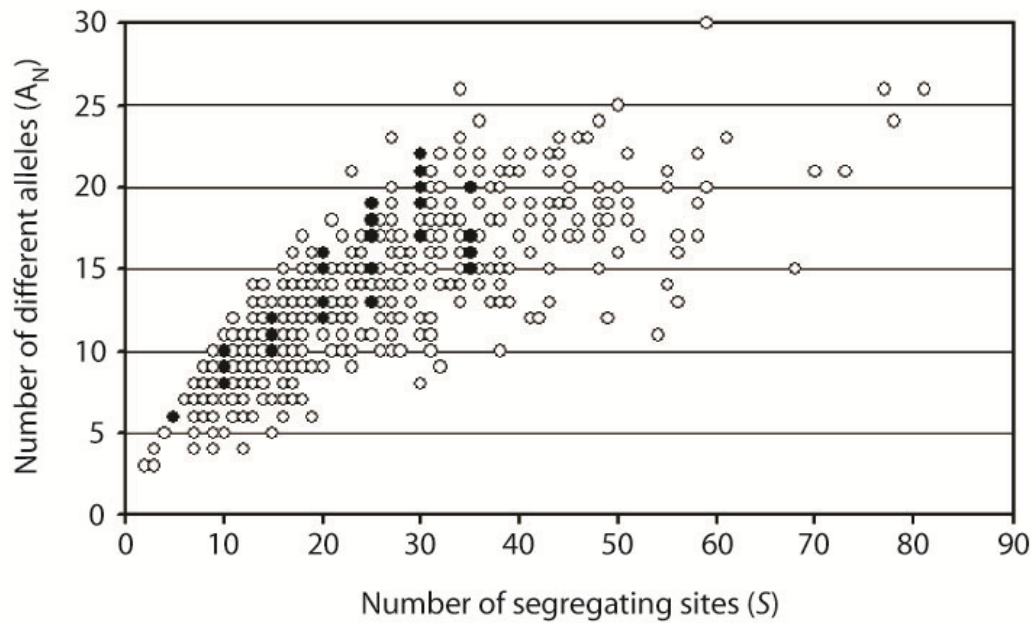




Additional file 2. *Frequency distribution of the relationship between number of segregating sites (S) and number of different alleles (A_N) in the 500 simulated datasets from which 35 (solid circles) were arbitrarily selected for further analysis using PHASE. Figure shows that none of the 35 datasets are atypical (i.e., outliers), and so the results presented in the main text are free from bias relating to the dataset selection procedure.*