GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.111.128355/DC1

Inference of Site Frequency Spectra From High-Throughput Sequence Data: Quantification of Selection on Nonsynonymous and Synonymous Sites in Humans

Peter D. Keightley and Daniel L. Halligan

DOI: 10.1534/genetics.111.128355

Table S1 Illumina sequencing error frequency distribution empirically estimated based on D. melanogaster MA line whole-genome sequence data.

		To			
		A	T	G	C
From	A	-	0.222	0.257	0.521
	T	0.222	-	0.519	0.260
	G	0.199	0.696	-	0.106
	C	0.695	0.199	0.105	-

We empirically estimated a matrix of nucleotide sequencing error frequencies from Illumina whole-genome sequences of three Drosophila melanogaster mutation accumulation (MA) lines aligned to the reference D. melanogaster genome by the MAQ aligner (Li, H.et al. 2008. Genome Res. 18: 1851-1858), which have been previously published (Keightley et al. 2009). The three initially isogenic lines were maintained for 262 generations by full-sib mating (Fernandez, J. and López-Fanjul, C. 1996. Genetics 143: 829 837), and are therefore close to 100% inbred. Almost all sites sequenced are therefore expected to be homozygous, and any base reads that are different from a consensus base are likely to be errors. We counted the frequency of putative errors at sites that had a depth of coverage of at least 10, and classified them according to the kind of nucleotide change involved. There were 1,788,344 putative errors in 1,717,965,680 base reads, giving an estimated mean error rate of 0.00104 per base read. The data indicate a substantial departure from uniform errors, with errors of type A/T → C/G and C/G → A/T predominating.

Table S2 Percentage of sites used and inferred error rate for different site classes, and numbers of individuals sampled, as a function of minimum base quality threshold (base quality), minimum map quality threshold (map quality) and the number of individuals sampled (n).

Base Quality	Map Quality	n	Class	% Sites Used	Error Rate
10	20	6	0-fold	90	0.0057
10	20	6	4-fold	87	0.0058
10	20	6	intron	83	0.0049
10	20	8	0-fold	86	0.0056
10	20	8	4-fold	83	0.0057
10	20	8	intron	76	0.0049
10	20	10	0-fold	80	0.0056
10	20	10	4-fold	77	0.0056
10	20	10	intron	68	0.0048
10	40	6	0-fold	87	0.0048
10	40	6	4-fold	85	0.0048
10	40	6	intron	79	0.0041
10	40	8	0-fold	82	0.0047
10	40	8	4-fold	79	0.0048
10	40	8	intron	72	0.0040
10	40	10	0-fold	75	0.0046
10	40	10	4-fold	72	0.0047
10	40	10	intron	62	0.0040
20	20	6	0-fold	84	0.0014
20	20	6	4-fold	80	0.0013
20	20	6	intron	77	0.0013
20	20	8	0-fold	75	0.0013

20	20	8	4-fold	71	0.0013
20	20	8	intron	66	0.0012
20	20	10	0-fold	63	0.0013
20	20	10	4-fold	59	0.0013
20	20	10	intron	53	0.0012
20	40	6	0-fold	81	0.0011
20	40	6	4-fold	77	0.0011
20	40	6	intron	73	0.0011
20	40	8	0-fold	71	0.0011
20	40	8	4-fold	67	0.0011
20	40	8	intron	62	0.0010
20	40	10	0-fold	58	0.0011
20	40	10	4-fold	55	0.0011
20	40	10	intron	49	0.0010