RESEARCH ARTICLE

Figures and figure supplements

Multiple introductions of multidrug-resistant typhoid associated with acute infection and asymptomatic carriage, Kenya

Samuel Kariuki *et al*

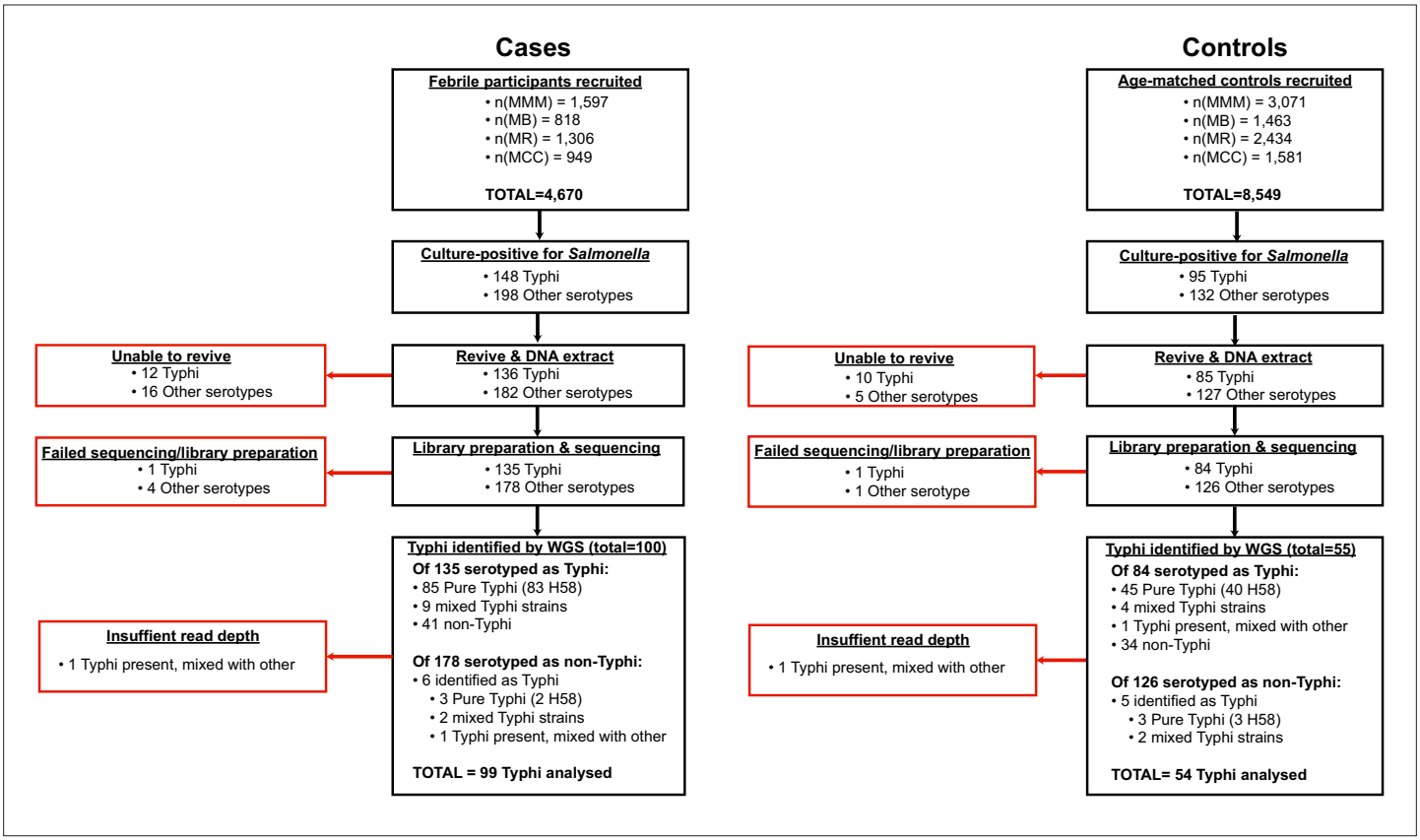

Figure 1. Flow chart of samples collected and analysed. Red boxes indicate bacterial isolates that could not be included in downstream genetic analyses, grouped by reason for exclusion.

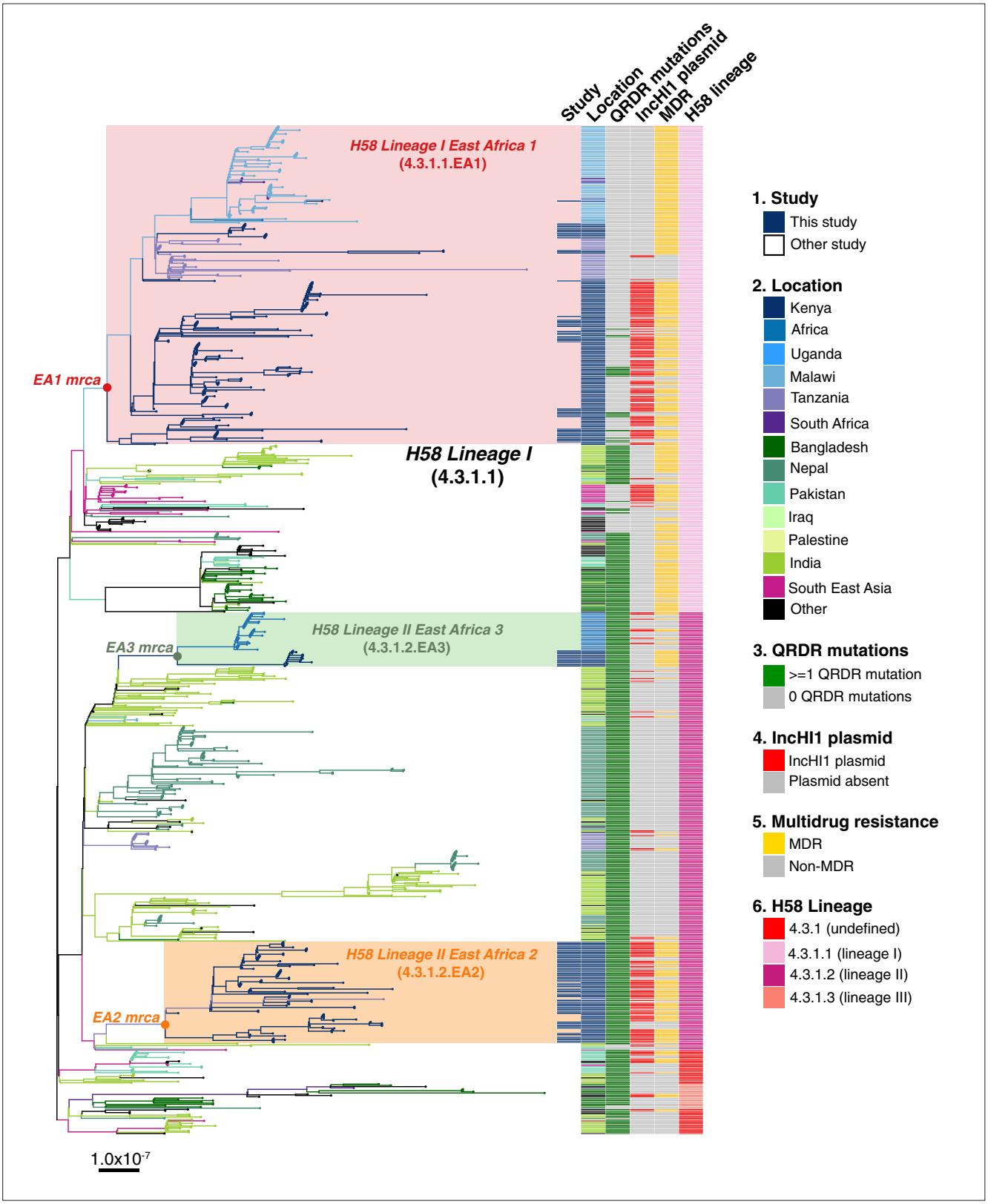

Figure 2. Global population structure of H58 (4.3.1) *S.Typhi* showing Kenyan isolates cluster into three East African clades. Whole genome phylogeny of 1,204 H58 isolates, including all available Kenyan genomes (n = 128 from this study, n=111 from prior studies) and globally distributed genomes for context (n=965, see Methods). Branch lengths are in substitutions per core-genome site, branches are coloured to indicate geographical origin (see inset legend), shaded boxes highlight the three East African H58 clades defined in this study. Colour bars to the right indicate (as per inset legend): 1,

Figure 2 continued on next page

Figure 2 continued

Kenyan strains isolated and sequenced during this study; 2, geographical location; 3, mutation(s) in the quinolone resistance determining region (QRDR) of genes *gyrA, gyrB,* and *parC*; 4, presence of multidrug resistance (MDR) IncHI1 plasmid; 5, presence of MDR genes; 6, H58 lineage. Interactive version available at https://microreactorg/project/wViqmaRdZuFVEb6yk4i1jU.

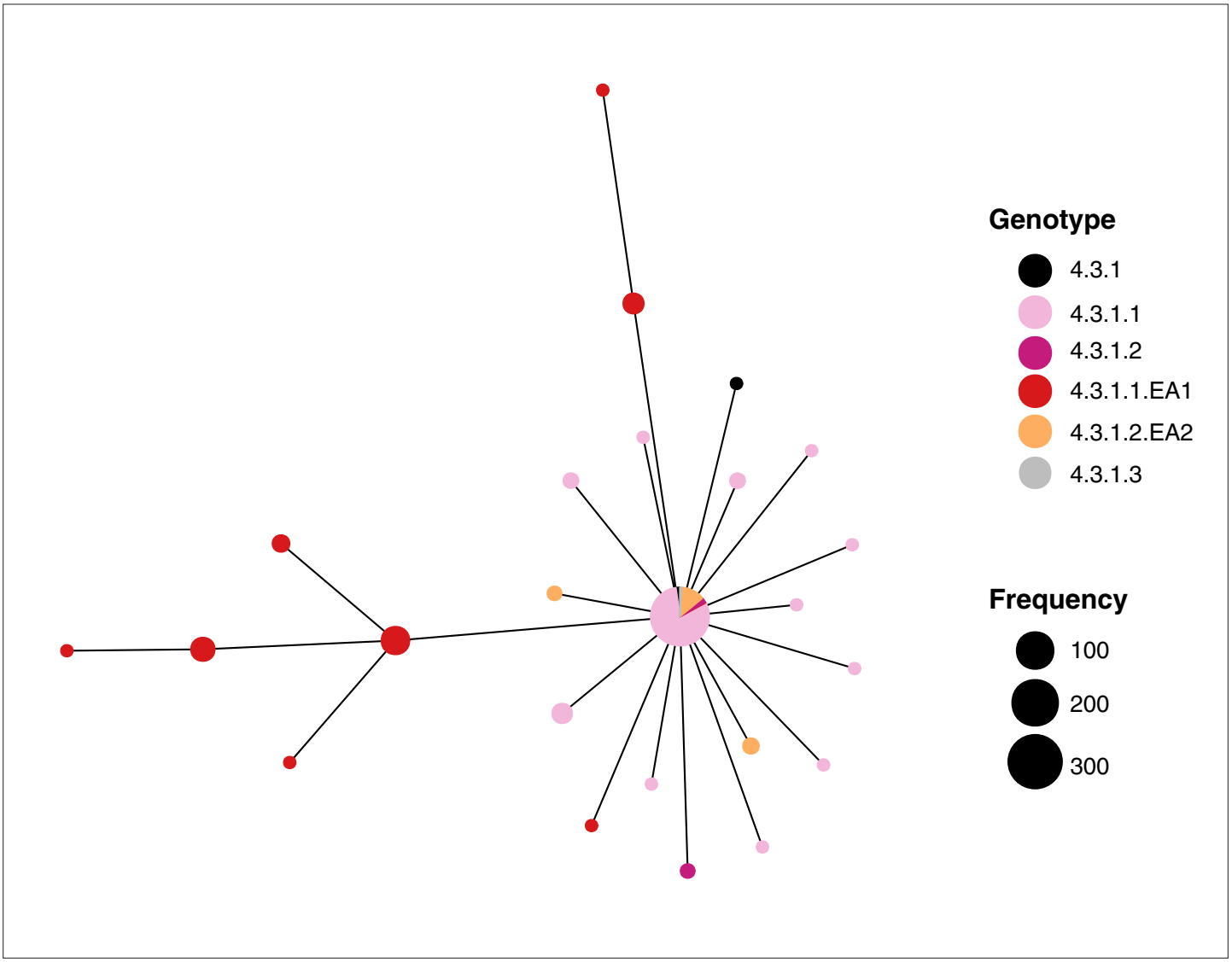

Figure 2—figure supplement 1. *S. Typhi* IncHI1 PST6 plasmid minimum spanning tree. Nodes indicate unique IncHI1 PST6 haplotypes observed among n=534 PST6 plasmid sequences. Nodes are pie charts sized by the number of sequences among which the different plasmid haplotypes were observed and are coloured by the genotype(s) of the host *S.* Typhi sequences carrying the plasmid haplotype.

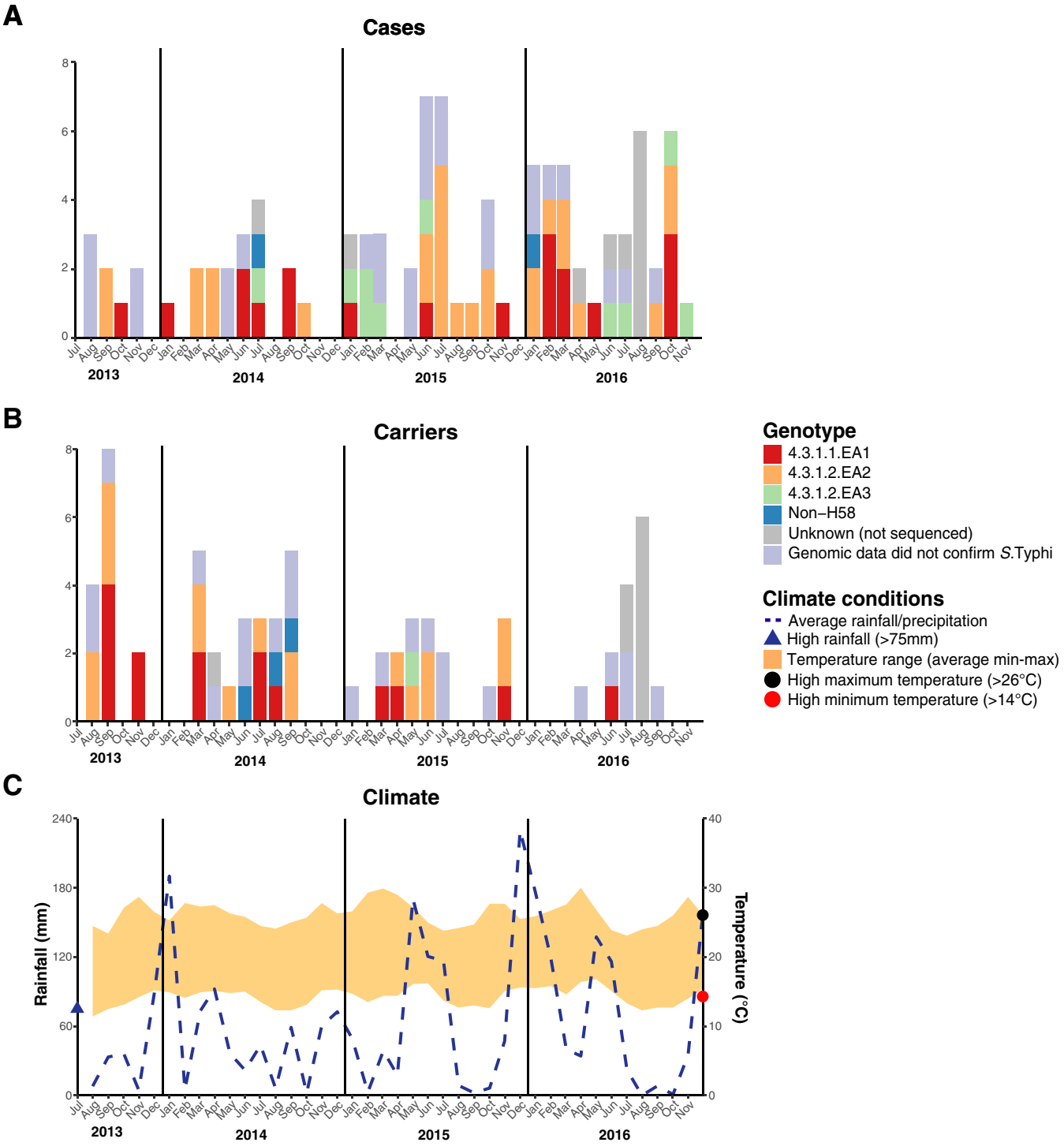

Figure 3. Epidemic curve of all *S. Typhi* cases and controls per month inside the DSS. (**A**) Monthly distribution of *S. Typhi* genotypes from cases. (**B**) Monthly distribution of *S. Typhi* genotypes from carriers. Note that the counts include all participants who were culture-positive for *S. Typhi* and also those who were culture-positive for other *Salmonella* but identified later by WGS as *S. Typhi*. (**C**) Weather conditions throughout the study period. Blue dashed line indicates precipitation level per month (rainfall), shaded orange polygon indicates the temperature range, red circle indicates threshold for high minimum temperature for statistical testing, black circle indicates threshold for high maximum temperature for statistical testing, blue triangle indicates threshold for high rainfall for statistical testing.

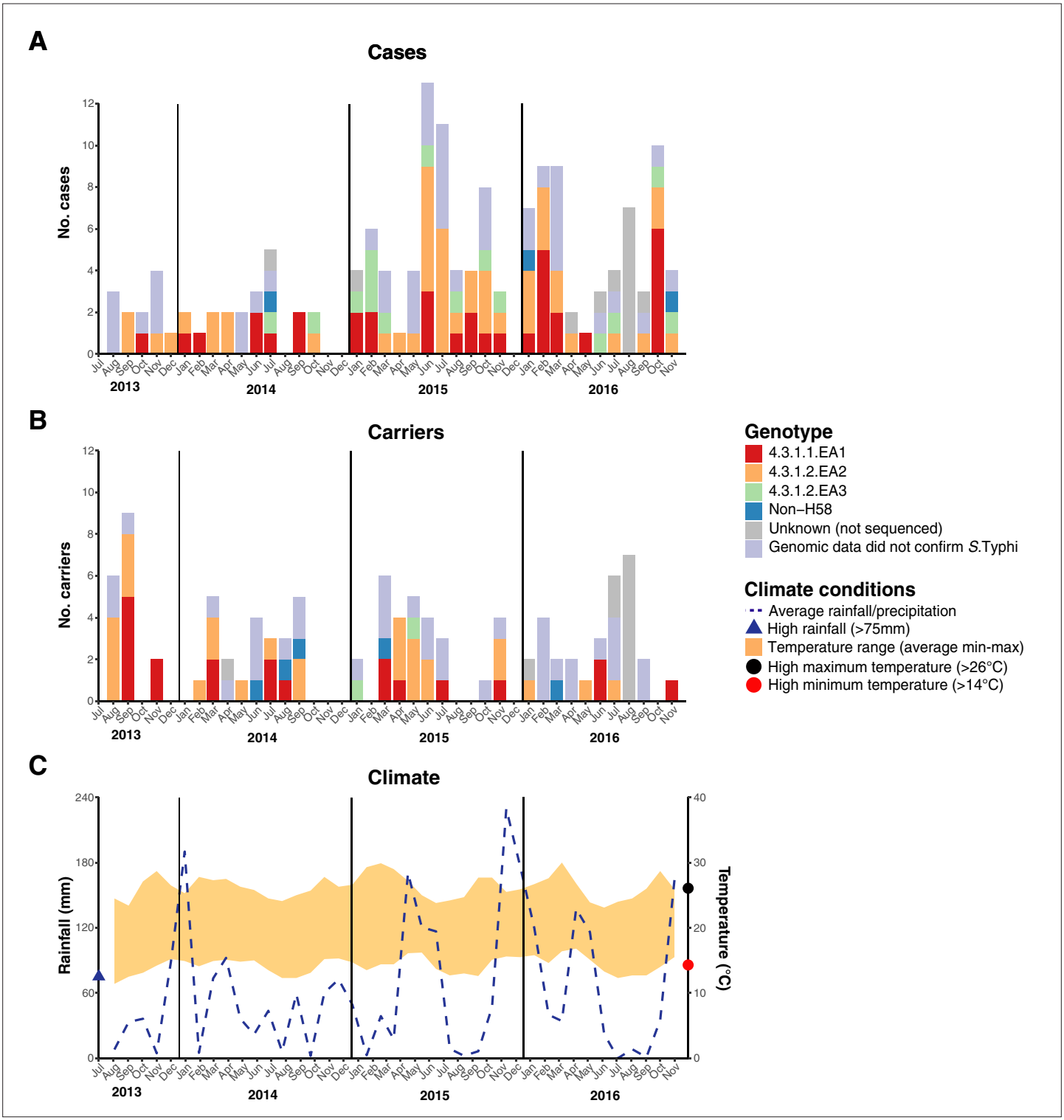

Figure 3—figure supplement 1. Epidemic curve of all *S.Typhi* cases and controls per month. (**A**) Monthly distribution of *S. Typhi* genotypes from cases. (**B**) Monthly distribution of *S. Typhi* genotypes from carriers. Note that the counts include all participants who were culture-positive for *S. Typhi* and also those who were culture-positive for other *Salmonella* but identified later by WGS as *S. Typhi*. Genotypes are indicated by colour as per inset legend. (**C**) Weather conditions throughout the study period. Blue dashed line indicates precipitation level per month (rainfall), shaded orange polygon indicates the temperature range, red circle indicates threshold for high minimum temperature for statistical testing, black circle indicates threshold for high maximum temperature for statistical testing, blue triangle indicates threshold for high rainfall for statistical testing.

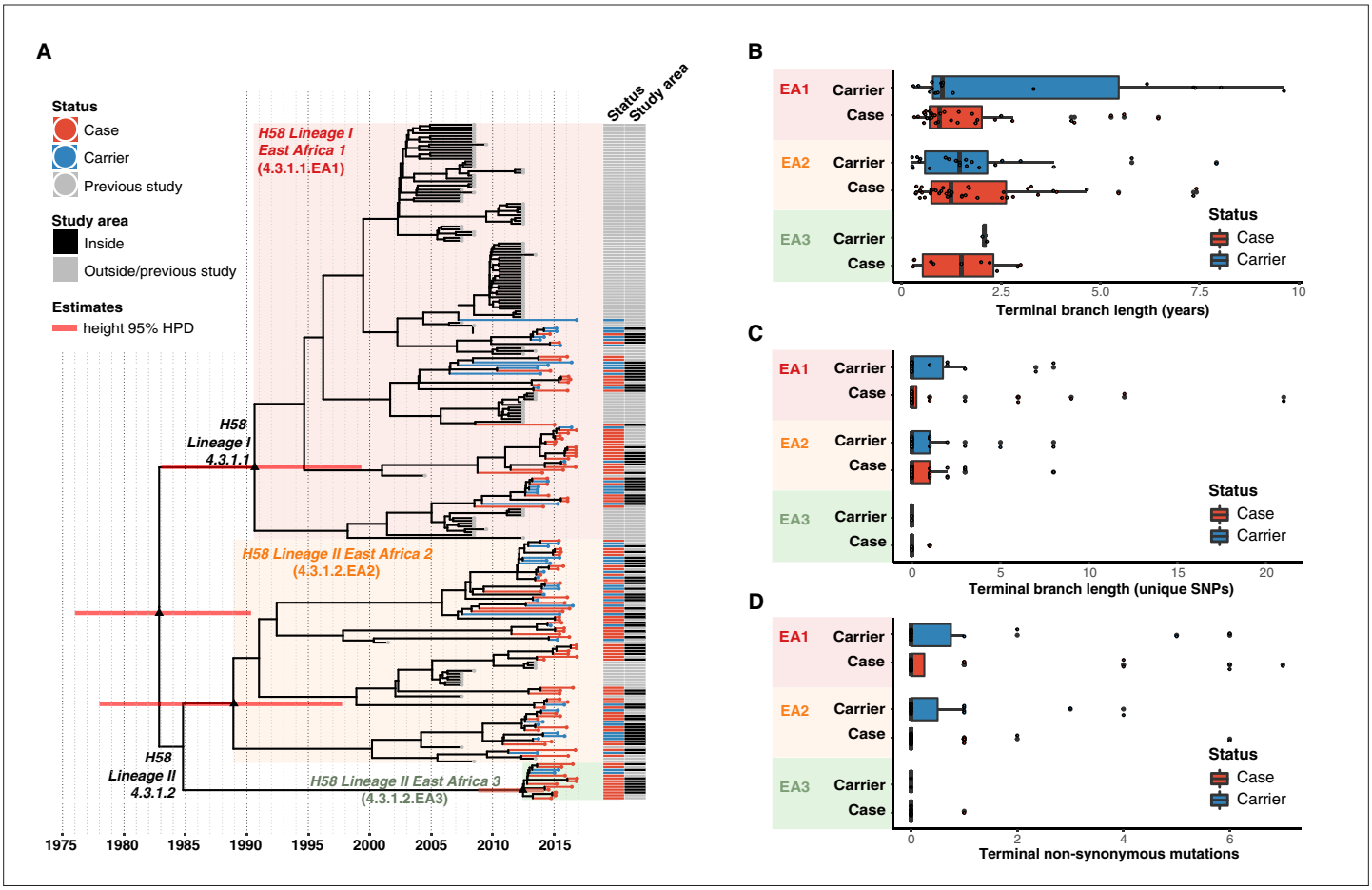

Figure 4. Temporal distribution of genotypes and among all cases and carriers. (**A**) Dated maximum-clade credibility phylogenetic tree of Kenyan *S.* Typhi genotype 4.3.1 (H58), including 128 isolated from this study. Tip colours & first colour bar indicate symptom status, second colour bar indicates those isolates from children living in the defined survey area. Black triangles demarcate nodes of interest, and the accompanying bars indicate 95% HPD of node heights. Interactive phylogeny available at https://microreactorg/project/I2KUoasUB. (**B**) Distribution of terminal branch lengths for all sequences, extracted from the Bayesian tree shown in (**A**). (**C**) Distribution of isolate-specific SNPs detected in sequences from all cases and controls. (**D**) Distribution of terminal non-synonymous mutations detected in sequences from all cases and controls. In the boxplots in panels B, C, and D, black bars indicate median values, boxes indicate interquartile range. Cases and carrier samples indicated as per the inset legend.

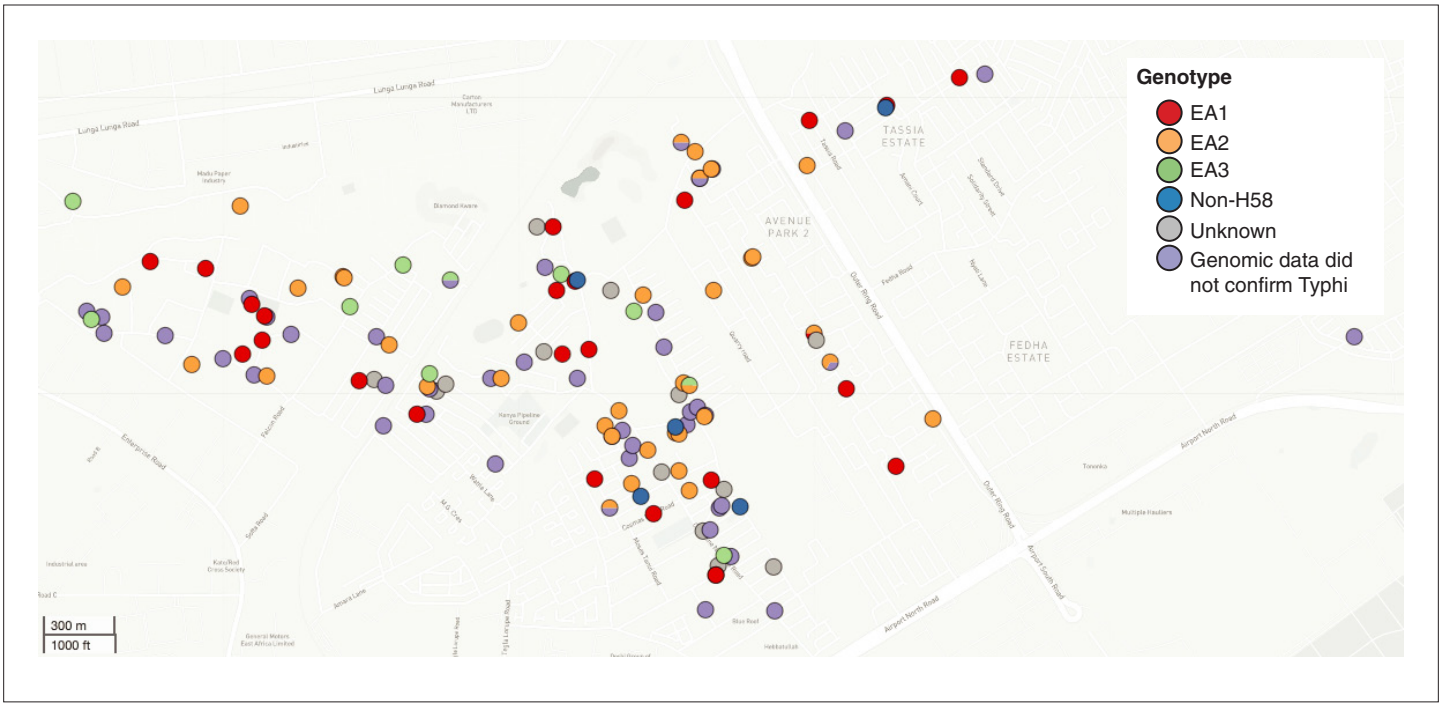

Figure 4—figure supplement 1. Spatial distribution of *S. Typhi* genotypes throughout the informal settlement. Each individual point represents an individual *S. Typhi* isolate obtained from the informal settlement which are coloured by genotype as per the inset legend.

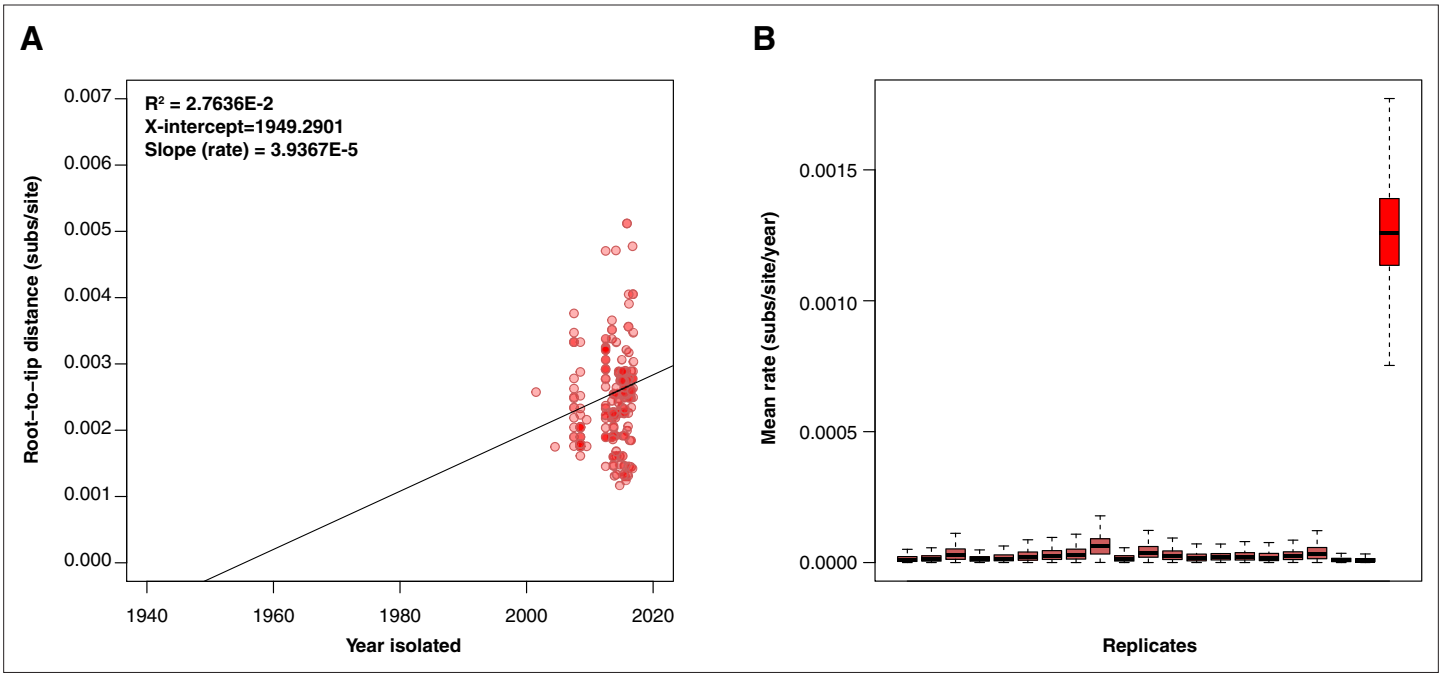

Figure 4—figure supplement 2. Tempest regressions & BEAST date randomisation testing. (**A**) Kenyan H58 *S.* Typhi tempest regression of root-to-tip distance as a function of sampling time, with the root of the tree selected using heuristic residual mean squared (each point represents a tip of the maximum likelihood tree). The slope is a crude estimate of the substitution rate for the SNP alignment, the x-intercept corresponds to the age of the root node, and the R^2 is a measure of clocklike behaviour (**B**) Kenyan H58 *S.* Typhi date randomisation test with the right most box plot showing the posterior substitution rate estimate from the SNP alignment of the data with the correct sampling times, and the remaining 20 boxplots showing the posterior distributions of the rate from replicate runs using randomised dates. The data are considered to have strong temporal signal if the estimate with the correct sampling times does not overlap with those from the randomisations.

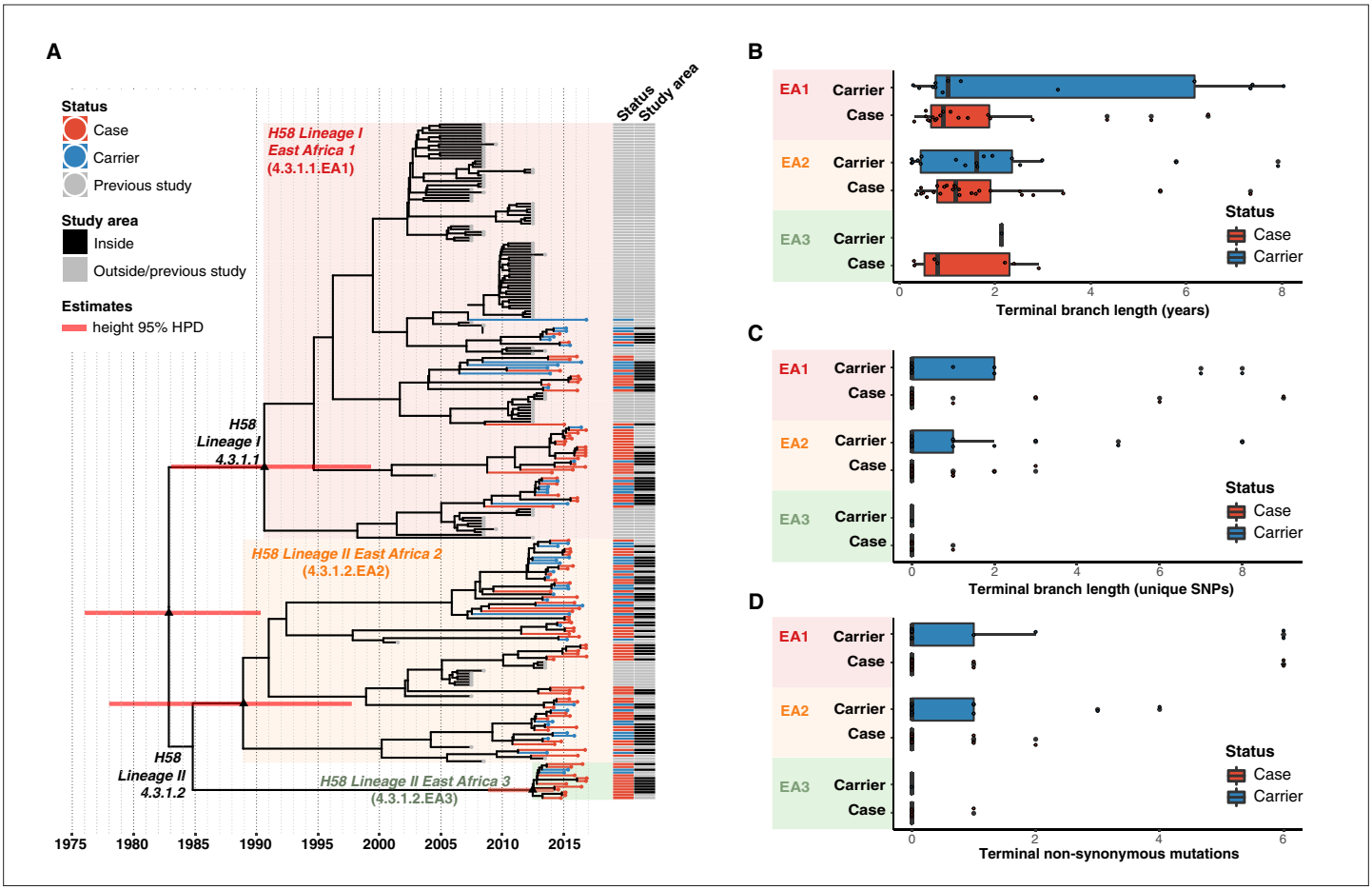

Figure 4—figure supplement 3. Temporal and age distribution of genotypes among cases and controls inside the survey site. (**A**) Dated maximum-clade credibility phylogenetic tree of Kenyan *S.* Typhi genotype 4.3.1 (H58), including 128 isolated from this study. Tip colours and first colour bar indicate symptom status, second colour bar indicates those isolates from children living in the defined survey area. Black triangles demarcate nodes of interest, and the accompanying bars indicate 95 % HPD of node heights. Interactive phylogeny available at https://microreactorg/project/I2KUoasUB. (**B**) Distribution of terminal branch lengths for sequences isolated from cases and controls within the survey area, extracted from the Bayesian tree shown in (**A**). (**C**) Distribution of isolate-specific SNPs detected in sequences from cases and controls resident the survey area. (**D**) Distribution of terminal non-synonymous mutations detected in sequences from cases and controls within the survey area. In the boxplots in panels **B, C, and D**, black bars indicate median values, boxes indicate interquartile range.

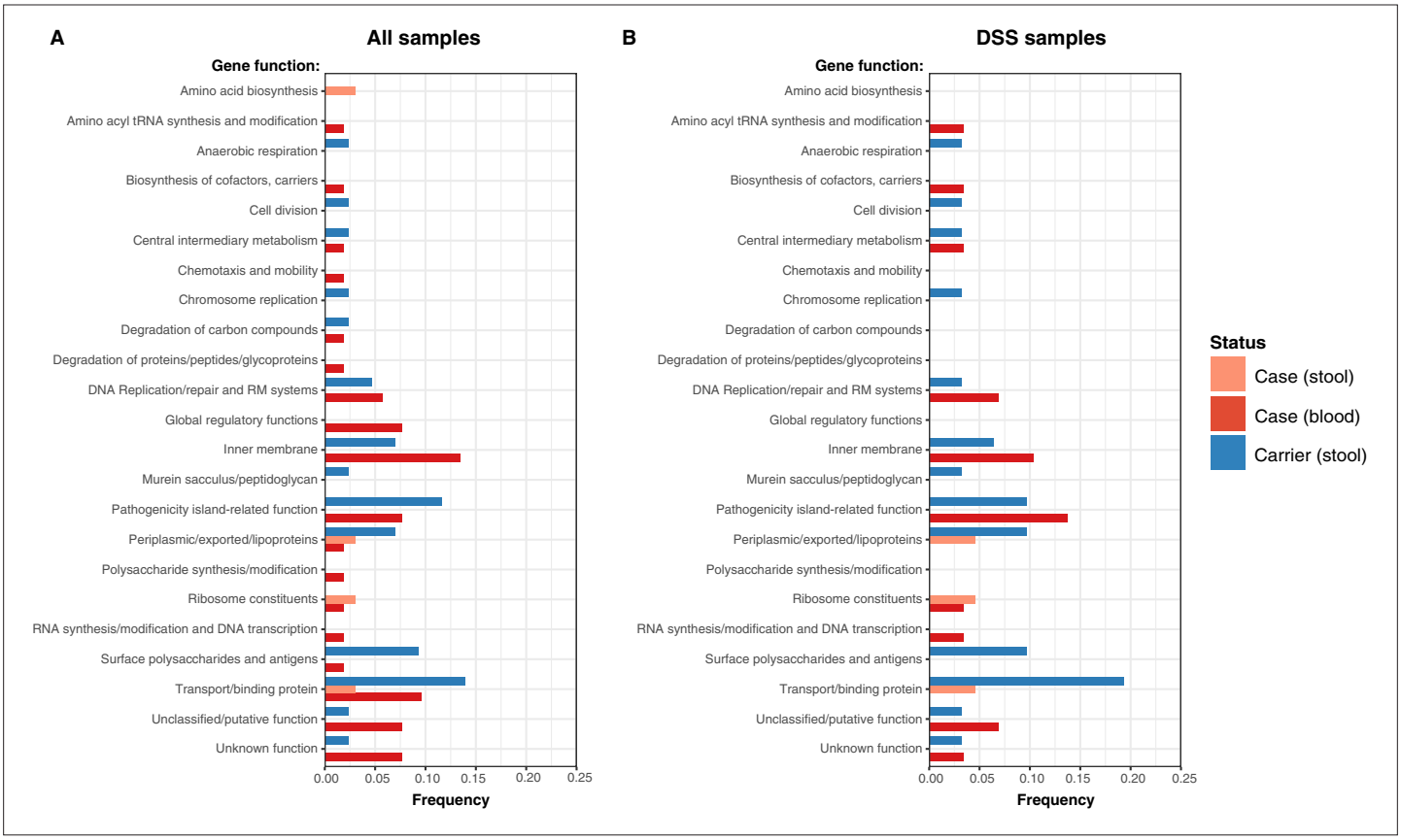

Figure 5. Frequency of terminal non-synonymous mutations in difference gene functional categories among cases and carriers. (**A**) Frequency of terminal non-synonymous mutations in all sequences collected. (**B**) Frequency of terminal non-synonymous mutations in sequences from within the DSS area. Red bars indicate the frequency non-synonymous mutations found in acute case samples from blood, peach bars indicate non-synonymous mutations found in acute case samples from stool, and blue bars indicate the frequency of mutations found in carrier samples from stool.