SEQUENCE-LEVEL SIMULATIONS WITH GENE CONVERSION, RECOMBINATION HOTSPOTS AND SELECTION

D.J. Balding[†1], T. Clark[1], C.J. Hoggart[1], J.C. Whittaker[2], M. De Iorio[1]

[1] *Imperial College London, UK;*
[2] *London School of Hygiene and Tropical Medicine, UK*

[†] E-mail: *d.balding@ic.ac.uk*

Population samples of DNA sequences can be simulated using coalescent-based simulation software, such as Hudsons MS. Coalescent methods work backwards in time, which is computationally efficient but is limited in the amount of recombination that can be incorporated, as well as the flexibility to include important features such as gene conversion and selection. With increased capacity of computers, it is now feasible to implement more flexible, forwards in- time simulation strategies. We have developed FREGENE, software for forwards-in-time, whole-population simulation of sequence-like data over large genomic intervals that can incorporate different demographic models, as well as recombination, both crossovers and gene conversions at highly variable rates, and several selection models. As illustrations of its potential uses, we (1) examine the performance of algorithms for identifying recombination hotspots, and (2) find approximations for genome-wide significance levels of genetic association studies using high marker density or resequence data, under different assumptions about demography and gene conversion. FREGENE is useful both for population genetics simulation studies, and for realistic power simulations for genetic epidemiology studies. We use it to investigate the relative merits of sequence data over dense marker data under various disease models.