

The Scale of Mutational Variation in the Murid Genome

Supplementary Material

Daniel J. Gaffney*¹ & Peter D. Keightley¹

April 22, 2005

*Daniel.Gaffney@ed.ac.uk

¹ Institute of Evolutionary Biology, Ashworth Laboratories, School
of Biological Sciences, University of Edinburgh West Mains Road,
Edinburgh EH9 3JT, United Kingdom

Scale of local similarity using alternate measures of substitution rate

A \leftrightarrow T and G \leftrightarrow C changes are the least likely to be affected by compositional change. Changes at non CpG-prone sites are the least likely to be influenced by CpG hypermutability. In order to assess the impact to these processes on our estimate of the scale of local similarity we calculated autocorrelation using both these types of changes only. The results are shown in Figure 1.

Between and within chromosome mutational variation using alternate measures of substitution rate

Similarly, in order to investigate the impact of compositional change and CpG hypermutability on the pattern of between and within chromosome variation we reanalysed our data using rates estimated counting only A \leftrightarrow T and G \leftrightarrow C changes and counting changes at non CpG-prone sites. The results of these analyses are presented in Figures 2 and 3 respectively.

Simulation protocol

Data were simulated according to the following protocol. Chromosome effects were simulated as random draws from a normal distribution, of mean = 0.1596 (the overall mean divergence estimated from ancestral repeats) and standard deviation = 0.0048 (the between chromosome standard deviation, estimated from fitting model 1 to the ancestral repeat data). In the null model (no regional effects) element substitution rates were simulated as random draws from a normal distribution of mean equal to the mean of the chromosome on which they were situated, and standard deviation = 0.0501 (residual from model 1). In the second model, ‘block’ regional effects were simulated as random draws from a normal distribution of mean equal to the mean of the chromosome on which they were located, and standard deviation = 0.0150 (the between block standard deviation estimated by fitting model 2, including a term for 1Mb blocks, to the ancestral repeat data). Within a ‘block’ element substitution rates were simulated as random draws from a normal distribution of mean equal to the simulated block substi-

tution rate, and standard deviation = 0.0487 (the residual from fitting model 2, including a term for 1Mb blocks, to the ancestral repeat data). We simulated blocks of 100kb, 1Mb and 5Mb in size.

First we analysed the null model simulation data with both Model 1 and Model 2, including a variety of block sizes in the latter (Figure 4). We then analysed the data with simulated regional effects using Model 2 again including a variety of block sizes (Figure 5). Results shown are averaged across 250 simulated replicates.

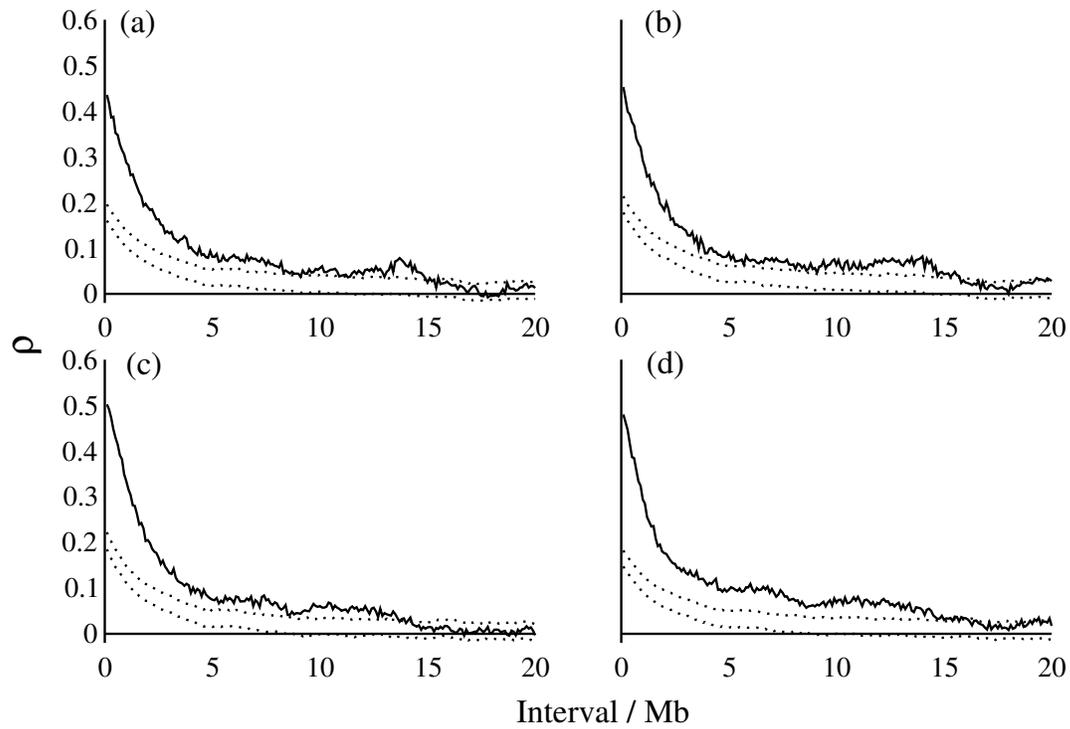


Figure 1: Autocorrelation of nucleotide substitution rates counting only A \leftrightarrow T and G \leftrightarrow C changes (a,c) and at non CpG-prone sites (b,d) in ancestral repeats (a,b) and flanking sequence (c,d) across 100kb bins. Dotted lines show the upper and lower bounds of the 95% confidence interval of autocorrelation under the null hypothesis of no dependence of rates between blocks. Blocks were permuted within common GC content intervals.

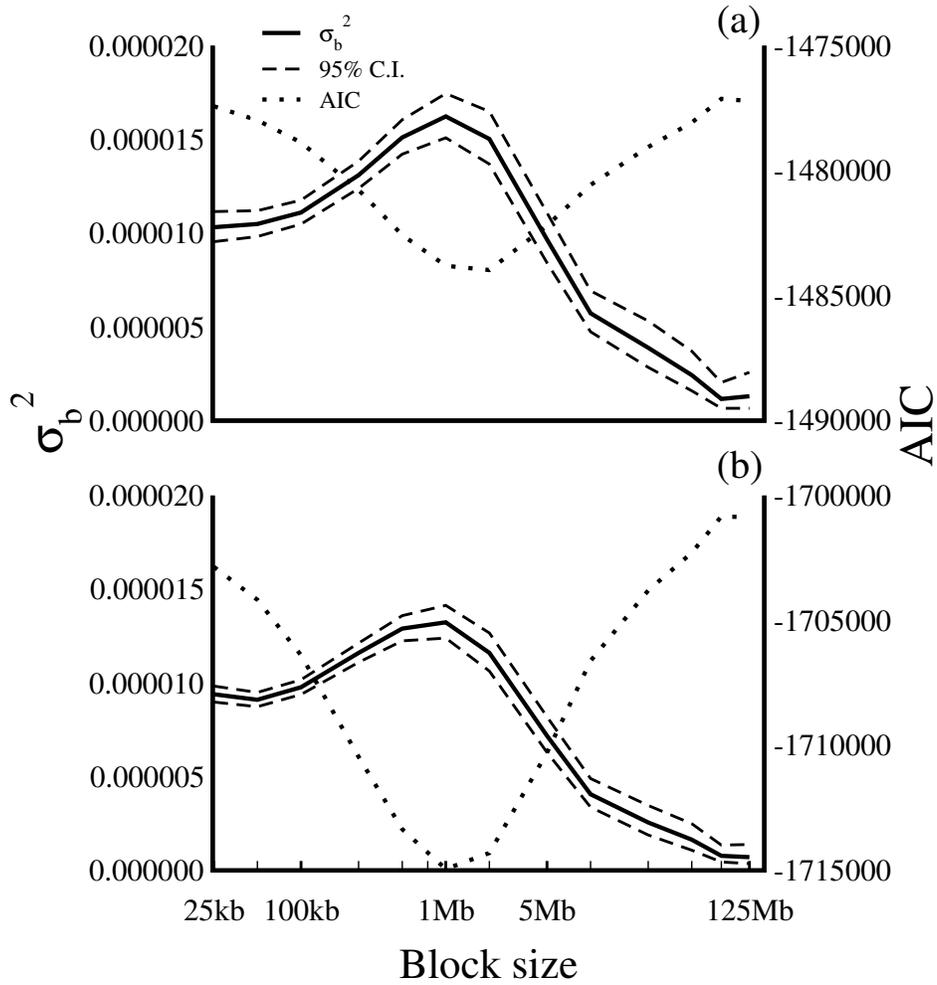


Figure 2: Between block variation (σ_b^2) in substitution rates within ancestral repeats (a) and flanking sequence (b). Substitution rates are estimated counting only A \leftrightarrow T and G \leftrightarrow C changes. Between block variances are estimated fitting chromosome as a fixed effect and block as a random effect across different block sizes, from 50kb to 125Mb. Block sizes are plotted on a \log_{10} scale. 95% confidence intervals of the between block variance were as estimated by the lme routine of the nlme package in R. Akaike's Information Criterion (AIC) is shown for each fitted model. The between chromosome variation estimated counting only A \leftrightarrow T and G \leftrightarrow C changes is 8.2×10^{-7} in ancestral repeats and 4.97×10^{-7} in flanking sequence.

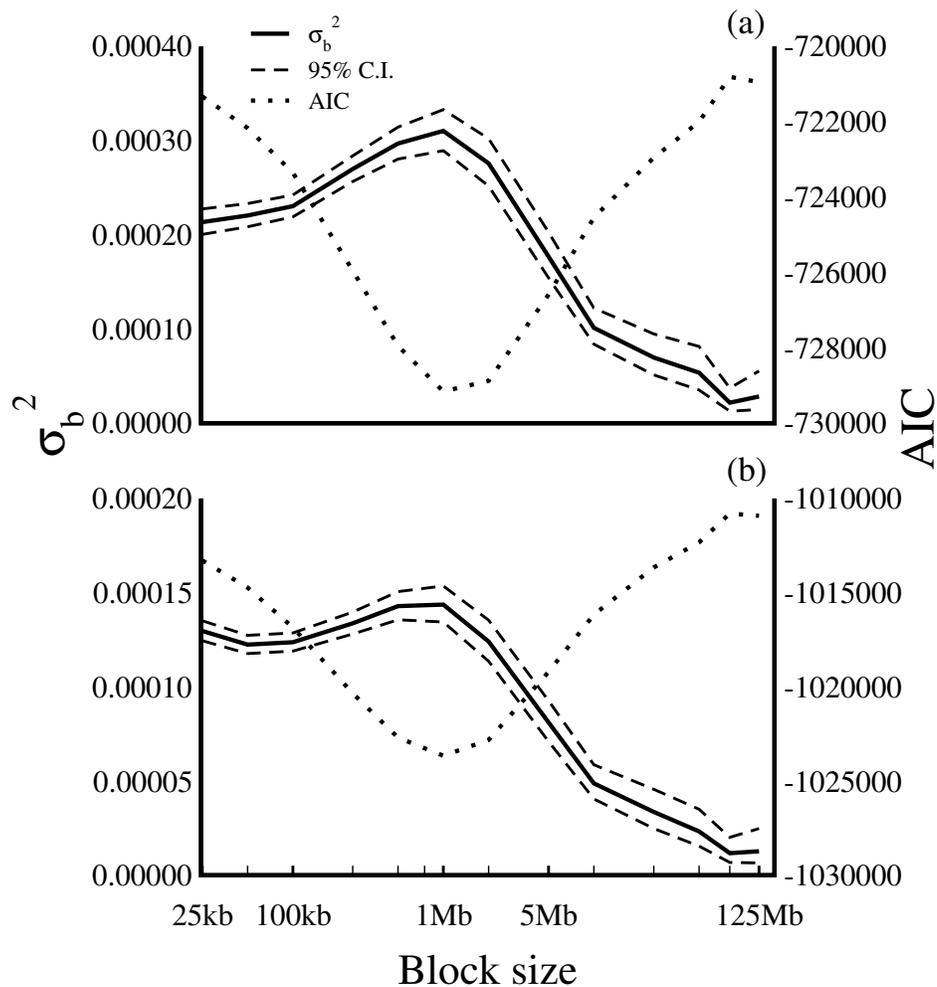


Figure 3: Between block variation (σ_b^2) in substitution rates within ancestral repeats (a) and flanking sequence (b). Substitution rates are estimated at non CpG-prone sites. Between block variances are estimated fitting chromosome as a fixed effect and block as a random effect across different block sizes, from 50kb to 125Mb. Block sizes are plotted on a \log_{10} scale. 95% confidence intervals of the between block variance were as estimated by the lme routine of the nlme package in R. Akaike's Information Criterion (AIC) is shown for each fitted model. The between chromosome variation estimated at non CpG-prone sites is 1.66×10^{-5} in ancestral repeats and 9.07×10^{-6} in flanking sequence.

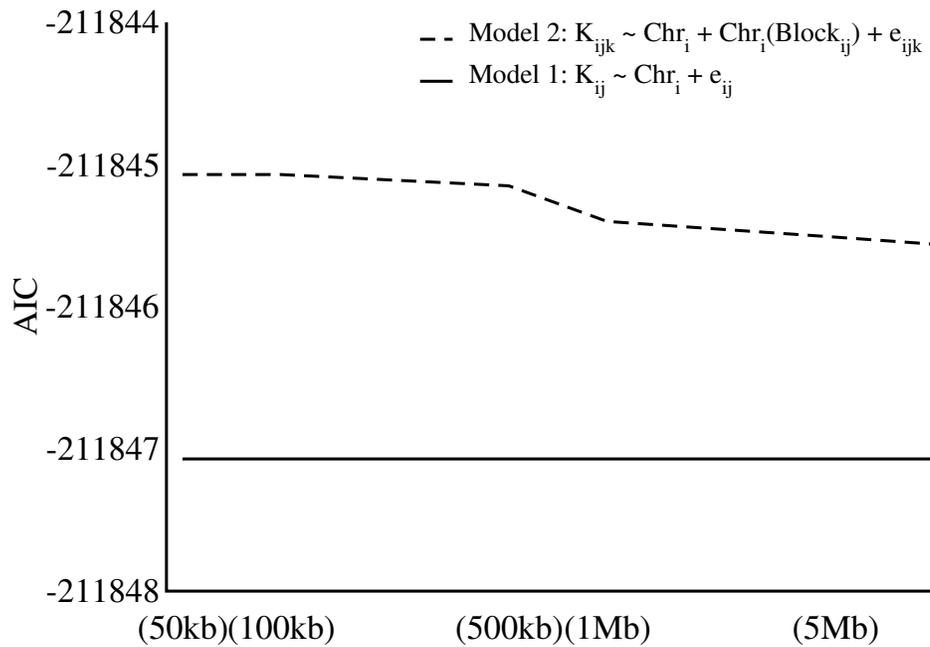


Figure 4: The AIC returned by the two linear models, model 1 fitting just chromosome fixed effects, model 2 fitting fixed chromosome and random regional effects to data with simulated chromosomal but no regional effects. AICs returned by model 2 are shown fitting a variety of different block sizes on the X axis.

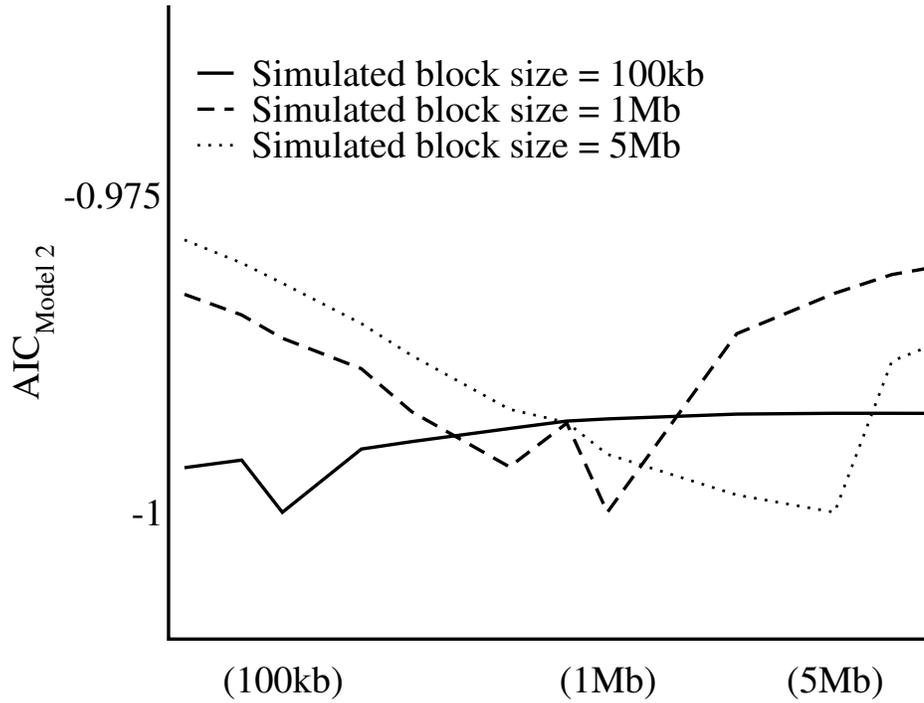


Figure 5: AIC returned by model 2, including terms for fixed chromosomal effects and random regional effects of a variety of sizes across the x axis, to data with simulated chromosomal and regional effects. All AIC values have been normalised to -1 for clarity. Regional effects were simulated in blocks of 100kb, 1Mb and 5Mb in size. The mean AIC (averaged across all replicates) returned fitting a model including only fixed effects were -209959.9, -209997.6 and -210215.7 for simulated block sizes of 100kb, 1Mb and 5Mb respectively. In contrast, the lowest mean AIC returned by any model fitting regional effects was -210210.2, -210700.4 and -211354.5 for simulated block sizes of 100kb, 1Mb and 5Mb respectively.

Site	X chromosome		Autosomes	
	All	Non CpG-prone	All	Non CpG-prone
Repeat				
SINE	0.154	0.163	0.160	0.162
LINE	0.146	0.148	0.157	0.154
LTR	0.150	0.155	0.162	0.160
DNA	0.135	0.135	0.151	0.145

Table 1: Nucleotide substitution rates at all and non CpG prone sites in the autosomes and X chromosome, by repeat family