BIOINFORMATICS APPLICATIONS NOTE Vol. 20 no. 15 2004, pages 2491–2492
doi:10.1093/bioinformatics/bth269

GenomeMixer: a complex genetic cross simulator

*Alexander G. Williams and Robert W. Williams**

Center of Genomics and Bioinformatics, Department of Anatomy and Neurobiology, University of Tennessee, 875 Monroe Avenue, Memphis, TN 38163, USA

Received on September 15, 2003; revised on February 21, 2004; accepted on April 1, 2004
Advance Access publication April 15, 2004

ABSTRACT

Summary: GenomeMixer is a cross-platform application that simulates meiotic recombination events for large and complex multigenerational genetic crosses among sexually reproducing diploid species and outputs simulated progeny to several standard mapping programs.

Availability: Documentation, C++ source, and binaries for Mac OS X and x86 Linux are freely available at http://www.nervenet.org/genome_mixer/. GenomeMixer can be compiled on any system with support for the Trolltech Qt toolkit, including Windows.

Contact: rwilliam@nb.utmem.edu

Mapping large sets of polymorphic genes that contribute to differences in phenotypes is a critical facet of modern genetics. Even in the post-sequencing era, <20% of genes have known functions. Mapping is essential in characterizing gene function. The utility of a population for mapping traits is dependent on both the underlying genetic variation and on accurate knowledge of population genotype and haplotype structure. For simple crosses, it is possible to directly compute the power and precision with which traits can be mapped. However, for novel and highly complex crosses, such as advanced intercrosses, heterogeneous stock crosses, and eight-way recombinant inbred crosses (Darvasi, 1998; Vogel, 2003; Winkler *et al*., 2003), it is difficult to estimate power or precision without extensive simulation. GenomeMixer is a simulation program that can handle complex breeding designs and output progeny genotype files for further analysis in mapping software packages.

USAGE

GenomeMixer allows the user to manage all relevant breeding design options from a single tabbed settings window (Fig. 1). The user can edit the number and length of chromosomes, the number of purebred parental strains, the number and location of markers and the breeding design itself. Markers may be automatically generated at specified intervals, imported from a tab-delimited text file or edited manually. The breeding design is entered in a spreadsheet-like table of cages (rows) by generations (columns), in which each cell represents the progeny from a particular mating (Fig. 1). The user specifies a particular mating by selecting a cell, selecting two parents (either purebreds or individuals from a previous mating) and then setting the desired number of progeny. The output of that cell is a set of progeny that can then be used as parents for subsequent generations.

Once the chromosome parameters, marker maps and breeding design have been specified, the simulation is run, and the resulting progeny are examined graphically at a marker-resolution level (Fig. 1). These simulated progeny can also be exported to formats readable by QGene (Nelson, 1997; http://www.qgene.org/), MapManager QTX (Manly *et al*., 2001; http://mapmgr.roswellpark.org/mmQTX.html) and R/qtl (Broman *et al*., 2003; http://www.biostat.jhsph.edu/~kbroman/qtl/), although these programs have limited support for crosses that involve more than two parental strains.

In order for simulations to be biologically accurate, it is necessary to allow non-uniform recombination frequency across chromosomes (Froenicke *et al*., 2002) and to simulate crossover interference (Broman *et al*., 2002). Recombination frequency can be controlled across each chromosome by specifying two parameters for each marker: the physical location of the marker in bases, and a recombination fraction expressed in centimorgans (cM). In this manner, recombination can be precisely controlled. A recombination hotspot, for example, can be modeled by several markers that are close physically but distant in cM.

ALGORITHM

GenomeMixer calculates the probability of recombination over a chromosome by examining the cM distance between pairs of adjacent markers (reviewed in Speed and Zhao, 2003. In the default model, a 1 cM interval has a 1% chance of containing a recombination before interference is taken into account. The default formula for a given inter-marker distance is given by Haldane's function, where the probability of a recombination in m (Morgans) is $0.5 \times (1 - e^{-2m})$.

*To whom correspondence should be addressed.

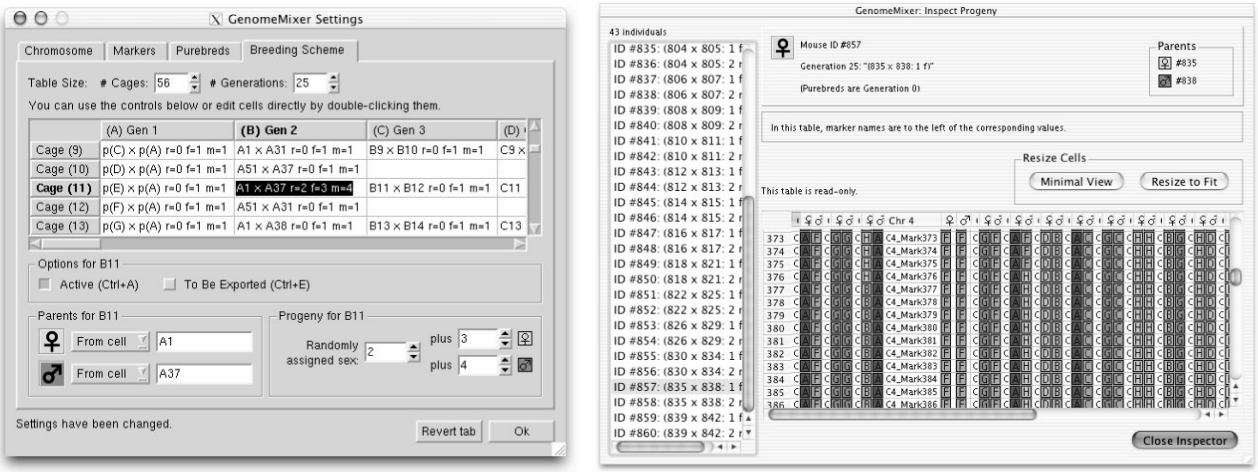

Fig. 1. Users design a breeding scheme in the spreadsheet window on the left (Linux version). The right-hand window displays the genotypes of simulated progeny (Mac OS X version). The columns have been resized so that the marker values for 12 chromosome pairs are visible, but the marker names are only shown for chromosome 4. The breeding scheme and results are both from an eight-way cross described in *Science* (Vogel, 2003).

The simulator never generates more than one recombination between two adjacent markers in a single meiosis, but as long as markers are more closely spaced than the interference distance, this limitation does not affect the final results, since interference would already forbid a second recombination in that interval. The minimum distance between recombinations is a user-specifiable parameter.

AVAILABILITY AND REQUIREMENTS

GenomeMixer is an open-source C++ project, released under the GPL. The source code, documentation, and compiled binaries for Mac OS X and x86 Linux can be downloaded from http://www.nervenet.org/genome_mixer/. A Perl command line version of GenomeMixer is also available for download, along with instructions for formatting the input files. Sample data files illustrating various types of crosses are included with the application. A version of Trolltech Qt (http://www.trolltech.com/) is required to compile the source. GenomeMixer should compile on any system that supports Qt, including Mac OS X, Linux and Windows.

Benchmarking. A 25-generation 8-way cross with 1000 mice and 3000 markers runs in ∼4.5 s on a 933 MHz Pentium Linux machine, and in ∼3.2 s on a 1.2 GHz Macintosh G4. With 30 000 markers, the same simulation takes ∼36 and ∼24 s, respectively.

FUTURE ADDITIONS

We plan to extend GenomeMixer to allow more sophisticated models of recombination and interference (Speed and Zhao, 2003), and selection coefficients for particular alleles or sets of alleles to simulate fitness and segregation distortion. Other potential additions include independent control of recombination in males and females, selfing for plant genetics, and routines to select progeny on the basis of genotype during simulation.

ACKNOWLEDGEMENTS

This project was funded jointly by the National Institute of Mental Health, National Institute on Drug Abuse, and the National Science Foundation (award P20-MH 62009), and by the National Institute on Alcohol Abuse and Alcoholism (INIA grants U01AA13499, U24AA13513).

REFERENCES

Broman,K.W., Rowe,L.B., Churchill,G.A. and Paigen,K. (2002) Crossover interference in the mouse. *Genetics*, **160**, 1123–1131.

Broman,K.W., Wu,H., Sen,S. and Churchill,G.A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**, 889–890.

Darvasi,A. (1998) Experimental strategies for the genetic dissection of complex traits in animal models. *Nat. Genet.*, **18**, 19–24.

Froenicke,E., Anderson,L.K., Wienberg,J. and Ashley,T. (2002) Male mouse recombination maps for each autosome identified by chromosome painting. *Am. J. Hum. Genet.*, **71**, 1353–1368.

Manly,K.F., Cudmore,R.H.,Jr and Meer,J.M. (2001). Map Manager QTX, cross-platform software for genetic mapping. *Mamm. Genome*, **12**, 930–932.

Nelson,J.C. (1997) QGene: software for marker-based genomic analysis and breeding. *Mol. Breed.*, **3**, 239–245.

Speed,T.P. and Zhao,H. (2003) Chromosome maps. In *Handbook of Statistical Genetics*, 2nd edn., Vol. 1, pp. 3–38.

Vogel,G. (2003) Genetics: scientists dream of 1001 complex mice. *Science*, **301**, 456–457.

Winkler,C.R., Jensen,N.M., Cooper,M., Podlich,D.W. and Smith,O.S. (2003) On the determination of recombination rates in intermated recombinant inbred populations. *Genetics*, **164**, 741–745.