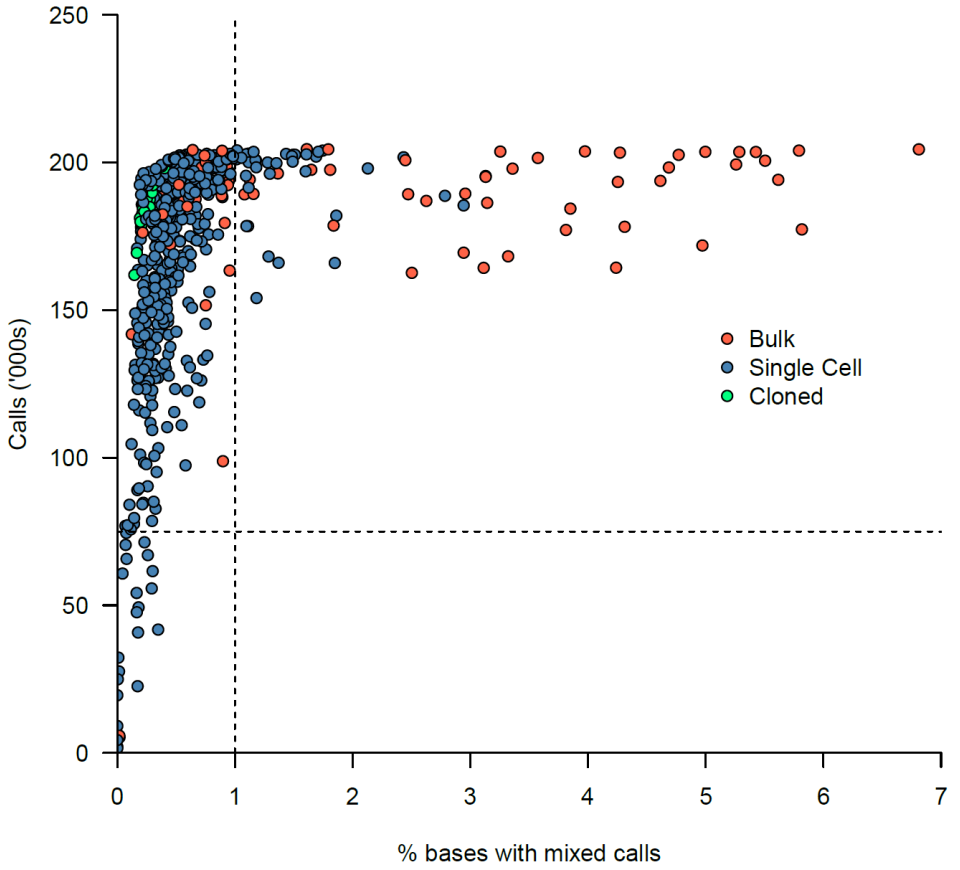
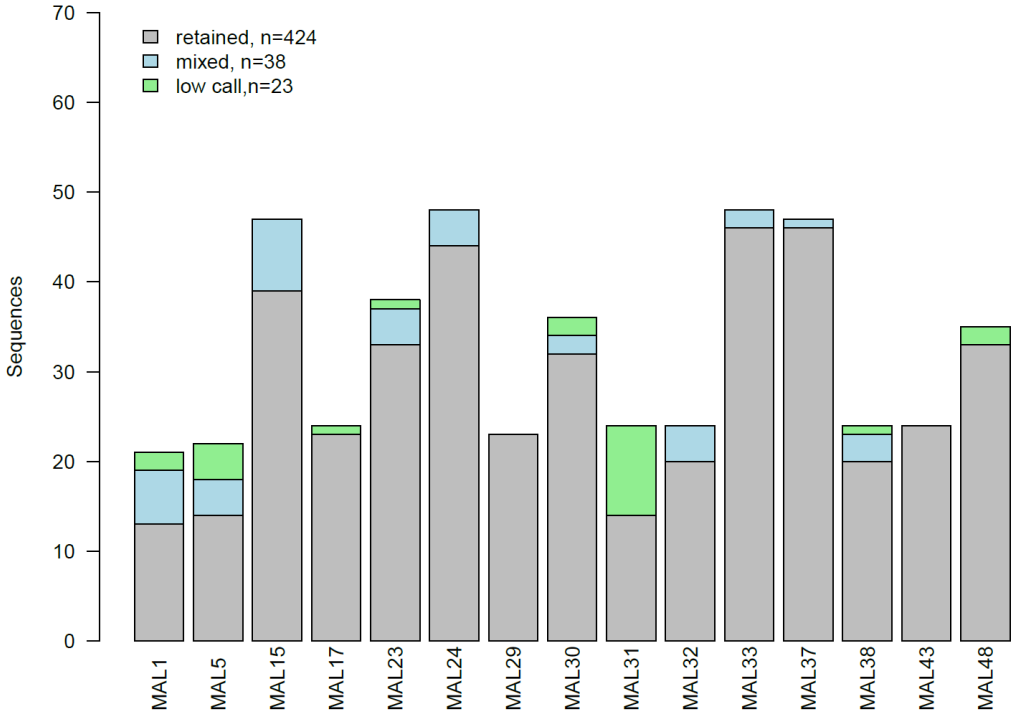
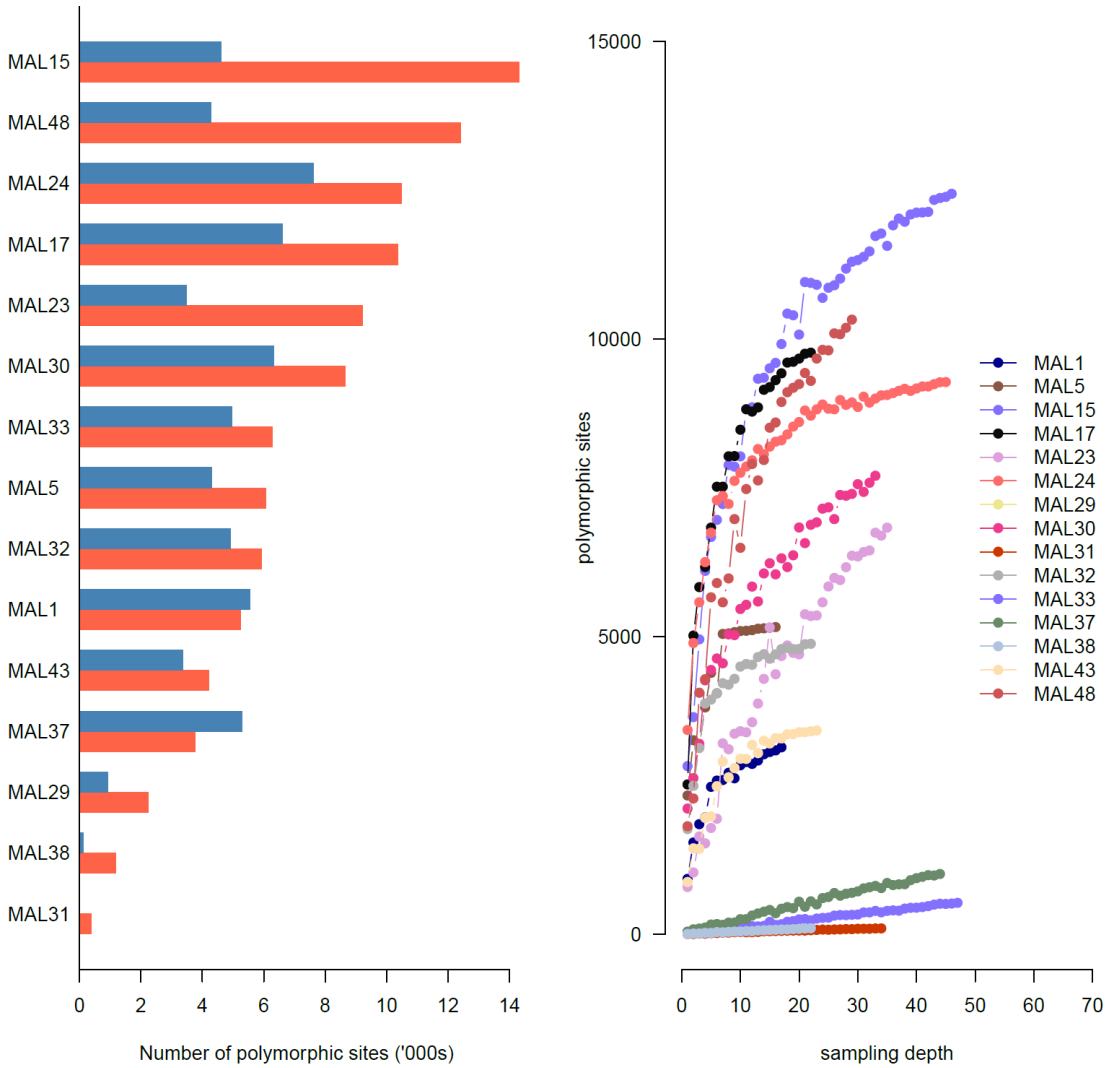
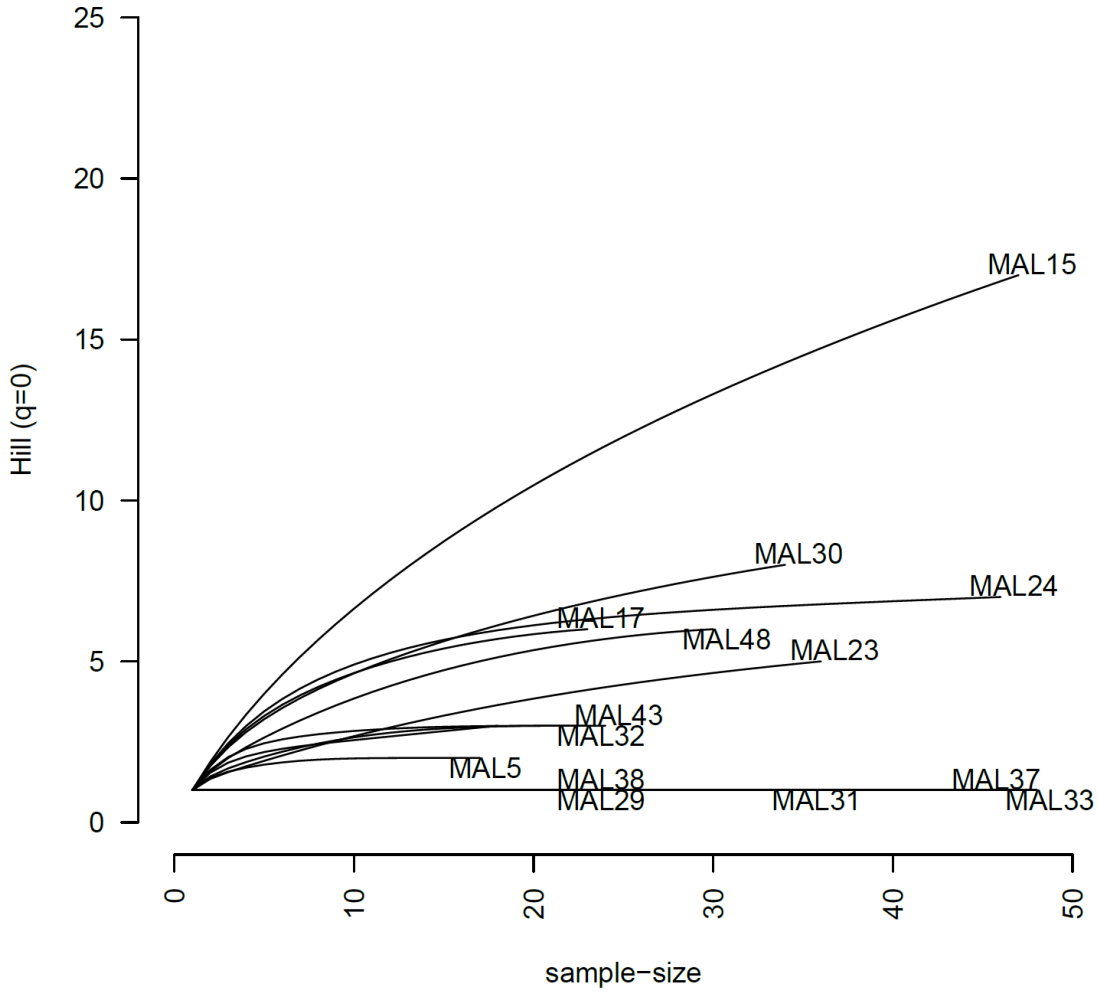




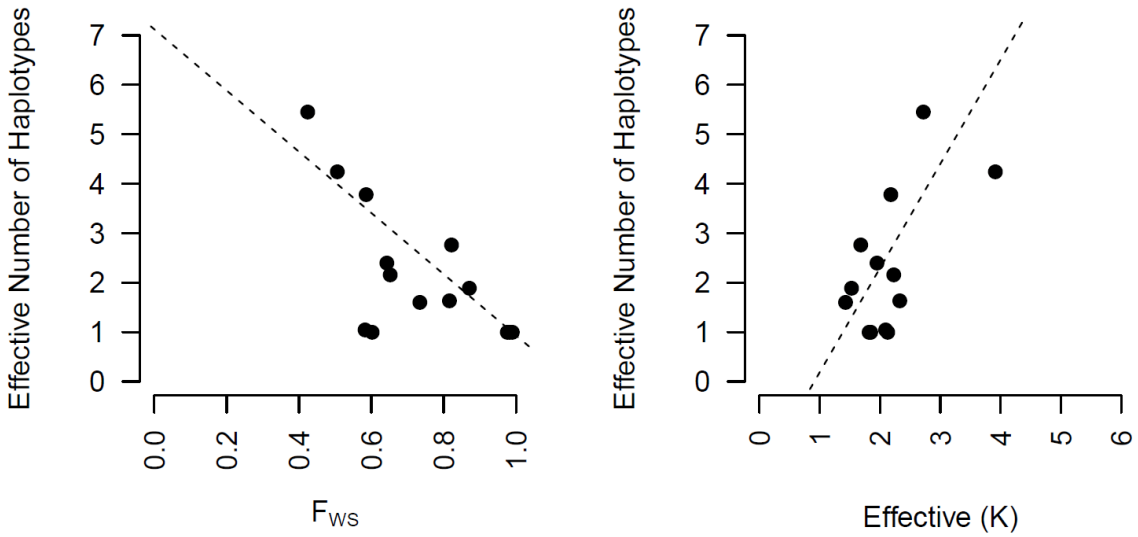
Extended Data Figure 1. Quality Control filtering of single cell sequencing. (A) Assessing the purity and coverage of each single cell sequence, single cells in the top right quadrant of the graph were retained. (B) A summary of retained samples for each infection.



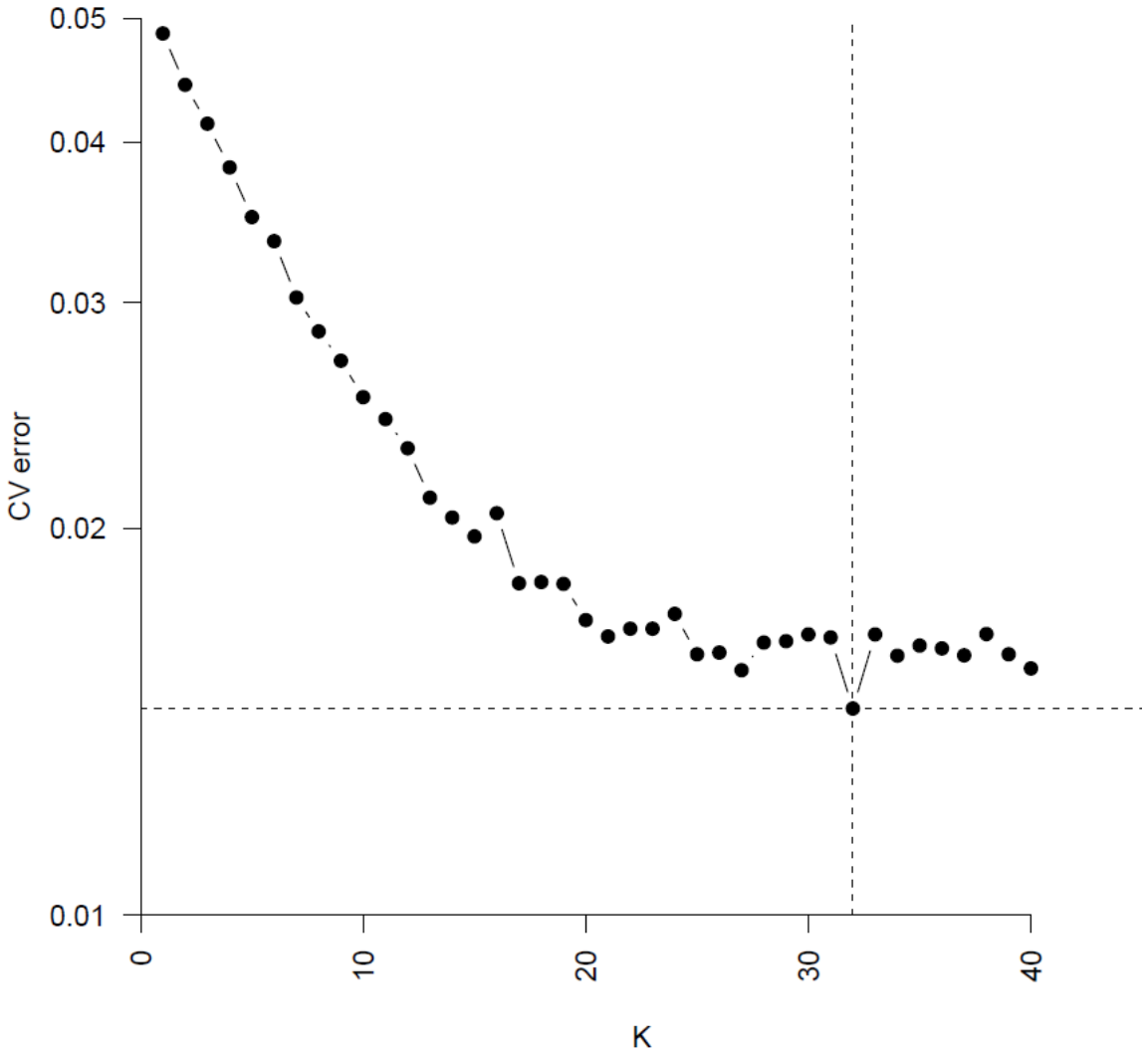
Extended Data Figure 2. Single cell sequencing captures more polymorphic sites than bulk sequencing. (A) the number of polymorphic sites detected by bulk sequencing (blue bars) and single cell sequencing (red bars). (B) Subsampling of single cell sequences reveals a plateau in the discovery of new sites as increasing numbers of samples are sequenced. This suggests most polymorphism in the infections have been effectively sampled.



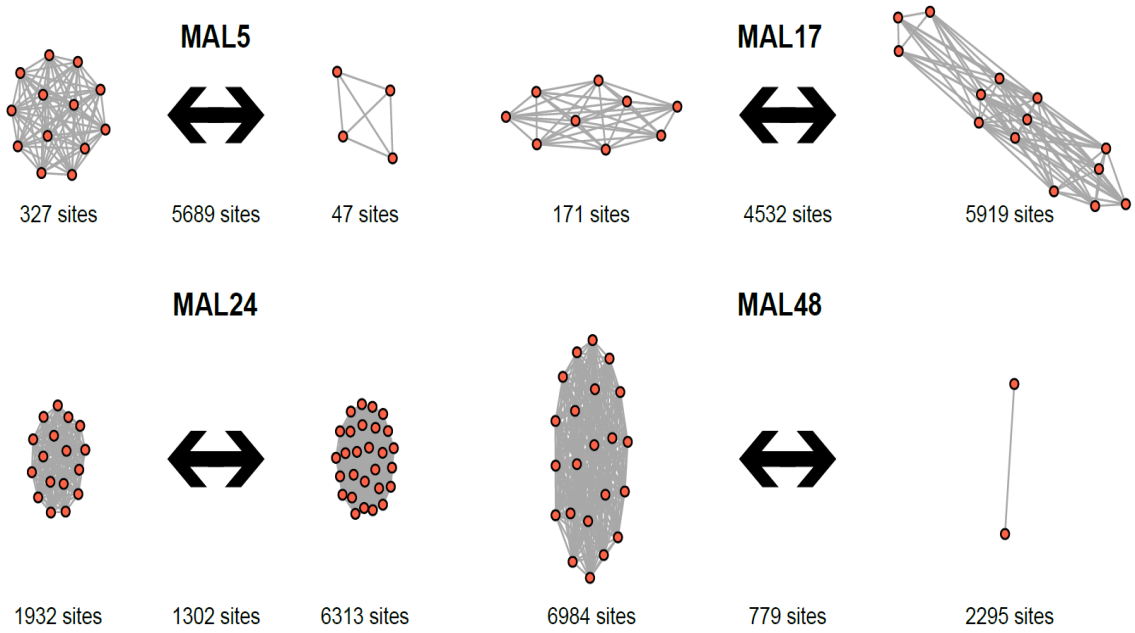
Extended Data Figure 3. Rarefaction of each sample suggest comprehensive capture of most unique haplotypes from the 15 infections.



Extended Data Figure 4. Correlation between the effective number of haplotypes inferred from single cell sequencing and either F_{WS} (A) or Effective K inferred from Deploid.



Extended Data Figure 5. Cross validation error score for ADMIXTURE identifies K=32 as the optimum value of K.



Extended Data Figure 6. Genetic diversity of coinfecting and superinfecting lineages. For each infection the sequences showing >15% of the genome shared IBD are clustered together. Under each cluster is the number of sites which were polymorphic in that cluster (and monomorphic in the other). Under the arrow for each infection is the number of sites which were polymorphic between the clusters, but not within either cluster.