

TABLE S1: Probability  $p_{\epsilon}(j)$  of observed nucleotide read  $j$  at site  $\alpha$  with two-locus genotype  $g$  as a function of the error rate  $\epsilon$ .

Genotype	Nucleotide read		
	1 ( $A$ )	2 ( $a$ )	3 ( $e$ )
1 ( $AB/AB$ )	$1 - \epsilon$	$\frac{\epsilon}{3}$	$\frac{2\epsilon}{3}$
2 ( $Ab/Ab$ )	$1 - \epsilon$	$\frac{\epsilon}{3}$	$\frac{2\epsilon}{3}$
3 ( $aB/aB$ )	$\frac{\epsilon}{3}$	$1 - \epsilon$	$\frac{2\epsilon}{3}$
4 ( $ab/ab$ )	$\frac{\epsilon}{3}$	$1 - \epsilon$	$\frac{2\epsilon}{3}$
5 ( $AB/Ab$ )	$1 - \epsilon$	$\frac{\epsilon}{3}$	$\frac{2\epsilon}{3}$
6 ( $aB/ab$ )	$\frac{\epsilon}{3}$	$1 - \epsilon$	$\frac{2\epsilon}{3}$
7 ( $AB/aB$ )	$\frac{1}{2} \cdot (1 - \epsilon) + \frac{1}{2} \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} + \frac{1}{2} \cdot (1 - \epsilon)$	$\frac{2\epsilon}{3}$
8 ( $Ab/ab$ )	$\frac{1}{2} \cdot (1 - \epsilon) + \frac{1}{2} \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} + \frac{1}{2} \cdot (1 - \epsilon)$	$\frac{2\epsilon}{3}$
9 ( $AB/ab$ )	$\frac{1}{2} \cdot (1 - \epsilon) + \frac{1}{2} \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} + \frac{1}{2} \cdot (1 - \epsilon)$	$\frac{2\epsilon}{3}$
10 ( $Ab/aB$ )	$\frac{1}{2} \cdot (1 - \epsilon) + \frac{1}{2} \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} + \frac{1}{2} \cdot (1 - \epsilon)$	$\frac{2\epsilon}{3}$

$A$  and  $a$  denote candidate alleles (the two most abundant nucleotide reads in the population, e.g., C and T) and  $e$  denotes other nucleotide reads (e.g., in this case A and G) at site  $\alpha$ .  $B$  and  $b$  denote candidate alleles at site  $\beta$ . In the two-locus genotype notation, the slash separates haplotypes.