

Estimating the scaled mutation rate and mutation bias

Claus Vogl*

^aInstitute of Animal Breeding and Genetics, Veterinärmedizinische Universität Wien, Veterinärplatz 1, A-1210 Vienna, Austria

Abstract

One of the simplest population genetic models is the biallelic mutation-drift model. After scaling, only two parameters are necessary: the overall scaled mutation rate θ and the mutation bias α . Nevertheless, the model is quite successful for modeling (putatively) unselected nucleotide sites in the genome. In equilibrium, the population allelic proportion x of a biallelic locus was shown to be beta distributed. With a sample of size M a binomial distribution conditional on x can be assumed, such that a beta-binomial compound distribution results. In spite of the simplicity of the model, maximum likelihood estimation of the two model parameters has so far been possible only for small scaled mutation rates θ , using the first order term in a Taylor series expansion of θ . For general θ , an iterative approach to maximum likelihood estimators is presented. An urn model that results in the beta-binomial distribution provides auxiliary variables. With these auxiliary variables, an expectation-maximization algorithm is derived for estimation of both parameters, i.e., the overall scaled mutation rate θ and the mutation bias α . The method is applied to real data from the fruitfly, the best-studied population genetic model species.

Keywords: biallelic mutation-drift model, equilibrium density, expectation-maximization algorithm, fruitfly.

*Corresponding author

Email address: `claus.vogl@vetmeduni.ac.at` (Claus Vogl)

Preprint submitted to Elsevier

February 25, 2015