

Bayesian Modeling of Haplotype Effects in Multiparent Populations

Zhaojun Zhang,* Wei Wang,[†] and William Valdar^{*,1}

*Department of Computer Science and [†]Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, and [†]Department of Computer Science, University of California, Los Angeles, California 90095

ABSTRACT A general Bayesian model, Diploffect, is described for estimating the effects of founder haplotypes at quantitative trait loci (QTL) detected in multiparental genetic populations; such populations include the Collaborative Cross (CC), Heterogeneous Socks (HS), and many others for which local genetic variation is well described by an underlying, usually probabilistically inferred, haplotype mosaic. Our aim is to provide a framework for coherent estimation of haplotype and diplotype (haplotype pair) effects that takes into account the following: uncertainty in haplotype composition for each individual; uncertainty arising from small sample sizes and infrequently observed haplotype combinations; possible effects of dominance (for noninbred subjects); genetic background; and that provides a means to incorporate data that may be incomplete or has a hierarchical structure. Using the results of a probabilistic haplotype reconstruction as prior information, we obtain posterior distributions at the QTL for both haplotype effects and haplotype composition. Two alternative computational approaches are supplied: a Markov chain Monte Carlo sampler and a procedure based on importance sampling of integrated nested Laplace approximations. Using simulations of QTL in the incipient CC (pre-CC) and Northport HS populations, we compare the accuracy of Diploffect, approximations to it, and more commonly used approaches based on Haley–Knott regression, describing trade-offs between these methods. We also estimate effects for three QTL previously identified in those populations, obtaining posterior intervals that describe how the phenotype might be affected by diplotype substitutions at the modeled locus.

AN increasingly well-established paradigm for investigating the genetics of complex traits is the multiparent cross (Churchill *et al.* 2004; Cavanagh *et al.* 2008; Kover *et al.* 2009; Johannesson *et al.* 2009; Huang *et al.* 2012; King *et al.* 2012a; Svenson *et al.* 2012; Bandillo *et al.* 2013; Mackay *et al.* 2014), and an increasingly popular target for genetic and phenotypic analysis is its result, the multiparent population (*e.g.*, Talbot *et al.* 1999; Valdar *et al.* 2006a,b; Huang *et al.* 2009; Aylor *et al.* 2011; Huang *et al.* 2011; Collaborative Cross Consortium 2012; Baud *et al.* 2013; Marriage *et al.* 2014; Tsaih *et al.* 2014). In a multiparent cross, a select set of known founders is combined and bred to an advanced generation to create a population of individuals whose genomes are mosaics of the original founder haplotypes; this multiparent population

is then well suited for detection of quantitative trait loci (QTL) mapping through linkage disequilibrium (LD) mapping—that is, QTL mapping based on inferred descent. The development of statistical methods for LD mapping in these populations has largely focused on QTL detection (*e.g.*, Mott *et al.* 2000; Valdar *et al.* 2009; Durrant and Mott 2010; Huang and George 2011; Yuan *et al.* 2011; Zhang *et al.* 2012). Methods to characterize the effects at detected QTL, however, namely those estimating how inheritance of alternate founder haplotypes drives phenotypic outcome (a key step in the design of follow-up studies), are relatively underdeveloped. This is particularly so for populations in which the identity of the haplotypes at the QTL is probabilistically inferred, where estimation of haplotype effects must proceed despite the fact that haplotype composition is itself uncertain.

Uncertainty in haplotype composition arises because haplotypes are not themselves the direct target of genotyping or sequencing assays. In a multiparent population, where the number of founders is $J > 2$, the markers used to genotype individuals will not be fully informative for descent at all (or often any) loci, and so the underlying haplotype mosaic of

Copyright © 2014 by the Genetics Society of America

doi: 10.1534/genetics.114.166249

Manuscript received May 14, 2014; accepted for publication July 7, 2014

Available freely online through the author-supported open access option.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.166249/-/DC1>.

¹Corresponding author: Department of Genetics, University of North Carolina, Campus Box 7264, Chapel Hill, NC 27599. E-mail: william.valdar@unc.edu

each individual must be inferred. Inference of the haplotype mosaic is typically performed using a suitably constructed hidden Markov model (HMM) (e.g., Mott *et al.* 2000; Liu *et al.* 2010; King *et al.* 2012b; Gatti *et al.* 2014) and produces for each individual at each locus a list of probabilities describing likely descent. For a diploid organism, this list enumerates the posterior probability of having inherited each haplotype pair (diplotype), given available genotype information on the individual and its founders.

Use of inferred haplotype composition rather than observed genotypes when testing for genetic association confers important advantages, including: automatic modeling of all ungenotyped, incompletely determined, or entirely uncharacterized genetic variants; increased robustness to genotyping errors due to borrowing of information across markers; implicit modeling of LD decay, leading to a clearer picture of available mapping resolution; and automatic modeling of local epistasis. This last benefit arises from the fact that the heritable range of allelic combinations of all variants within a QTL region is concisely circumscribed by local haplotype descent; this provides a more comprehensive account of genetic variation and thereby allows QTL to be detected with increased power.

Strictly speaking, QTL mapping based on inferred haplotypes should account for haplotype uncertainty (Lander and Botstein 1989), an idealized framework inferring both association and composition simultaneously (Lin and Zeng 2006). For large-scale detection of QTL, however, such combined inference is usually computationally impractical; it is also largely unnecessary given the existence of fast and powerful regression-based alternatives. Specifically, rapid and highly flexible detection of QTL can be achieved by treating the inferred diplotype probabilities, or transformations thereof, as fixed covariates in an ordinary regression (Haley and Knott 1992; Mott *et al.* 2000; Zaykin *et al.* 2002). In particular, under additive genetic assumptions, the $J(J - 1)$ diplotype probabilities (or J^2 if phase is distinguished) can be condensed into a $(J - 1)$ -d.f. vector of expected haplotype counts; these “haplotype dosages” can then be used to detect additive “haplotype effects.” This general approach, which we term regression on probabilities (ROP), is not only highly scalable but also, due to the ubiquity and flexibility of linear modeling software, relatively simple to implement, at least once diplotype probabilities are provided. Indeed, for LD mapping in multiparent populations, ROP has become the dominant approach (Kover *et al.* 2009; Aylor *et al.* 2011; Svenson *et al.* 2012).

Although ROP is powerful for QTL detection, it is problematic when applied to subsequent characterization of QTL effects. Treating probabilities of unobserved diplotypes as if they were observed fractional dosages means that uncertainty about haplotype composition fails to propagate into haplotype effect estimation; although this has minimal (or at least controllable) impact on significance testing (e.g., Broman and Sen 2009; Rönnegård and Valdar 2011), it can lead to estimates of effects that are inflated, unstable, and

misleading. An artificial example is given in Table 1, in which at a given QTL each of two individuals has one of two diplotypes, A or B. Regression on the diplotypes (if known) would estimate the substitution effect as 1; regression on inferred diplotype probabilities, which in this example are highly uncertain but nonetheless accurate in placing more probability on the correct answer, leads to the difference in the estimated effects of diplotype A and B being 50. Both estimations fit their input data equally well; applied to new inputs of the *same form* (specifically, the same degree of uncertainty), they would give equally accurate predictions. However, using the ROP estimate of 50 to predict phenotype for individuals where diplotype is known (e.g., by design), or even where it is inferred with greater certainty, would clearly result in poor accuracy.

As the number of possible diplotype states grows (e.g., $J = 8$ founders implies 36 states, ignoring phase), the problem of inflated estimates increases and is compounded with additional problems of multicollinearity, whereby higher-order confounding in diplotype inference leads to linear dependence that in turn reduces the effective number of parameters estimable by ordinary regression. Although this multicollinearity can be circumvented by full rank factorization (e.g., appendix A of Valdar *et al.* 2009) or ridge-type penalization (e.g., Solberg Woods *et al.* 2012), such fixes severely limit the interpretability and/or validity of downstream inference.

Estimating effects of latent genetic states has been considered by a number of authors. For example, Sillanpää and Arjas (1998, 1999) advanced a fully Bayesian treatment for multilocus interval mapping in inbred and outbred populations derived from two founders. More recently, and directly relevant to multiparent populations, Kover *et al.* (2009), after using ROP to detect QTL in the *Arabidopsis* multiparent recombinant inbred population, estimated additive haplotype effects using multiple imputation: Sampling unobserved diplotypes from the inferred diplotype probabilities and then averaging least-squares estimates of haplotype effects from the imputed data sets. That approach was extended by Durrant and Mott (2010), who describe a partially Bayesian mixed model of QTL mapping: By focusing on additive effects of QTL for normally distributed traits with no additional covariates or population structure, they provided an efficient method for combined multiple imputation and shrinkage estimation via complete factorization of a pseudo-posterior.

Here we build on work of Kover *et al.* (2009), Durrant and Mott (2010), and others, developing a flexible framework for estimating haplotype-based additive and dominance effects at QTL detected in multiparent populations in which haplotype descent has been previously inferred. Our Bayesian hierarchical model, Diploffect, induces variable shrinkage to obtain full posterior distributions for additive and dominance effects that take account of both uncertainty in the haplotype composition at the QTL and confounding factors such as polygenic or sibship effects. In basing our model around existing, extendable software, we describe a flexible framework that accommodates non-normal phenotypes. In addition, by using a model

Table 1 Illustrative example of true diplotype state vs. inferred diplotype probabilities for two individuals at a QTL

Individual	True diplotype		Inferred diplotype probability		Phenotype
	A	B	A	B	
1	1	0	0.51	0.49	1
2	0	1	0.49	0.51	0

that is fully Bayesian, at least once conditioning on HMM-inferred diplotype probabilities, we exploit an opportunity untapped by earlier methods: The potential, when phenotypes and uncertain haplotypes are modeled jointly, for phenotypic data to inform and improve inference about haplotype configuration at the QTL as well as vice versa. To provide practical solutions and perspectives on relative trade-offs, we demonstrate two implementations of our model and compare their performance in terms of accuracy and running time to simpler procedures.

Statistical Models and Methods

We consider the following increasingly common scenario. A complex trait $\mathbf{y} = (y_1, \dots, y_n)$ has been measured in n individuals $i = 1, \dots, n$ from a multiparent population derived from $J \geq 2$ founders $j = 1, \dots, J$. Both the individuals and founders have been genotyped at high density, and, based on this information, for each individual descent across the genome has been probabilistically inferred. A one-dimensional genome scan of the trait has been performed using a variant of Haley–Knott regression, whereby a linear model (LM) or, more generally, a generalized linear mixed model (GLMM) tests at each locus $m = 1, \dots, M$ for a significant association between the trait and the inferred probabilities of descent. (Note that it is assumed that the GLMM could be controlling for multiple experimental covariates and effects of genetic background and that its repeated application for large M , both during association testing and in establishment of significance thresholds, may incur an already substantial computational burden.) This scan identifies one or more QTL; and for each such detected QTL, initial interest then focuses on reliable estimation of its marginal effects—specifically, the effect on the trait of substituting one type of descent for another, this being most relevant to follow-up experiments in which, for example, haplotype combinations may be varied by design.

To address estimation in this context, we start by describing a haplotype-based decomposition of QTL effects under the assumption that descent at the QTL is known. We then describe a Bayesian hierarchical model, Diploffect, for estimating such effects when descent is unknown but is available probabilistically. To estimate the parameters of this model, two alternate procedures are presented, representing different trade-offs between computational speed, required expertise of use, and modeling flexibility. A selection of alternative estimation approaches is then described, including a partially Bayesian approximation to Diploffect

and several non-Bayesian estimators that use regression on probabilities. (A summary list of all estimation procedures evaluated is given in Table 2.)

Haplotypes and diplotype states

The genetic state at locus m in each individual of a multiparent population can be described in terms of the pair of founder haplotypes present, that is, the diplotype state. We encode the diplotype state for individual i at locus m , using the $J \times J$ indicator matrix $\mathbf{D}_i(m)$, defined as follows. For maternally inherited founder haplotype $j \in \{1, \dots, J\}$ and paternally inherited haplotype $k \in \{1, \dots, J\}$, which together correspond to diplotype jk , the entry in the j th row and the k th column of $\mathbf{D}_i(m)$ is $\{\mathbf{D}_i(m)\}_{jk} = 1$, with all other elements being zero. Diplotype jk is defined as homozygous when $j = k$ and heterozygous when $j \neq k$. Under the heterozygote diplotype, when parent of origin is unknown or disregarded, $jk \equiv kj$ and it is assumed that $\{\mathbf{D}_i(m)\}_{jk} + \{\mathbf{D}_i(m)\}_{kj} = 1$.

Haplotype effects, diplotype effects, and dominance deviations

The effect at locus m of substituting one diplotype for another on the trait value can be expressed using a GLMM of the form $y_i \sim \text{Target}(\text{Link}^{-1}(\eta_i), \boldsymbol{\xi})$, where Target is the sampling distribution, Link is the link function, η_i models the expected value of y_i and in part depends on diplotype state, and $\boldsymbol{\xi}$ represents other parameters in the sampling distribution; for example, with a normal target distribution and identity link, $y_i \sim N(\eta_i, \sigma^2)$, and $E(y_i) = \eta_i$. In what follows, it is assumed that effects of other known influential factors, including other QTL, polygenes, and experimental covariates, are modeled to an acceptable extent within the GLMM itself, either implicitly in the sampling distribution or explicitly through additional terms in η_i .

Under the assumption that haplotype effects combine additively to influence the phenotype, the linear predictor can be minimally modeled as

$$\eta_i = \mu + \boldsymbol{\beta}^T \mathbf{add}(\mathbf{D}_i(m)), \quad (1)$$

where $\mathbf{add}(\mathbf{X}) = \mathbf{1}^T(\mathbf{X} + \mathbf{X}^T)$ such that $\boldsymbol{\beta}$ is a zero-centered J -vector of (additive) *haplotype effects*, and μ is an intercept term. The assumption of additivity can be relaxed to admit effects of dominance by introducing a *dominance deviation*:

$$\eta_i = \mu + \boldsymbol{\beta}^T \mathbf{add}(\mathbf{D}_i(m)) + \boldsymbol{\gamma}^T \mathbf{dom}(\mathbf{D}_i(m)). \quad (2)$$

The definitions of $\mathbf{dom}(\mathbf{X})$ and $\boldsymbol{\gamma}$ depend on whether the reciprocal heterozygous diplotypes jk and kj are modeled to have equivalent effects. If so, then dominance is *symmetric*: $\mathbf{dom}(\mathbf{X})$ is defined as $\mathbf{dom.sym}(\mathbf{X}) = \text{vec}(\text{upper.tri}(\mathbf{X} + \mathbf{X}^T))$, where upper.tri() returns only elements above the diagonal of a matrix, and zero-centered effects vector $\boldsymbol{\gamma}$ has length $J(J - 1)/2$. Otherwise, if diplotype effects are modeled to differ by parent of origin, then dominance is *asymmetric*: $\mathbf{dom}(\mathbf{X})$ is defined as $\mathbf{dom.asym}(\mathbf{X}) = \text{vec}(\text{off-diag}(\mathbf{X}))$,

Table 2 Summary of the estimation procedures evaluated in this article

Model	Description	ROP
partial.lm	Each haplotype (additive) effect estimated by single predictor regression	Yes
ridge.add	Ridge regression modeling additive effects	Yes
ridge.dom	Ridge regression modeling both additive and dominance effects	Yes
DF.IS.noweight	Approx. Diploffect using INLA and multiple imputation	No
DF.IS	Diploffect using INLA and importance sampling	No
DF.IS.kinship	DF.IS including a polygenic effect	No
DF.MCMC.pseudo	Approx. Diploffect model using partial MCMC and multiple imputation	No
DF.MCMC	Diploffect using MCMC	No

where off-diag returns all off-diagonal elements, and zero-centered $\boldsymbol{\gamma}$ has length $J^2 - J$. Throughout the remainder of the article, for simplicity, dominance is modeled as symmetric. For notational convenience, we further define the *diploftype effects*, $\boldsymbol{\delta}$, as composed of

$$\delta_{jk} = \beta_j + \beta_k + I(j \neq k)\gamma_{(jk)} \quad (3)$$

for all distinguishable jk . Definitions of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ (if instead estimated directly) are analogous to “additive” and “full” models defined in Yalcin *et al.* (2005).

Haplotype inference and diplotype probabilities

In practice, the diplotype state at a locus m cannot be observed directly, but it can be inferred probabilistically from genotype data (*e.g.*, Broman and Sen 2009). Denoting available genotype data on individuals as $\mathcal{G} = \{\mathcal{G}_1, \dots, \mathcal{G}_n\}$ and genotype information on the founders as $\mathcal{H} = \{\mathcal{H}_1, \dots, \mathcal{H}_J\}$, haplotype reconstruction algorithms based on a hidden Markov model typically seek to estimate for each individual i at each locus $m = 1, \dots, M$ a $J \times J$ matrix of inferred diplotype probabilities,

$$\mathbf{P}_i(m) = p(\mathbf{D}_i(m) | \mathcal{G}_i, \mathcal{H}), \quad (4)$$

where each element $\{\mathbf{P}_i(m)\}_{jk}$ contains the probability that diplotype jk is present and where in more sophisticated algorithms additional terms may be present in the conditioning statement (*e.g.*, \mathcal{G} in place of \mathcal{G}_i). Diplotype state is therefore modeled as if drawn from a categorical distribution with probability parameter $\mathbf{P}_i(m)$, *i.e.*,

$$\mathbf{D}_i(m) \sim \text{Cat}(\mathbf{P}_i(m)), i = 1, \dots, n. \quad (5)$$

In the HAPPY formulation of Mott *et al.* (2000), which we adopt here, element $\{\mathbf{P}_i(m)\}_{jk}$ is the HMM-derived Baum–Welsh probability of diplotype jk , averaged over the interval between two adjacent markers m and $m + 1$. In other words, $\{\mathbf{P}_i(m)\}_{jk}$ is approximately the probability that a randomly chosen point within the interval inherits from the diplotype jk . When descent is both constant within the interval and unambiguous, $\mathbf{P}_i(m) = \mathbf{D}_i(m)$; otherwise $\mathbf{P}_i(m)$ represents a hedged bet on which diplotype is present, and typically becomes less informed as a function of marker sparsity, recombination density, and genotyping error.

Diploffect: Bayesian inference of QTL effects conditional on diplotype probabilities

When haplotype composition at a QTL, $\boldsymbol{\Delta} = \{\mathbf{D}_1(m), \dots, \mathbf{D}_n(m)\}$, is available only as inferred diplotype probabilities, $\boldsymbol{\Psi} = \mathbf{P}_1(m), \dots, \mathbf{P}_n(m)$, estimation of QTL effects can be naturally framed as a Bayesian integration. Defining for convenience all parameters relating to the QTL effect (and not to diplotype state) as $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \dots\}$, diplotype states $\boldsymbol{\Delta}$ and effects $\boldsymbol{\theta}$ can be considered latent variables in the joint posterior

$$p(\boldsymbol{\theta}, \boldsymbol{\Delta} | \boldsymbol{\Psi}, \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\Delta}) p(\boldsymbol{\Delta} | \boldsymbol{\Psi}) p(\boldsymbol{\theta}). \quad (6)$$

Here, the phenotype is modeled in the likelihood $p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\Delta})$ as a function of the effects $\boldsymbol{\theta}$ and the diplotypes $\boldsymbol{\Delta}$; the diplotype states $\boldsymbol{\Delta}$ are in turn modeled as latent categorical variables with prior $p(\boldsymbol{\Delta} | \boldsymbol{\Psi})$. Two important consequences follow. First, the posterior distribution of effects

$$p(\boldsymbol{\theta} | \boldsymbol{\Psi}, \mathbf{y}) = \int p(\boldsymbol{\theta}, \boldsymbol{\Delta} | \boldsymbol{\Psi}, \mathbf{y}) d\boldsymbol{\Delta}, \quad (7)$$

from which can be obtained all estimates and intervals of haplotype effects, dominance deviations, and diplotype effects, averages over plausible diplotype configurations; this leads to effect estimates of $\boldsymbol{\theta}$, including interval estimates, that incorporate uncertainty in diplotype state. Second, a posterior distribution conditional on the phenotype \mathbf{y} is generated for the diplotype state $\boldsymbol{\Delta}$:

$$p(\boldsymbol{\Delta} | \boldsymbol{\Psi}, \mathbf{y}) = \int p(\boldsymbol{\theta}, \boldsymbol{\Delta} | \boldsymbol{\Psi}, \mathbf{y}) d\boldsymbol{\theta}. \quad (8)$$

This posterior is a Bayesian update of prior $\boldsymbol{\Psi} = p(\boldsymbol{\Delta} | \mathcal{G}, \mathcal{H})$ from Equation 4. Specifically, since the prior of diplotypes is a categorical distribution, the marginal posterior of the diplotype is also categorical,

$$p(\mathbf{D}_i(m) | \boldsymbol{\theta}, \boldsymbol{\Psi}, \mathbf{y}) \sim \text{Cat}(Q(\mathbf{D}_i(m)_{11}), Q(\mathbf{D}_i(m)_{12}), \dots, Q(\mathbf{D}_i(m)_{JJ})), \quad (9)$$

where

$$Q(\mathbf{D}_i(m)_{jk}) = p(\mathbf{D}_i(m)_{jk} = 1 | y_i, \boldsymbol{\theta}, \mathbf{P}_i(m)) \propto \underbrace{\mathbf{P}_i(m)_{jk}}_{\text{prior}} \times \underbrace{p(y_i | \boldsymbol{\theta}, \mathbf{D}_i(m)_{jk} = 1)}_{\text{likelihood}}. \quad (10)$$

This updating of diplotype state for each individual $i = 1, \dots, n$ in light of phenotypic information reflects the following intuition: Suppose that prior to observing \mathbf{y} , diplotype probabilities $\mathbf{P}_1, \dots, \mathbf{P}_{n-1}$ are well informed but \mathbf{P}_n is not; if analysis with \mathbf{y} reveals a clear pattern of effects (e.g., high phenotype values associated with particular diplotype states), then y_n provides information to update \mathbf{P}_n . Moreover, it implies that different phenotypes could in theory promote different underlying diplotype states Δ —a useful feature when locus m is defined broadly enough to contain multiple recombinants and therefore multiple configurations of Δ , of which only one is relevant to the interrogated QTL.

The likelihood conditional on diplotype states, $p(\mathbf{y}|\boldsymbol{\theta}, \Delta)$, is specified in terms of a GLMM with linear predictor

$$\eta_i = \mu + \underbrace{\boldsymbol{\alpha}^T \mathbf{x}_i}_{\substack{\text{fixed covariates} \\ \text{(optional)}}} + \underbrace{\sum_{r \in \mathcal{R}} \left(\mathbf{z}_i^{(r)} \right)^T \mathbf{u}^{(r)}}_{\substack{\text{random effects} \\ \text{(optional)}}} + \underbrace{\boldsymbol{\beta}^T \text{add}(\mathbf{D}_i(m))}_{\text{additive haplotype effects}} + \underbrace{\boldsymbol{\gamma}^T \text{dom}(\mathbf{D}_i(m))}_{\text{dominance deviation}}, \quad (11)$$

where terms are included above for fixed effects $\boldsymbol{\alpha}$ based on covariates in \mathbf{x}_i and for an arbitrary set of variance components \mathcal{R} , each with incidence vector $\mathbf{z}_i^{(r)}$ and effect vector modeled as $\mathbf{u}^{(r)} \sim N(\mathbf{0}, \mathbf{R}^{(r)} \tau_r^2)$. Such additional terms are here defined loosely because our model is deliberately presented within established, extensible software that allows bespoke specification. In the current study, we demonstrate use of the following variance components: cage, where for c cages, $\mathbf{R}^{(\text{cage})} = \mathbf{I}_c$; sibship, where for s sibships, $\mathbf{R}^{(\text{sibship})} = \mathbf{I}_s$; and polygenic effects approximated by an animal model (hereafter, a “kinship” effect), where $\mathbf{R}^{(\text{kinship})} = \mathbf{K}$ is the $s \times s$ additive relationship matrix estimated from the pedigree (Kennedy *et al.* 1992). Haplotype effects $\boldsymbol{\beta}$ and dominance deviations $\boldsymbol{\gamma}$ are modeled hierarchically as $\boldsymbol{\beta} \sim N(0, \mathbf{I} \tau_{\text{add}}^2)$ and $\boldsymbol{\gamma} \sim N(0, \mathbf{I} \tau_{\text{dom}}^2)$; this hierarchical modeling benefits estimation in two ways—it allows information to be borrowed across effects on related scales, and it provides robustness to data sets where the sampling of diplotypes is sparse. The remaining parameters are given weak, conjugate priors: Fixed effects ($\mu, \boldsymbol{\alpha}$) are given independent normal priors $N(0, c)$, where c is large relative to the phenotype scale [e.g., $c = 1000$ for $\text{Var}(y) = 1$]; and dispersions $\sigma^2, \tau_{\text{add}}^2, \tau_{\text{dom}}^2$, and each τ_r^2 are given inverse-gamma priors as in, e.g., Lenarcic *et al.* (2012). The complete Diploffect model, shown with a polygenic effect, is summarized using plate notation in Figure 1.

The posterior of effects integrated in Equation 7 involves integrating over a J^{2n} -dimensional space. We consider two alternatives for sampling from this posterior below.

Diploffect estimation by MCMC: DF.MCMC

Posteriors for all parameters in the Diploffect model can be estimated by Markov chain Monte Carlo (MCMC) by

iterating two basic Gibbs sampling steps. At each iteration, $k = 1, \dots, K$:

1. Sample all effect variables $\boldsymbol{\theta}^{(k+1)}$ given the previous iteration’s diplotypes $\Delta^{(k)}$,

$$\boldsymbol{\theta}^{(k+1)} \sim p(\boldsymbol{\theta}|\mathbf{y}, \Delta^{(k)}). \quad (12)$$

2. Sample diplotypes $\Delta^{(k+1)}$ given effect variables $\boldsymbol{\theta}^{(k+1)}$,

$$\Delta^{(k+1)} \sim p(\Delta^{(k+1)}|\boldsymbol{\Psi}, \boldsymbol{\theta}^{(k+1)}, \mathbf{y}). \quad (13)$$

Initial values for $k = 1$ are randomly sampled from their priors.

Although relatively efficient Gibbs sampling schemes for step 1 are well established (we use those provided in Plummer 2003; see *Implementation details*), step 2 requires special consideration. A straightforward approach is to sample from the full conditional, evaluating all diplotypes’ posterior probabilities in $\mathbf{D}_i(m)$ by Equation 10 and drawing a diplotype state for each individual in turn. Per individual, however, this incurs $O(J^2)$ computational time because it requires evaluating the function Q for all diplotypes. For the sake of efficiency, we develop an optimization, discrete slice sampling with prior reordering, described in *Appendix A*, which makes this sampling more efficient. Hereafter we refer to this method as Diploffect estimation by MCMC (DF.MCMC).

Diploffect estimation by importance sampling: DF.IS and DF.IS.kinship

Seeking a noniterative estimation procedure that is more efficient for standard GLMMs, we also provide a strategy based on Importance Sampling (IS) of integrated nested Laplace approximations (INLA). INLA provides a deterministic estimate of the multivariate posterior distribution of a GLMM (Rue *et al.* 2009), giving analytic approximations for effects and sampling approximations for variances. In our IS procedure, these posteriors are estimated conditional on diplotype for many possible diplotype configurations; they are then combined through reweighting to give a final mixture distribution that resembles more closely the integration of the full posterior in Equation 7. Specifically, the procedure is

1. Sample diplotypes $\Delta^{(k)}$ from their prior, $\Delta^{(k)} \sim p(\boldsymbol{\Psi})$.
2. Obtain an INLA estimate of posterior $p(\boldsymbol{\theta}|\mathbf{y}, \Delta^{(k)})$ for effect variables $\boldsymbol{\theta}^{(k)}$.
3. Obtain an INLA estimate of the marginal likelihood $w^{(k)} = p(\mathbf{y}|\Delta^{(k)})$.
4. Repeat steps 1–3 K times.
5. Estimate the posterior of any statistic of interest $T(\boldsymbol{\theta})$, using the weighted mixture

$$\hat{T}_{\text{IS}}(\boldsymbol{\theta}) = \frac{\sum_k T(\boldsymbol{\theta}^{(k)}) w^{(k)}}{\sum_k w^{(k)}}, \quad (14)$$

where for each k , statistic $T(\boldsymbol{\theta}^{(k)})$ is calculated from the corresponding posterior $p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\Delta}^{(k)})$ calculated in step 2.

Calculation of the weighting function $w^{(1)}, \dots, w^{(K)}$ uses the marginal likelihood obtained from INLA and is described more fully in *Appendix B*. The statistic $T(\boldsymbol{\theta})$ is defined in this study according to the following requirements: for point estimation is needed, we use the posterior mean $T(\boldsymbol{\theta}) = E(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\Delta})$; for obtaining highest posterior density (HPD) intervals of effects parameters, $T(\boldsymbol{\theta})$ records the analytic approximation of $p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\Delta})$; and for estimating the additive vs. dominance proportion, $p(\pi_{\text{add}}|\mathbf{y})$, where $\pi_{\text{add}} = \tau_{\text{add}}^2 / (\tau_{\text{add}}^2 + \tau_{\text{dom}}^2)$, $T(\boldsymbol{\theta})$ records 1000 posterior samples from $p(\pi_{\text{add}}|\mathbf{y}, \boldsymbol{\Delta})$.

Importance sampling of the above mixture model can be highly inefficient and lead to unstable results when the mixture prior $p(\boldsymbol{\Phi})$ is uninformed; in particular, a large number of samples drawn from the prior may, after reweighting, translate into a comparatively tiny number of samples from the posterior (Carlin and Louis 2008). To track sampling efficiency, we therefore use the effective sample size (ESS) metric of Liu *et al.* (2001) equivalent to $\text{ESS} = [\sum_k w^{(k)}]^2 / [\sum_k (w^{(k)})^2]$ (Robert and Casella 2009).

Application of the IS procedure is denoted Diploffect estimation by importance sampling (DF.IS) in all cases except when the GLMM includes a kinship effect, in which case it is denoted DF.IS.kinship. This distinction is made in light of the fact that fitting kinship effects (as opposed to approximating them with, *e.g.*, sibship effects) can incur significant additional computation.

Partial Bayesian approximation: DF.MCMC.pseudo and DF.IS.noweight

In cases where it can be assumed that $p(\boldsymbol{\Delta}|\boldsymbol{\Psi}) \approx p(\boldsymbol{\Delta}|\boldsymbol{\Psi}, \mathbf{y})$, for example, where the QTL effect is weak or the posteriors of the diplotypes are consistently vague across individuals, integration of the joint posterior in Equation 7 can be approximated as

$$p(\boldsymbol{\theta}|\boldsymbol{\Psi}, \mathbf{y}) \approx \int p(\boldsymbol{\theta}|\boldsymbol{\Delta}, \boldsymbol{\Psi}, \mathbf{y}) p(\boldsymbol{\Delta}|\boldsymbol{\Psi}) d\boldsymbol{\Delta}. \quad (15)$$

This approximation, essentially a form of unweighted multiple imputation, is used by Durrant and Mott (2010) to estimate haplotype effects at QTL in populations of recombinant inbreds. By restricting attention to normally distributed traits with no covariates or structure, they develop a method to sample from the above pseudoposterior directly. To explore the utility of this approximation, we implement versions of it based on both DF.MCMC and DF.IS. In DF.MCMC.pseudo, the sampling of the posterior of $\boldsymbol{\Delta}$ conditional on the current value of $\boldsymbol{\theta}$ in Equation 13 is replaced by a draw from the prior, $\boldsymbol{\Delta}^{(k)} \sim p(\boldsymbol{\Psi})$; this method was recently used by us in the analysis of immune phenotypes in the pre-CC (Phillippi *et al.* 2014). In DF.IS.noweight, DF.IS is modified so that

weights are uniform, *i.e.*, $w^{(k)} = 1$ for all k ; this approach is similar to that used in the *Arabidopsis* study of Kover *et al.* (2009), who instead estimate $\boldsymbol{\beta}$ via a fixed-effects regression.

Non-Bayesian approximations using regression on probabilities: partial.lm, ridge.add, and ridge.dom

ROP methods model the phenotype in a regression where diplotype states (and functions thereof) are represented by their corresponding probabilities of being present. Use of probabilities in this way, substituting functions of $\mathbf{D}_i(m)$ with functions of $\mathbf{P}_i(m)$, can provide stable estimation when $\mathbf{P}_i(m)$ probabilities are well informed; otherwise, when uncertainty is present, the design matrix can become multicollinear, making some effects nonidentifiable and thus ineligible for a fair comparison with the truth. Rather than providing a comprehensive survey of ROP, we therefore consider three illustrative examples that at least guarantee identifiability under all possible levels of haplotype uncertainty: a marginal estimator, partial.lm; and two ridge regression estimators, ridge.add and ridge.dom.

The marginal estimator partial.lm uses a single predictor linear model to estimate, for each founder haplotype in turn, the effect of that haplotype's dosage on the phenotype; *i.e.*,

$$\eta_i = \mu_j + \beta_j \{\text{add}(\mathbf{P}_i(m))\}_j,$$

where β_j is the j th element of $\boldsymbol{\beta}$. This method was used to estimate effects in the pre-CC study of Aylor *et al.* (2011).

In ridge.add, ridge regression (Hoerl and Kennard 1970) is applied to a ROP form of the additive model of Equation 1,

$$\eta_i = \mu + \boldsymbol{\beta}^T \text{add}(\mathbf{P}_i(m)), \quad (16)$$

with $\boldsymbol{\beta}$ estimated by minimizing $\sum_i (y_i - \eta_i)^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{\beta}$, where tuning parameter λ is set by 10-fold cross-validation. In ridge.dom, an analogous model is fitted based on the additive plus dominance model of Equation 2,

$$\eta_i = \mu + \boldsymbol{\beta}^T \text{add}(\mathbf{P}_i(m)) + \boldsymbol{\gamma}^T \text{dom}(\mathbf{P}_i(m)). \quad (17)$$

Implementation details

MCMC-based approaches (DF.MCMC and DF.MCMC.pseudo) were implemented in R (R Development Core Team 2011), JAGS (Plummer 2003), and rjags (Plummer 2014). JAGS is an open-source general MCMC sampling package; we implemented add-on code to support the partially Bayesian prior sampling of DF.MCMC.pseudo (see code in File S4). MCMC was performed for 5000 time steps, of which the first 1000 were discarded as burn-in, and the remaining 4000 were thinned at 1/10 to give 400 usable samples. Importance sampling approaches (DF.IS, DF.IS.noweight, and DF.IS.kinship) were implemented using the R package INLA (Rue *et al.* 2009). In each application of the IS methods we used 1000 independent samples directly drawn from the haplotype probabilities inferred by HAPPY (Mott *et al.* 2000; Mott 2009). Estimation of the additive relationship

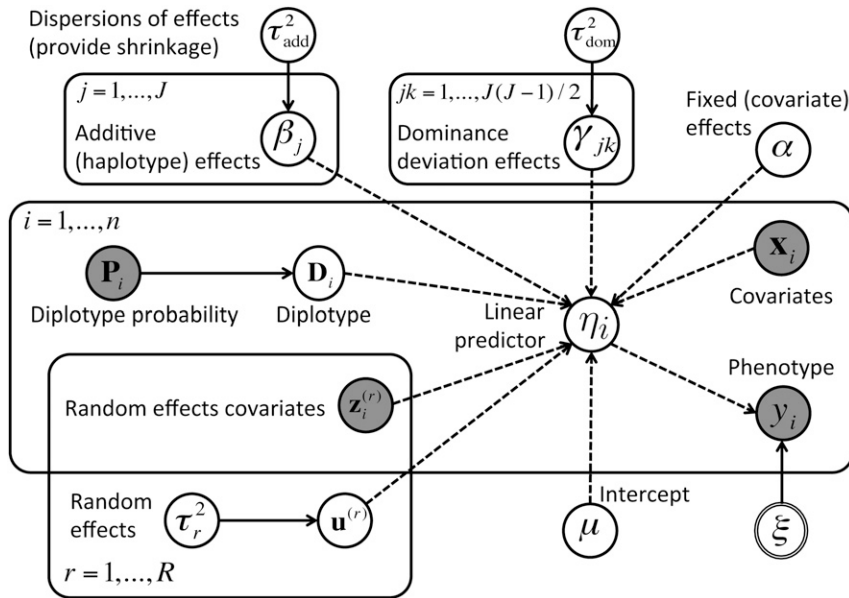


Figure 1 The Diploffect model depicted as a directed acyclic graph. Dashed arrows indicate deterministic relationships and solid arrows indicate stochastic relationships. Shaded nodes are observed variables, and open nodes are unobserved variables, with a double circle representing the remaining parameters; priors are omitted. The number of instances of each variable is shown using plate notation.

matrix was performed using the R package *pedigreemm* (Vazquez *et al.* 2010). Ridge regression was performed using the R package *GLMNet* (Friedman *et al.* 2010), with tuning parameters selected by 10-fold cross-validation. All other analysis was performed in R.

Data and Simulations

We use simulation to evaluate the ability of our Diploffect model to estimate haplotype and diplotype effects at a single QTL segregating in a multiparent population. It is assumed that the QTL location has been determined already and phenotype information per individual is available, but diplotype state at the QTL for each individual is available only as inferred diplotype probabilities. For methods in Table 2, we assess subsequent estimation in terms of both numerical accuracy and ability to rank effects under a range of QTL effect sizes and in different genetic contexts. Practical use of the Diploffect model is then illustrated through application to real, previously mapped QTL. Both simulation and application use data from two real populations: the incipient strains of the Collaborative Cross (pre-CC) (Aylor *et al.* 2011) and the Northport HS mice (Valdar *et al.* 2006a). These data sets are described below.

Pre-CC data set

The Collaborative Cross (Collaborative Cross Consortium 2012) is a large panel of recombinant inbred lines bred from a set of eight inbred founder mouse strains (abbreviated names in parentheses): 129S1/SvImJ (129S1), A/J (AJ), C57BL/6J (B6), NOD/ShiLtJ (NOD), NZO/HILtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB). Breeding of the CC is an ongoing effort, and at the time of this writing a relatively small number of finalized lines are available. Nonetheless, partially inbred lines taken from an

early stage of the CC breeding process, the so-called pre-CC population, have been studied and used for QTL identification (Aylor *et al.* 2011; Kelada *et al.* 2012; Ferris *et al.* 2013; Phillippi *et al.* 2014). The pre-CC data set analyzed here is that from the study of Aylor *et al.* (2011). This comprises data for 184 mice from 184 independent pre-CC lines (*i.e.*, one replicate per line); these lines had attained on average 6.7 generations of inbreeding following the initial eight-way cross and as a result have genomes with $\sim 16\%$ residual heterozygosity. Aylor *et al.* (2011) used HAPPY (Mott *et al.* 2000) to generate diplotype probability matrices for all 184 mice based on genotype information for 16,159 markers across the genome. For simulation purposes, we use the originally analyzed probability matrices for a subset of 50 loci spaced approximately evenly throughout the genome (provided in [Supporting Information, File S1](#), and [File S2](#)). For data analysis, we consider the white head-spotting phenotype mapped by Aylor *et al.* (2011) to a QTL with a peak at 95.0 Mb on chromosome 10. This QTL data set comprises a binary phenotype value (presence or absence of a white head spot) defined for 110 nonalbino mice and diplotype probability matrices for the QTL peak.

HS data set

The heterogeneous stocks are an outbred population of mice also derived from eight inbred strains: AJ, AKR/J (AKR), BALBc/J (BALB), CBA/J (CBA), C3H/HeJ (C3H), B6, DBA/2J (DBA), and LP/J (LP). We used data from the study of Valdar *et al.* (2006a), which includes mice from approximately generation 50 of the cross and comprises genotypes and phenotypes for 1762 mice from 180 families, with family sizes varying from 8 to 48. Valdar *et al.* (2006a) also used HAPPY to generate diplotype probability matrices based on 10,148 markers across the genome. For simulation purposes, we use the originally analyzed probability matrices

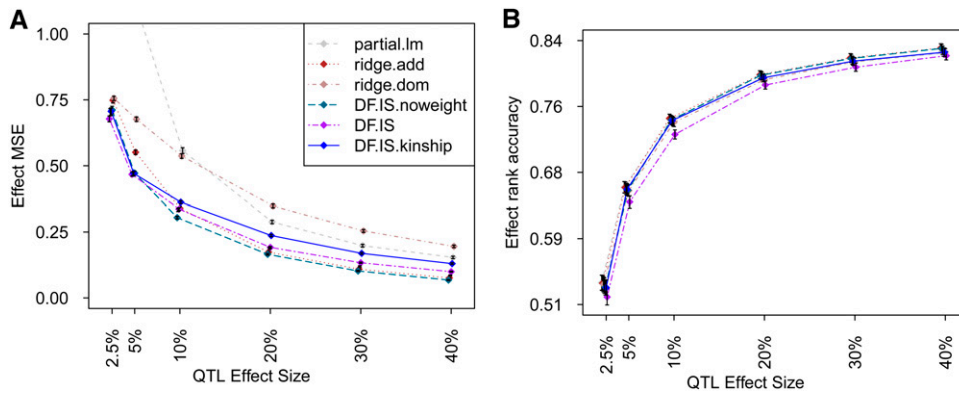


Figure 2 (A and B) Estimation of additive effects for a simulated additive-acting QTL in the pre-CC population, judged by (A) prediction error and (B) rank accuracy. For a given combination of QTL effect size and estimation method, each point indicates the mean of the evaluation metric based on 2500 simulation trials, and each vertical line indicates the 95% confidence interval of that mean. Points and lines are grouped by the corresponding QTL effect sizes and also are shifted slightly to avoid overlap. At the same QTL effect size, left to right jittering of the methods reflects relative performance from better to worse.

for a subset of 50 loci spaced approximately evenly throughout the genome (provided in File S1). For data analysis, we consider two phenotypes: total cholesterol (CHOL: 1656 observations), mapped by Valdar *et al.* (2006a) to a QTL at 171.5–172.0 Mb on chromosome 1; and the total startle time to a loud noise [fear potentiated startle (FPS): 1508 observations], which was mapped to a QTL at 91.37–92.62 Mb on chromosome 15. In each case, we use the original probability matrices defined at the peak loci; partial pedigree information; per-individual values for phenotype; and per-individual values for pre-determined covariates (defined in Valdar *et al.* 2006b)—sibship, cage, sex, testing chamber (FPS only), and date of birth (CHOL only) (all provided in File S3).

Simulating QTL effects

The ability of the Diploffect-based methods to estimate and rank haplotype and diplotype effects is assessed by simulation: We apply those methods, and their competitors listed in Table 2, to simulated single QTL for which the true effects are known. This is performed first using pre-CC data, which emphasizes estimation of haplotype (*i.e.*, additive) effects, potentially in the presence of dominance from residual heterozygotes, and then separately using the HS data, which emphasizes estimation of diplotype effects that could arise from both additive and dominance genetics. In either population, simulation of QTL involves four basic steps: selecting a locus; assigning true diplotypes; assigning QTL effects;

and simulating a phenotype based on the QTL effect, polygenic factors, and noise. This is described in detail below.

Let \mathcal{B} be a set of 50 representative haplotype effects (listed in File S1): 25 of these are binary alleles distributed among the eight founders [*e.g.*, (0, 0, 0, 0, 1, 1, 1, 1), (0, 1, 1, 1, 1, 1, 1, 1)]; the remaining 25 were drawn from $N(\mathbf{0}, \mathbf{I}_8)$. Let $\mathcal{V} = \{2.5, 5, 10, 20, 30, 40\}$ be the set of percentages of variance explained considered to be attributable to the QTL effect. Simulations are performed in the following (factorial) manner: For each data set (pre-CC or HS), for each locus m from the 50 defined in that data set, for $\beta \in \mathcal{B}$, and for dominance effects being either included or excluded, we perform the following simulation trial for every QTL effect size $v \in \mathcal{V}$:

1. For each individual $i \in \{1, \dots, n\}$, assign a true diplotype state by sampling $\mathbf{D}_i(m) \sim p(\mathbf{P}_i(m))$.
2. If including dominance effects, draw $\gamma \sim N(\mathbf{0}, \mathbf{I})$; otherwise, set $\gamma = \mathbf{0}$.
3. Calculate QTL contribution for each individual i as $q_i = \beta^T \text{add}(\mathbf{D}_i(m) + \gamma^T \text{dom}(\mathbf{D}_i(m)))$.
4. If HS, draw polygenic effect as n -vector $\mathbf{u} \sim N(\mathbf{0}, \mathbf{K}_{\text{IBS}})$ (see below); otherwise, if pre-CC, set $\mathbf{u} = \mathbf{0}$.
5. Draw individual-specific noise as n -vector $\epsilon \sim N(\mathbf{0}, \mathbf{I}_n)$.
6. Calculate the phenotype for each individual i as $y_i = a q_i + b u_i + c \epsilon_i$, where a , b , and c are constants used to adjust relative contributions of each term to the total phenotypic variance, ensuring that the QTL accounts for $v\%$, polygenic effects account for 50% (in HS) or 0%

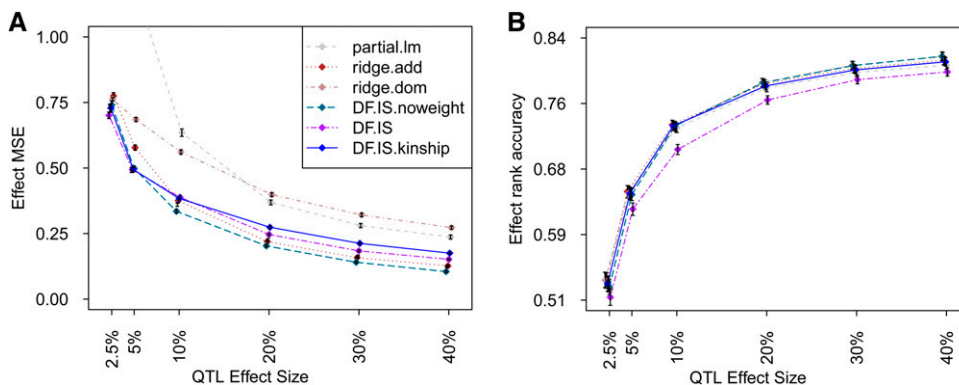


Figure 3 (A and B) Estimation of additive effects for a QTL simulated to have both additive and dominant effects in the pre-CC population. Symbols are defined as in Figure 2.

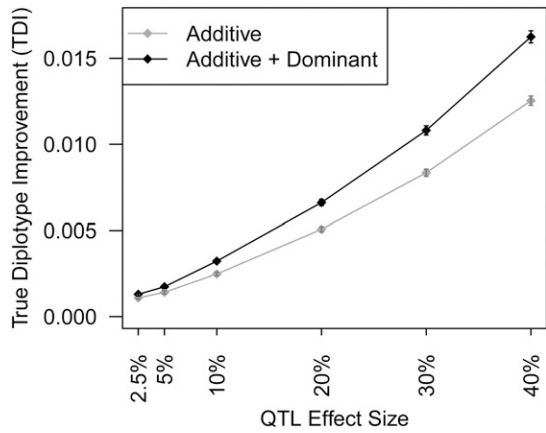


Figure 4 Increased posterior probability placed on the true diplotype at QTL simulated in the pre-CC, as analyzed using DF.MCMC.

(in pre-CC), and the remainder is attributed to individual-specific noise.

- Assess the ability of each method to estimate QTL effects given only \mathbf{y} and $\mathbf{P}_1(m), \dots, \mathbf{P}_n(m)$.

In step 4, \mathbf{K}_{IBS} is the realized genomic relationship matrix calculated using EMMA (Kang *et al.* 2008), applied to the entire set of HS genotypes. This polygenic effect, which represents potentially confounding effects of other QTL, is simulated only for the HS; the pre-CC lines are (in expectation) genetically exchangeable, and it is therefore assumed (for simulation purposes) that polygenic effects in the pre-CC would be indistinguishable from individual-specific noise. Also, because of this, in the pre-CC simulations we do not evaluate method DF.IS.kinship.

The above simulation scheme describes $2 \times 50 \times 50 \times 6 \times 2 = 60 \times 10^3$ distinct experimental conditions; this makes evaluating some methods in some populations impractical—specifically, DF.MCMC and DF.MCMC.pseudo are not evaluated in simulations involving the HS.

Evaluating estimation of QTL effects

Methods are evaluated based on their accuracy in determining two types of information: (1) the relative shift in phenotype value associated with a given diplotype substitution, which we measure across all such substitutions using an adjusted mean square error (the “effect MSE”), and (2) the relative ordering of the diplotype effect series, which we measure by agreement of ranks (the “effect rank accuracy”). Both types of information we consider highly relevant to subsequent experimental and/or bioinformatic characterization of QTL detected in multiparent populations.

Effect MSE and effect rank accuracy are defined as follows. Let $\boldsymbol{\theta}$ denote the p -vector of simulated effects (the target) after mean centering, and let $\hat{\boldsymbol{\theta}}$ be the p -vector of point estimates for the target, also after mean centering. Note that mean centering is performed both to adjust for the fact that the estimation models are naturally overspecified (*e.g.*, the intercepts are redundant) and to reflect the fact that scientific

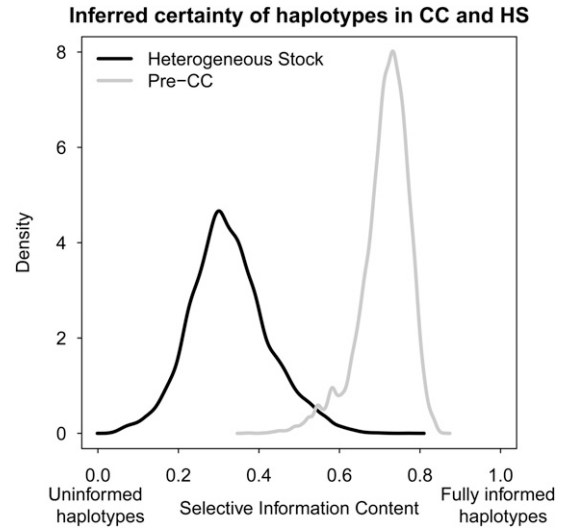


Figure 5 Certainty of inferred diplotype assignments across all marker loci in the pre-CC and HS.

interest is more meaningfully focused on substitution effects relative to each other (both in magnitude and in rank) than in absolute terms. The estimator $\hat{\boldsymbol{\theta}}$ is defined according to the method used: For Bayesian or partially Bayesian methods in Table 2 (DF.IS, DF.IS.kinship, DF.IS.noweight, DF.MCMC, and DF.MCMC.pseudo) it is defined as the posterior mean; for the remaining methods (partial.lm, ridge.add, and ridge.dom) it is the standard point estimate (*i.e.*, that maximizing the likelihood or penalized likelihood). The effect MSE is then defined as the average squared difference between parameters in target and estimate, normalized by the variance of the target; *i.e.*,

$$\text{Effect MSE} = \frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{p \times \text{Var}(\boldsymbol{\theta})}. \quad (18)$$

The effect rank accuracy is measured by Spearman’s rank correlation of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$.

The set of effects included in the target $\boldsymbol{\theta}$ differs according to the population. For the pre-CC, which is almost inbred, the target includes only the haplotype (additive) effects, *i.e.*, $\boldsymbol{\theta} = \boldsymbol{\beta}$; dominance effects may be present, but the infrequency of heterozygotes in the pre-CC precludes their meaningful estimation, making them akin to a type of confounding noise in the pre-CC. For the HS, many heterozygous diplotype states will be present at a given QTL, although overall some diplotype states may be absent; the target is therefore the $J \times (J + 1)/2$ vector of diplotype effects; *i.e.*, $\boldsymbol{\theta} = \boldsymbol{\delta}$.

Evaluating improvement in diplotype inference

Application of the Diploffect model leads not only to posteriors for QTL effects but also to updated probabilities for each individual’s diplotype state (see *Statistical Models and Methods*, Equation 8, and Figure A1); because the simulations

Table 3 Running time in seconds of different methods applied to simulated QTL in the pre-CC and HS data sets

Population	Pre-CC	Heterogeneous stock
partial.lm	0.056	2.72
ridge.add	0.124	2.80
ridge.dom	0.151	2.92
DF.MCMC.pseudo	27.96	NA
DF.MCMC	92.77	NA
DF.IS.noweight	580.7	3,727
DF.IS	580.9	3,724
DF.IS.kinship	NA	16,231

described above actually generate diplotype states for each individual (step 1), in postsimulation analysis it is possible to observe to what extent these updated probabilities are closer to the truth. We quantify this using a summary statistic, the true diplotype improvement (TDI). Letting $jk[i]$ index the true diplotype for individual i , such that $\{D_i(m)\}_{jk[i]} \equiv \mathbf{D}_i(m)_{jk[i]} = 1$, the TDI measures, over all true diplotypes, the average increase in probability from prior to posterior, *i.e.*,

$$\text{TDI} = \frac{1}{n} \sum_{i=1}^n [p(\mathbf{D}_i(m)_{jk[i]} | \Psi, \mathbf{y}) - \mathbf{P}_i(m)_{jk[i]}]. \quad (19)$$

Results

Pre-CC simulations

Methods were evaluated on their ability to estimate additive (haplotype) effects at simulated QTL in the pre-CC. Figure 2 shows performance when the QTL effect simulated was entirely additive; Figure 3 shows performance when the simulated effect also included dominance. In the latter case, dominance was not the target of inference as such—the mostly inbred pre-CC can manifest dominance only through residual heterozygosity; rather, it was included as a form of noise, potentially disrupting additive effect estimation.

Comparison of Figure 2 and Figure 3 reveals this disruptive effect to be slight (at least for dominance effects simulated here), causing an overall increase in effect MSE and a decrease in effect rank accuracy; otherwise there were few qualitative differences in relative performance across methods.

For both types of QTL effect, strong performance was seen across all effect sizes and metrics by implementations of Diploffect DF.MCMC (best) and DF.IS and ridge estimator ridge.add. Other methods produced competitive performance only in some settings and by some criteria. In particular, although the MSE of ridge.dom was worse than that of most other methods, and the MSE of partial.lm was worst of all (especially for QTL effect sizes <10%), both achieved competitive rank accuracy; this suggests that under this setting, and relative to a full probabilistic model, these ROP methods give estimates that are inflated (and therefore strongly penalized by MSE) but not necessarily poorly ordered.

Comparison of Bayesian methods DF.MCMC and DF.IS with their partially Bayesian approximations DF.MCMC.pseudo and DF.IS.noweight reflected some advantages of the Bayesian methods. DF.IS outperformed its approximation DF.IS.noweight in terms of MSE when the QTL effect was strong and matched it when the QTL effect was weak; similarly, DF.MCMC consistently outperformed approximation DF.MCMC.pseudo in rank accuracy for all QTL effect sizes. Both these observations can be explained in part by the fact that the Diploffect model when implemented fully makes more efficient use of phenotype data—not only using those data to inform the QTL effects, but also then using those QTL effects to help resolve uncertainty in diplotype state. This phenomenon is reflected in Figure 4, which uses results from DF.MCMC to demonstrate that as the strength of the QTL is increased, the posterior distribution of diplotype state at the QTL is moved consistently closer to the truth.

Heterogeneous Stock simulations

The HS population represents a far more challenging target for inference of QTL effects than the pre-CC largely because its haplotype composition, inferred from a smaller number

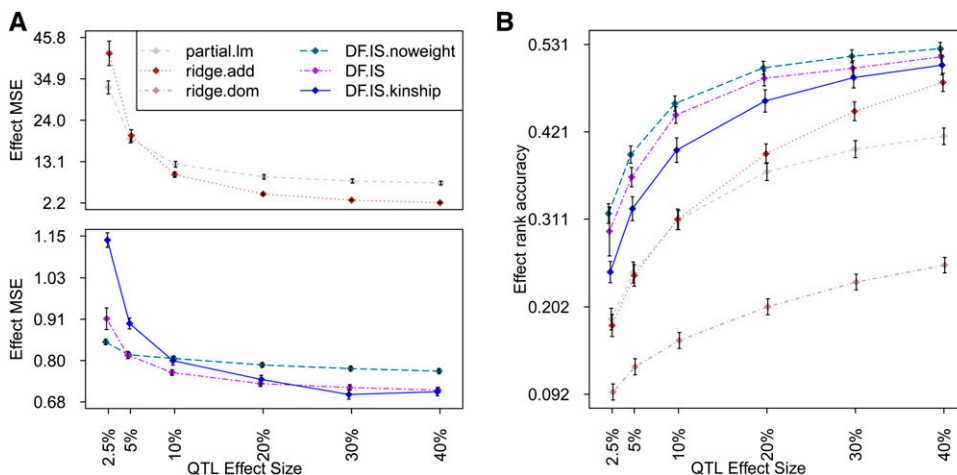


Figure 6 (A and B) Estimation of diplotype effects for an additive-only QTL simulated in the HS. Symbols are defined as in Figure 2.

of genotyped markers on a more highly recombinant genome, is far more uncertain. This increased uncertainty is illustrated in Figure 5, which plots the distribution of the scaled selective information content (SIC) (a rescaling of the negated Shannon entropy, previously used for this purpose in, e.g., Rönnegård and Valdar 2011). Although locus information varies in both populations, and the pre-CC manifests uncertainty for some individuals even at those loci that are overall most informed, the HS has many individuals whose diplotype state is almost uninformed.

Methods were evaluated on their ability to estimate simultaneously 36 diplotype effects for QTL simulated in the HS, with separate simulation studies for QTL with effects of additive vs. effects of additive plus dominance. Excluded from these simulations were the MCMC methods, owing to their impractically slow mixing: The time required for acceptable MCMC convergence on this relatively large data set (1762 individuals) made performing 5000 trials at each QTL effect size unfeasible (see Table 3 and *Data and Simulations* section). The Diploffect model was therefore represented by importance samplers DF.IS, DF.IS.noweight, and DF.IS.kinship. Of these, genetic background effects arising from unequal relatedness were represented in two ways: DF.IS.kinship, which used a pedigree-derived animal model; and DF.IS and DF.IS.noweight, which used instead a simple random intercept for sibship—an approximation that reduced their running time by more than fourfold (Table 3). We note that the polygenic effect used to generate the simulated phenotypes corresponded to neither of these estimation models, but was instead based on realized genomic relationships in the HS (see *Data and Simulations*).

As shown in Figure 6 and Figure 7, methods based on the Diploffect model strongly outperformed ROP-based competitors. This is somewhat expected: In the face of uncertainty, ROP will lead to often highly inflated estimates and resulting poor MSE (see discussion of Table 1 in *Statistical Models and Methods*). In this context, ridge regression should improve considerably over standard least-squares multivariable regression. Nonetheless, ridge.add produced effect estimates that were considerably dispersed relative to their true values, and ridge.dom performed so poorly (prediction error >2000 for 2.5%; see *File S1*) that it had to be excluded from Figure 6A and Figure 7A for legibility.

The addition of dominance effects to the simulated QTL reduced rank accuracy among all methods (Figure 8B) but did not alter relative performance. The plots showing MSE in the HS for different types of QTL effect (*i.e.*, Figure 6A and Figure 8A) are not directly comparable in scale because of the way in which this metric is normalized (*i.e.*, the denominator in Equation 18 is greater owing to greater variance of the target estimand); nonetheless, the pattern of performance among methods changes little between the additive setting and additive plus dominance—this is perhaps most surprising for ridge.dom, whose performance benefits so little from its additional dominance parameterization.

For QTL effect sizes of $\geq 10\%$, MSE was controlled best by Diploffect models DF.IS and DF.IS.kinship; for smaller

effect QTL, however, the less informed, partially Bayesian method DF.IS.noweight was seen to retain its accuracy while that of the more fully Bayesian models deteriorated. As with the pre-CC simulations above, this may reflect a trade-off between modeling sophistication and computational stability: DF.IS allows phenotype to inform descent, but to do so it must reweight each of its 1000 prior samples using an importance function, which, when the posterior is far from the prior, can drastically reduce sampling efficiency, leading to erratic results when sampling is capped, as in these simulations; DF.IS.noweight does not allow phenotype to inform descent but also does not require its samples to be reweighted, leading to pseudo-posteriors that are nonetheless well estimated. The need for using many more importance samples when applying DF.IS to the HS rather than to the pre-CC is illustrated in Figure 8, which shows for DF.IS analysis of 10% QTL the distribution of the effective sample size (ESS) statistic (see *Statistical Models and Methods*).

Finally, we note that in estimating both types of QTL effects, modeling genetic background using a kinship-specified polygenic effect, as in DF.IS.kinship, was not clearly superior to using a sibship-based approximation—indeed, at least in this context, it performed worse on rank accuracy while requiring substantially more computation. We speculate that this relative robustness of the sibship approximation could reflect either the circular-mated breeding scheme of the HS leading to kinship being well approximated by sibship, or/and computational efficiencies associated with estimating random effects with more complex covariance structures (which may lead to, for example, smaller ESS values), rather than any advantage of sibship approximations in general.

Haplotype effects on a binary outcome: white head spotting in the pre-CC

The pre-CC study of Aylor *et al.* (2011) identified a Mendelian trait locus on chromosome 10 (at 92.0 Mb) for white head spotting. White head spotting is a characteristic of the inbred CC founder strain WSB, and this phenotype was visibly present in 6 of the 110 nonalbino pre-CC mice. Because the identified locus was dominant Mendelian, associated with the presence of either one or two WSB haplotypes, it was straightforward to identify by LD mapping using a ROP model as in Equation 16. Estimating meaningful strain effects was not, in this case, necessary, because the effect was obvious. Such estimation would, however, have been awkward statistically because proper treatment of the binary outcome is most naturally modeled as a binary logistic regression, which in a standard maximum-likelihood estimation would have quickly become problematic due to separation (see, e.g., Gelman and Hill 2007). We considered estimation using Diploffect, which, as a GLMM with hierarchical shrinkage on the strain effects, models the white spot data without further development. Figure 9 plots 95% highest posterior density (HPD) intervals for all haplotype effects at the QTL estimated by both DF.MCMC and DF.IS. Both models report a similar result for this QTL: The non-WSB posteriors are similar to

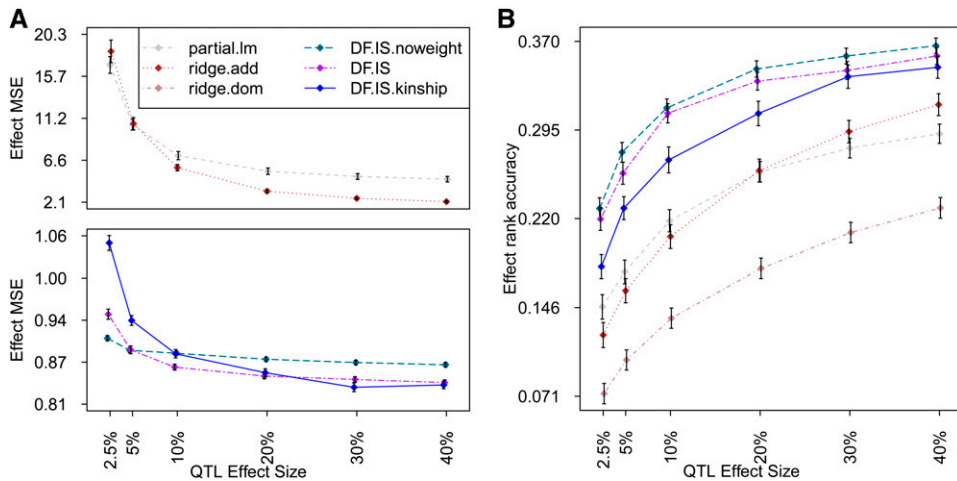


Figure 7 (A and B) Estimation of diplo-type effects for QTL simulated to have both additive and dominance effects in the HS. Symbols are defined as in Figure 2.

each other and broad, reflecting high uncertainty about the relative effects of these strains; the WSB posterior distribution is shifted above the others, reflecting its positive effect. The HPD of the contrast WSB vs. the other strains, calculated by applying this contrast to each MCMC sample from DF.MCMC, is 1.35–36.18, further reflecting the positive effect of the WSB haplotype but also the fact that uncertainty about this effect remains because the sample size is not infinite.

Haplotype and diplo-type effects at QTL in the HS: FPS and CHOL

To demonstrate Diploffect-based estimation of additive and dominance effects, we examined two previously detected QTL from the HS mapping study of Valdar *et al.* (2006a). The first QTL examined was for fear potentiated startle (a conditioned test of anxiety), located between 91.37 and 92.62 Mb on chromosome 15. To the central marker interval of the FPS QTL (rs3722990–rs3716673), we fitted a Diploffect linear mixed model (LMM) using DF.IS that included fixed effects of sex and testing chamber and random effects of cage and sibship, and where FPS was subject to a cube root transformation; this model specification follows that described in the parallel HS study of Valdar *et al.* (2006b). The results of the DF.IS analysis are shown in Figure 10 and Figure 11. Figure 10A shows HPD intervals for the haplotype effects. Dominance effects, which comprise 28 deviations from the additive haplotype model, are harder to graph intuitively; instead we plot the posterior predictive means of the 36 possible diplo-type effects, *i.e.*, $E(\delta|y)$, as a symmetric grayscale matrix in Figure 10B. Both plots suggest that effects are driven by C57BL/6J, and the consistent banding pattern of the diplo-type effect plot suggests these effects are mainly additive. Figure 11 more directly quantifies the extent to which the QTL effects are additive vs. dominant: It plots the posterior of the proportion of hyperparameter variance attributable to additive effects (see *Statistical Models and Methods*). The mode of this distribution is close to 1 but, as expected for a quantity defined using hyperparameters, the distribution itself is very broad (posterior mean 0.546), reflecting a high degree of uncertainty.

The second QTL examined in the HS was for total cholesterol concentration, located between 171.15 and 171.51 Mb on chromosome 1. To the central marker interval of the CHOL QTL (rs13476229–rs3657320), we fitted a Diploffect LMM using DF.IS that included fixed effects of sex and birth month, and random intercepts for cage and sibship (again following Valdar *et al.* 2006b). Results of this analysis are shown in Figure 11 and Figure 12. Unlike the FPS QTL, the HPD intervals for CHOL (Figure 12A) cluster into three different groups: the highest effect from LP, a second group comprising C3H and CBA with positive mean effects, and the remaining five strains having negative effects. This pattern is consistent with a multiallelic QTL, potentially arising through multiple, locally epistatic biallelic variants. In the diplo-type effect plot (Figure 12B), although most of the effects are additive, off-diagonal patches provide some evidence of

Comparison of Effective Sample Sizes

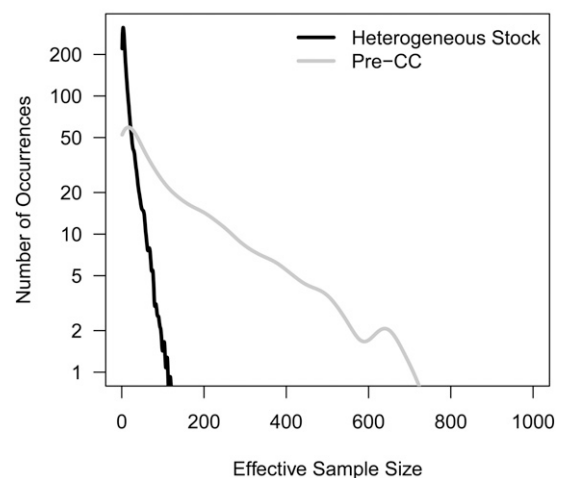


Figure 8 Density plot of the effective sample size (ESS) of posterior samples for the DF.IS method (maximum possible is 1000) applied to HS and pre-CC when analyzing a 10% QTL with additive and dominance effects. The plot shows that ESS is more efficient in the pre-CC data set than in the HS, reflecting the much larger dimension of the posterior in modeling QTL for the larger and less informed HS population.

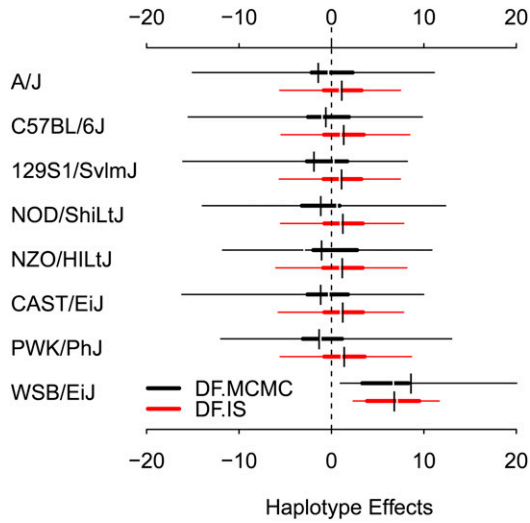


Figure 9 Highest posterior density intervals (95%, 50% and mean) for the haplotype effects of the binary trait white spotting in the pre-CC.

dominance effects—in particular, the haplotype combinations $AKR \times DBA$ and $C3H \times CBA$ deviate from the banding otherwise expected under additive genetics. The fraction of additive QTL effect variance for CHOL in Figure 11 is, however, strongly skewed toward additivity (posterior mean 0.717, with a sharp peak near 1), suggesting that additive effects predominate.

Discussion

We present here a statistical model and associated computational techniques for estimating the marginal effects of alternating haplotype composition at QTL detected in multi-parent populations. Our statistical model is intuitive in its construction, connecting phenotype to underlying diplotype state through a standard hierarchical regression model. Its

chief novelty, and the source of greatest statistical challenge, is that diplotype state, although efficiently encapsulating multiple facets of local genetic variation, cannot be observed directly and is typically available only probabilistically—meaning that statistically coherent and predictively useful description of QTL action requires estimating effects of haplotype composition from data where composition is itself uncertain. We frame this problem as a Bayesian integration, in which both diplotype states and QTL effects are latent variables to be estimated, and provide two computational approaches to solving it: one based on MCMC, which provides great flexibility but is also heavily computationally demanding, and the other using importance sampling and noniterative Bayesian GLMM fits, which is less flexible but more computationally efficient. Importantly, in theory and simulation, we describe how simpler, approximate methods for estimating haplotype effects relate to our model and how the trade-offs they make can affect inference.

An important comparison is made between Diploffect and approaches based on Haley–Knott regression, which regress on the diplotype probabilities themselves (or functions of them, such as the haplotype dosage) rather than the latent states those probabilities represent. In the context of QTL detection, where the need to scan potentially large numbers of loci makes fast computation essential, we believe that such ROP-based approaches are typically well justified and often the only practical solution. But for estimating effects at detected QTL, where the number of loci interrogated will be fewer by several orders of magnitude and the amount of time and energy devoted to interpretation will be far greater, there is room for a different trade-off.

We do expect ROP to provide accurate effect estimates under some circumstances. When, for example, descent can

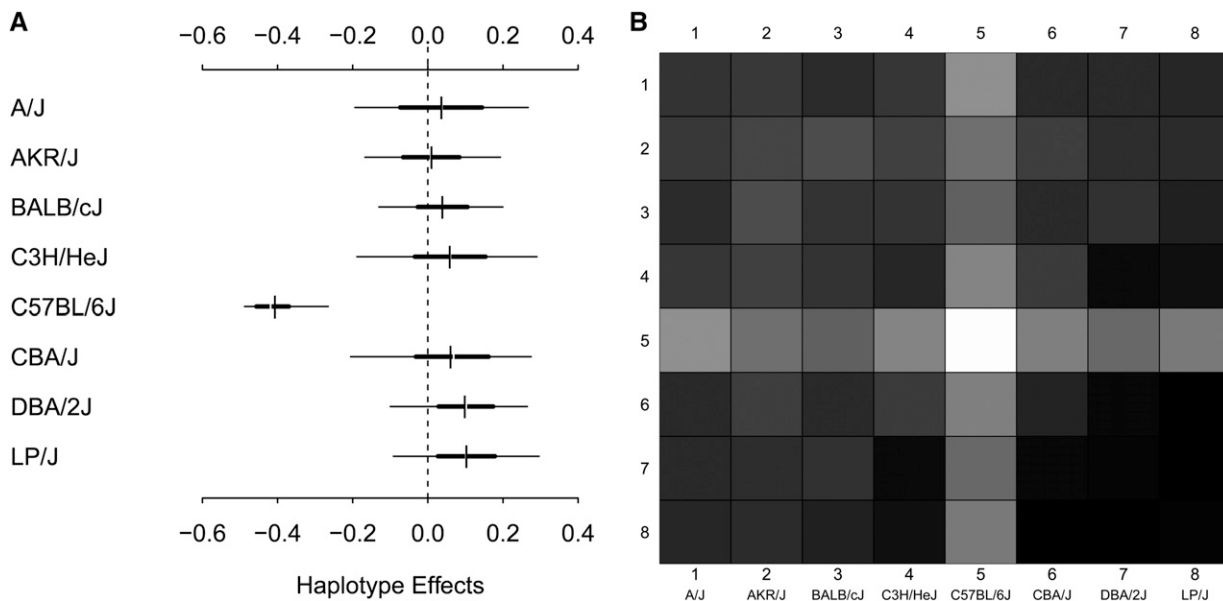


Figure 10 (A and B) Haplotype (A) and diplotype (B) effects estimated by DF.IS for phenotype FPS in the HS.

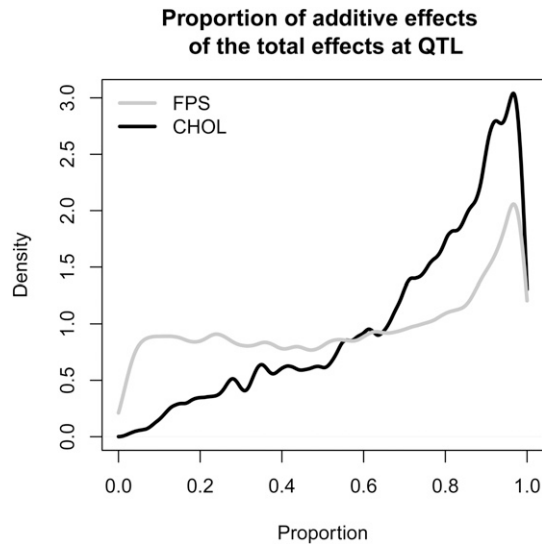


Figure 11 Posteriors of the fraction of effect variance due to additive rather than dominance effects at QTL for phenotypes FPS and CHOL in the HS data set.

be determined with near certainty (as may become more common as marker density is increased), a design matrix of diplotype probabilities (and haplotype dosages) will reduce to zeros and ones (and twos); in this case, although hierarchical modeling of effects would induce useful shrinkage, modeling diplotypes as latent variables would produce comparatively little benefit. This is demonstrated in the results of ridge regression (ridge.add) on the pre-CC: In this context, with only moderate uncertainty for most individuals at most loci, the performance of a simple ROP-based eight-allele ridge model (which we consider an optimistic equivalent to an unpenalized regression of the same model) approaches that of the best Diploffect-based method. Adding dominance effects to this ridge regression (which again we consider a more stable equivalent to doing so

with an ordinary regression) produces effect estimates that are far more dispersed. Applying these stabilized ROP approaches to the HS data set, whose higher ratio of recombination density to genotype density implies a less certain haplotype composition, leads to effect estimates that can be erratic; indeed, such point estimates should not be taken at face value without substantial caveats or examining (if possible) likely estimator variance. In populations and studies where this ratio is lower, and haplotype reconstruction is more advanced (e.g., in the DO population of Svenson *et al.* 2012 and Gatti *et al.* 2014), or where the number of founders is small relative to the sample size, we expect that additive ROP models will often be adequate, if suboptimal. Only in extreme cases, however, do we expect that reliable estimation of additive plus dominance effects will not require some form of hierarchical shrinkage.

A strong motivation for developing Diploffect, and in particular to use a Bayesian approach to its estimation, is to facilitate design of follow-up studies—in particular, the ability to obtain for any future combination of haplotypes, covariates, and concisely specified genetic background effects a posterior-predictive distribution for some function of the phenotype. This could be, for example, a cost or utility function whose posterior predictive distribution can inform decisions about how to prioritize subsequent experiments. Such predictive distributions are easily obtained from our MCMC procedure and can also be extracted with only slightly more effort [via specification of $T(\theta)$ in Equation 14] from our importance sampling methods. We anticipate that, applied to (potentially multiple) independent QTL, Diploffect models could provide more robust out-of-sample predictions of the phenotype value in, e.g., proposed crosses of multiparental recombinant inbred lines than would be possible using ROP-based models.

Use of the Bayesian procedures proposed here nonetheless has several potential drawbacks, foremost among which is computation time: Although our modified slice sampler

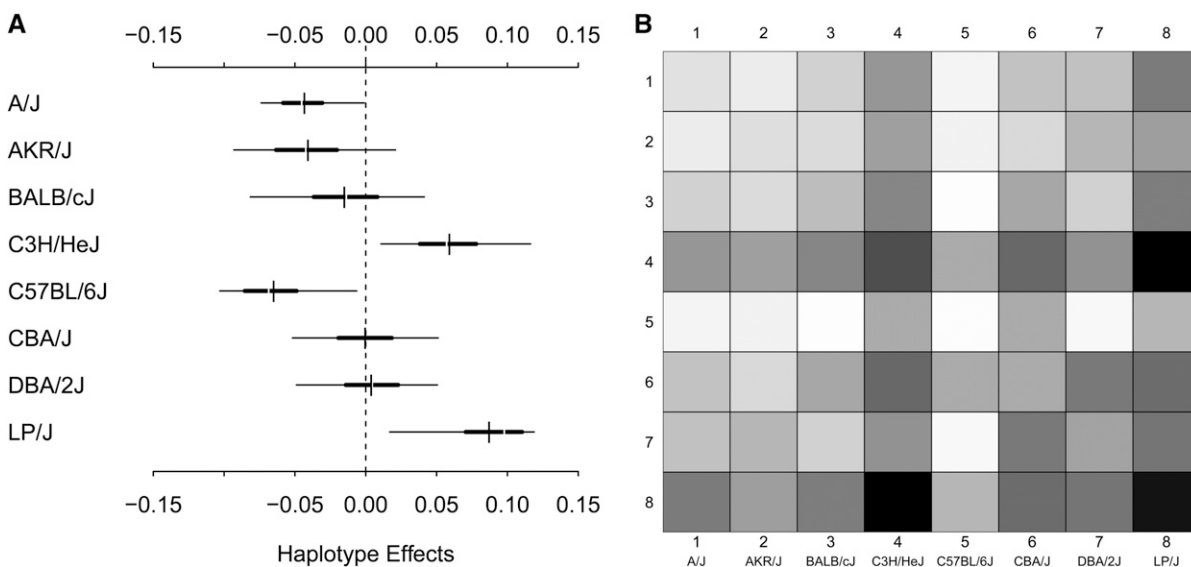


Figure 12 (A and B) Haplotype (A) and diplotype (B) effects estimated by DF.IS for phenotype CHOL in the HS.

(DF.MCMC; Appendix A) makes MCMC sampling of both diplotypes and effects feasible, it is highly computationally intensive. For large outbred populations, especially those with a high degree of diplotype uncertainty, we therefore prefer our importance sampler (DF.IS). For either method, however, a high degree of diplotype uncertainty and weak QTL effects result in computational inefficiency, since the posterior distribution that must be traversed (in MCMC) or sampled (in IS) is much more diffuse: For DF.MCMC this means convergence must be carefully monitored; for DF.IS, this means many more samples must be taken to achieve a reasonable picture of the posterior.

In light of the additional computational costs incurred by jointly modeling diplotypes and effects, it is worth considering the utility of partially Bayesian approaches in which diplotypes are multiply imputed, as in, for example, Kover *et al.* (2009) or Durrant and Mott (2010). Indeed, in discussing their partially Bayesian but highly computationally efficient random haplotype effects model, Durrant and Mott (2010) warn that Bayesian updating of the joint model described here would likely suffer from the *label-switching problem* (Stephens 1997). We consider this somewhat pessimistic: The label-switching problem usually occurs when the prior on the mixture components (in this case, the set of diplotype probabilities in Ψ) is uniform or nearly uniform; in practice, diplotype probabilities from modern haplotype reconstructions tend to be well informed enough for most individuals (even in the HS data set reported here) that label switching will be minimal, negligibly impact inference. Nonetheless, although our more fully Bayesian modeling adds value to inference when QTL effect sizes are large, when QTL effect sizes are small ($\leq 10\%$), the partially Bayesian approximations DF.MCMC.pseudo and DF.IS.noweight become more competitive. Indeed, we observe that when analyzing small effect QTL ($\leq 10\%$) in the high-dimensional/low-information setting of the HS data set, DF.IS.noweight outperformed its fully Bayesian counterpart, reflecting a potential trade-off between statistical and computational efficiency.

At greater computational cost, our modeling of QTL effects could be more comprehensive. At one extreme, we could consider a complete probabilistic treatment, for example in the spirit of Lin and Zeng (2006), whereby QTL effects and diplotypes are estimated conditional on raw genotype data, rather than, as here, conditional on diplotype probabilities that have been inferred previously and independently. Alternatively, and more realistically, we could attempt to model diplotype states explicitly at all contributing QTL, rather than, as here, focusing on marginal effects at a single QTL and presuming that all other effects can be well approximated by covariates and structured noise. Instead we provide a starting point—one that, while somewhat computationally demanding, relies on previously computed results (HMM output) and standard simplifying assumptions. In implementing Diploffect through an adaptation of existing, flexible modeling software (JAGS and INLA), we further aim that other researchers will be able to extend the model to better suit the specifics of a given phenotype, covariate structure, or experimental setting.

Acknowledgments

The authors thank the associate editor and an anonymous reviewer for helpful feedback. Research reported in this article was supported by the following: the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under award R01GM104125 (primary support to Z.Z. and W.V.), the National Science Foundation (NSF) under award IIS-1162369 (to W.W.), and the University of North Carolina Lineberger Comprehensive Cancer Center (partial support to Z.Z. and W.V.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the NSF.

Literature Cited

- Aylor, D. L., W. Valdar, W. Foulds-Mathes, R. J. Buus, R. A. Verdugo *et al.*, 2011 Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res.* 21(8): 1213–1222.
- Bandillo, N., C. Raghavan, P. A. Muyco, M. A. L. Sevilla, I. T. Lobina *et al.*, 2013 Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6(1): 11.
- Baud, A., R. Hermsen, V. Guryev, P. Stridh, D. Graham *et al.*, 2013 Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat. Genet.* 45(7): 767–775.
- Broman, K. W., and S. Sen, 2009 *A Guide to QTL Mapping with R/qlt*. Springer-Verlag, New York.
- Carlin, B. P., and T. A. Louis, 2008 *Bayesian Methods for Data Analysis*, Ed. 3. Chapman & Hall/CRC, Boca Raton, FL.
- Cavanagh, C., M. Morell, I. Mackay, and W. Powell, 2008 From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11(2): 215–221.
- 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. *Nat. Genet.* 36(11): 1133–1137.
- Collaborative Cross Consortium, 2012 The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics* 190: 389–401.
- Durrant, C., and R. Mott, 2010 Bayesian quantitative trait locus mapping using inferred haplotypes. *Genetics* 184: 839–852.
- Ferris, M. T., D. L. Aylor, D. Bottomly, A. C. Whitmore, L. D. Aicher *et al.*, 2013 Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. *PLoS Pathog.* 9(2): e1003196.
- Friedman, J., T. Hastie, and R. Tibshirani, 2010 Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33(1): 1.
- Gatti, D. M., K. L. Svenson, A. Shabalina, L. Wu, W. Valdar *et al.*, 2014 Quantitative trait locus mapping methods for diversity outbred mice. *G3* 9: 1623–1633.
- Gelman, A., and J. Hill, 2007 *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge University Press, New York.
- Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69(4): 315–324.
- Hoerl, A. E., and R. W. Kennard, 1970 Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12(1): 55–67.
- Huang, B. E., and A. W. George, 2011 R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics* 27(5): 727–729.
- Huang, B. E., A. W. George, K. L. Forrest, A. Kilian, M. J. Hayden *et al.*, 2012 A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol. J.* 10(7): 826–839.

- Huang, G.-J., S. Shifman, W. Valdar, M. Johannesson, B. Yalcin *et al.*, 2009 High resolution mapping of expression QTLs in heterogeneous stock mice in multiple tissues. *Genome Res.* 19(6): 1133–1140.
- Huang, X., M.-J. Paulo, M. Boer, S. Effgen, P. Keizer *et al.*, 2011 Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proc. Natl. Acad. Sci. USA* 108(11): 4488–4493.
- Johannesson, M., R. Lopez-Aumatell, P. Stridh, M. Diez, J. Tuncel *et al.*, 2009 A resource for the simultaneous high-resolution mapping of multiple quantitative trait loci in rats: the NIH heterogeneous stock. *Genome Res.* 19(1): 150–158.
- Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman *et al.*, 2008 Efficient control of population structure in model organism association mapping. *Genetics* 178: 1709–1723.
- Kelada, S. N. P., D. L. Aylor, B. C. E. Peck, J. F. Ryan, U. Tavares *et al.*, 2012 Genetical analysis of hematological parameters in incipient lines of the Collaborative Cross. *G3* 2: 157–165.
- Kennedy, B. W., M. Quinton, and J. A. van Arendonk, 1992 Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* 70(7): 2000–2012.
- King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hooper, S. Sen *et al.*, 2012a Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Res.* 22: 1558–1566.
- King, E. G., S. J. Macdonald, and A. D. Long, 2012b Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. *Genetics* 191: 935–949.
- 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 Genet.* 5(7): e1000551.
- Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- Lenarcic, A. B., K. L. Svenson, G. A. Churchill, and W. Valdar, 2012 A general Bayesian approach to analyzing diallel crosses of inbred strains. *Genetics* 190: 413–435.
- Lin, D. Y., and D. Zeng, 2006 Likelihood-based inference on haplotype effects in genetic association studies. *J. Am. Stat. Assoc.* 101(473): 89–104.
- Liu, E. Y., Q. Zhang, L. McMillan, F. Pardo-Manuel de Villena, and W. Wang, 2010 Efficient genome ancestry inference in complex pedigrees with inbreeding. *Bioinformatics* 26(12): i199–i207.
- Liu, J., R. Chen, and T. Logvinenko, 2001 A theoretical framework for sequential importance sampling with resampling, pp. 225–246 in *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. Freitas, and N. Gordon. Springer-Verlag, New York.
- Mackay, I. J., P. Bansept-Basler, T. Barber, A. R. Bentley, J. Cockram *et al.*, 2014 An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3 (Bethesda)* 4: 1603–1610.
- Marriage, T. N., E. G. King, A. D. Long, and S. J. Macdonald, 2014 Fine-mapping nicotine resistance loci in *Drosophila* using a multiparent advanced generation intercross population. *Genetics* 198: 45–57.
- Mott, R., 2009 happy.hbrem: quantitative trait locus genetic analysis in heterogeneous stocks. R package version 2.3. <http://www.r-project.org>, <http://www.well.ox.ac.uk/happy>.
- Mott, R., C. J. Talbot, M. G. Turri, A. C. Collins, and J. Flint, 2000 A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci. USA* 97(23): 12649–12654.
- Neal, R., 2003 Slice sampling. *Ann. Stat.* 31(3): 705–741.
- Phillippi, J., Y. Xie, D. R. Miller, T. A. Bell, Z. Zhang *et al.*, 2014 Using the emerging Collaborative Cross to probe the immune system. *Genes Immun.* 15(1): 38–46.
- Plummer, M., 2003 JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria. ISSN 1609-395X.
- Plummer, M., 2014 rjags: Bayesian graphical models using MCMC. R package version 3–13. <http://CRAN.R-project.org/package=rjags>.
- R Development Core Team, 2011 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Robert, C. P., and G. Casella, 2009 *Introducing Monte Carlo Methods with R*. Springer-Verlag, New York.
- Rönnegård, L., and W. Valdar, 2011 Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics* 188: 435–447.
- Rue, H. v., S. Martino, and N. Chopin, 2009 Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. B* 71(2): 319–392.
- Sillanpaa, M. J., and E. Arjas, 1998 Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* 148: 1373–1388.
- Sillanpaa, M. J., and E. Arjas, 1999 Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* 151: 1605–1619.
- Solberg Woods, L. C., K. L. Holl, D. Oreper, Y. Xie, S.-W. Tsaih *et al.*, 2012 Fine-mapping diabetes-related traits, including insulin resistance, in heterogeneous stock rats. *Physiol. Genomics* 44(21): 1013–1026.
- Stephens, M., 1997 Bayesian methods for mixtures of normal distributions. Ph.D. Thesis, University of Oxford, Oxford.
- Svenson, K. L., D. M. Gatti, W. Valdar, C. E. Welsh, R. Cheng *et al.*, 2012 High-resolution genetic mapping using the mouse Diversity Outbred population. *Genetics* 190: 437–447.
- Talbot, C. J., A. Nicod, S. S. Cherny, D. W. Fulker, C. Collins *et al.*, 1999 High-resolution mapping of quantitative trait loci in outbred mice. *Nat. Genet.* 21(3): 305–308.
- Tsaih, S., K. Holl, S. Jia, M. Kaldunski, M. Tschannen *et al.*, 2014 Identification of a Novel Gene for Diabetic Traits in Rats, Mice, and Humans. *Genetics* 198: 17–29.
- Valdar, W., L. C. Solberg, D. Gauguier, S. Burnett, P. Klenerman *et al.*, 2006a Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* 38: 879–887.
- Valdar, W., L. C. Solberg, D. Gauguier, W. O. Cookson, J. N. P. Rawlins *et al.*, 2006b Genetic and environmental effects on complex traits in mice. *Genetics* 174: 959–984.
- Valdar, W., C. C. Holmes, R. Mott, and J. Flint, 2009 Mapping in structured populations by resample model averaging. *Genetics* 182: 1263–1277.
- Vazquez, A. I., D. M. Bates, G. J. M. Rosa, D. Gianola, and K. A. Weigel, 2010 Technical note: an R package for fitting generalized linear mixed models in animal breeding. *J. Anim. Sci.* 88(2): 497–504.
- Yalcin, B., J. Flint, and R. Mott, 2005 Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. *Genetics* 171: 673–681.
- Yuan, Z., F. Zou, and Y. Liu, 2011 Bayesian multiple quantitative trait loci mapping for recombinant inbred intercrosses. *Genetics* 188: 189–195.
- Zaykin, D. V., P. H. Westfall, S. S. Young, M. a. Karnoub, M. J. Wagner *et al.*, 2002 Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum. Hered.* 53(2): 79–91.
- Zhang, Z., X. Zhang, and W. Wang, 2012 HTreeQA: using semi-perfect phylogeny trees in quantitative trait loci study on genotype data. *G3* 2: 175–189.

Communicating editor: F. van Eeuwijk

Appendix A

Discrete Slice Sampling with Reordering of Prior Probabilities

To help efficiently traverse the space of possible diplotype states during sampling of each $\mathbf{D}_i(m)$ in Equation 13, we propose an optimization of the discrete slice sampling algorithm of Neal (2003). First, we describe the original algorithm, which adapts slice sampling on a continuous density to the discrete case. Let $T(jk)$ represent diplotype jk 's order in the range $1, \dots, J \times (J + 1)/2$, and define two boundary diplotypes for $T(L) = 0$ and $T(R) = J \times (J + 1)/2 + 1$, with posterior probabilities for these set to zero. For the previous diplotype sampled, x' , first evaluate $S = Q(\mathbf{D}_i(m)_{T^{-1}(x')})$, and then sample an auxiliary variable $q \sim U(0, S)$. After that, expand a region $[l, u]$ satisfying $Q(l) \geq q$ and $Q(l - 1) < q$ and $Q(u) < q$ and $Q(u + 1) \geq q$. From the uniform distribution defined on $[l, u]$, continue sampling the new diplotype status x^{new} until one is reached for which $Q(\mathbf{D}_i(m)_{T^{-1}(x)}) \geq q$. The diplotype posterior in Equation 10 is evaluated only a few times during each iteration, and thus it is much faster than full sampling.

Directly applying discrete slice sampling as above with the diplotype states ordered arbitrarily (e.g., alphabetically) will typically produce a prior and a posterior that are highly multimodal; this leads to extremely slow mixing, with posterior updates remaining stuck for many MCMC iterations within islands of higher-probability diplotypes separated by barrier-forming groups of lower-probability diplotypes (Figure A1A). To improve mixing, we introduce a small modification: Before applying slice sampling, we first reorder all $J \times (J + 1)/2$ entries in matrix $\mathbf{D}_i(m)$ according to their prior probabilities in $\mathbf{P}_i(m)$. As shown in Figure A1B, this creates a prior that is monotone decreasing, removes the barrier, and results in posteriors that are more efficiently mixed (Figure A1C).

Appendix B

Importance Sampling Using the Diplotype Prior and the Conditional Effects Posterior

Method DF.IS approximates statistics on the joint posterior $p(\mathbf{\Delta}, \mathbf{\beta} | \mathbf{y})$ by importance sampling. The IS procedure it uses owes its efficiency to INLA (Rue *et al.* 2009) and, in particular, to deterministic results for $p(\mathbf{\beta} | \mathbf{\Delta}, \mathbf{y})$ and $p(\mathbf{y} | \mathbf{\Delta})$. This approach is explained in more detail here. Importance sampling, when used to estimate posterior statistics on disjoint parameter sets \mathbf{a} and \mathbf{b} whose joint posterior density $p(\mathbf{a}, \mathbf{b} | \mathbf{y})$ is hard to sample from directly, describes the following general strategy: Sample a large number of times from a more tractable density $g(\mathbf{a}, \mathbf{b})$ (the importance function) and then reweight each sample relative to the others using importance weight $w(\mathbf{a}, \mathbf{b}) = p(\mathbf{a}, \mathbf{b} | \mathbf{y})/g(\mathbf{a}, \mathbf{b})$. In this article, $\mathbf{a} \equiv \mathbf{\Delta}$, $\mathbf{b} \equiv \mathbf{\beta}$, with independent priors $p(\mathbf{a}, \mathbf{b}) = p(\mathbf{a})p(\mathbf{b})$; additionally, due to INLA, deterministic approximations are available for both the conditional effects posterior $p(\mathbf{b} | \mathbf{y}, \mathbf{a})$ and the conditional marginal likelihood $p(\mathbf{y} | \mathbf{a})$. Considered together, these features suggest an importance function $g(\mathbf{a}, \mathbf{b}) = p(\mathbf{b} | \mathbf{a}, \mathbf{y})p(\mathbf{a})$, whereby each importance sample $(\mathbf{a}_{\text{IS}}^{(k)}, \mathbf{b}_{\text{IS}}^{(k)})$ for $k = 1, \dots, K$ is generated by first sampling $\mathbf{a}_{\text{IS}}^{(k)}$ from diplotype prior $p(\mathbf{a})$ and then sampling $\mathbf{b}_{\text{IS}}^{(k)}$ from conditional effects posterior $p(\mathbf{b} | \mathbf{a}_{\text{IS}}^{(k)}, \mathbf{y})$. Since only the difference between sampling from $g(\mathbf{a}, \mathbf{b})$ and sampling from $p(\mathbf{a}, \mathbf{b} | \mathbf{y})$ is the use of $p(\mathbf{a})$ rather than $p(\mathbf{a} | \mathbf{y})$, the importance weight for each $(\mathbf{a}_{\text{IS}}^{(k)}, \mathbf{b}_{\text{IS}}^{(k)})$ pair is proportional only to $p(\mathbf{y} | \mathbf{a}_{\text{IS}}^{(k)})$; that is,

$$\begin{aligned} w(\mathbf{a}, \mathbf{b}) &= \frac{p(\mathbf{a}, \mathbf{b} | \mathbf{y})}{g(\mathbf{a}, \mathbf{b})} = \frac{p(\mathbf{a}, \mathbf{b} | \mathbf{y})}{p(\mathbf{b} | \mathbf{y}, \mathbf{a})p(\mathbf{a})} = \frac{p(\mathbf{a}, \mathbf{b} | \mathbf{y})}{\frac{p(\mathbf{a}, \mathbf{b} | \mathbf{y})p(\mathbf{a})}{p(\mathbf{a} | \mathbf{y})}} \\ &= \frac{p(\mathbf{a} | \mathbf{y})}{p(\mathbf{a})} \times \frac{p(\mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{a}, \mathbf{y})}{p(\mathbf{a})p(\mathbf{y})} = \frac{p(\mathbf{y} | \mathbf{a})}{p(\mathbf{y})} \propto p(\mathbf{y} | \mathbf{a}). \end{aligned}$$

These proportionally defined weights can then be renormalized by their sum and used to appropriately reweight statistics on each sampled $(\mathbf{a}_{\text{IS}}^{(k)}, \mathbf{b}_{\text{IS}}^{(k)})$, as in Equation 14 (refer to *Statistical Models and Methods*).

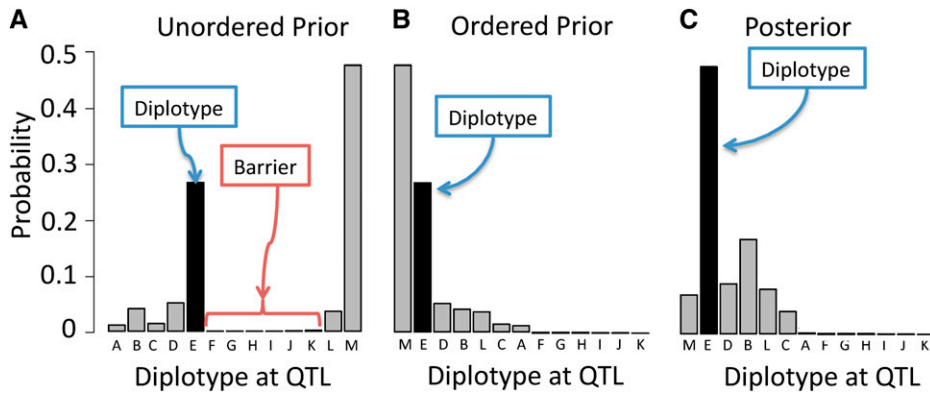


Figure A1 Reordering of prior probabilities in the discrete slice sampler, using as an example the diplotype probabilities from haplotype reconstruction on the pre-CC. Diplotypes are represented by different letters, and 23 diplotypes with very low probabilities are omitted. The true diplotype, selected during simulation, is shaded black. An arbitrary ordering of diplotypes is shown in A and illustrates the problem to be addressed: If the initially sampled diplotype is M, the slice sampler cannot easily cross the barrier region to sample other high-probability diplotypes. Reordering the diplotypes by their prior probabilities to create a smoother distribution, as in B, removes this barrier region and allows the sampler to move easily between its initial value and all other values of high to moderate probability. C shows the posterior of this distribution given phenotype data (from the DF.MCMC procedure), in which the true diplotype's posterior probability is increased.

this barrier region and allows the sampler to move easily between its initial value and all other values of high to moderate probability. C shows the posterior of this distribution given phenotype data (from the DF.MCMC procedure), in which the true diplotype's posterior probability is increased.

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.166249/-/DC1>

Bayesian Modeling of Haplotype Effects in Multiparent Populations

Zhaojun Zhang, Wei Wang, and William Valdar

File S1

Supplementary Methods and Results for the paper
“Bayesian modeling of haplotype effects in multiparent populations”
July 14, 2014

Zhaojun Zhang, Wei Wang, and William Valdar

July 14, 2014

- Table 1 shows the 50 simulated effect vectors used in the simulation study, with each row being a single effect vector for eight founders. For CC population, the effects to the founders from A to H are assigned to 129S1/SvImJ (129S1), A/J (AJ), C57BL/6J (B6), NOD/ShiLtJ (NOD), NZO/HILtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB), respectively. And For HS population, the effects are assigned to A/J (AJ), AKR/J (AKR), BALBc/J (BALB), CBA/J (CBA), C3H/HeJ (C3H), C57BL/6J (B6), DBA/2J (DBA) and LP/J (LP), respectively.
- Table 2 and 3 show the simulation result based on Pre-CC population with mean values of two metrics and their 95% confident intervals. The data of the tables is used for generating Figure 3 and 4 in the main manuscript.
- Table 4 and 5 show the simulation result based on HS population with mean values of two metrics and their 95% confident intervals. The data of the tables is used for generating Figure 5 and 6 in the main manuscript.

Strain	A	B	C	D	E	F	G	H
	2	1	1	1	1	1	1	1
	1	2	1	1	1	1	1	1
	1	1	2	1	1	1	1	1
	1	1	1	2	1	1	1	1
	1	1	1	1	2	1	1	1
	1	1	1	1	1	2	1	1
	1	1	1	1	1	1	2	1
	1	1	1	1	1	1	1	2
	2	2	1	1	1	1	1	1
	1	2	2	1	1	1	1	1
	1	1	2	2	1	1	1	1
	1	1	1	2	2	1	1	1
	1	1	1	1	2	2	1	1
	1	1	1	1	1	2	2	1
	2	1	1	1	1	1	1	2
	1	1	1	1	2	2	2	1
	2	1	1	1	1	2	2	1
	2	2	1	1	1	1	2	1
	1	1	1	2	2	1	2	1
	1	2	2	1	1	1	1	2
	2	2	2	2	1	1	1	1
	1	2	2	2	2	1	1	1
	1	1	2	2	2	2	1	1
	1	1	1	2	2	2	2	1
	2	2	2	1	1	1	1	2
	1.78	0.72	-1.54	0.74	-0.68	-0.17	-1.08	1.28
	-1.505	0.059	1.868	1.156	0.206	-0.514	-0.049	-0.909
	-1.31	-1.26	0.66	-0.52	-1.56	-0.79	-0.72	0.09
	-0.739	-1.915	0.081	1.060	-0.380	0.361	0.348	0.153
	2.70	0.29	0.60	-0.72	-1.06	-1.21	1.32	-0.19
	0.88	-0.21	0.85	2.08	-3.52	-0.34	-0.55	-0.57
	0.781	-1.886	0.099	0.687	0.539	0.755	-0.752	-1.159
	0.428	-1.052	0.459	0.049	-0.731	0.406	-0.601	-0.972
	-2.40	0.13	0.94	0.85	0.34	-0.23	1.60	0.29
	0.237	0.022	0.428	-0.390	0.103	-0.551	-0.835	-0.113
	-0.67	-1.76	1.05	0.29	0.50	-0.59	-0.54	-1.27
	-1.22	-1.97	1.81	-0.26	-0.18	-0.64	-0.72	-0.84
	-0.048	0.974	1.048	0.536	-1.349	-0.290	0.307	0.522
	1.090	-0.552	-0.731	0.067	-0.693	-0.459	-0.629	-0.012
	0.064	-0.678	1.154	1.477	-0.118	-0.947	0.954	0.778
	-0.700	0.023	-1.199	-1.168	-0.637	-0.515	0.059	1.332
	-0.287	-0.092	-1.238	0.388	-0.980	-0.605	-0.608	0.832
	-0.133	-0.907	-0.342	0.024	1.542	-0.013	2.274	2.435
	-2.211	-1.598	-0.617	-0.738	0.774	-1.006	-0.073	-0.467
	0.073	0.331	-0.027	-0.136	-0.863	0.105	1.065	-0.088
	0.235	-0.637	-0.326	1.200	0.024	-1.819	1.384	-0.759
	0.44028	-0.46124	1.24133	-0.57630	0.74145	0.00036	-1.55446	0.37339
	0.62	1.58	1.07	-0.25	0.44	-0.16	-0.56	0.68
	-0.71	0.86	1.73	-1.02	-0.24	-1.78	-1.11	-0.32
	-1.139	0.442	-1.779	-0.739	0.545	-1.192	0.023	1.230

Table 1: A table of all effect vectors used in the simulation study.

Simulated Effects	Metric	Model	QTL effect size		
			2.5%	5%	10%
Additive	Effect MSE	<code>partial.lm</code>	2.174 (0.042)	1.096 (0.021)	0.557 (0.011)
		<code>ridge.add</code>	0.748 (0.010)	0.551 (0.009)	0.340 (0.006)
		<code>ridge.dom</code>	0.756 (0.008)	0.677 (0.009)	0.537 (0.009)
		<code>DF.MCMC.pseudo</code>	0.678 (0.011)	0.467 (0.007)	0.337 (0.006)
		<code>DF.MCMC</code>	0.712 (0.014)	0.472 (0.008)	0.305 (0.005)
		<code>DF.IS.noweight</code>	0.706 (0.014)	0.471 (0.008)	0.362 (0.006)
		<code>DF.IS</code>	0.703 (0.014)	0.469 (0.008)	0.333 (0.006)
	Effect Rank Accuracy	<code>partial.lm</code>	0.537 (0.009)	0.656 (0.007)	0.741 (0.005)
		<code>ridge.add</code>	0.537 (0.009)	0.656 (0.007)	0.742 (0.005)
		<code>ridge.dom</code>	0.535 (0.009)	0.652 (0.007)	0.738 (0.005)
		<code>DF.MCMC.pseudo</code>	0.519 (0.010)	0.638 (0.008)	0.722 (0.006)
		<code>DF.MCMC</code>	0.531 (0.010)	0.653 (0.007)	0.741 (0.005)
		<code>DF.IS.noweight</code>	0.531 (0.009)	0.654 (0.007)	0.740 (0.005)
		<code>DF.IS</code>	0.534 (0.010)	0.654 (0.007)	0.739 (0.006)
Additive + Dominant	Effect MSE	<code>partial.lm</code>	2.242 (0.048)	1.170 (0.025)	0.635 (0.014)
		<code>ridge.add</code>	0.775 (0.012)	0.578 (0.010)	0.376 (0.008)
		<code>ridge.dom</code>	0.762 (0.007)	0.685 (0.008)	0.561 (0.008)
		<code>DF.MCMC.pseudo</code>	0.701 (0.013)	0.493 (0.008)	0.388 (0.007)
		<code>DF.MCMC</code>	0.740 (0.016)	0.498 (0.009)	0.335 (0.006)
		<code>DF.IS.noweight</code>	0.729 (0.015)	0.494 (0.009)	0.384 (0.006)
		<code>DF.IS</code>	0.728 (0.016)	0.492 (0.009)	0.360 (0.006)
Additive + Dominant	Effect Rank Accuracy	<code>partial.lm</code>	0.535 (0.010)	0.646 (0.008)	0.727 (0.006)
		<code>ridge.add</code>	0.535 (0.010)	0.647 (0.007)	0.731 (0.006)
		<code>ridge.dom</code>	0.535 (0.010)	0.645 (0.007)	0.729 (0.006)
		<code>DF.MCMC.pseudo</code>	0.514 (0.010)	0.624 (0.008)	0.700 (0.006)
		<code>DF.MCMC</code>	0.527 (0.010)	0.643 (0.008)	0.729 (0.006)
		<code>DF.IS.noweight</code>	0.531 (0.010)	0.645 (0.007)	0.729 (0.006)
		<code>DF.IS</code>	0.530 (0.010)	0.643 (0.008)	0.728 (0.006)

Table 2: Performance of the methods applied on CC population with simulated phenotypes. The table shows the mean values of either prediction error or rank accuracy with halved 95% confidence interval size in parentheses.

Simulated Effects	Metric	Model	QTL effect size		
			20%	30%	40%
Additive	Effect MSE	<code>partial.lm</code>	0.288 (0.006)	0.199 (0.005)	0.154 (0.005)
		<code>ridge.add</code>	0.174 (0.003)	0.109 (0.002)	0.076 (0.001)
		<code>ridge.dom</code>	0.348 (0.007)	0.254 (0.006)	0.196 (0.005)
		<code>DF.MCMC.pseudo</code>	0.192 (0.004)	0.134 (0.003)	0.100 (0.002)
		<code>DF.MCMC</code>	0.166 (0.003)	0.102 (0.002)	0.068 (0.001)
		<code>DF.IS.noweight</code>	0.237 (0.005)	0.169 (0.004)	0.130 (0.003)
		<code>DF.IS</code>	0.188 (0.004)	0.120 (0.003)	0.086 (0.002)
	Effect Rank Accuracy	<code>partial.lm</code>	0.793 (0.005)	0.813 (0.005)	0.825 (0.005)
		<code>ridge.add</code>	0.797 (0.005)	0.818 (0.005)	0.831 (0.005)
		<code>ridge.dom</code>	0.791 (0.005)	0.814 (0.005)	0.826 (0.005)
		<code>DF.MCMC.pseudo</code>	0.784 (0.005)	0.807 (0.005)	0.821 (0.005)
		<code>DF.MCMC</code>	0.797 (0.005)	0.818 (0.005)	0.830 (0.005)
		<code>DF.IS.noweight</code>	0.793 (0.005)	0.814 (0.005)	0.825 (0.005)
		<code>DF.IS</code>	0.796 (0.005)	0.815 (0.005)	0.828 (0.005)
Additive + Dominant	Effect MSE	<code>partial.lm</code>	0.369 (0.009)	0.281 (0.008)	0.237 (0.007)
		<code>ridge.add</code>	0.219 (0.005)	0.158 (0.004)	0.127 (0.003)
		<code>ridge.dom</code>	0.399 (0.008)	0.321 (0.007)	0.273 (0.006)
		<code>DF.MCMC.pseudo</code>	0.246 (0.005)	0.184 (0.004)	0.152 (0.003)
		<code>DF.MCMC</code>	0.203 (0.004)	0.140 (0.004)	0.105 (0.004)
		<code>DF.IS.noweight</code>	0.274 (0.005)	0.213 (0.005)	0.175 (0.004)
		<code>DF.IS</code>	0.227 (0.005)	0.163 (0.004)	0.123 (0.003)
Additive + Dominant	Effect Rank Accuracy	<code>partial.lm</code>	0.777 (0.005)	0.797 (0.005)	0.806 (0.005)
		<code>ridge.add</code>	0.782 (0.005)	0.802 (0.005)	0.812 (0.005)
		<code>ridge.dom</code>	0.784 (0.005)	0.805 (0.005)	0.817 (0.005)
		<code>DF.MCMC.pseudo</code>	0.762 (0.005)	0.788 (0.005)	0.797 (0.005)
		<code>DF.MCMC</code>	0.784 (0.005)	0.805 (0.005)	0.817 (0.006)
		<code>DF.IS.noweight</code>	0.780 (0.005)	0.800 (0.005)	0.810 (0.005)
		<code>DF.IS</code>	0.782 (0.005)	0.801 (0.005)	0.811 (0.005)

Table 3: Performance of the methods applied on CC population with simulated phenotypes. The table shows the mean values of either prediction error or rank accuracy with halved 95% confidence interval size in parentheses.

Simulated Effects	Metric	Model	QTL effect size		
			2.5%	5%	10%
Additive	Effect MSE	<code>partial.lm</code>	32.678 (1.697)	19.145 (1.085)	12.427 (0.763)
		<code>ridge.add</code>	41.639 (3.217)	19.961 (1.556)	9.692 (0.715)
		<code>ridge.dom</code>	2293 (603)	1156 (289)	599 (156.9)
		<code>DF.IS.noweight</code>	0.850 (0.007)	0.814 (0.007)	0.803 (0.006)
		<code>DF.IS.kinship</code>	1.136 (0.020)	0.902 (0.015)	0.797 (0.013)
		<code>DF.IS</code>	0.915 (0.031)	0.813 (0.010)	0.764 (0.008)
	Effect Rank Accuracy	<code>partial.lm</code>	0.186 (0.014)	0.245 (0.013)	0.310 (0.012)
		<code>ridge.add</code>	0.178 (0.014)	0.241 (0.014)	0.311 (0.013)
		<code>ridge.dom</code>	0.095 (0.010)	0.126 (0.010)	0.160 (0.010)
		<code>DF.IS.noweight</code>	0.318 (0.012)	0.393 (0.011)	0.457 (0.009)
		<code>DF.IS.kinship</code>	0.246 (0.013)	0.325 (0.015)	0.398 (0.015)
		<code>DF.IS</code>	0.296 (0.030)	0.364 (0.012)	0.442 (0.010)
Additive + Dominant	Effect MSE	<code>partial.lm</code>	16.954 (0.909)	10.486 (0.602)	7.111 (0.444)
		<code>ridge.add</code>	18.434 (1.222)	10.580 (0.666)	5.805 (0.343)
		<code>ridge.dom</code>	1085 (256)	557 (142)	275 (68.40)
		<code>DF.IS.noweight</code>	0.910 (0.004)	0.892 (0.003)	0.887 (0.003)
		<code>DF.IS.kinship</code>	1.051 (0.011)	0.936 (0.007)	0.887 (0.006)
		<code>DF.IS</code>	0.946 (0.008)	0.893 (0.006)	0.867 (0.004)
	Effect Rank Accuracy	<code>partial.lm</code>	0.146 (0.010)	0.176 (0.010)	0.218 (0.009)
		<code>ridge.add</code>	0.122 (0.011)	0.160 (0.010)	0.205 (0.010)
		<code>ridge.dom</code>	0.073 (0.009)	0.101 (0.008)	0.137 (0.009)
		<code>DF.IS.noweight</code>	0.229 (0.009)	0.276 (0.008)	0.313 (0.008)
		<code>DF.IS.kinship</code>	0.180 (0.010)	0.229 (0.010)	0.270 (0.011)
		<code>DF.IS</code>	0.220 (0.010)	0.258 (0.009)	0.309 (0.008)

Table 4: Performance of the methods applied on HS population with simulated phenotypes. The table shows the mean values of either prediction error or rank accuracy with halved 95% confidence interval size in parentheses.

Simulated Effects	Metric	Model	QTL effect size		
			20%	30%	40%
Additive	Effect MSE	<code>partial.lm</code>	9.103 (0.590)	8.008 (0.529)	7.467 (0.498)
		<code>ridge.add</code>	4.566 (0.296)	2.959 (0.171)	2.293 (0.146)
		<code>ridge.dom</code>	270 (74.24)	164 (46.05)	111 (29.70)
		<code>DF.IS.noweight</code>	0.785 (0.006)	0.775 (0.007)	0.768 (0.007)
		<code>DF.IS.kinship</code>	0.744 (0.012)	0.702 (0.013)	0.709 (0.011)
		<code>DF.IS</code>	0.733 (0.008)	0.721 (0.008)	0.714 (0.008)
	Effect Rank Accuracy	<code>partial.lm</code>	0.371 (0.011)	0.399 (0.011)	0.416 (0.010)
		<code>ridge.add</code>	0.394 (0.012)	0.447 (0.012)	0.483 (0.011)
		<code>ridge.dom</code>	0.202 (0.010)	0.233 (0.010)	0.254 (0.010)
		<code>DF.IS.noweight</code>	0.502 (0.008)	0.516 (0.008)	0.526 (0.008)
		<code>DF.IS.kinship</code>	0.460 (0.014)	0.489 (0.013)	0.505 (0.012)
		<code>DF.IS</code>	0.488 (0.009)	0.501 (0.009)	0.516 (0.009)
Additive + Dominant	Effect MSE	<code>partial.lm</code>	5.414 (0.354)	4.845 (0.320)	4.558 (0.301)
		<code>ridge.add</code>	3.249 (0.164)	2.471 (0.115)	2.116 (0.110)
		<code>ridge.dom</code>	124 (30.90)	78 (19.99)	55 (13.51)
		<code>DF.IS.noweight</code>	0.879 (0.003)	0.874 (0.003)	0.870 (0.003)
		<code>DF.IS.kinship</code>	0.859 (0.006)	0.837 (0.007)	0.840 (0.005)
		<code>DF.IS</code>	0.854 (0.004)	0.849 (0.004)	0.844 (0.004)
	Effect Rank Accuracy	<code>partial.lm</code>	0.260 (0.009)	0.280 (0.008)	0.292 (0.008)
		<code>ridge.add</code>	0.261 (0.010)	0.293 (0.009)	0.316 (0.009)
		<code>ridge.dom</code>	0.179 (0.009)	0.208 (0.009)	0.229 (0.009)
		<code>DF.IS.noweight</code>	0.346 (0.007)	0.357 (0.007)	0.366 (0.006)
		<code>DF.IS.kinship</code>	0.309 (0.010)	0.340 (0.010)	0.348 (0.009)
		<code>DF.IS</code>	0.336 (0.008)	0.345 (0.008)	0.357 (0.007)

Table 5: Performance of the methods applied on HS population with simulated phenotypes. The table shows the mean values of either prediction error or rank accuracy with halved 95% confidence interval size in parentheses.

Simulated Effects	2.5% QTL	5% QTL	10% QTL	20% QTL	30% QTL	40% QTL
Additive	0.0010 (0.0001)	0.0014 (0.0001)	0.0025 (0.0001)	0.0051 (0.0001)	0.0083 (0.0002)	0.0125 (0.0003)
Additive + Dominant	0.0013 (0.0001)	0.0017 (0.0001)	0.0032 (0.0001)	0.0066 (0.0001)	0.0108 (0.0002)	0.0163 (0.0003)

Table 6: The table shows the mean values of TDI with halved 95% confidence interval size in parentheses.

Files S2-S4

Available for download as .gz files at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.166249/-/DC1>

File S2 Computer code, including simulation and example code

File S3 Data used in the simulations

File S4 Real data examples used