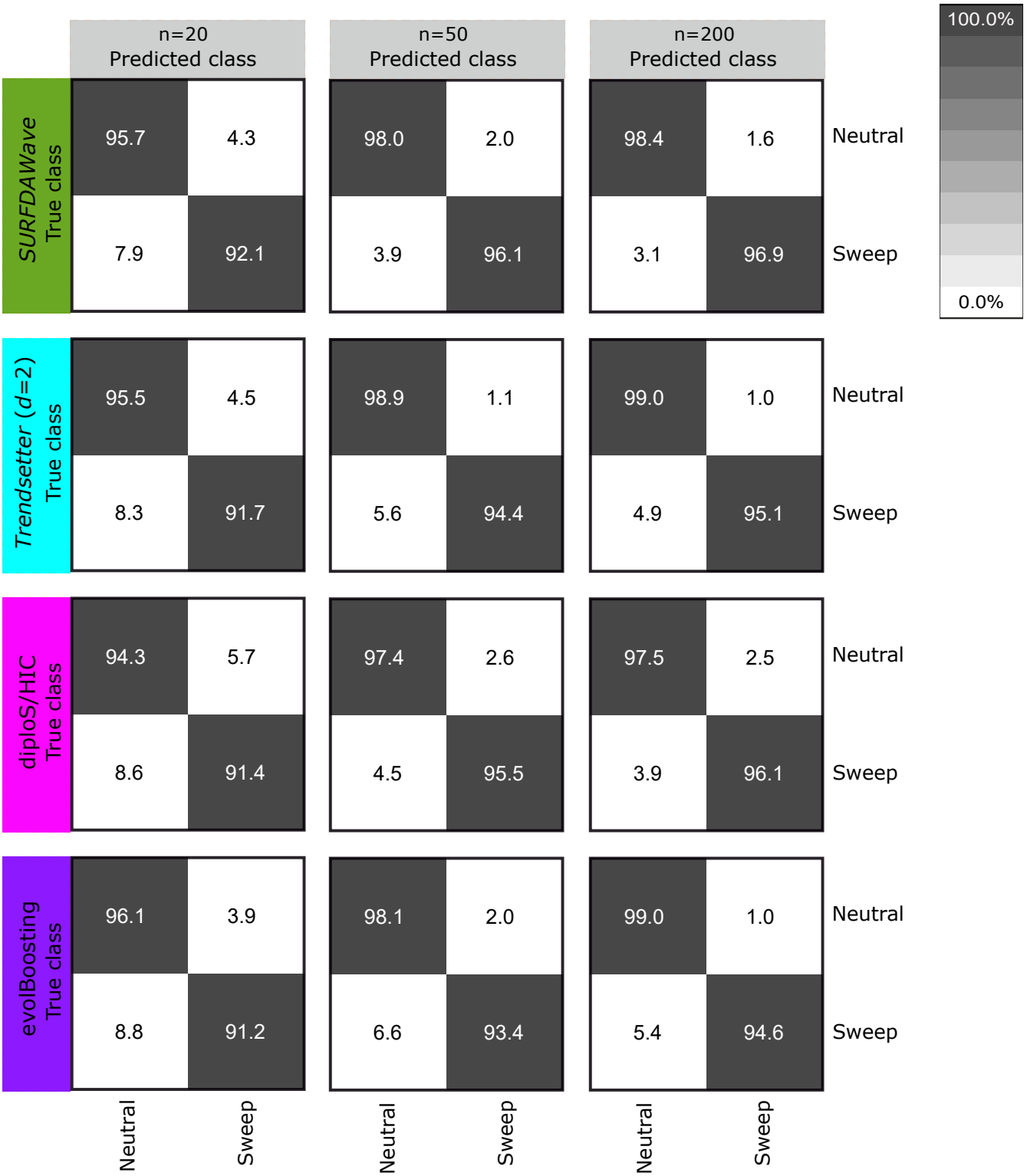

Figure S19: Confusion matrices showing the effect sample size has on classification rates. We train and test *SURFDAWave*, *Trendsetter*, diploS/HIC, and evolBoosting classifiers to differentiate sweeps and neutrality using sample sizes of $n = 20$, 50, and 200 haploid genomes. *SURFDAWave* results shown are using Daubechies' least-asymmetric wavelets to estimate spatial distributions of summary statistics and γ and levels are chosen through cross validation (see *Training the models*). Summary statistics $\widehat{\pi}$, H_1, H_{12}, H_2/H_1, and frequency of the first, second, third, fourth, and fifth most common haplotypes used by both *Trendsetter* and *SURFDAWave*.