



Figure S1: Predicted sampling standard deviation of ML estimates of LD under a fixed total budget, as a function of the mean depth of coverage μ , when library preparation is three times more expensive than sequencing. Figure S1A shows the predicted sampling standard

deviation of the ML estimates of the LD coefficient D when all sequence reads cover both sites of interest (given by the square root of the theoretical sampling variance from Equations 9 and 13). Figure S1B shows that when all sequence reads cover just single polymorphic sites (given by the square root of the theoretical sampling variance from Equations 10 and 15). The major allele frequencies at two sites of interest are $p = q = 0.9$, and $D = 0.01$.

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

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 α . B and b denote candidate alleles at site β . In the two-locus genotype notation, the slash separates haplotypes.

TABLE S2: Probability $p_g(j)$ of observed dinucleotide read j at the two sites of interest with two-locus genotype g as a function of the error rate ϵ .

Genotype	Dinucleotide read								
	1 (AB)	2 (Ab)	3 (Ae)	4 (aB)	5 (ab)	6 (ae)	7 (eB)	8 (eb)	9 (ee)
1 (AB/AB)	$(1-\epsilon)^2$	$(1-\epsilon) \cdot \frac{\epsilon}{3}$	$(1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{\epsilon}{3} \cdot (1-\epsilon)$	$\left(\frac{\epsilon}{3}\right)^2$	$\frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot (1-\epsilon)$	$\frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\left(\frac{2\epsilon}{3}\right)^2$
2 (Ab/Ab)	$(1-\epsilon) \cdot \frac{\epsilon}{3}$	$(1-\epsilon)^2$	$(1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\left(\frac{\epsilon}{3}\right)^2$	$\frac{\epsilon}{3} \cdot (1-\epsilon)$	$\frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot (1-\epsilon)$	$\left(\frac{2\epsilon}{3}\right)^2$
3 (aB/aB)	$\frac{\epsilon}{3} \cdot (1-\epsilon)$	$\left(\frac{\epsilon}{3}\right)^2$	$\frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$(1-\epsilon)^2$	$(1-\epsilon) \cdot \frac{\epsilon}{3}$	$(1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot (1-\epsilon)$	$\frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\left(\frac{2\epsilon}{3}\right)^2$
4 (ab/ab)	$\left(\frac{\epsilon}{3}\right)^2$	$\frac{\epsilon}{3} \cdot (1-\epsilon)$	$\frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$(1-\epsilon) \cdot \frac{\epsilon}{3}$	$(1-\epsilon)^2$	$(1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot (1-\epsilon)$	$\left(\frac{2\epsilon}{3}\right)^2$
5 (AB/Ab)	$\frac{1}{2}(1-\epsilon)^2 + \frac{1}{2} \cdot (1-\epsilon) \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{\epsilon}{3} + \frac{1}{2}(1-\epsilon)^2$	$(1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot (1-\epsilon) + \frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$	$\frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2 + \frac{1}{2} \cdot \frac{\epsilon}{3} \cdot (1-\epsilon)$	$\frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot (1-\epsilon) + \frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3} + \frac{1}{2} \cdot \frac{2\epsilon}{3}$	$\left(\frac{2\epsilon}{3}\right)^2$
6 (aB/ab)	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot (1-\epsilon) + \frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$	$\frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2 + \frac{1}{2} \cdot \frac{\epsilon}{3}$	$\frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{1}{2}(1-\epsilon)^2 + \frac{1}{2} \cdot (1-\epsilon) \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{\epsilon}{3} + \frac{1}{2}(1-\epsilon)^2$	$(1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot (1-\epsilon) + \frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3} + \frac{1}{2} \cdot \frac{2\epsilon}{3}$	$\left(\frac{2\epsilon}{3}\right)^2$

7 (AB/aB)	$\frac{1}{2}(1-\epsilon)^2 + \frac{1}{2} \cdot \frac{\epsilon}{3} \cdot (1-\epsilon)$	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{\epsilon}{3}$ $+ \frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$ $+ \frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot (1-\epsilon)$ $+ \frac{1}{2}(1-\epsilon)^2$	$\frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2 + \frac{1}{2}$ $\cdot (1-\epsilon) \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3} + \frac{1}{2}$ $\cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot (1-\epsilon)$	$\frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\left(\frac{2\epsilon}{3}\right)^2$
8 (Ab/ab)	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{\epsilon}{3}$ $+ \frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$	$\frac{1}{2}(1-\epsilon)^2 + \frac{1}{2} \cdot \frac{\epsilon}{3} \cdot (1-\epsilon)$	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$ $+ \frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2 + \frac{1}{2}$ $\cdot (1-\epsilon) \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot (1-\epsilon)$ $+ \frac{1}{2}(1-\epsilon)^2$	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3} + \frac{1}{2}$ $\cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\frac{2\epsilon}{3} \cdot (1-\epsilon)$	$\left(\frac{2\epsilon}{3}\right)^2$
9 (AB/ab)	$\frac{1}{2}(1-\epsilon)^2 + \frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$	$(1-\epsilon) \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$ $+ \frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{\epsilon}{3} \cdot (1-\epsilon)$ $+ \frac{1}{2}(1-\epsilon)^2$	$\frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3} + \frac{1}{2}$ $\cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot (1-\epsilon)$ $+ \frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3} + \frac{1}{2} \cdot \frac{2\epsilon}{3}$ $\cdot (1-\epsilon)$	$\left(\frac{2\epsilon}{3}\right)^2$
10 (Ab/aB)	$(1-\epsilon) \cdot \frac{\epsilon}{3}$	$\frac{1}{2}(1-\epsilon)^2 + \frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$	$\frac{1}{2} \cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$ $+ \frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \left(\frac{\epsilon}{3}\right)^2$ $+ \frac{1}{2}(1-\epsilon)^2$	$\frac{\epsilon}{3} \cdot (1-\epsilon)$	$\frac{1}{2} \cdot \frac{\epsilon}{3} \cdot \frac{2\epsilon}{3} + \frac{1}{2}$ $\cdot (1-\epsilon) \cdot \frac{2\epsilon}{3}$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3} + \frac{1}{2}$ $\cdot \frac{2\epsilon}{3} \cdot (1-\epsilon)$	$\frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot (1-\epsilon)$ $+ \frac{1}{2} \cdot \frac{2\epsilon}{3} \cdot \frac{\epsilon}{3}$	$\left(\frac{2\epsilon}{3}\right)^2$

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 α . B and b denote candidate alleles at site β . In the two-locus genotype notation, the slash separates haplotypes.

TABLE S3: Comparison of LD estimates by the proposed ML estimator to those by an imputation-based method when the major allele frequencies at two sites of interest, p and q , are intermediate.

D	μ	(p, q)	Φ	Method	\hat{D} (mean \pm SD)	\hat{D}^2 (mean \pm SD)	Asymptotic theoretical SD(\hat{D})	Theoretical $E[\hat{D}^2]$	RMSD(\hat{D})	RMSD(\hat{D}^2)
0.01	2	(0.6,0.7)	0	ML	0.007 \pm 0.039	0.0016 \pm 0.0024	0.023	0.0006	0.039	0.0028
0.01	2	(0.6,0.7)	0	Imputation	0.002 \pm 0.018	0.0003 \pm 0.0007	0.023	0.0006	0.020	0.0007
0.01	2	(0.6,0.7)	1/3	ML	0.008 \pm 0.025	0.0007 \pm 0.0008	0.016	0.0004	0.025	0.0010
0.01	2	(0.6,0.7)	1/3	Imputation	0.002 \pm 0.016	0.0003 \pm 0.0006	0.016	0.0004	0.018	0.0006
0.01	2	(0.6,0.7)	1	ML	0.009 \pm 0.020	0.0005 \pm 0.0008	0.016	0.0004	0.020	0.0008
0.01	2	(0.6,0.7)	1	Imputation	0.009 \pm 0.025	0.0007 \pm 0.0019	0.016	0.0004	0.025	0.0020
0.01	10	(0.6,0.7)	0	ML	0.007 \pm 0.025	0.0006 \pm 0.0008	0.023	0.0006	0.025	0.0010
0.01	10	(0.6,0.7)	0	Imputation	0.006 \pm 0.026	0.0007 \pm 0.0014	0.023	0.0006	0.026	0.0016
0.01	10	(0.6,0.7)	1/3	ML	0.009 \pm 0.016	0.0003 \pm 0.0005	0.016	0.0004	0.016	0.0005
0.01	10	(0.6,0.7)	1/3	Imputation	0.007 \pm 0.017	0.0003 \pm 0.0004	0.016	0.0004	0.017	0.0005
0.01	10	(0.6,0.7)	1	ML	0.009 \pm 0.019	0.0004 \pm 0.0007	0.016	0.0004	0.018	0.0007
0.01	10	(0.6,0.7)	1	Imputation	0.008 \pm 0.026	0.0007 \pm 0.0017	0.016	0.0004	0.026	0.0018
0.1	2	(0.6,0.7)	0	ML	0.092 \pm 0.042	0.0103 \pm 0.0073	0.023	0.0105	0.043	0.0073
0.1	2	(0.6,0.6)	0	Imputation	0.036 \pm 0.029	0.0022 \pm 0.0026	0.023	0.0105	0.070	0.0083
0.1	2	(0.6,0.7)	1/3	ML	0.095 \pm 0.031	0.0101 \pm 0.0049	0.016	0.0101	0.031	0.0049
0.1	2	(0.6,0.7)	1/3	Imputation	0.043 \pm 0.031	0.0028 \pm 0.0028	0.016	0.0101	0.065	0.0077
0.1	2	(0.6,0.7)	1	ML	0.099 \pm 0.030	0.0107 \pm 0.0042	0.016	0.0101	0.030	0.0042
0.1	2	(0.6,0.7)	1	Imputation	0.082 \pm 0.034	0.0078 \pm 0.0062	0.016	0.0101	0.039	0.0066
0.1	10	(0.6,0.7)	0	ML	0.098 \pm 0.030	0.0104 \pm 0.0039	0.023	0.0105	0.030	0.0039
0.1	10	(0.6,0.7)	0	Imputation	0.107 \pm 0.026	0.0121 \pm 0.0045	0.023	0.0105	0.027	0.0049
0.1	10	(0.6,0.7)	1/3	ML	0.100 \pm 0.016	0.0103 \pm 0.0031	0.016	0.0101	0.016	0.0031
0.1	10	(0.6,0.7)	1/3	Imputation	0.110 \pm 0.021	0.0125 \pm 0.0040	0.016	0.0101	0.023	0.0047

0.1	10	(0.6,0.7)	1	ML	0.099±0.028	0.0106±0.0037	0.016	0.0101	0.027	0.0037
0.1	10	(0.6,0.7)	1	Imputation	0.108±0.026	0.0123±0.0047	0.016	0.0101	0.027	0.0052

Sample means and standard deviations of the LD coefficient \hat{D} and its square \hat{D}^2 estimated from simulated data by the ML estimator and the imputation-based method are compared with different values of the parameters. Root mean square deviations (RMSD) of the LD estimates are also compared. The comparisons are made when the mean depth of coverage, μ is low (2) or moderately high (10). Furthermore, comparisons with different values of the probability that an informative read covers both polymorphic sites, ϕ , are made. The theoretical prediction of the asymptotic standard deviation, which is calculated as a square root of the sampling variance given by Equation 9 (when $\phi > 0$) or 10 (when $\phi = 0$), is the expected level of achievement when individual genotypes are known without errors. Sample size $N = 100$, error rate $\varepsilon = 0.01$. A total of 100 simulation replicates were run for each set of parameter values.

TABLE S4: Comparison of LD estimates by the proposed ML estimator to those by an imputation-based method when major allele frequencies at two sites of interest, p and q , are high.

D	μ	(p, q)	Φ	Method	\hat{D} (mean \pm SD)	\hat{D}^2 (mean \pm SD)	Asymptotic theoretical SD(\hat{D})	Theoretical $E[\hat{D}^2]$	RMSD(\hat{D})	RMSD(\hat{D}^2)
0.01	2	(0.9,0.9)	0	ML	0.012 \pm 0.014	0.00034 \pm 0.00049	0.011	0.00021	0.014	0.00054
0.01	2	(0.9,0.9)	0	Imputation	0.003 \pm 0.005	0.00003 \pm 0.00005	0.011	0.00021	0.009	0.00008
0.01	2	(0.9,0.9)	1/3	ML	0.011 \pm 0.014	0.00030 \pm 0.00043	0.008	0.00017	0.014	0.00047
0.01	2	(0.9,0.9)	1/3	Imputation	0.003 \pm 0.006	0.00005 \pm 0.00015	0.008	0.00017	0.009	0.00016
0.01	2	(0.9,0.9)	1	ML	0.011 \pm 0.011	0.00022 \pm 0.00031	0.008	0.00017	0.011	0.00033
0.01	2	(0.9,0.9)	1	Imputation	0.004 \pm 0.006	0.00005 \pm 0.00009	0.008	0.00017	0.009	0.00010
0.01	10	(0.9,0.9)	0	ML	0.009 \pm 0.010	0.00017 \pm 0.00024	0.011	0.00021	0.010	0.00025
0.01	10	(0.9,0.9)	0	Imputation	0.005 \pm 0.008	0.00008 \pm 0.00013	0.011	0.00021	0.010	0.00013
0.01	10	(0.9,0.9)	1/3	ML	0.010 \pm 0.009	0.00018 \pm 0.00023	0.008	0.00017	0.009	0.00024
0.01	10	(0.9,0.9)	1/3	Imputation	0.006 \pm 0.008	0.00009 \pm 0.00013	0.008	0.00017	0.009	0.00013
0.01	10	(0.9,0.9)	1	ML	0.010 \pm 0.008	0.00016 \pm 0.00018	0.008	0.00017	0.008	0.00019
0.01	10	(0.9,0.9)	1	Imputation	0.006 \pm 0.008	0.00011 \pm 0.00025	0.008	0.00017	0.009	0.00025
0.05	2	(0.9,0.9)	0	ML	0.048 \pm 0.021	0.00280 \pm 0.00245	0.016	0.00271	0.021	0.00245
0.05	2	(0.9,0.9)	0	Imputation	0.011 \pm 0.010	0.00022 \pm 0.00044	0.016	0.00271	0.041	0.00232
0.05	2	(0.9,0.9)	1/3	ML	0.048 \pm 0.020	0.00270 \pm 0.00204	0.014	0.00266	0.020	0.00204
0.05	2	(0.9,0.9)	1/3	Imputation	0.015 \pm 0.013	0.00039 \pm 0.00054	0.014	0.00266	0.037	0.00218
0.05	2	(0.9,0.9)	1	ML	0.049 \pm 0.018	0.00274 \pm 0.00204	0.014	0.00266	0.018	0.00204
0.05	2	(0.9,0.9)	1	Imputation	0.028 \pm 0.017	0.00106 \pm 0.00104	0.014	0.00266	0.027	0.00177
0.05	10	(0.9,0.9)	0	ML	0.051 \pm 0.015	0.00284 \pm 0.00153	0.016	0.00271	0.015	0.00156
0.05	10	(0.9,0.9)	0	Imputation	0.046 \pm 0.022	0.00257 \pm 0.00189	0.016	0.00271	0.023	0.00188
0.05	10	(0.9,0.9)	1/3	ML	0.051 \pm 0.015	0.00282 \pm 0.00158	0.014	0.00266	0.015	0.00161
0.05	10	(0.9,0.9)	1/3	Imputation	0.044 \pm 0.022	0.00242 \pm 0.00189	0.014	0.00266	0.023	0.00188

0.05	10	(0.9,0.9)	1	ML	0.052±0.014	0.00284±0.00152	0.014	0.00266	0.014	0.00156
0.05	10	(0.9,0.9)	1	Imputation	0.045±0.021	0.00248±0.00186	0.014	0.00266	0.021	0.00185

Sample means and standard deviations of the LD coefficient \hat{D} and its square \hat{D}^2 estimated from simulated data by the ML estimator and the imputation-based method are compared with different values of the parameters. Root mean square deviations (RMSD) of the LD estimates are also compared. The comparisons are made when the mean depth of coverage, μ is low (2) or moderately high (10). Furthermore, comparisons with different values of the probability that an informative read covers both polymorphic sites, ϕ , are made. The theoretical prediction of the asymptotic standard deviation, which is calculated as a square root of the sampling variance given by Equation 9 (when $\phi > 0$) or 10 (when $\phi = 0$), is the expected level of achievement when individual genotypes are known without errors. Sample size $N = 100$, error rate $\varepsilon = 0.01$. A total of 100 simulation replicates were run for each set of parameter values.