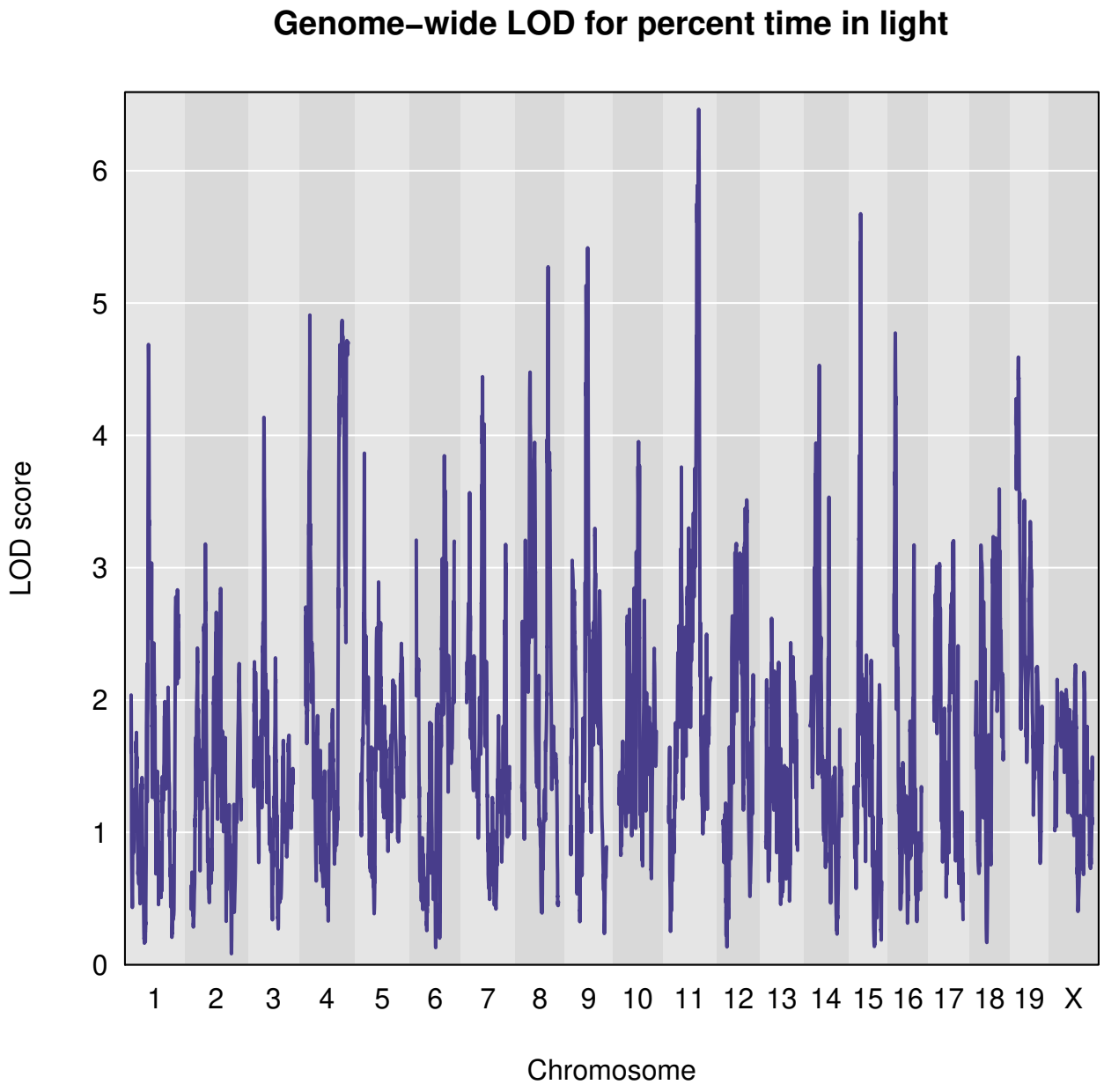

Figure 6 Genome-wide QTL scan for percent time in light reveals multiple QTL, including one on Chromosome 8.

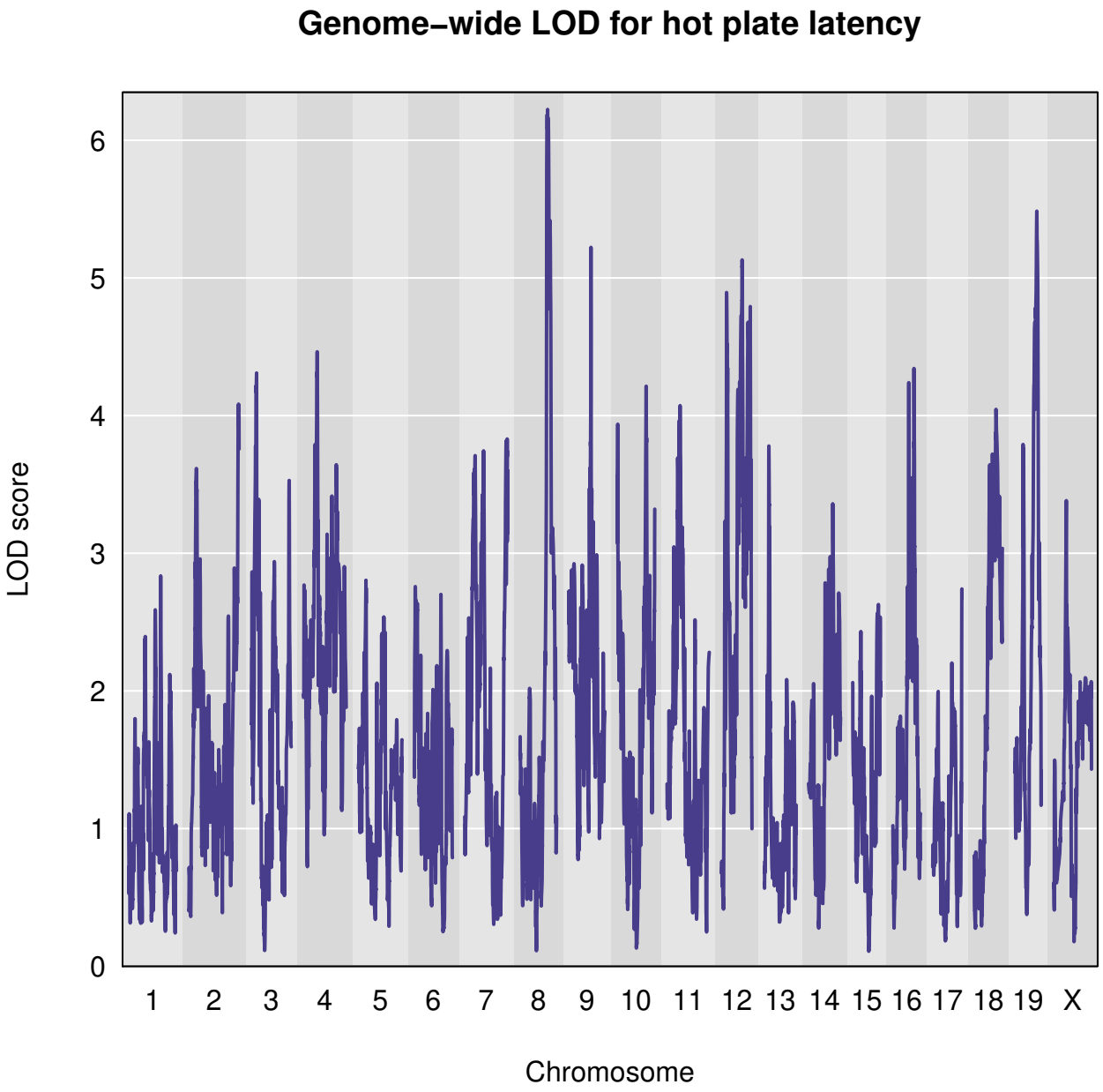

Figure 7 Genome-wide QTL scan for hot plate latency reveals multiple QTL, including a peak on Chromosome 8.