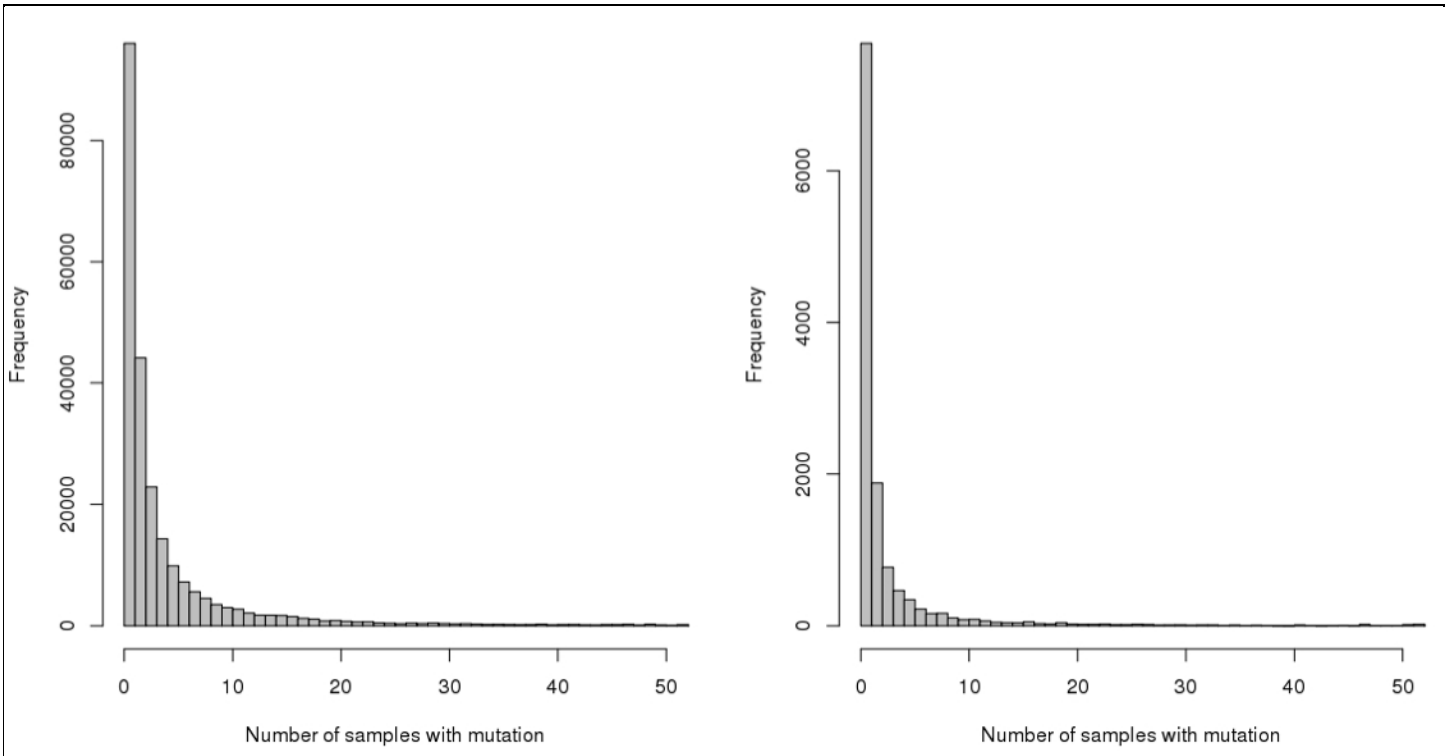
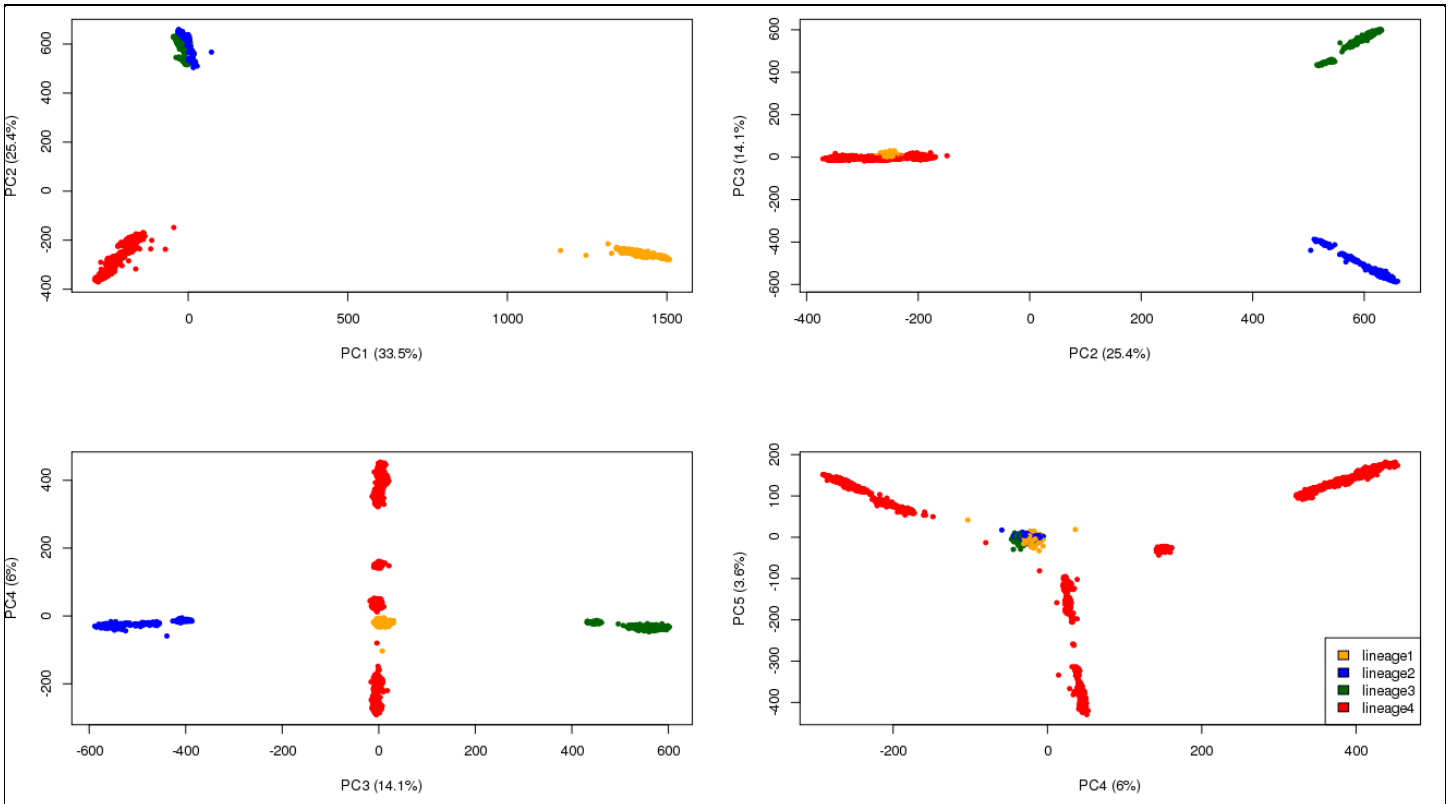


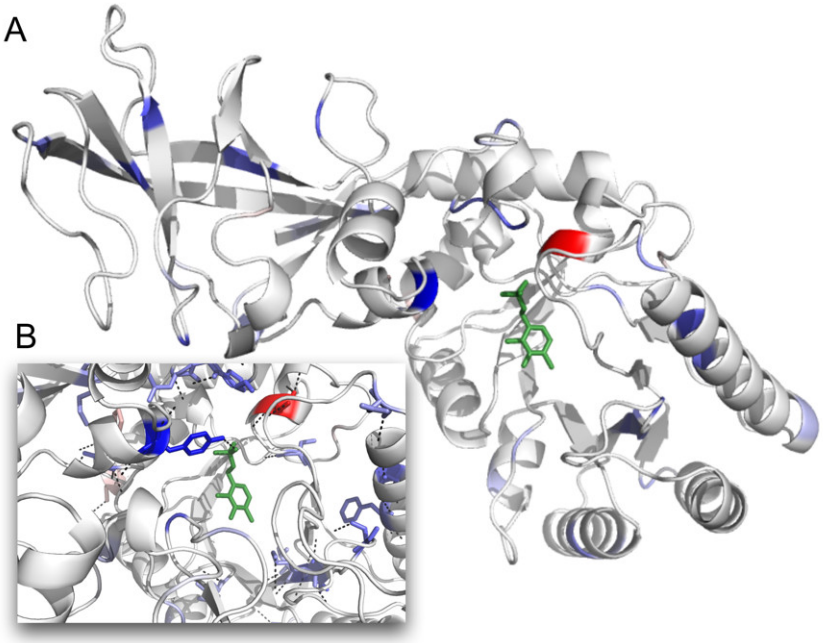
Supplementary Figure 1

Allele frequency spectra for SNPs (left) and small insertions and deletions (indels, right)



Supplementary Figure 2

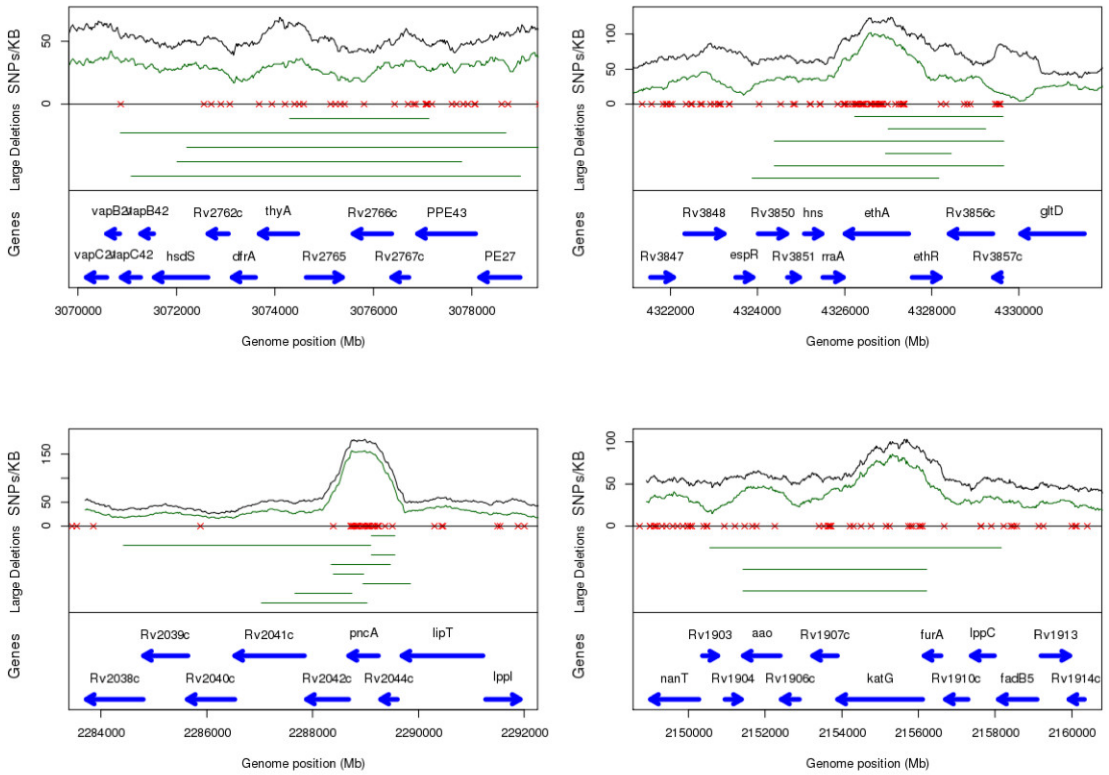
Principal component (PC) analysis confirms lineage and sub-lineage based population structure (total variation explained across five components is 82.7%)



Supplementary Figure 3

Protein structure for *alr*

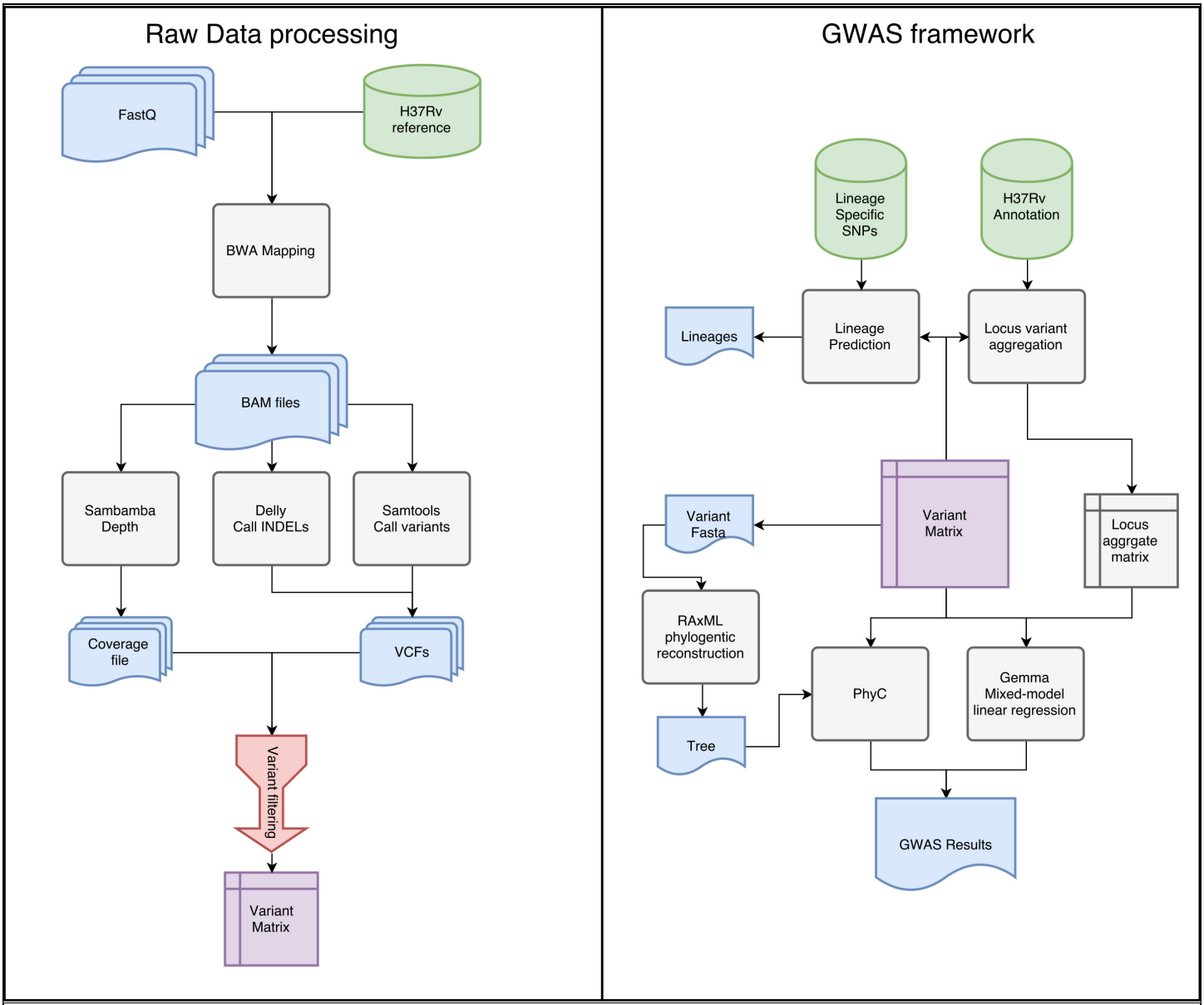
Alanine racemase mutational map showing position and effect of mutations based on measure of protein stability by DUET. Unfavourable mutations are depicted in blue and favourable mutations in red, where colour intensity reflects extent of effect. The PLP co-factor shown as a stick representation in green. (A) shows the protomer structure of alanine racemase depicted as a cartoon with the PLP co-factor shown as sticks. Insert (B) shows the active site with residues that have been identified in the GWAS depicted as sticks and their hydrogen bonding propensity shown as dashed black lines.



Supplementary Figure 4

Polymorphisms in regions surrounding *ethA* (top left), *thyA* (top right), *pncA* (bottom left), and *katG* (bottom right) using the complete dataset (n=6,465)

The top panel in the figures shows the density of SNPs per Kb (green – non-synonymous, black – all). The red crosses show the location of the small indels. The middle panel shows the location of the large deletions found in samples used in this study. The lower panel shows the location of the candidate regions and flanking genes.



Supplementary Figure 5

The analytical workflow, including procedures adopted for raw sequence data processing and the genome-wide association study (GWAS) approach