generic HMM for multi-parent populations

Karl Broman

Biostatistics & Medical Informatics, UW–Madison

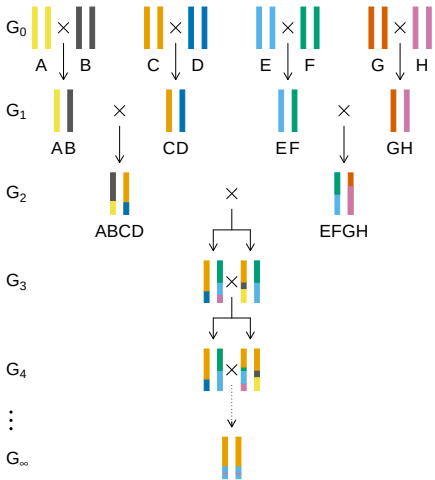
@kwbroman

kbroman.org

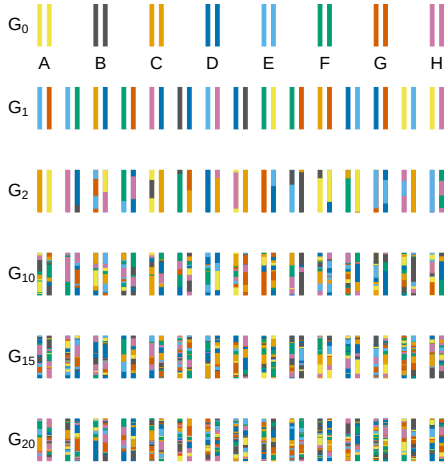
github.com/kbroman

kbroman.org/Talk_GenericHMM

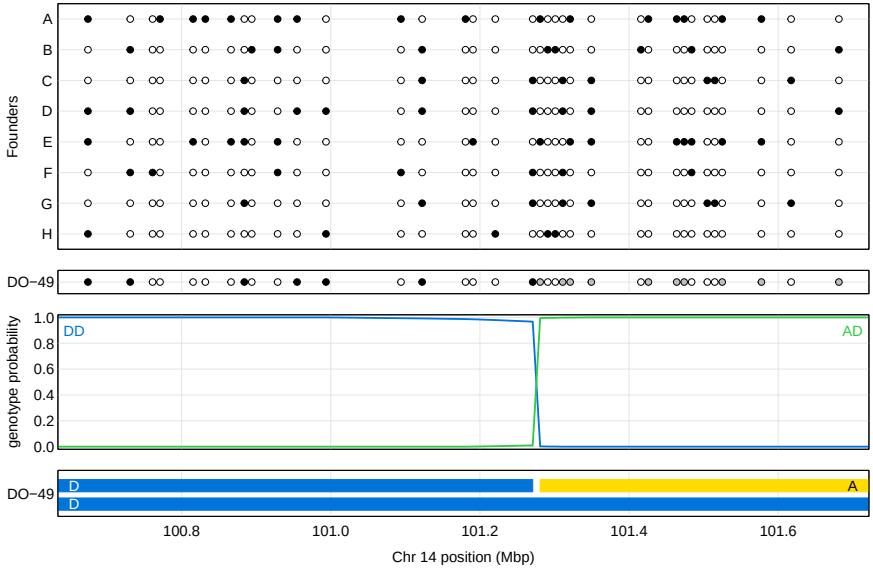
Recombinant Inbred Lines



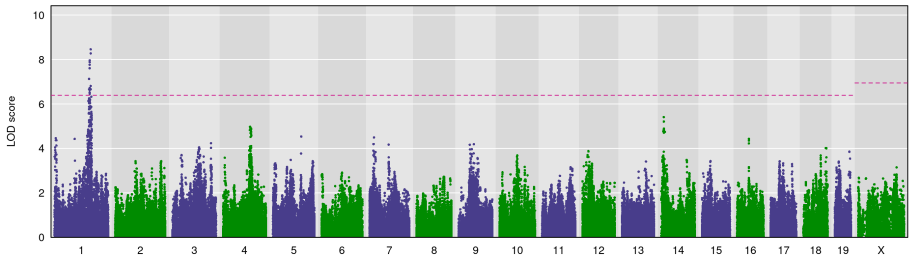
Advanced Intercross Population



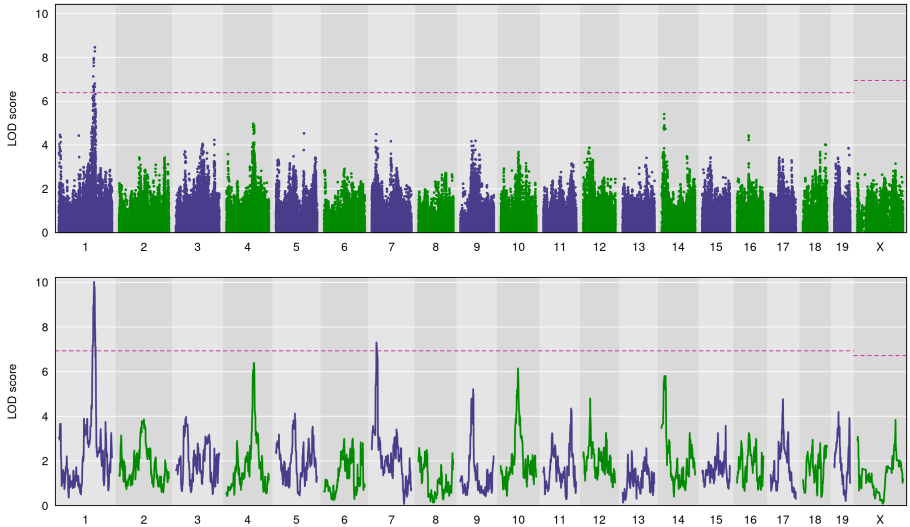
Genome reconstruction



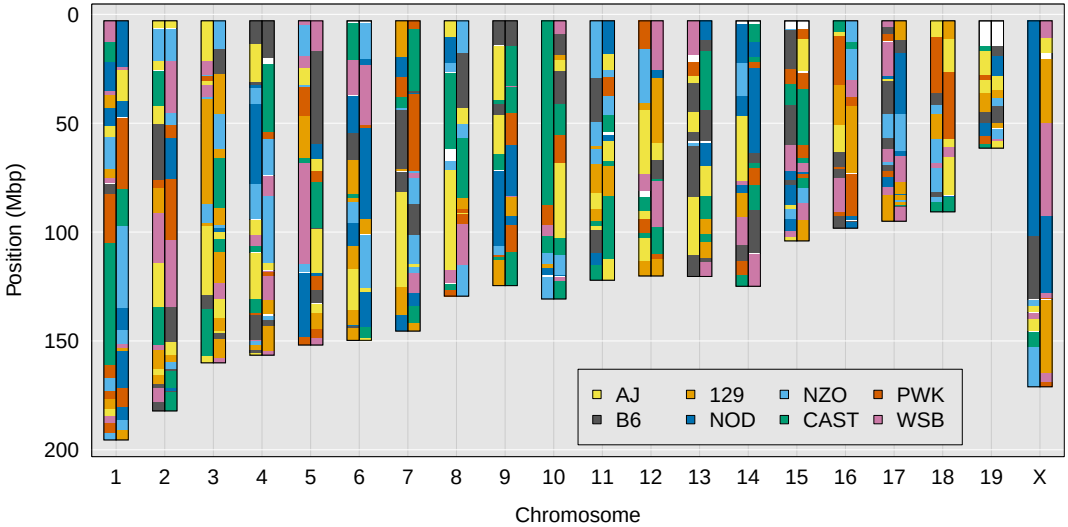
QTL genome scan



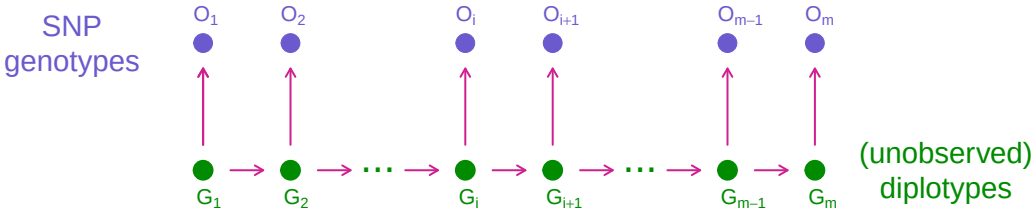
QTL genome scan



DO genome


0
50
100
150
200
Position (Mbp)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
X
Chromosome
AJ
B6
129
NOD
NZO
CAST
PWK
WSB

Hidden Markov model



Initial	$\pi(g) = \Pr(G_1 = g)$
Transition	$t_i(g, g') = \Pr(G_{i+1} = g' \mid G_i = g)$
Emission	$e_i(g) = \Pr(O_i \mid G_i = g)$

Exact probabilities

The Genomes of Recombinant Inbred Lines

Karl W. Broman[1]

Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205

Manuscript received August 20, 2004
Accepted for publication November 5, 2004

ABSTRACT

Recombinant inbred lines (RILs) can serve as powerful tools for genetic mapping. Recently, members of the Complex Trait Consortium proposed the development of a large panel of eight-way RILs in the mouse, derived from eight genetically diverse parental strains. Such a panel would be a valuable community resource. The use of such eight-way RILs will require a detailed understanding of the relationship between alleles at linked loci on an RI chromosome. We extend the work of Haldane and Waddington on two-way RILs and describe the map expansion, clustering of breakpoints, and other features of the genomes of multiple-strain RILs as a function of the level of crossover interference in meiosis.

Exact probabilities

The Genomes of Recombinant Inbred Lines

Karl W. Broman[1]

Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205

Manuscript received August 20, 2004
Accepted for publication November 5, 2004

ABSTRACT

Recombinant inbred lines (RILs) can serve as powerful tools for genetic mapping. Recently, members of the Complex Trait Consortium proposed the development of a large panel of eight-way RILs in the mouse, derived from eight genetically diverse parental strains. Such a panel would be a valuable community resource. The use of such eight-way RILs will require a detailed understanding of the relationship between alleles at linked loci on an RI chromosome. We extend the work of Haldane and Waddington on two-way RILs and describe the map expansion, clustering of breakpoints, and other features of the genomes of multiple-strain RILs as a function of the level of crossover interference in meiosis.

Haplotype Probabilities for Multiple-Strain Recombinant Inbred Lines

Friedrich Teuscher* and Karl W. Broman†,[1]

**Research Unit Genetics and Biometry, Research Institute for the Biology of Farm Animals (FBN), Dummerstorf, Germany 18196 and* †*Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205*

Manuscript received July 28, 2006
Accepted for publication November 28, 2006

ABSTRACT

Recombinant inbred lines (RIL) derived from multiple inbred strains can serve as a powerful resource for the genetic dissection of complex traits. The use of such multiple-strain RIL requires a detailed knowledge of the haplotype structure in such lines. Broman (2005) derived the two- and three-point haplotype probabilities for 2^n-way RIL; the former required hefty computation to infer the symbolic results, and the latter were strictly numerical. We describe a simpler approach for the calculation of these probabilities, which allowed us to derive the symbolic form of the three-point haplotype probabilities. We also extend the two-point results for the case of additional generations of intermating, including the case of 2^n-way intermated recombinant inbred populations (IRIP).

Genotype Probabilities at Intermediate Generations in the Construction of Recombinant Inbred Lines

Karl W. Broman[1]

Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin 53706

ABSTRACT The mouse Collaborative Cross (CC) is a panel of eight-way recombinant inbred lines: eight diverse parental strains are intermated, followed by repeated sibling mating, many times in parallel, to create a new set of inbred lines whose genomes are random mosaics of the genomes of the original eight strains. Many generations are required to reach inbreeding, and so a number of investigators have sought to make use of phenotype and genotype data on mice from intermediate generations during the formation of the CC lines (so-called pre-CC mice). The development of a hidden Markov model for genotype reconstruction in such pre-CC mice, on the basis of incompletely informative genetic markers (such as single-nucleotide polymorphisms), formally requires the two-locus genotype probabilities at an arbitrary generation along the path to inbreeding. In this article, I describe my efforts to calculate such probabilities. While closed-form solutions for the two-locus genotype probabilities could not be derived, I provide a prescription for calculating such probabilities numerically. In addition, I present a number of useful quantities, including single-locus genotype probabilities, two-locus haplotype probabilities, and the fixation probability and map expansion at each generation along the course to inbreeding.

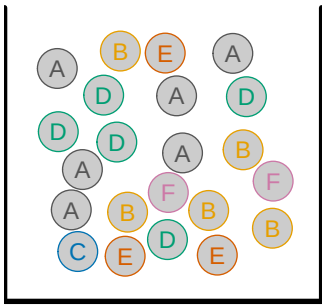
Haplotype Probabilities in Advanced Intercross Populations

Karl W. Broman[1]

Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, Madison, Wisconsin 53706

ABSTRACT Advanced intercross populations, in which multiple inbred strains are mated at random for many generations, have the advantage of greater precision of genetic mapping because of the accumulation of recombination events across the multiple generations. Related designs include heterogeneous stock and the diversity outcross population. In this article, I derive the two-locus haplotype probabilities on the autosome and X chromosome with these designs. These haplotype probabilities provide the key quantities for developing hidden Markov models for the treatment of missing genotype information. I further derive the map expansion in these populations, which is the frequency of recombination breakpoints on a random chromosome.

Generic model



k founders in proportions $\{\alpha_i\}$

n generations of random mating

Random chromosome:

$$\pi_i = \alpha_i$$

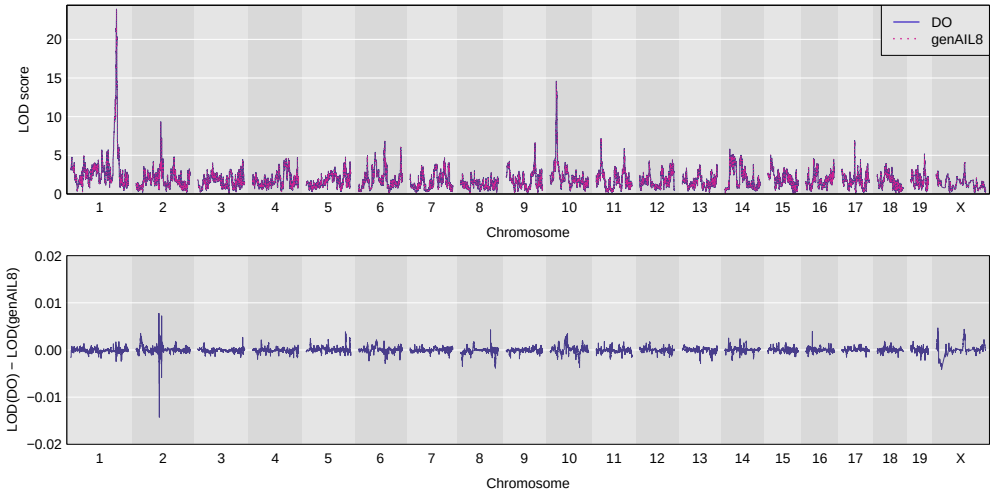
$$t_{ij} = \alpha_j\,[1 - (1 - r)^n] \quad \text{when } i \neq j$$

Map expansion:

$$n(1 - \sum \alpha_i^2)$$

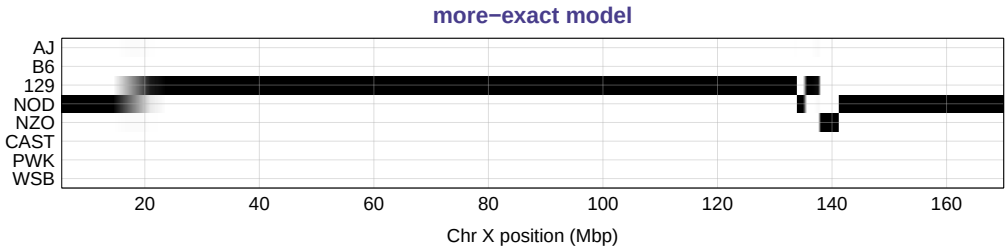
$$= n\left(\tfrac{k-1}{k}\right) \quad \text{if } \alpha_i \equiv 1/k$$

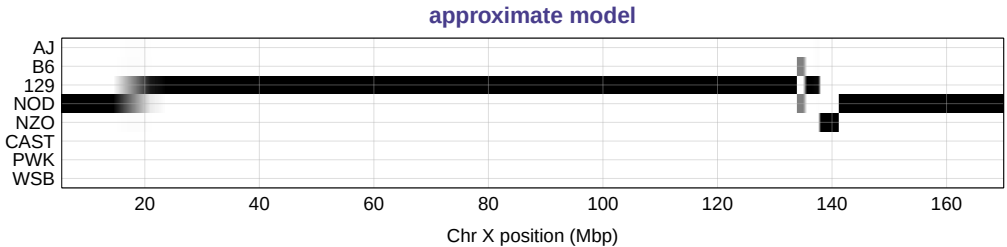
DO application



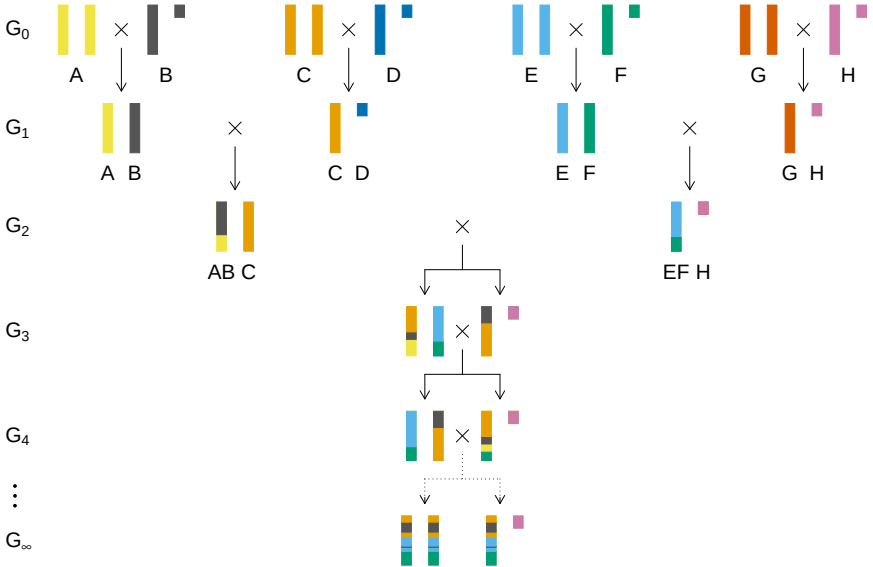
data from Al-Barghouthi et al (2021) doi.org/gkf64n

CC038 X chr


more−exact model
AJ
B6
129
NOD
NZO
CAST
PWK
WSB
20
40
60
80
100
120
140
160
Chr X position (Mbp)


approximate model
AJ
B6
129
NOD
NZO
CAST
PWK
WSB
20
40
60
80
100
120
140
160
Chr X position (Mbp)

X chr in CC



Summary

- Generic model for genome reconstruction in multi-parent populations
- Specify relative proportions of founders
  + effective number of generations of random mating
- Basic conclusion: HAPPY is effective
- Implemented in R/qtl2 as cross types `genril`n and `genail`n
  (replacing n with the number of founders)
- bioRxiv manuscript: `doi.org/gswx`

Slides: kbroman.org/Talk_GenericHMM

bioRxiv manuscript: doi.org/gswx

kbroman.org

github.com/kbroman

@kwbroman

kbroman.org/qtl2