

**Figure S1: Alternative model for predicting the proportion of informative reads: incorporating the finite length of mRNA sequence .**

Equations described in the main text treat mRNA length as infinite despite the fact that real mRNA molecules have a finite length. To examine the impact of this assumption on the accuracy of predictions, we replaced the binomial distribution used for sampling SNPs within an mRNA with a hypergeometric distribution that allows the length of an mRNA to be considered explicitly. Black solid lines show expected proportions of

informative reads predicted by Eq. 3 with  $\Pr(X \geq x) = \sum_{k=x}^l \frac{\binom{l_g d}{k} \binom{l_g(1-d)}{l-k}}{\binom{l_g}{l}}$ , where  $l_g$  is the

mRNA length. The black dashed lines are expected proportions of informative reads predicted by Eq. 3 using Eq. 1 (assuming an infinite mRNA length). The gray lines are the 95% confidence intervals of the expected proportions.  $l_g = 2000$  bp (see Fig. 2 for details).

**Figure S2: The location of a SNP within a sequence affects its probability of being sampled in a random sampling strategy.**

The sampling probability of each nucleotide position in a 2000 bp mRNA sequence is shown. (A) The figure shows edge effects of read sampling: bases close to the 5' or 3' ends of transcripts had a lower probability of being observed among 10,000 reads, each 50 bp length. The red line shows the average probability of sampling a SNP at any site.

The green line shows the expected maximum probability (see text). Panels (B), (C), and (D) show variable sampling schemes.

**Figure S3: The number of informative reads per gene, as well as the magnitude of AI, affect the accuracy of AI estimates.**

**Figure S4: Distribution of significant AI in F<sub>1</sub> hybrids.**

Genes showing greater abundance of *D. melanogaster* and *D. simulans* alleles are positive and negative, respectively. The X-axis shows the AI values after a log (base 2) transformation.

Table S1: Genes and intergenic regions with significant AI (binomial test,  $p < 0.05$ ).

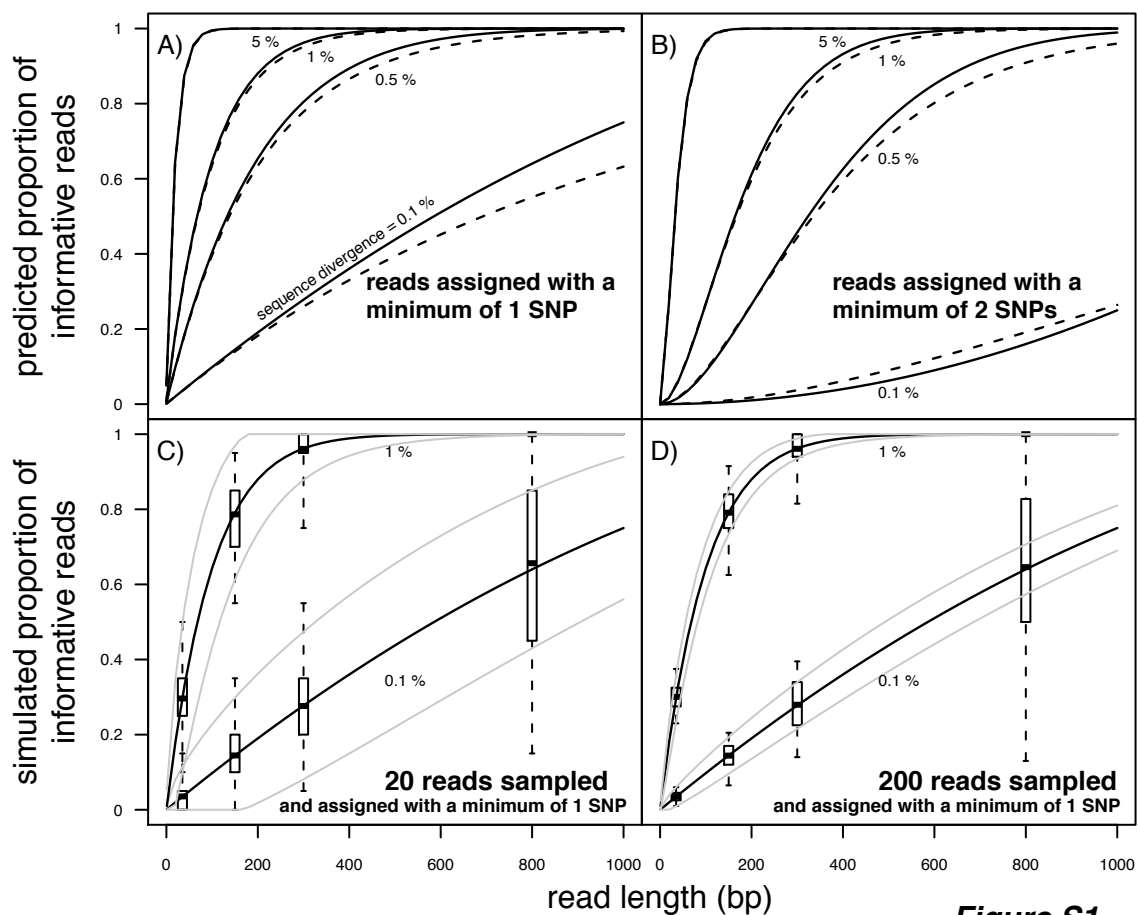
gene or intergenic FlyBase names	# of <i>D. melanogaster</i> reads	# of <i>D. simulans</i> reads	# of uninformative reads	AI	gene or intergenic FlyBase names	# of <i>D. melanogaster</i> reads	# of <i>D. simulans</i> reads	# of uninformative reads	AI	gene or intergenic FlyBase names	# of <i>D. melanogaster</i> reads	# of <i>D. simulans</i> reads	# of uninformative reads	AI
FBgn0010019	81	121	79	0.67	FBgn0040827	29	13	1	2.23	FBgn0034751	18	3	6	6
FBgn0005391	82	155	12	0.53	FBgn0028396	29	10	2	2.9	FBgn0001092	18	4	4	4.5
FBgn0031176	24	107	1	0.22	FBgn0005533	2	15	23	0.13	FBgn0041605	17	2	6	8.5
FBgn0066084	64	32	27	2	FBgn0044812	31	8	0	3.88	FBgn0040309	5	16	3	0.31
FBgn0058317/FBgn0069962	11	76	31	0.14	FBgn0002590	20	5	13	4	FBgn0003432	16	5	2	3.2
FBgn0003357	69	26	4	2.65	FBgn0004403	27	9	1	3	FBgn0003075	16	3	3	5.33
FBgn0014184	23	47	24	0.49	FBgn0010412	24	3	10	8	FBgn0036766	4	17	1	0.24
FBgn0035664	55	35	2	1.57	FBgn0023170	0	23	14	0	FBgn0000253	10	2	9	5
FBgn0004427	49	28	9	1.75	FBgn0010352	25	7	4	3.57	FBgn0015766	15	3	3	5
FBgn0001219	38	19	26	2	FBgn0015745	10	23	2	0.43	FBgn0026753	11	2	8	5.5
FBgn0039830	48	15	13	3.2	FBgn0032833	7	18	9	0.39	FBgn0030487	4	13	4	0.31
FBgn0024733	21	44	6	0.48	FBgn0030999	6	22	5	0.27	FBgn0039737	13	4	4	3.25
FBgn0039713	38	16	16	2.38	FBgn0000078	22	5	5	4.4	FBgn0029669/FBgn0052792	11	1	7	11
FBgn0031653	42	15	6	2.8	FBgn0015221	21	8	3	2.62	FBgn0032049	14	4	1	3.5
FBgn0052473/FBgn0038139	21	40	0	0.52	FBgn0040606	7	22	3	0.32	FBgn0034497	10	1	7	10
FBgn0027571	18	33	7	0.55	FBgn0015288	19	6	4	3.17	FBgn0002868	15	1	1	15
FBgn0028697	20	36	2	0.56	FBgn0052373	18	5	6	3.6	FBgn0003462	13	3	1	4.33
FBgn0026372	9	25	15	0.36	FBgn0002719	16	5	6	3.2	FBgn0035978	12	3	1	4
FBgn0029897	25	9	14	2.78	FBgn0004426	0	23	4	0	FBgn0033246	12	2	0	6
FBgn0003517	12	27	6	0.44	FBgn0004907	15	2	10	7.5	FBgn0036007	11	2	1	5.5

gene or intergenic FlyBase names	# of D. melanogaster reads	# of D. simulans reads	# of uninformative reads	AI	gene or intergenic FlyBase names	# of D. melanogaster reads	# of D. simulans reads	# of uninformative reads	AI	gene or intergenic FlyBase names	# of D. melanogaster reads	# of D. simulans reads	# of uninformative reads	AI
FBgn0040529	11	1	2	11	FBgn0039801	1	8	0	0.12	FBgn0029172	6	0	0	Inf
FBgn0040733	10	2	1	5	FBgn0003979	7	0	1	Inf	FBgn0029676	0	6	0	0
FBgn0050438	2	11	0	0.18	FBgn0004828	7	0	1	Inf	FBgn0032229	6	0	0	Inf
FBgn0024807	1	11	0	0.09	FBgn0014869	7	0	1	Inf	FBgn0034229	6	0	0	Inf
FBgn0033522	1	8	3	0.12	FBgn0028336	7	0	1	Inf	FBgn0038318	6	0	0	Inf
FBgn0033624	2	10	0	0.2	FBgn0028415	8	0	0	Inf	FBgn0039857	0	6	0	0
FBgn0040660	10	1	1	10	FBgn0034471	0	7	1	0	FBgn0044323	0	6	0	0
FBgn0046689/FBgn0038827	8	1	3	8	FBgn0052626	8	0	0	Inf					
FBgn0001125	8	1	2	8	FBgn0012037	6	0	1	Inf					
FBgn0001186	8	1	2	8	FBgn0024288	6	0	1	Inf					
FBgn0005536	7	0	3	Inf	FBgn0025741	6	0	1	Inf					
FBgn0027932	8	1	1	8	FBgn0031457	0	7	0	0					
FBgn0029687	10	0	0	Inf	FBgn0034259	7	0	0	Inf					
FBgn0036672	8	1	1	8	FBgn0035424/FBgn0035425	6	0	1	Inf					
FBgn0053937/FBgn0037683	1	8	1	0.12	FBgn0036357	6	0	1	Inf					
FBgn0058002	9	1	0	9	FBgn0040734	6	0	1	Inf					
FBgn0025454	8	1	0	8	FBgn0051072	7	0	0	Inf					
FBgn0036533	8	0	1	Inf	FBgn0053126	0	6	1	0					
FBgn0036663	7	0	2	Inf	FBgn0011754	6	0	0	Inf					
FBgn0039761	1	8	0	0.12	FBgn0015778/FBgn0010258	6	0	0	Inf					

Table S2: Comparisons between 454 sequencing and Pyrosequencing AI estimates.

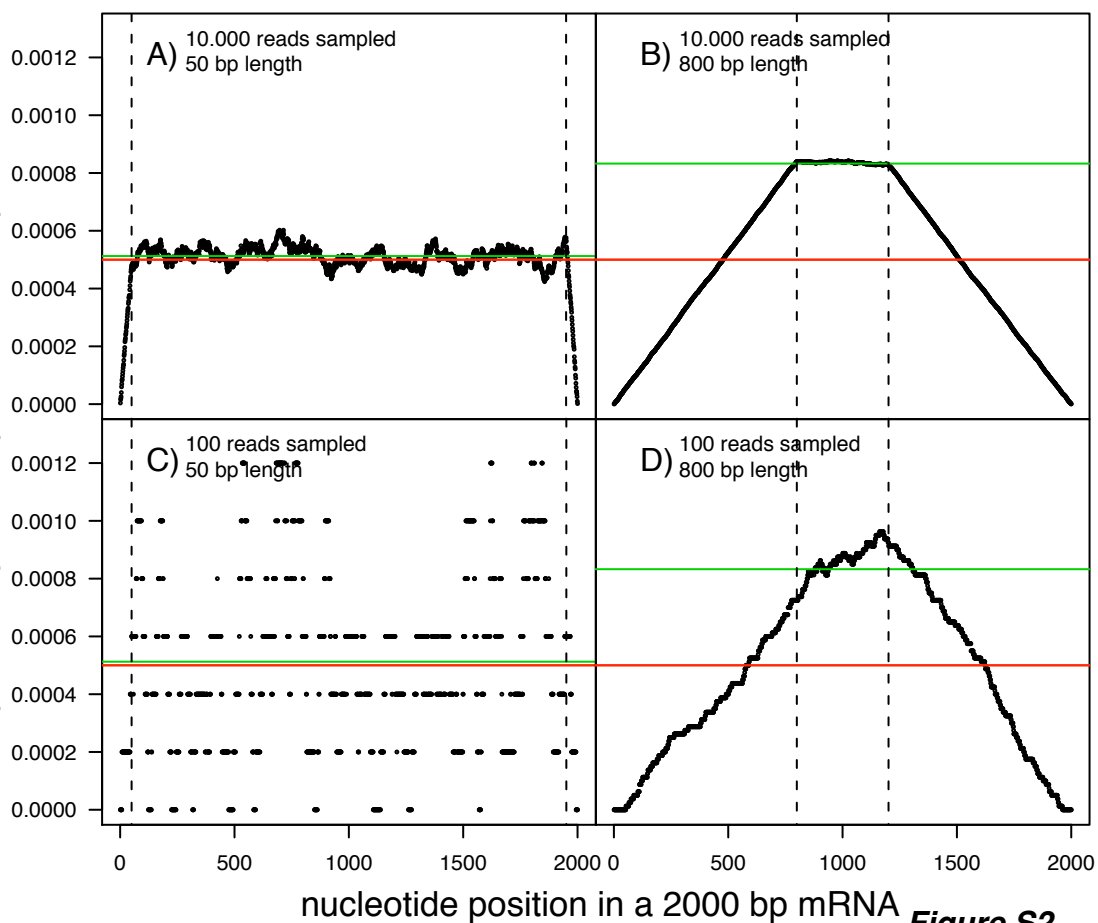
'read only' means that reads mapped to only one reference genome.

	total # of reads	# of <i>D. melanogaster</i> reads	# of <i>D. simulans</i> reads	# of uninformative reads	# of <i>D. melanogaster</i> reads only	# of <i>D. simulans</i> reads only	454 sequencing AI p-value < 0.05	454 sequencing log <sub>2</sub> (dmel/dsim)	pyrosequencing log <sub>2</sub> (dmel/dsim)	pyrosequencing ms
FBgn0016684	6	4	0	0	2	0	yes	2	0.26	1.19
FBgn0033065	7	6	1	0	0	0	no	2.58	0.6	1.51
FBgn0003076	15	3	1	0	11	0	yes	1.58	1.08	2.12
FBgn0002868	17	15	1	1	0	0	yes	3.91	1.58	2.99
FBgn0028920	19	9	9	1	0	0	no	0	-0.79	0.58
FBgn0015031	20	0	0	0	20	0	yes	-Inf	-0.19	0.88
FBgn0012034	21	8	9	4	0	0	no	-0.17	0.51	1.42
FBgn0026753	21	11	2	8	0	0	yes	2.46	-0.07	0.95
FBgn0036766	22	4	17	1	0	0	yes	-2.09	-0.95	0.52
FBgn0000406	27	12	15	0	0	0	no	-0.32	-0.61	0.66
FBgn0033820	39	14	24	1	0	0	no	-0.78	1.3	2.46
FBgn0005533	40	2	15	23	0	0	yes	-2.91	-0.24	0.85
FBgn0024733	71	21	44	6	0	0	yes	-1.07	-0.82	0.57
FBgn0031176	133	24	107	1	1	0	yes	-2.16	-1.79	0.29

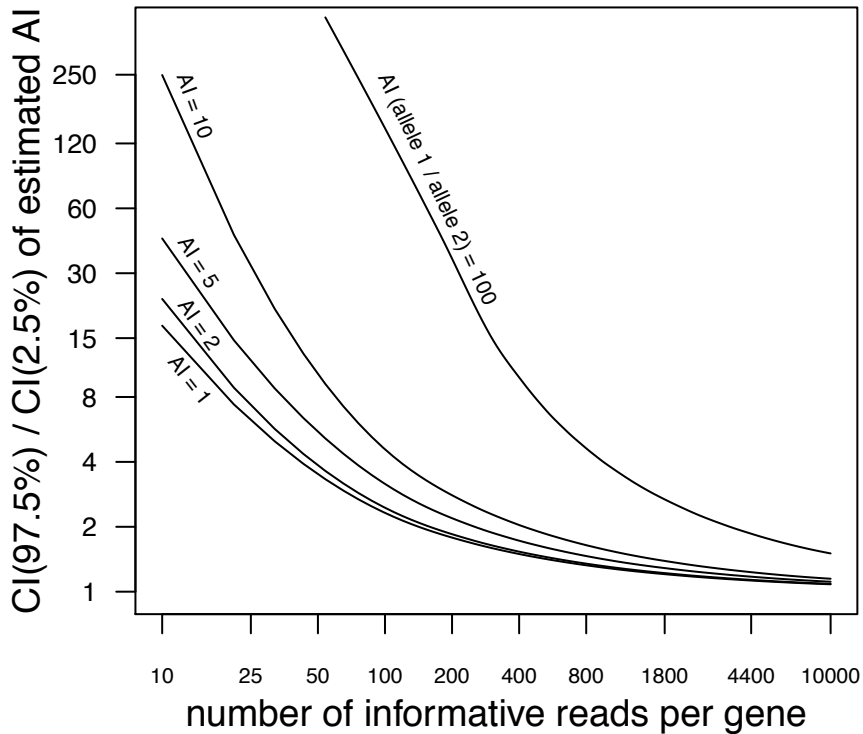


**Figure S1**

probability to sample nucleotide position



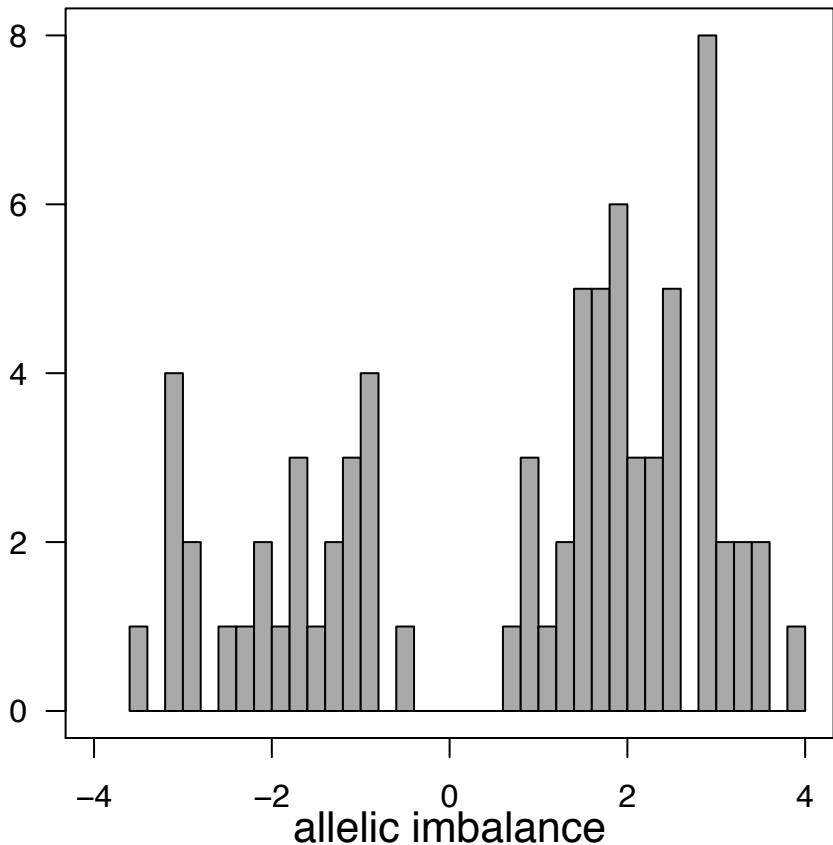
**Figure S2**



**Figure S3**



number of genes



*(D. melanogaster allelic expression / D. simulans allelic expression)*

**Figure S4**