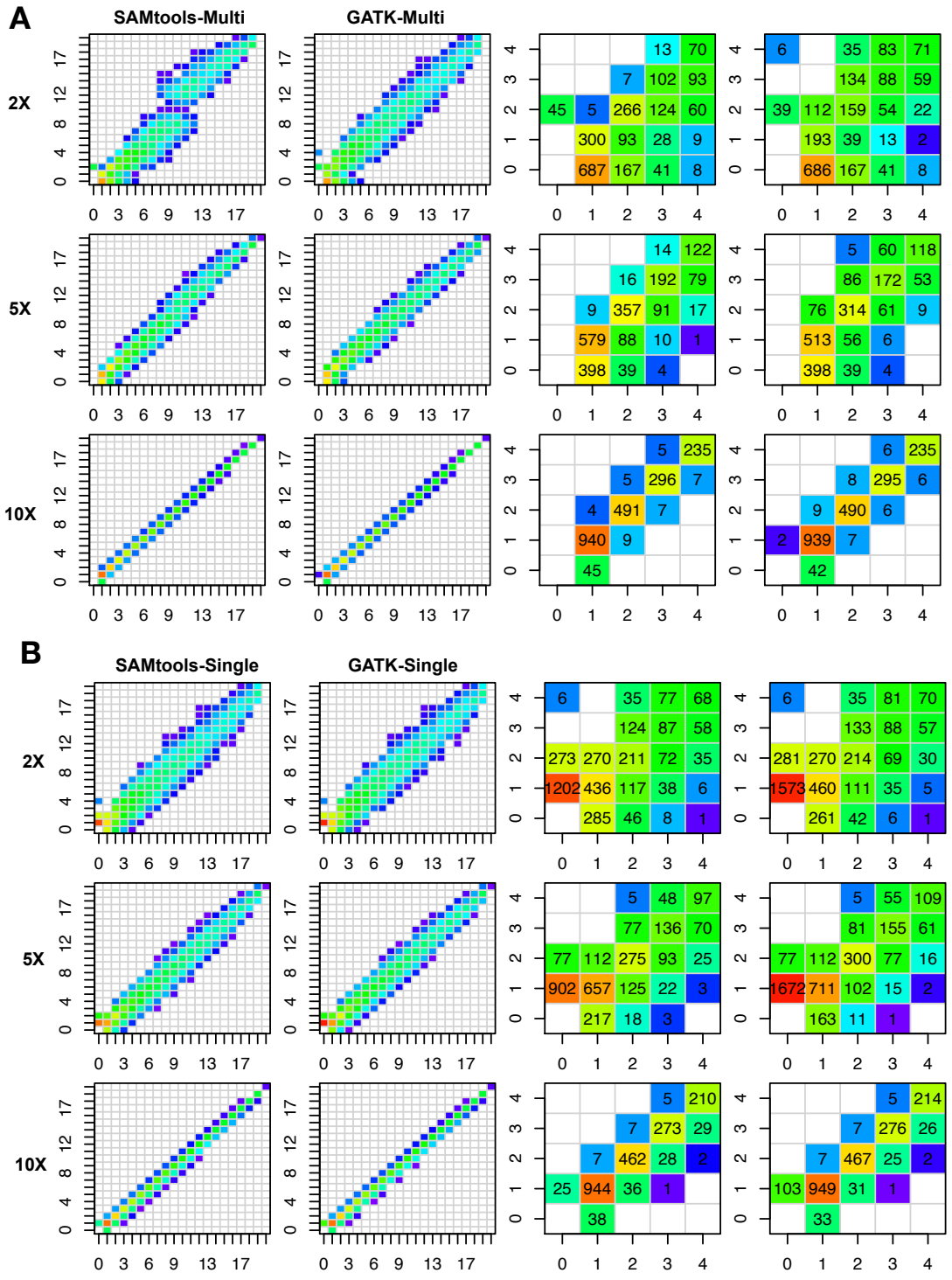
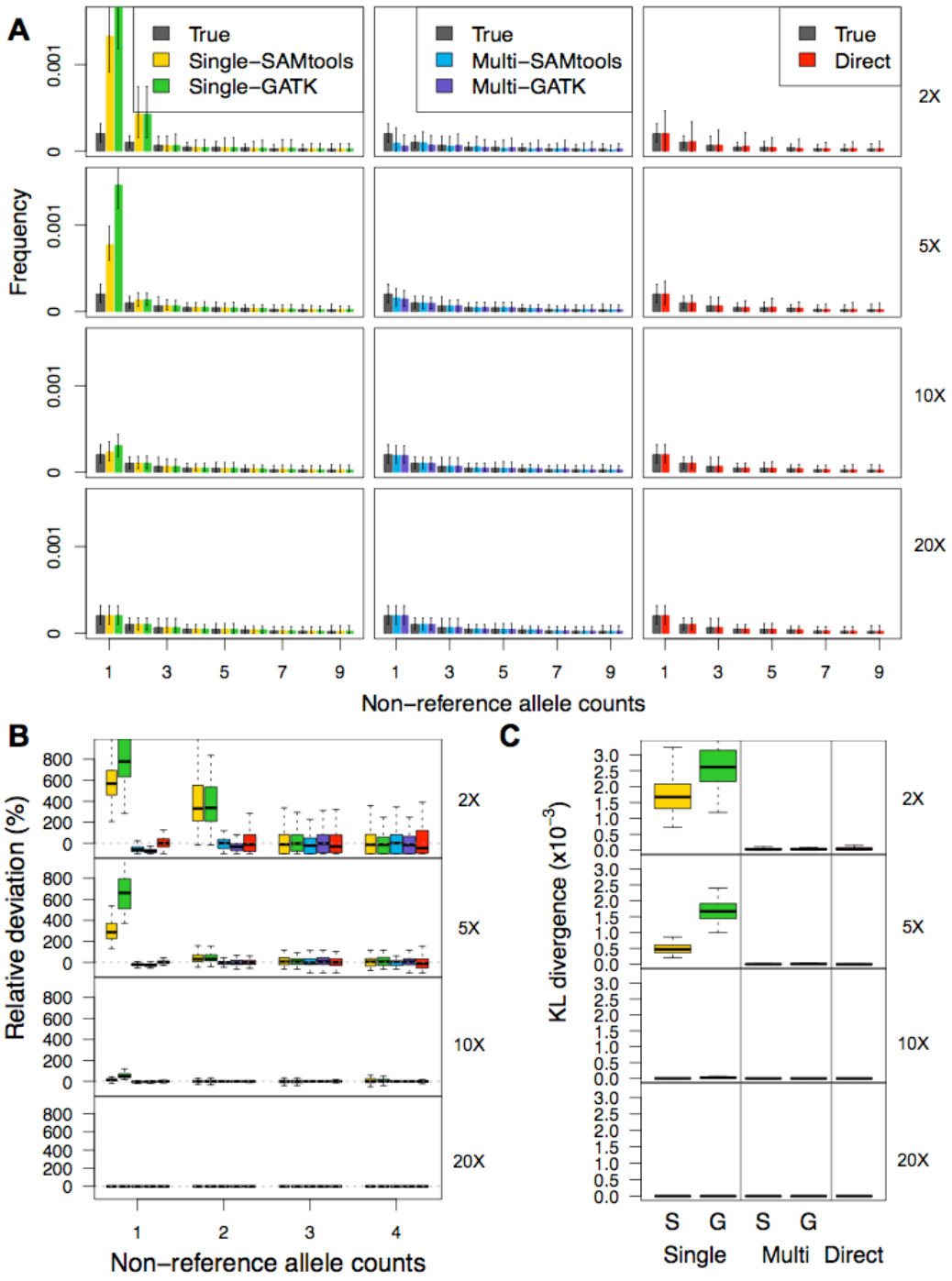
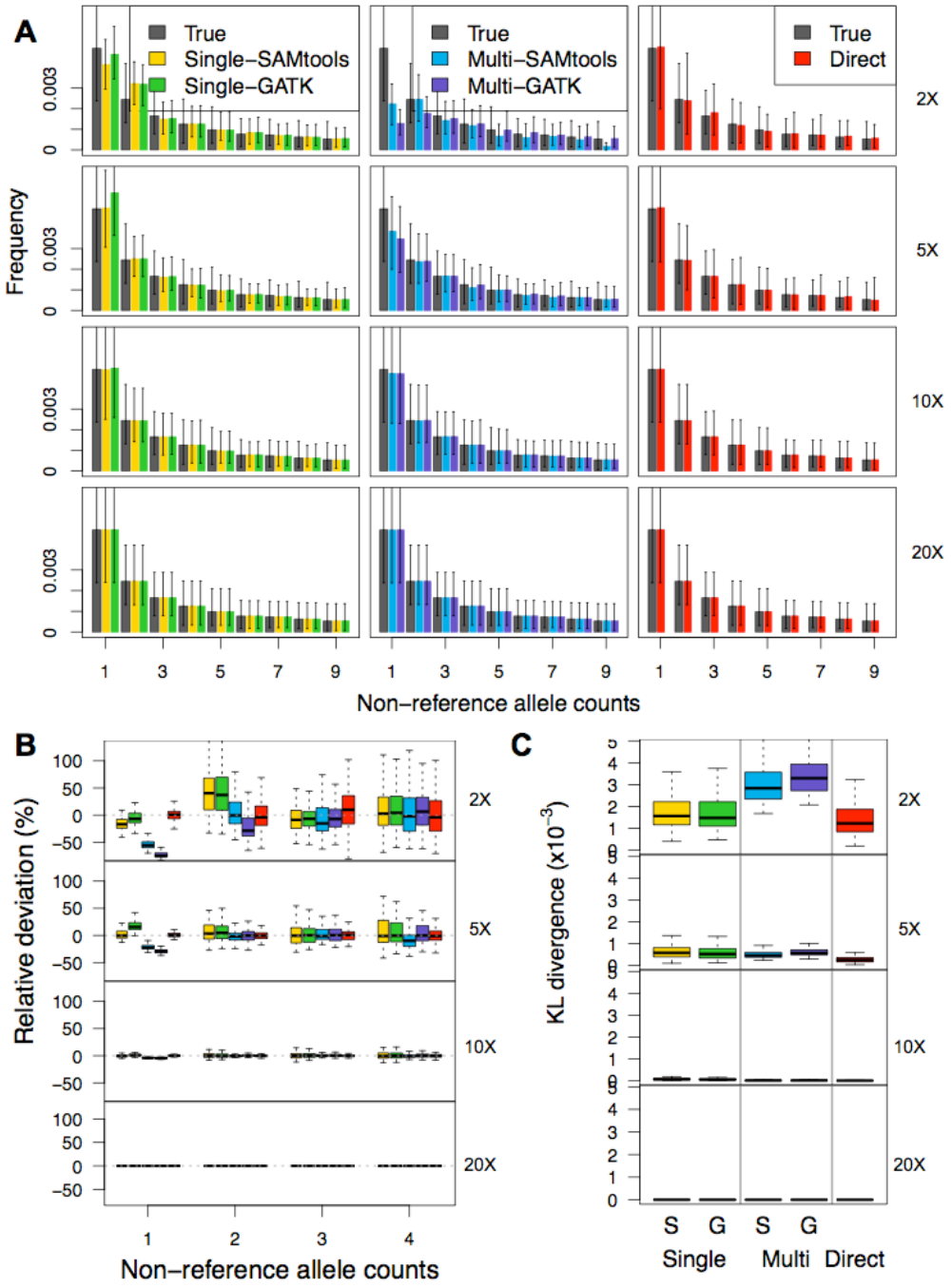


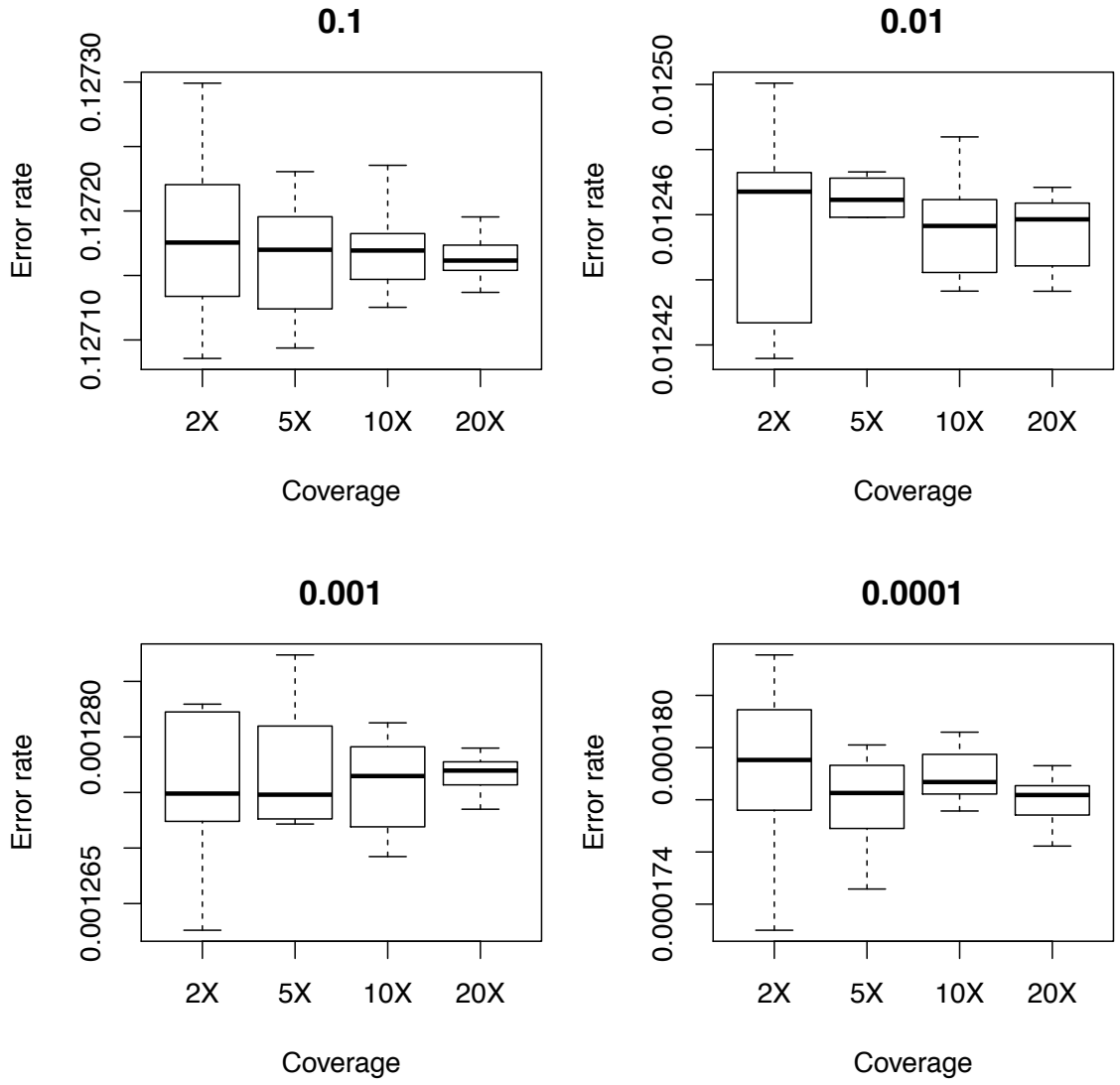
Supp. Figure 1. Error matrices where true allele counts are shown on x-axis and observed allele counts based on genotype calls are shown on y-axis. The SFS is inferred by either GATK or SAMtools with the multi-sample calling pipeline (A) or the single-sample calling pipeline (B). The first two columns show the error matrix with all allele counts and the next two columns show the error matrix with first five allele counts.



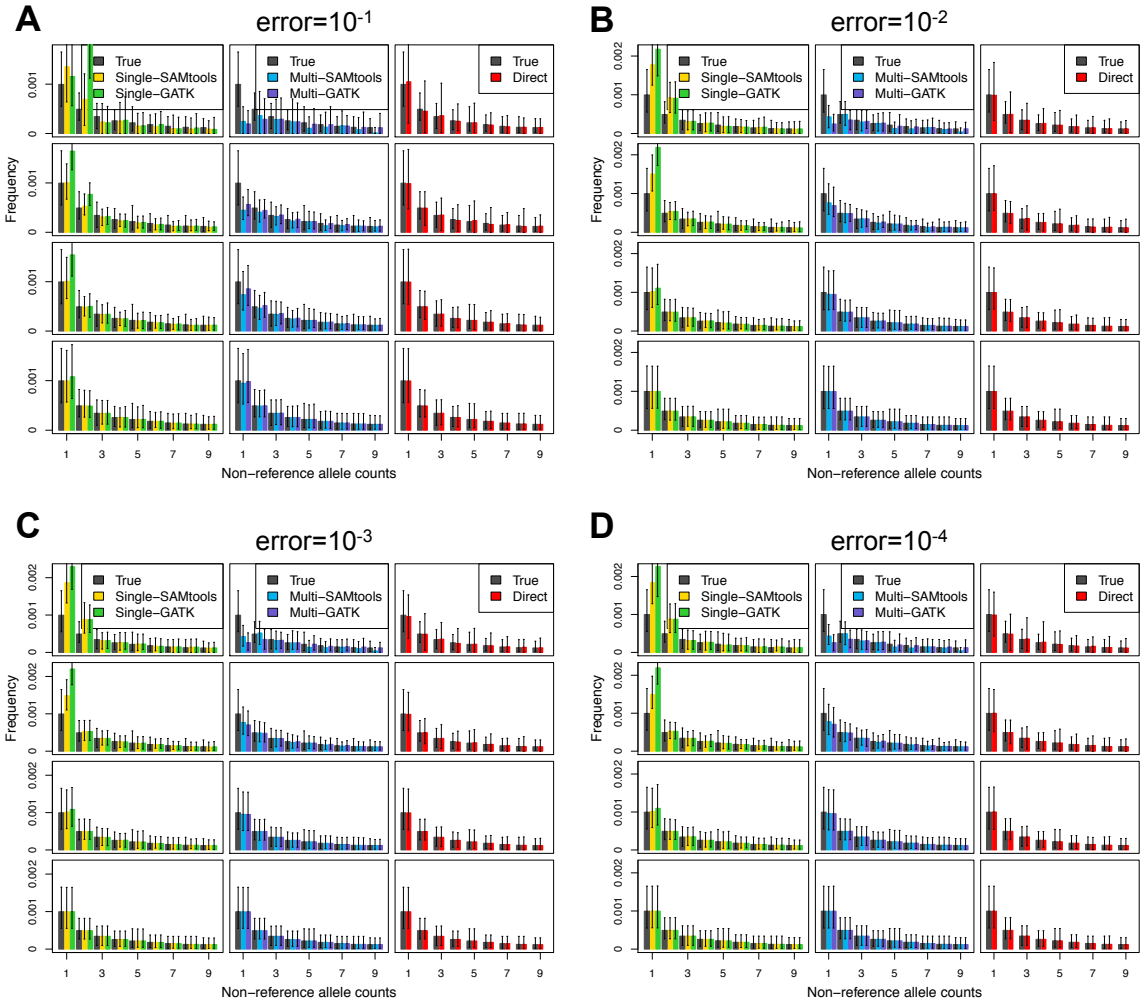
Supp. Figure 2. Evaluation of accuracy of inferred SFS by the call-based and direct estimation approach based on 100 replicates of genomic regions of length 100Kb when expected nucleotide diversity is $2 * 10^{-4}$ and sequencing error rate is 10^{-3}. A. Shapes of the inferred SFS (shown in colors in legend) compared to the ground-truth SFS (shown in grey) for coverage 2X, 5X, 10X, 20X, B. relative deviation of a fraction of sites with the non-reference allele counts of 1-4, C. a measure of a distance between the inferred and ground-truth SFS (KL divergence).



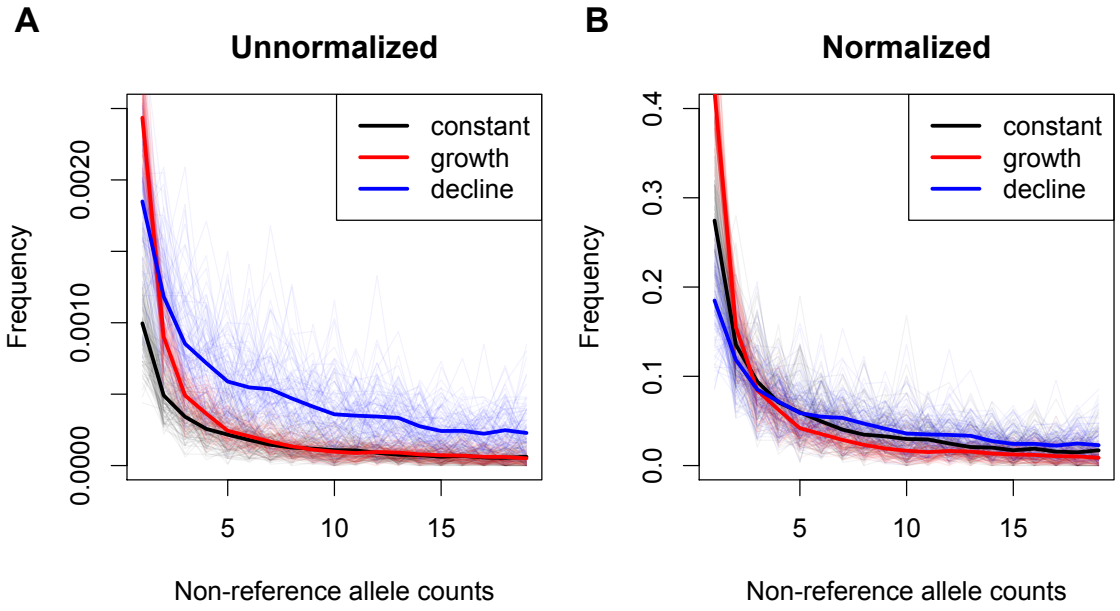
Supp. Figure 3. Evaluation of accuracy of inferred SFS by the call-based and direct estimation approach based on 100 replicates of genomic regions of length 100Kb when expected nucleotide diversity is $5 * 10^{-3}$ and sequencing error rate is 10^{-3}. A. Shapes of the inferred SFS (shown in colors in legend) compared to the ground-truth SFS (shown in grey) for coverage 2X, 5X, 10X, 20X, B. relative deviation of a fraction of sites with the non-reference allele counts of 1-4, C. a measure of a distance between the inferred and ground-truth SFS (KL divergence).



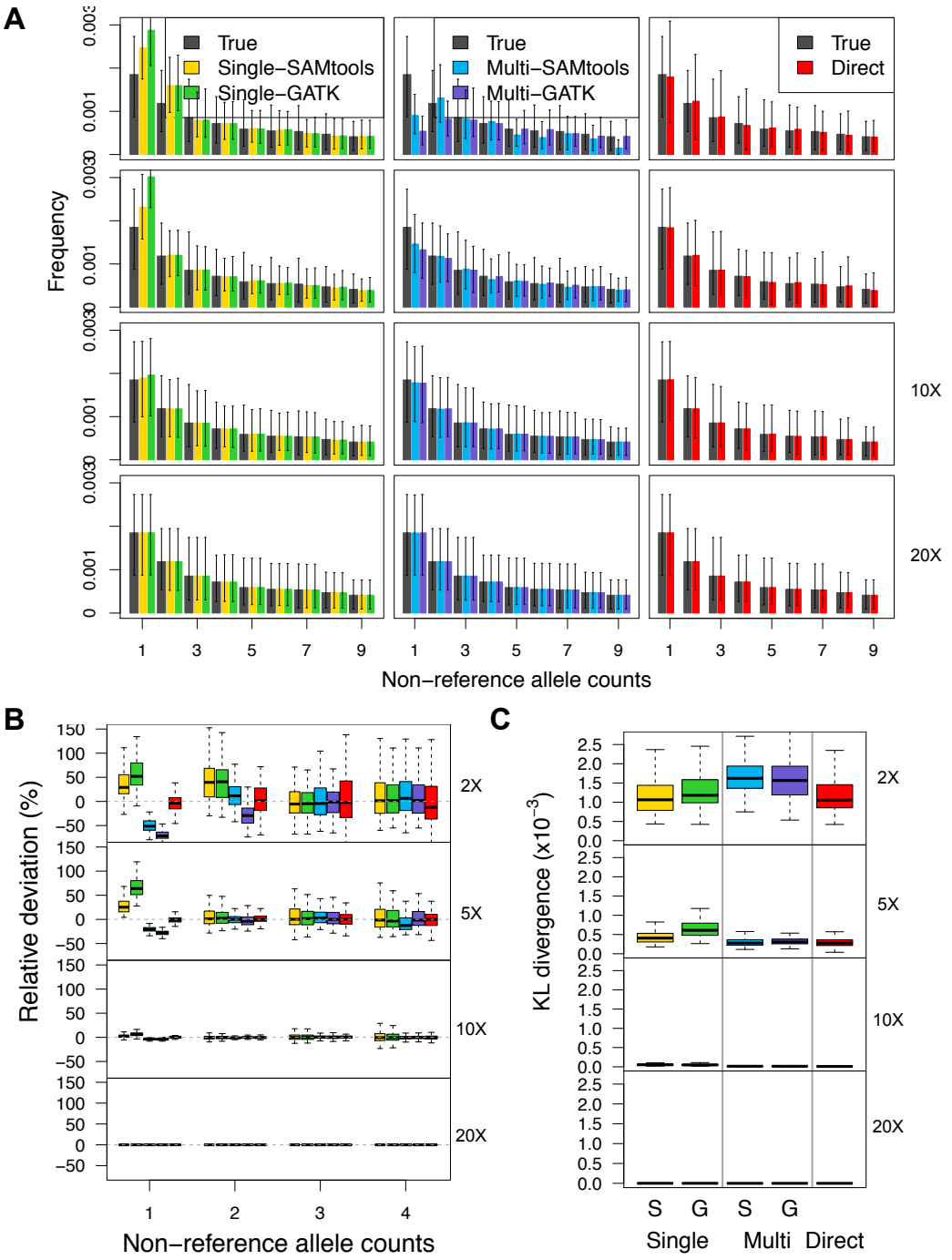
Supp. Figure 4. The distribution of observed error rates from sequencing experiment simulations given the sequencing error rate we used for simulations (shown in the title).



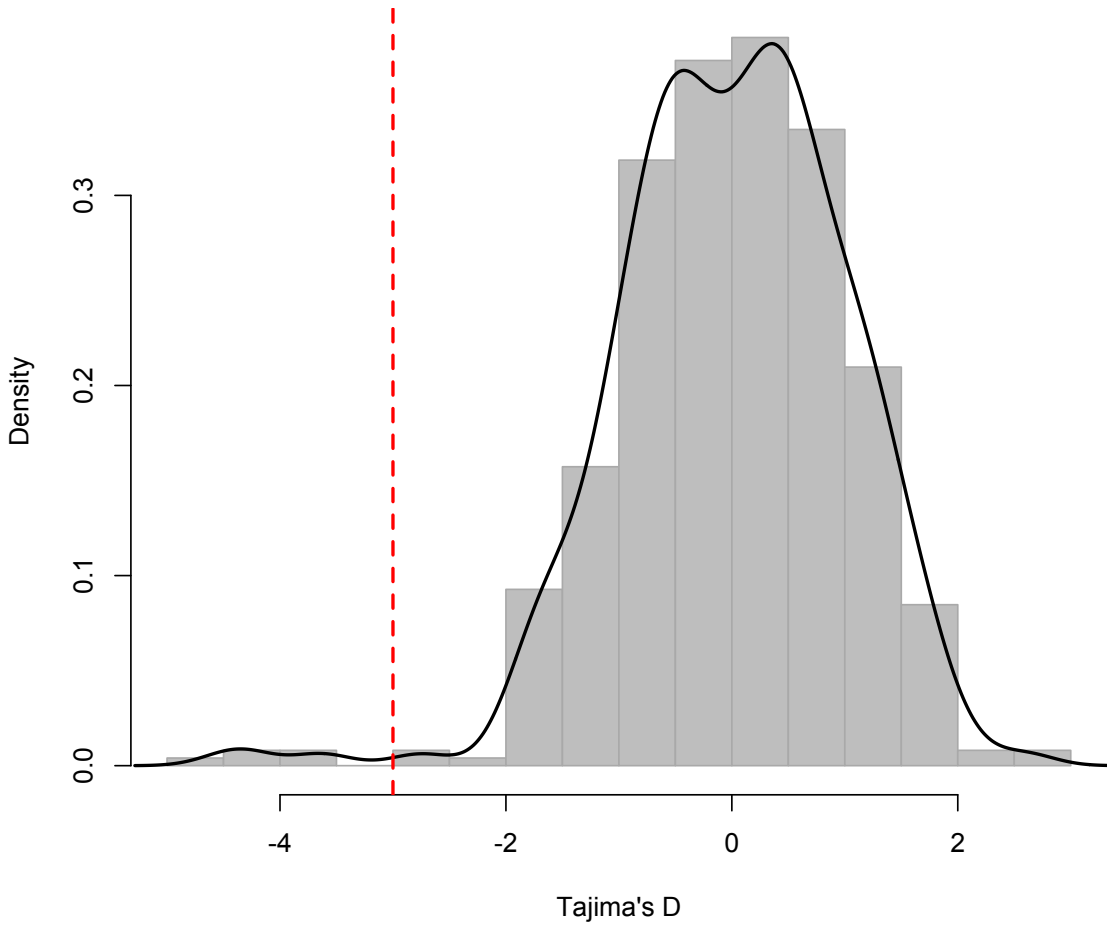
Supp. Figure 5. Shapes of the inferred SFS (shown in colors in legend) compared to the ground-truth SFS (shown in grey) for coverage 2X, 5X, 10X, 20X



Supp. Figure 6. The SFS from simulated genotype data of length 100kb. Data are simulated with a constant population size (blue), an exponential population growth with a rate of 0.01% (red), and an exponential population decline with a rate of -0.01% (blue). For each scenario, the SFS from 100 replicates are shown with thin lines and the mean SFS over 100 replicates is shown with a bold line.



Supp. Figure 7. . Evaluation of accuracy of inferred SFS by the call-based and direct estimation approach based on 100 replicates of genomic regions of length 100Kb under an exponential population decline model (rate=-0.01%). A. Shapes of the inferred SFS (shown in colors in legend) compared to the ground-truth SFS (shown in grey) for coverage 2X, 5X, 10X, 20X, B. relative deviation of a fraction of sites with the non-reference allele counts of 1-4, C. a measure of a distance between the inferred and ground-truth SFS (KL divergence).



Supp. Figure 8. The empirical distribution of Tajima's D associated with sliding windows of length 100kb in a 10Mb genomic region. The dotted red line indicates top 1% (ranked in an increasing order).