A general hidden Markov model for multi-parent populations

Karl W. Broman[*,1]

[*]Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison,

Madison, Wisconsin 53706

2021-08-03

Running head: General HMM for MPPs

Key words: quantitative trait loci, QTL, HMM, Collaborative Cross, MAGIC, diversity outbred mice, heterogeneous stock

[1]**Corresponding author:**

Karl W Broman

Department of Biostatistics and Medical Informatics

University of Wisconsin–Madison

2126 Genetics-Biotechnology Center

425 Henry Mall

Madison, WI 53706

Phone: 608–262–4633

Email: broman@wisc.edu

Abstract

A common step in the analysis of multi-parent populations is genotype reconstruction: identifying the founder origin of haplotypes from dense marker data. This process often makes use of a model for the exchange process along chromosomes, including the relative frequency of founder alleles and of the probability of exchanges among them. While the precise experimental design used to generate the population may be used to derive a precise characterization of the exchange process, this can be tedious, particularly given the great variety of experimental designs that have been proposed. We describe an approximate model that can be applied for a variety of multi-parent populations. We have implemented the approach in the R/qtl2 software, and we illustrate its use in applications to publicly available data on Diversity Outbred and Collaborative Cross mice.

Introduction

Multi-parent populations (MPPs) are a valuable resource for the analysis of complex traits (de Koning and McIntyre 2017), including the mapping of quantitative trait loci (QTL). A wide variety of MPPs have been developed, including heterogeneous stock (HS) in mice (Mott *et al.* 2000) and rats (Solberg Woods *et al.* 2010), eight-way recombinant inbred lines (RIL) in mice (The Complex Trait Consortium 2004) and Drosophila (King *et al.* 2012), and multi-parent advanced generation intercross (MAGIC) populations in a variety of plant species including arabidopsis (Kover *et al.* 2009), wheat (Cavanagh *et al.* 2008), maize (Dell'Acqua *et al.* 2015), and rice (Bandillo *et al.* 2013).

QTL mapping in MPPs can be performed through statistical tests at individual single nucleotide polymorphisms (SNPs), as used in genome-wide association studies. However, many investigators first seek to reconstruct the mosaic of founder haplotypes along the chromosomes of MPP individuals and use this reconstruction to test for association between founder alleles and the quantitative phenotype. This approach was first introduced by Mott *et al.* (2000) for the analysis of HS mice, implemented in the HAPPY software, and has been continued in packages such as R/mpMap (Huang and George 2011), DOQTL (Gatti *et al.* 2014), and R/qtl2 (Broman *et al.* 2019a).

The process of genotype reconstruction in an MPP individual is illustrated in Fig. **1**. The genotypes in the founder strains (Fig. **1a**) and the MPP offspring (Fig. **1b**) are used to calculate the probability of each possible founder genotype at each position along the chromosome (Fig. **1c**). Thresholding of these probabilities can be used to infer the founder genotypes and the locations of recombination breakpoints (Fig. **1d**).

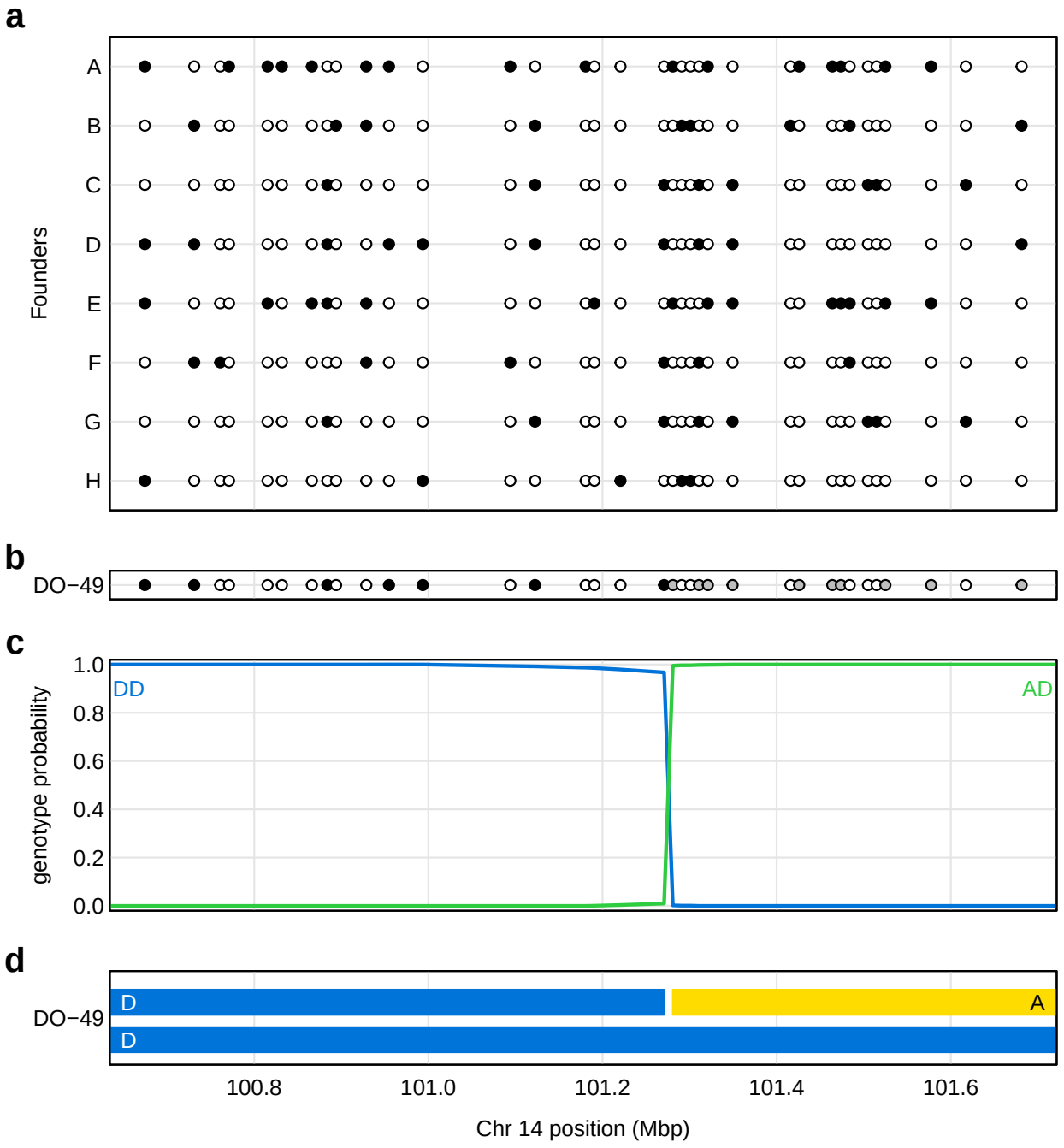

Figure 1 Illustration of genotype reconstruction in a 1 Mbp region in a single Diversity Outbred (DO) mouse. **a**. Genotypes of eight founder strains at a set of SNPs, with open and closed circles corresponding to being homozygous for the more-frequent and less-frequent allele, respectively. **b**. Genotype of the DO mouse at the SNPs, with gray indicating the mouse is heterozygous. **c**. Genotype probabilities for the DO mouse along the chromosome segment, given the observed data. Genotypes other than the two shown have negligible probability across the region. **d**. Inferred haplotypes in the DO mouse.

Such genotype reconstructions are valuable not just for QTL analysis but also for data diagnostics (Broman *et al.* 2019b). For example, the inferred number of recombination breakpoints is a useful diagnostic for sample quality. Further, the reconstructed genotypes can be used to derive predicted SNP genotypes; comparing these to the observed SNP genotypes can help to identify problems in both samples and SNPs.

The probability calculation in Fig. **1c** depends on a model for the process along MPP chromosomes in Fig. **1d**. In the HAPPY software for HS mice, Mott *et al.* (2000) used a model of random mating in a large population. Broman (2005) extended the work of Haldane and Waddington (1931) to derive transition probabilities along chromosomes in multi-strain recombinant inbred lines. This was later developed for the case of advanced intercross populations (Broman 2012a,b), including Diversity Outbred (DO) mice (Churchill *et al.* 2012).

Genotype reconstruction for a variety of MPP designs has been implemented in the R/qtl2 software (Broman *et al.* 2019a, `https://kbroman.org/qtl2`). But it can be tedious analytical work to derive the appropriate transition probabilities for each new MPP design that is proposed. An alternative is to develop a more general approach for genotype reconstruction, such as used in the software RABBIT (Zheng *et al.* 2015). However, this approach has a variety of parameters that can be difficult to specify.

Here we propose a similarly general method for genotype reconstruction in MPPs. We imagine that an MPP was derived from a population of homozygous founder strains at known proportions, α_i, followed by n generations of random mating among a large number of mating pairs. We can derive the exact transition probabilities for this situation. The α_i should be simple to specify from the MPP design, and the effective number of generations of random mating, n, can be determined by computer simulation, to match the expected density of recombination

breakpoints.

Our approach has been implemented in R/qtl2. While we currently focus on data with SNP genotype calls, such as from microarrays, our model could be incorporated into methods for genotype imputation from low-coverage sequencing, such as that of Zheng *et al.* (2018). We illustrate our approach through application to publicly-available datasets on DO (Al-Barghouthi *et al.* 2021) and Collaborative Cross mice (Srivastava *et al.* 2017).

Methods

For genotype reconstruction in a multi-parent population (MPP), we use a hidden Markov model (HMM; see Rabiner 1989). Our basic approach is as described in Broman and Sen (2009, App. D) for a biparental cross; the extension to an MPP is straightforward and described below.

Consider an MPP derived from k inbred lines. We focus on a single individual, and on a single chromosome with M marker positions (including pseudomarkers: positions between markers at which we have no data but would like to infer the underlying genotype). Let G_i be the underlying genotype at position i. In a homozygous population, such as RIL, the G_i take one of k possible values, the k homozygous genotypes. In a heterozygous population, such as advanced intercross lines (AIL), the G_i take one of $\binom{k}{2} + k$ possible values, the $\binom{k}{2}$ heterozygotes and k homozygotes. Let O_i be the observed SNP genotype at position i (possibly missing). We assume that the G_i form a Markov chain (that $G_1, \ldots, G_{i-1}$ are conditionally independent of $G_{i+1}, \ldots, G_M$, given G_i), and that O_i is conditionally independent of everything else, given G_i. The forward-backward algorithm (see Rabiner 1989) provides a simple way to take advantage of the conditional independence structure of the HMM to calculate $\Pr(G_i|\boldsymbol{O})$.

The key parameters in the model are the *initial* probabilities, $\pi_g = \Pr(G_1 = g)$, the *transition* probabilities, $t_i(g, g') = \Pr(G_{i+1} = g' \mid G_i = g)$, and the *emission* probabilities, $e_i(g) = \Pr(O_i|G_i = g)$. A particular advantage of the HMM for genotype reconstruction is the easy incorporation of a model for genotyping errors (Lincoln and Lander 1992), which is done through the emission probabilities, which condition on the founder SNP genotypes but allow some fixed probability ϵ that the observed SNP genotype in the MPP individual is in error and incompatible with the underlying genotype G_i and the SNP genotypes in the founder lines.

The initial and transition probabilities govern the underlying Markov chain, including the relative frequency of founder alleles and the frequency of recombination breakpoints along MPP chromosomes. In principle, these probabilities may be derived on the basis of the crossing design for the MPP (Broman 2005). In practice, the transition probabilities can be tedious to derive, and exact calculations may provide no real advantage for genotype reconstruction.

Here, we derive the transition probabilities for a generic MPP design, which may then be applied generally. We consider a founder population with k inbred lines in proportions α_i, and imagine subsequent generations are produced by random mating with a very large set of mating pairs.

Consider a pair of loci separated by a recombination fraction of r (assumed the same in both sexes) and let $p_{ij}^{(n)}$ be the probability of that a random haplotype at generation n has alleles i and j. At $n = 0$, we have just the founding inbred lines, and so $p_{ij}^{(0)} = \alpha_i$ if $i = j$ and $= 0$ if $i \neq j$.

The probabilities from one generation to the next are related by a simple recursion, as in Broman (2012b). Consider a random haplotype at generation n. It was either a random haplotype from generation $n - 1$ transmitted intact without recombination, or it is a recombinant haplotype bringing together two random alleles. Thus

$$p_{ij}^{(n)} = (1 - r)p_{ij}^{(n-1)} + r\alpha_i\alpha_j \tag{1}$$

Using the same techniques described in Broman (2012b), we find the solutions:

$$p_{ij}^{(n)} = \begin{cases} \alpha_i^2 + (1 - r)^n \alpha_i(1 - \alpha_i) & \text{if } i = j \\ \alpha_i\alpha_j[1 - (1 - r)^n] & \text{if } i \neq j \end{cases} \tag{2}$$

The transition probabilities along a haplotype are derived by dividing the above by the marginal probability, α_i. Thus if G_1 and G_2 are the genotypes at the two loci, we have the following transition probabilities.

$$\Pr(G_2 = j \mid G_1 = i) = \begin{cases} \alpha_i + (1-r)^n(1-\alpha_i) & \text{if } i = j \\ \alpha_j[1-(1-r)^n] & \text{if } i \neq j \end{cases} \tag{3}$$

For a heterozygous population (such as heterogeneous stock or Diversity Outbred mice), an individual will have two random such haplotypes. For homozygous population (such as MAGIC), we treat them like doubled haploids, by taking a single random chromosome and doubling it.

For the X chromosome, we use the same equations but replace n with $(2/3)n$, since recombination occurs only in females, so in 2/3 of the X chromosomes. This provides a remarkably tight approximation.

You can potentially use the expected number of crossovers to calibrate the number generations of random mating, or the map expansion, which is the relative increase in the number of crossovers. Let $R(r)$ be the chance that a random haplotype has an exchange of alleles across an interval with recombination fraction r, so $R(r) = 1 - \sum_i p_{ii}^{(n)}$. The map expansion is dR/dr evaluated at $r = 0$ (see Teuscher and Broman 2007). Using equation (2) above, we then get that the map expansion in this population is $n(1 - \sum \alpha_i^2)$. In the special case that $\alpha_i \equiv 1/k$ for all i, this reduces to $n(k-1)/k$.

The map expansion at generation s in DO mice on an autosome is $(7/8)(s-1) + M_1$ where M_1 is the weighted average of map expansion in the pre-CC founders (Broman 2012b), or about $(7s+37)/8$. Equating this with $(7/8)n$, we can thus take $n \approx s + 5$ when using this model to

approximate the DO. For the Collaborative Cross, Broman (2005) showed that R = 7r/(1+6r), and so the map expansion is 7. Thus we can take $n = 8$ as the effective number of generations of random mating.

Data and software availability

The R/qtl2 software is available at the Comprehensive R Archive Network (CRAN), `https://cran.r-project.org/package=qtl2`, as well as GitHub, `https://github.com/rqtl/qtl2`. Further documentation is available at the R/qtl2 website, `https://kbroman.org/qtl2`.

The Diversity Outbred mouse data from Al-Barghouthi *et al.* (2021) is available at Zenodo, `https://doi.org/10.5281/zenodo.4265417`. Also see their companion repository of analysis scripts at GitHub, `https://github.com/basel-maher/DO_project`, and archived at Zenodo, `https://doi.org/10.5281/zenodo.4718146`.

The Collaborative Cross mouse data from Srivastava *et al.* (2017) is available at Zenodo, `https://doi.org/10.5281/zenodo.377036`. Reorganized files in R/qtl2 format are at `https://github.com/rqtl/qtl2data/tree/master/CC`.

Our detailed analysis code is available at GitHub, `https://github.com/kbroman/Paper_GeneralHMM`, and archived at Zenodo, `https://doi.org/10.5281/zenodo.5156660`.

Applications

We illustrate our approach with application to datasets on Diversity Outbred mice (Al-Barghouthi *et al.* 2021) and Collaborative Cross (CC) mice (Srivastava *et al.* 2017). In both cases, the approach provided results that were generally equivalent to those from the more exact model, though with important differences in the results for the X chromosome in the CC application.

Diversity Outbred mice

The Diversity Outbred mouse data of Al-Barghouthi *et al.* (2021) concerns a set of 619 mice from DO generations 23–33, in 11 batches by generation and including 304 females and 315 males. The mice were genotyped on the GigaMUGA array (Morgan *et al.* 2016) and the cleaned data consist of genotypes at 109,427 markers. A wide variety of phenotypes are available; we focus on the 20 contributing to the results in Table 1 of Al-Barghouthi *et al.* (2021).

We performed genotype reconstruction using the transition matrices derived specifically for DO mice (Broman *et al.* 2019b) as well as by the approximate model proposed above. For the DO mice at generation n, we used the transition probabilities for a general 8-way advanced intercross lines (AIL) at $n + 5$.

Following Al-Barghouthi *et al.* (2021), we assumed a 0.2% genotyping error rate and used the Carter-Falconer map function (Carter and Falconer 1951). Calculations were performed in R (R Core Team 2021) with R/qtl2 (Broman *et al.* 2019a), on an 8-core Linux laptop with 64 GB RAM. The calculations with the DO-specific model took approximately 35 min, while those with the general AIL model took 27 min, an almost 25% reduction in computation time.

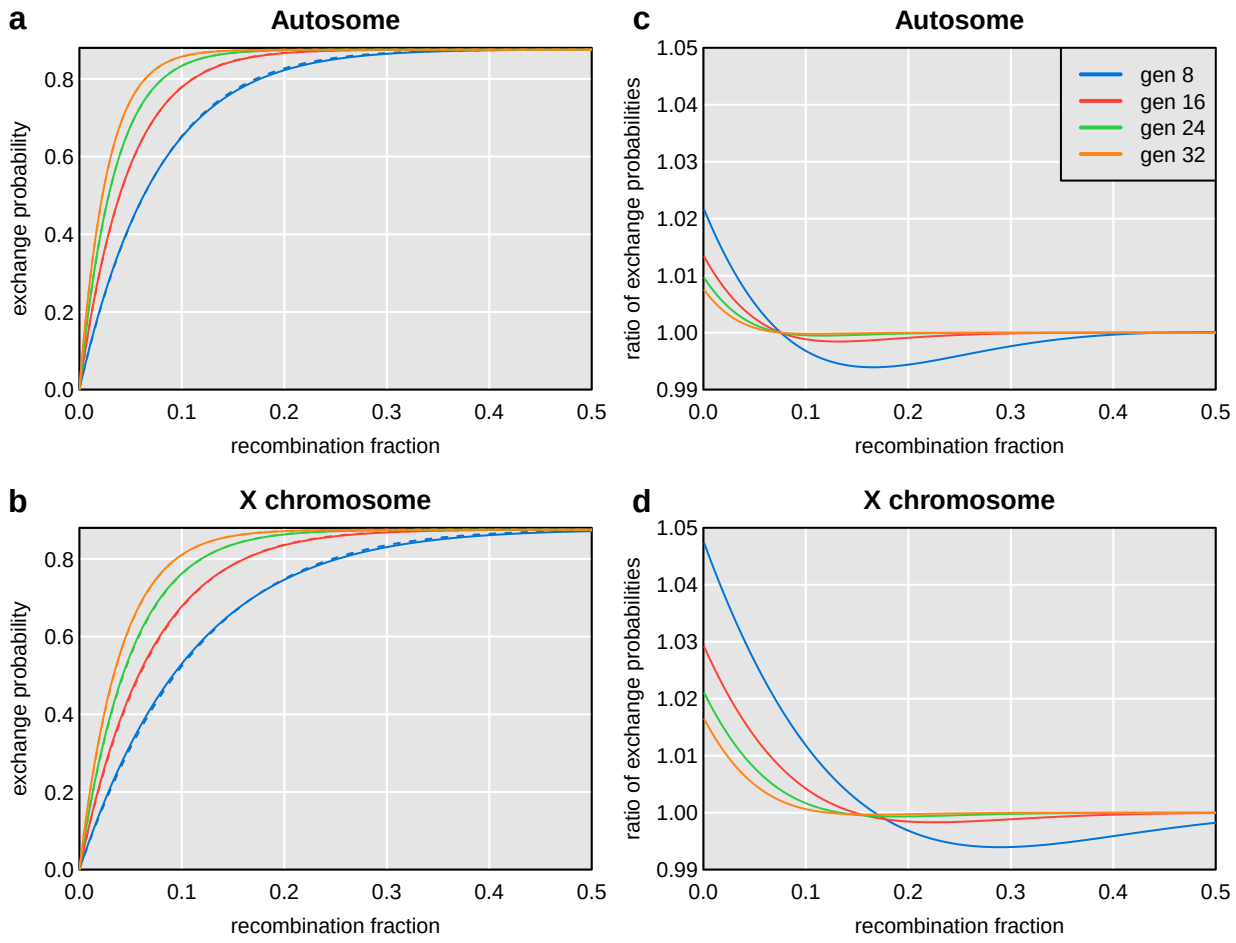

Figure 2 Differences in transition probabilities for Diversity Outbred mice from more-exact calculations and the proposed approximations. Probability of an exchange of alleles across an interval as a function of generation with the more-exact calculations (solid lines) and the proposed approximation (dashed lines) for autosomes (**a**) and the X chromosome (**b**). Ratio of the probabilities (more-exact versus approximation) for autosomes (**c**) and the X chromosome (**d**).

The transition probabilities used by the two models are only subtly different and become less different in later generations. The probability of an exchange across an interval on a random DO chromosome, as a function of the recombination fraction for the interval and the number of generations, is shown in Fig. **2**.

QTL analysis proceeded by the method described in Gatti *et al.* (2014) and also used by Al-Barghouthi *et al.* (2021). Namely, we fit a linear mixed model assuming an additive model for the founder haplotypes, with a residual polygenic effect to account for relationships among

individuals with kinship matrices calculated using the "leave-one-chromosome-out" (LOCO) method (see Yang *et al.* 2014), and with a set of fixed-effect covariates defined in Al-Barghouthi *et al.* (2021).

The genotype probabilities were almost indistinguishable. The maximum difference was 0.011 on the X chromosome followed by a difference of 0.007 on chromosome 8. For that reason, the QTL mapping results were hardly different. Across all 20 traits considered, the maximum difference in LOD scores in the two sets of results was 0.02.

The LOD curves by the two methods for tissue mineral density (TMD) and the differences between them are shown in Fig. **3**. The QTL on chromosomes 1 and 10 have LOD scores of 23.9 and 14.6, respectively, but the maximum difference in LOD, genome-wide, between the two methods is just 0.014.

Collaborative Cross mice

As a second application of our approach, we consider the data for a set of 69 Collaborative Cross (CC) lines (Srivastava *et al.* 2017). These are eight-way recombinant inbred lines (RIL) derived from the same eight founders as the DO mice, as the DO was formed from 144 partially-inbred lines from the process of developing the CC (Svenson *et al.* 2012).

Each CC line was formed from a separate "funnel," bringing the eight founder genomes together as rapidly as possible, for example [(A×B)×(C×D)]×[(E×F)×(G×H)], where the female parent is listed first in each cross. Inbreeding was accomplished by repeated mating between siblings.

The recombination probabilities for the autosomes in the CC do not depend on the order of the

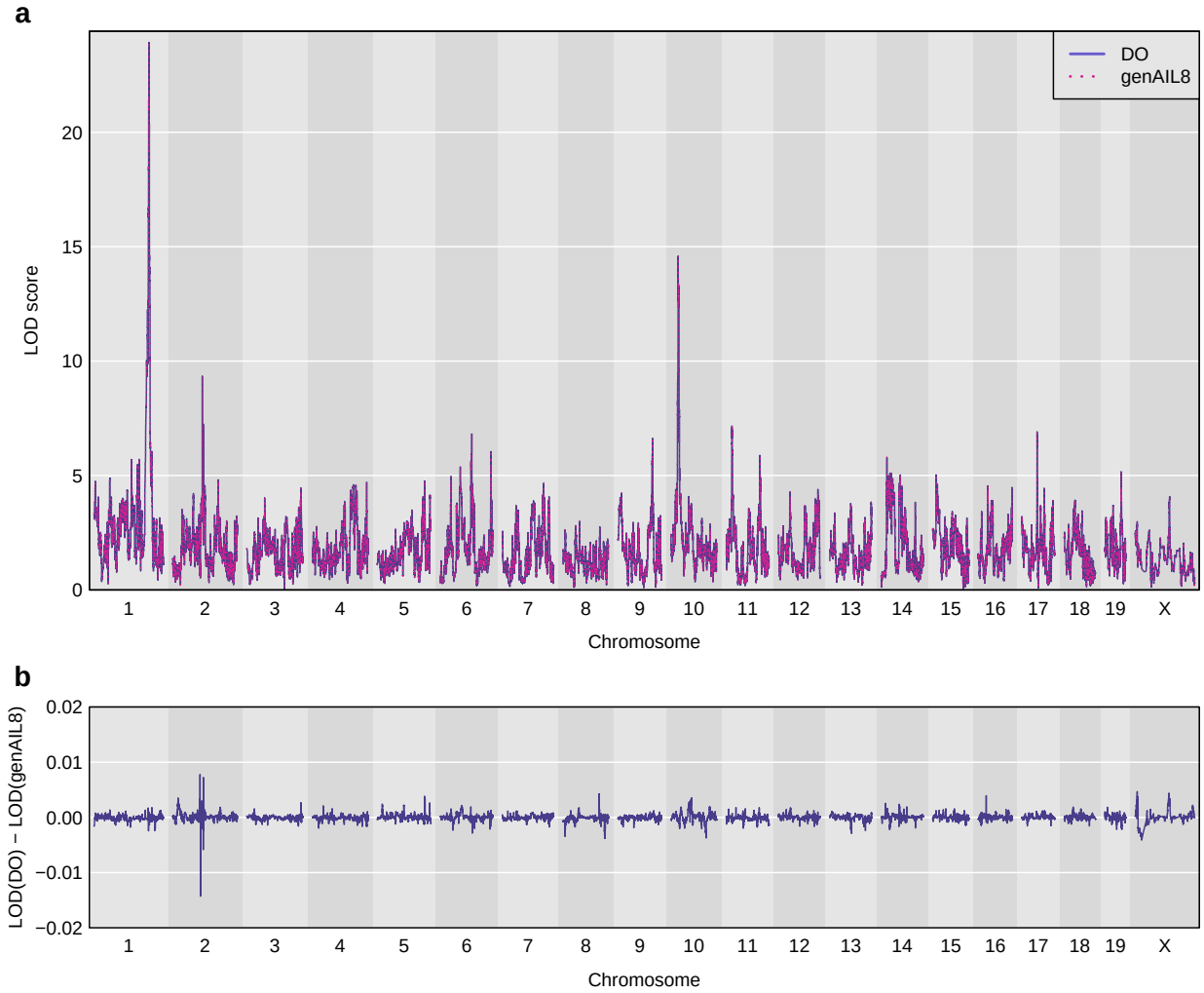

Figure 3 Genome scan for tissue mineral density (TMD) for the DO mouse data from Al-Barghouthi *et al.* (2021). **a**. LOD curves across the genome using the genotype probabilities from the DO-specific model (solid blue curves) and the proposed general model (dotted pink curves). **b**. Differences between the two sets of LOD curves.

founders in the funnel for a line (Broman 2005). This is in contrast with the case of 8-way RIL by selfing (see Broman 2005, Table 2). For the X chromosome, however, the cross order is important, as only 5 of the 8 founders can contribute. For example, in a line derived from the cross [(A×B)×(C×D)]×[(E×F)×(G×H)], the single-locus genotype probabilities on the X chromosome are 1/6 each for alleles A, B, E, and F, and 1/3 for allele C, while alleles D, G, and H will be absent. And note that the mitochondrial DNA will come from founder A, while the Y chromosome will be from founder H.

The cross funnel information was missing for 14 of the 69 CC lines. While the sources of the mitochondria and Y chromosome were provided for all lines, there were several inconsistencies in these data: line CC013/GeniUnc has the same founder listed as the source for its mitochondria and Y chromosome, and for three lines (CC031/GeniUnc, CC037/TauUnc, and CC056/GeniUnc) the founder on the Y chromosome is also seen contributing to the X chromosome. We used the genotype probabilities reported in Srivastava *et al.* (2017) to construct compatible cross funnels, with small modifications to handle the inconsistent information.

We performed genotype reconstruction using the transition matrices derived specifically for CC mice (Broman 2005) as well as by the approximate model proposed above, using $n = 8$ generations of random mating, chosen to match the expected frequency of recombination breakpoints.

The resulting probabilities were nearly identical on all autosomes in all CC lines. The maximum difference in probabilities on the autosomes was just 0.0006.

There were some important differences on the X chromosome, however. There were no cases with high probability pointing to different founder alleles by the two models, but there were several cases where two or more founders cannot be distinguished, but some would be excluded

by the assumed cross design.

For example, in Figure **4**, we show the genotype probabilities along the X chromosome for strain CC038/GeniUnc, as calculated with the more-exact model (Figure **4a**) and with the approximate model (Figure **4b**). We also include the results for the case that the more-exact model but with an incorrect cross design was used (Figure **4c**). Note the segment near 135 Mbp, which is inferred to be from founder NOD with the more-exact model but is equally likely B6 or NOD with the approximate model; the B6 and NOD founder strains are identical in the region, but the assumed cross design for the CC038/GeniUnc strain excluded B6. For the results using the incorrect cross design (which excluded not just B6 but also 129 and NOD), the results across the entire chromosome become a chopped-up mess, with an apparent 39 recombination breakpoints, versus 5 when the correct cross information is used.

Overall, there were seven strains where the maximum difference in the probabilities from the more-exact model and the proposed approximate model were in the range 0.25 – 0.50, and another eight strains with maximum difference in the range 0.10 – 0.25. All of the differences concern cases where multiple founders are identical for a region and either some would be excluded by the cross design, or where the difference in prior frequencies affects the results, in that in the cross [(A×B)×(C×D)]×[(E×F)×(G×H)], the frequency of the C allele on the X chromosome is twice that of A, B, E, and F.

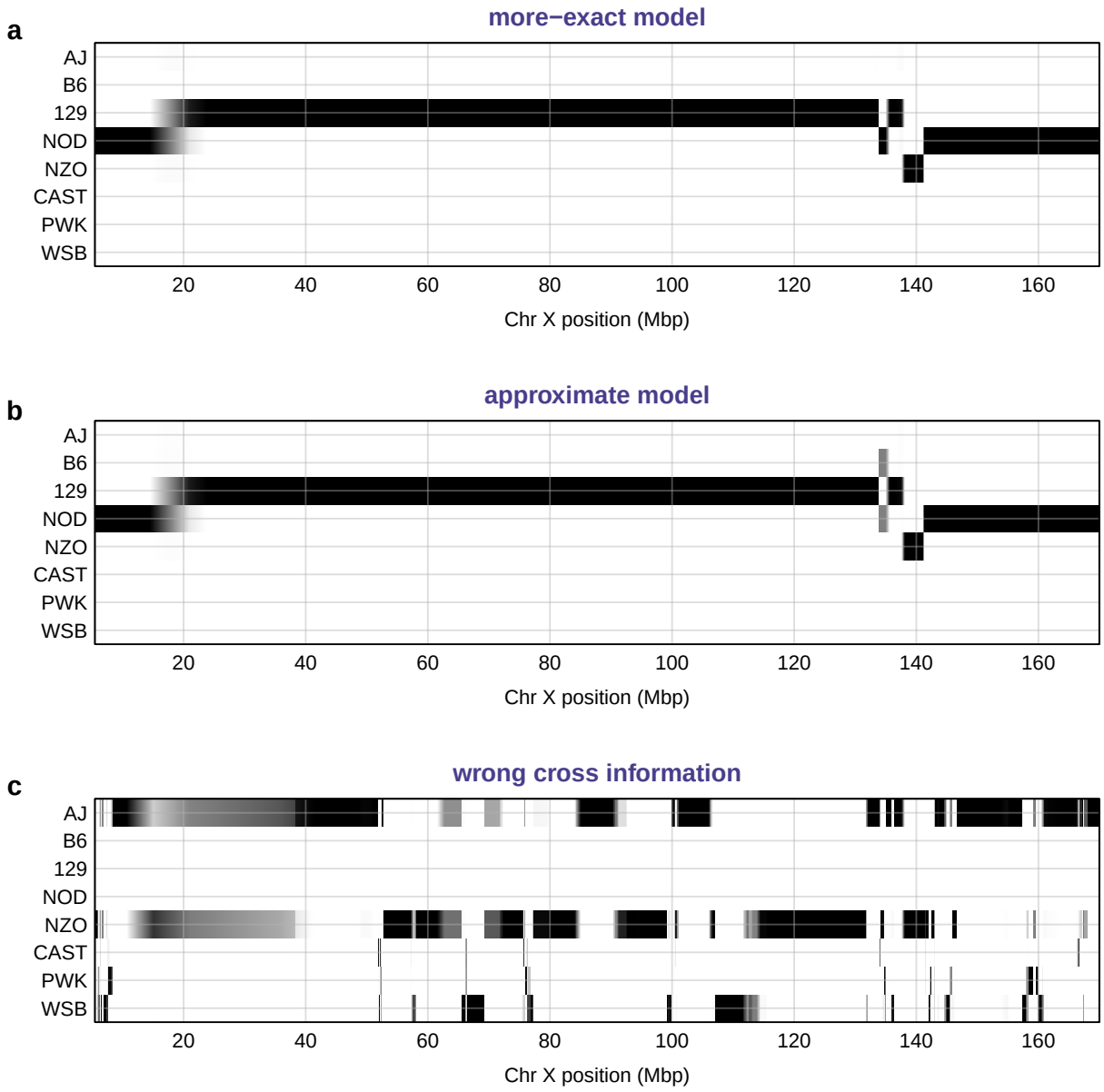

Figure 4 Genotype probabilities along the X chromosome for Collaborative Cross strain CC038/GeniUnc. **a**. Results using the more-exact model that excludes founders B6, CAST, and WSB. **b**. Results using the proposed approximate model. **c**. Results using the more-exact model but with the wrong cross information, excluding founders B6, 129, and NOD.

Discussion

We have proposed an approximate model for use with genotype reconstruction in multi-parent populations (MPPs). We derived the two-point probabilities on autosomes in the case of random mating in large, discrete generations, derived from a founder population of a set of inbred lines in known proportions. We use the same frequencies for the X chromosome, but with 2/3 the number of generations. The approach is shown to give equivalent results for the mouse DO and CC populations, though with important differences for the X chromosome in CC lines, where some founder alleles can be excluded based on the cross design. The more-exact model for the X chromosome in the CC excludes three of the eight founders based on the cross design. This is particularly useful in cases that multiple founders are identical by descent across a region. However, the approximate model is robust to errors in the specified cross design.

The value of this generic model points towards the general usefulness of the original software for multi-parent populations, HAPPY (Mott *et al.* 2000), developed for the analysis of mouse heterogeneous stock. Here, we have extended that approach to allow differences in the frequencies of founders. While in the two applications considered here, the founders were combined in equal proportions, there are other multi-parent populations where that is not the case. For example, the maize MAGIC lines of Dell'Acqua *et al.* (2015) were largely derived from eight founders, but breeding difficulties led to the introduction of a ninth founder at somewhat varying proportion.

The results are dependent on marker density and informativeness. With a dense set of informative markers, the choice of model used for the reconstruction has less influence on the results.

The hidden Markov model itself is an approximation. Meiosis generally exhibits positive crossover interference, but the Markov property is closer to being correct in multi-parent populations with multiple generations of mating, because nearby recombination events come from independent generations. This was apparent in the 3-point probabilities derived by Haldane and Waddington (1931) for two-way RIL and was further explored in Broman (2005) for multi-way RIL.

The proposed method has been implemented in the R/qtl2 software (Broman *et al.* 2019a). It requires specification of the founder proportions and one other parameter (the number of generations of random mating) which governs the frequency of recombination breakpoints. The founder proportions should be straightforward from the cross design; the effective number of generations of random mating may require some calibration, such as through computer simulation to match the expected frequency of recombination breakpoints.

Acknowledgments

This work was supported in part by National Institutes of Health grant R01GM070683.

Literature Cited

Al-Barghouthi, B. M., L. D. Mesner, G. M. Calabrese, D. Brooks, S. M. Tommasini *et al.*, 2021 Systems genetics in Diversity Outbred mice inform BMD GWAS and identify determinants of bone strength. Nat Commun **12**: 1–19.

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**: 1–15.

Broman, K. W., 2005 The genomes of recombinant inbred lines. Genetics **169**: 1133–1146.

Broman, K. W., 2012a Genotype probabilities at intermediate generations in the construction of recombinant inbred lines. Genetics **190**: 403–412.

Broman, K. W., 2012b Haplotype probabilities in advanced intercross populations. G3 (Bethesda) **2**: 199–202.

Broman, K. W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins *et al.*, 2019a R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics **211**: 495–502.

Broman, K. W., D. M. Gatti, K. L. Svenson, Ś. Sen, and G. A. Churchill, 2019b Cleaning genotype data from Diversity Outbred mice. G3 (Bethesda) **9**: 1571–1579.

Broman, K. W., and S. Sen, 2009 *A guide to QTL mapping with R/qtl*. Springer, New York.

Carter, T., and D. Falconer, 1951 Stocks for detecting linkage in the mouse, and the theory of their design. Journal of Genetics **50**: 307–323.

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**: 215–221.

Churchill, G. A., D. M. Gatti, S. C. Munger, and K. L. Svenson, 2012 The diversity outbred mouse population. Mamm Genome **23**: 713–718.

de Koning, D., and L. McIntyre, 2017 Back to the future: multiparent populations provide the key to unlocking the genetic basis of complex traits. Genetics **206**: 527–529.

Dell'Acqua, M., D. M. Gatti, G. Pea, F. Cattonaro, F. Coppens *et al.*, 2015 Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome Biol **16**: 167.

Gatti, D. M., K. L. Svenson, A. Shabalin, L.-Y. Wu, W. Valdar *et al.*, 2014 Quantitative trait locus mapping methods for diversity outbred mice. G3 (Bethesda) **4**: 1623–1633.

Haldane, J. B. S., and C. H. Waddington, 1931 Inbreeding and linkage. Genetics **16**: 357–374.

Huang, B. E., and A. W. George, 2011 R/mpmap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. Bioinformatics **27**: 727–729.

King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen *et al.*, 2012 Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res **22**: 1558–1566.

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**: e1000551.

Lincoln, S. E., and E. S. Lander, 1992 Systematic detection of errors in genetic linkage data. Genomics **14**: 604–610.

Morgan, A. P., C.-P. Fu, C.-Y. Kao, C. E. Welsh, J. P. Didion *et al.*, 2016 The Mouse Universal Genotyping Array: from substrains to subspecies. G3 (Bethesda) **6**: 263–279.

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**: 12649–12654.

R Core Team, 2021 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabiner, L. R., 1989 A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**: 257–286.

Solberg Woods, L. C., C. Stelloh, K. R. Regner, T. Schwabe, J. Eisenhauer *et al.*, 2010 Heterogeneous stock rats: a new model to study the genetics of renal phenotypes. Am J Physiol-Renal **298**: F1484–F1491.

Srivastava, A., A. P. Morgan, M. L. Najarian, V. K. Sarsani, J. S. Sigmon *et al.*, 2017 Genomes of the mouse Collaborative Cross. Genetics **206**: 537–556.

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.

Teuscher, F., and K. W. Broman, 2007 Haplotype probabilities for multiple-strain recombinant inbred lines. Genetics **175**: 1267–1274.

The Complex Trait Consortium, 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet **36**: 1133–1137.

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

Zheng, C., M. P. Boer, and F. A. van Eeuwijk, 2015 Reconstruction of genome ancestry blocks in multiparental populations. Genetics **200**: 1073–1087.

Zheng, C., M. P. Boer, and F. A. van Eeuwijk, 2018 Accurate genotype imputation in multiparental populations from low-coverage sequence. Genetics **210**: 71–82.