GenetSim: Software for Simulation of Familial Data in Genetics and Epidemiology

Michael B. Miller PhD

Assistant Professor, Epidemiology

University of Minnesota - Twin Cities

Abstract:

GenetSim provides flexible simulations of family data within an easy-to-use, high-level programming language. GenetSim has been developed so far within the MATLAB-like environment of the free software package Octave (Eaton, 1997), but forthcoming versions will be designed to work within the R statistical language. GenetSim has no limit on pedigree sizes or structures (these can be imported from LINKAGE-format files), or number of families, no limit on number of marker or trait loci, no limit on number of chromosomes (nonhuman diploid species can be handled). Genetic transmission is modeled by first generating the locations of recombination junctions (according to nearly any multilocus feasible model desired - Haldane, Sturt, etc., or a user-specified model), and then performing gene dropping according to the given recombination pattern. Any pattern of missing data can be specified and genotyping errors (or other kinds of errors) can be simulated. GenetSim can simulate multiple QTLs with pleiotropic effects, multivariate polygenic background and any number of environmental factors, age effects, sex effects, epistasis and variable expression. Traits could be quantitative (continuous) or one could use penetrance functions and/or liability threshold models for affection-status (binary) traits. Users can also select families based on ascertainment schemes by repeating simulations (e.g., retain only families where at least two members have trait values exceeding 140). As a proof of concept, we generated multivariate genetic data with 6 QTLs on 10 chromosomes under the model used in GAW10. All 400 data sets (all markers and phenotypes) for 200 sets of 23 extended families and 200 sets of 239 nuclear families were completed in a total of less than 20 minutes on a dual Intel Xeon 2.8 GHz machine running Linux. GenetSim is freely available under the GNU General Public License.