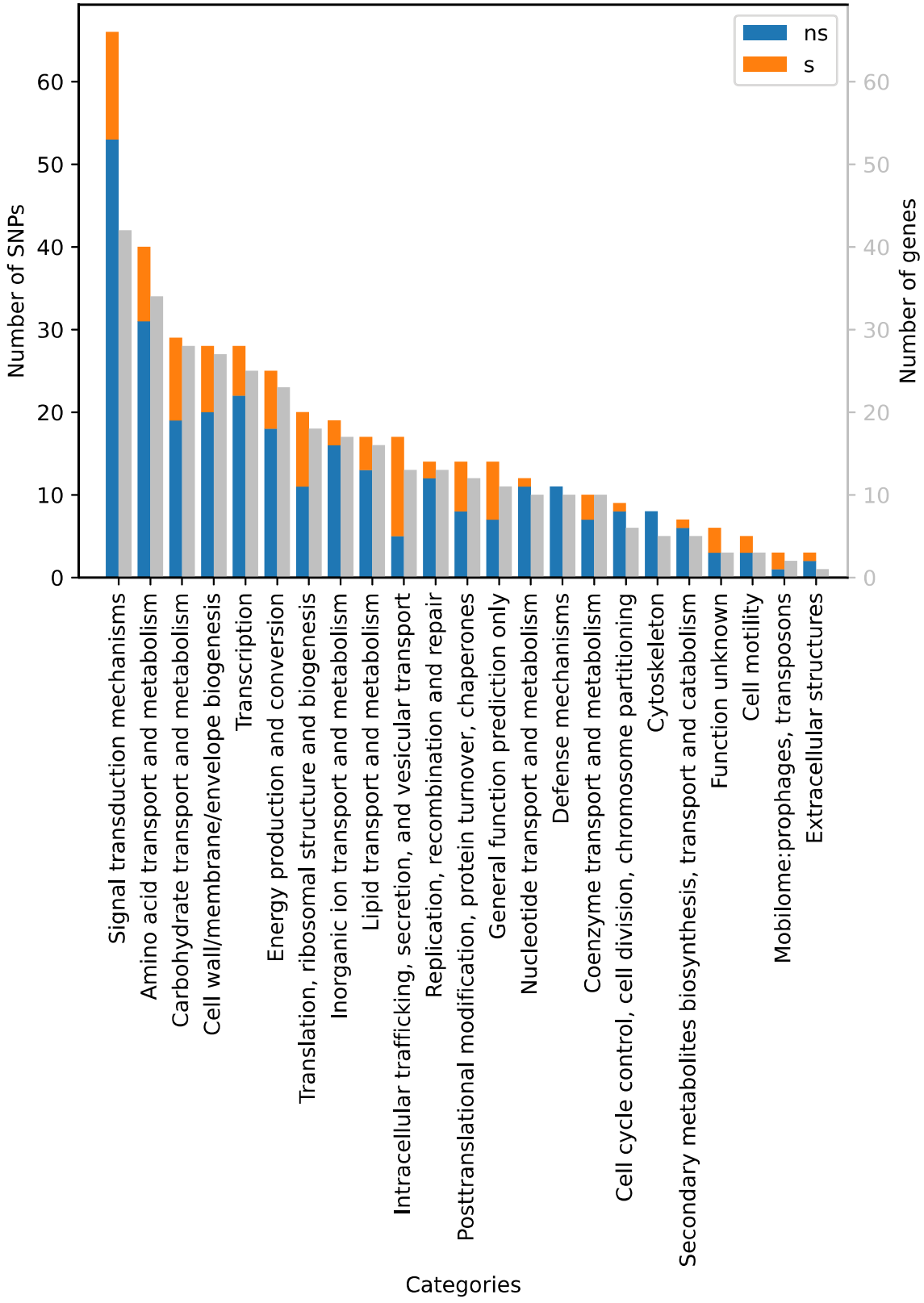
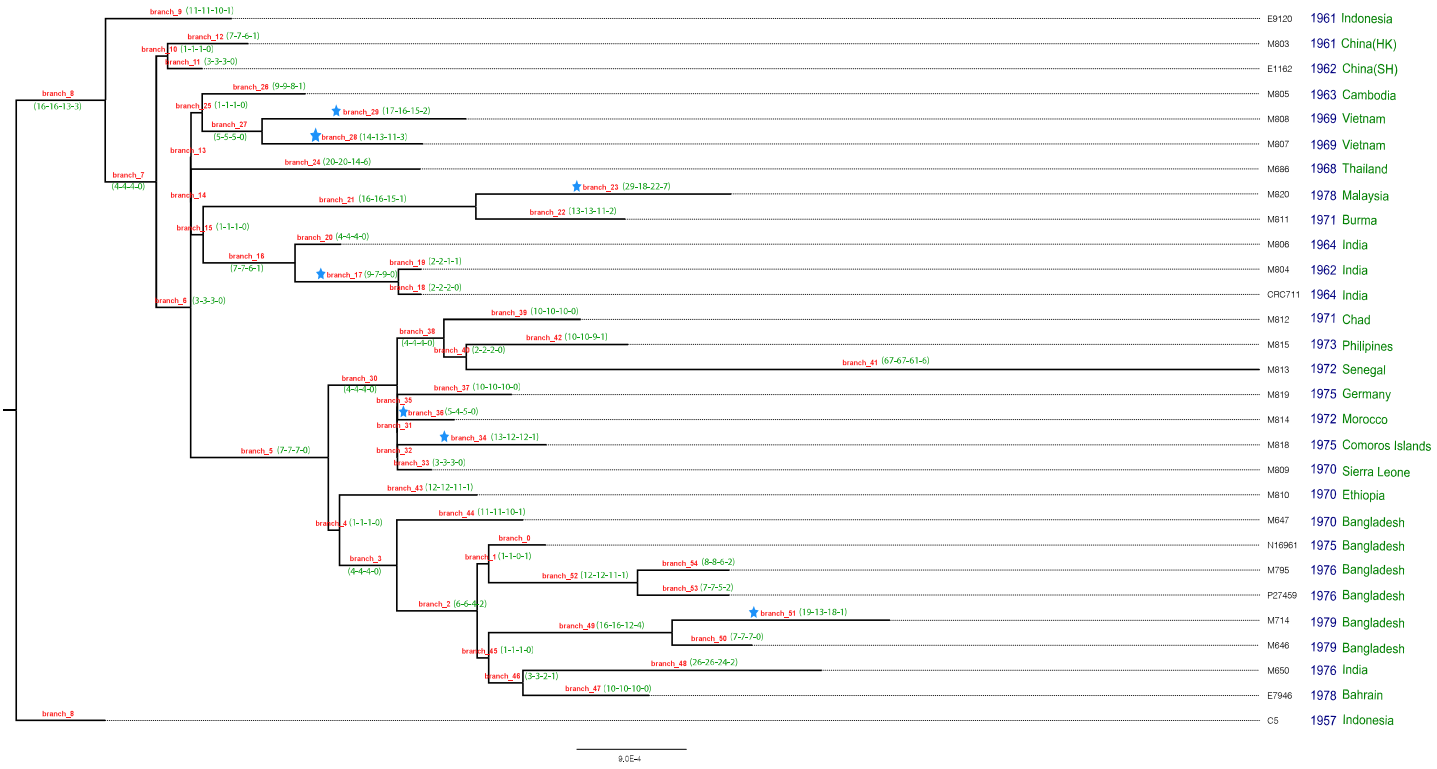


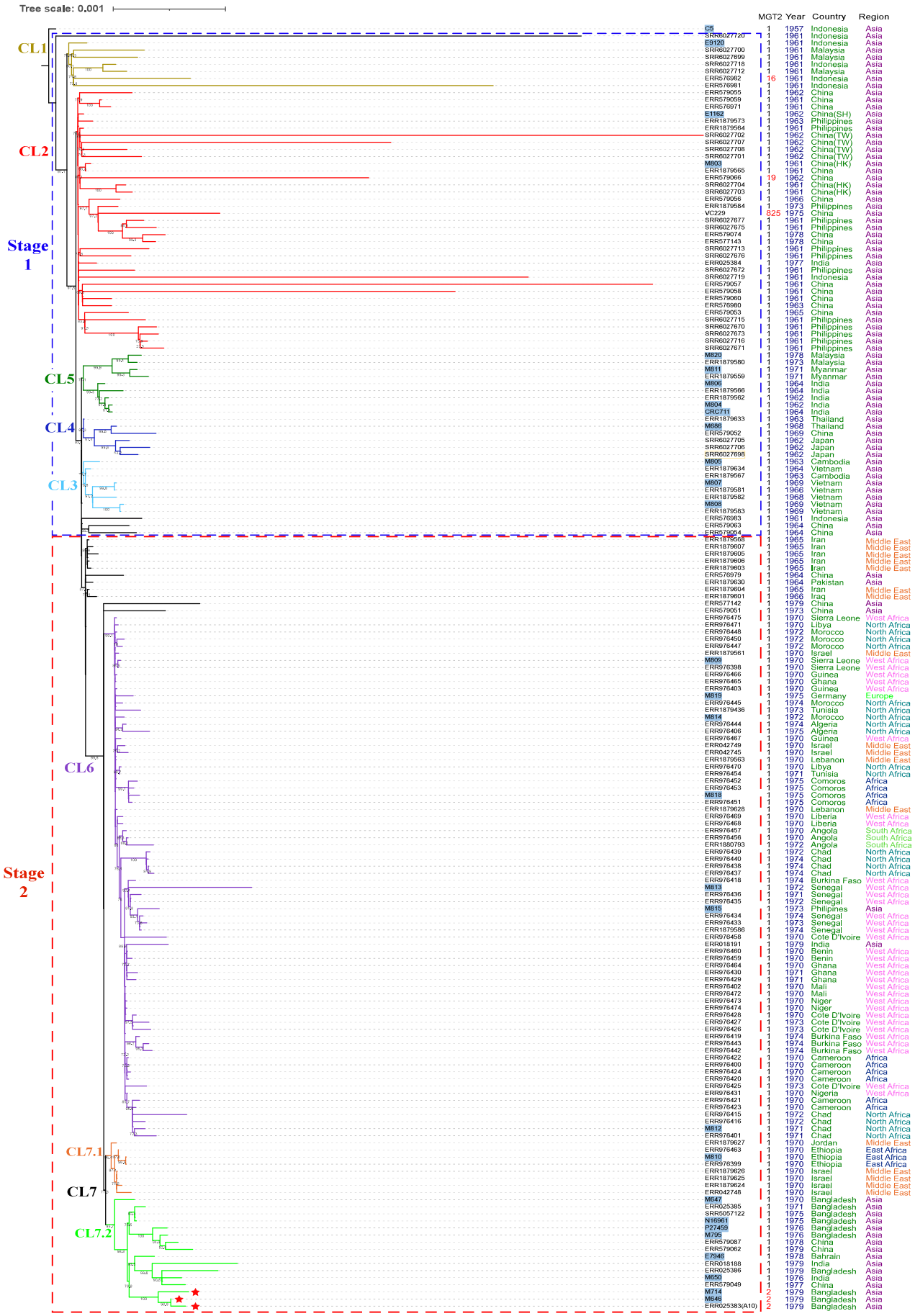
Supplementary Fig S1. Functional categories of genes containing non-synonymous mutations in the early seventh pandemic

The colour bars corresponding to the x-axis represent the number of SNPs in each category with blue and orange bars representing the number of non-synonymous SNPs (NS) and synonymous SNPs (S) respectively. The grey bar corresponding to y-axis represents the number of genes with SNPs in each category.



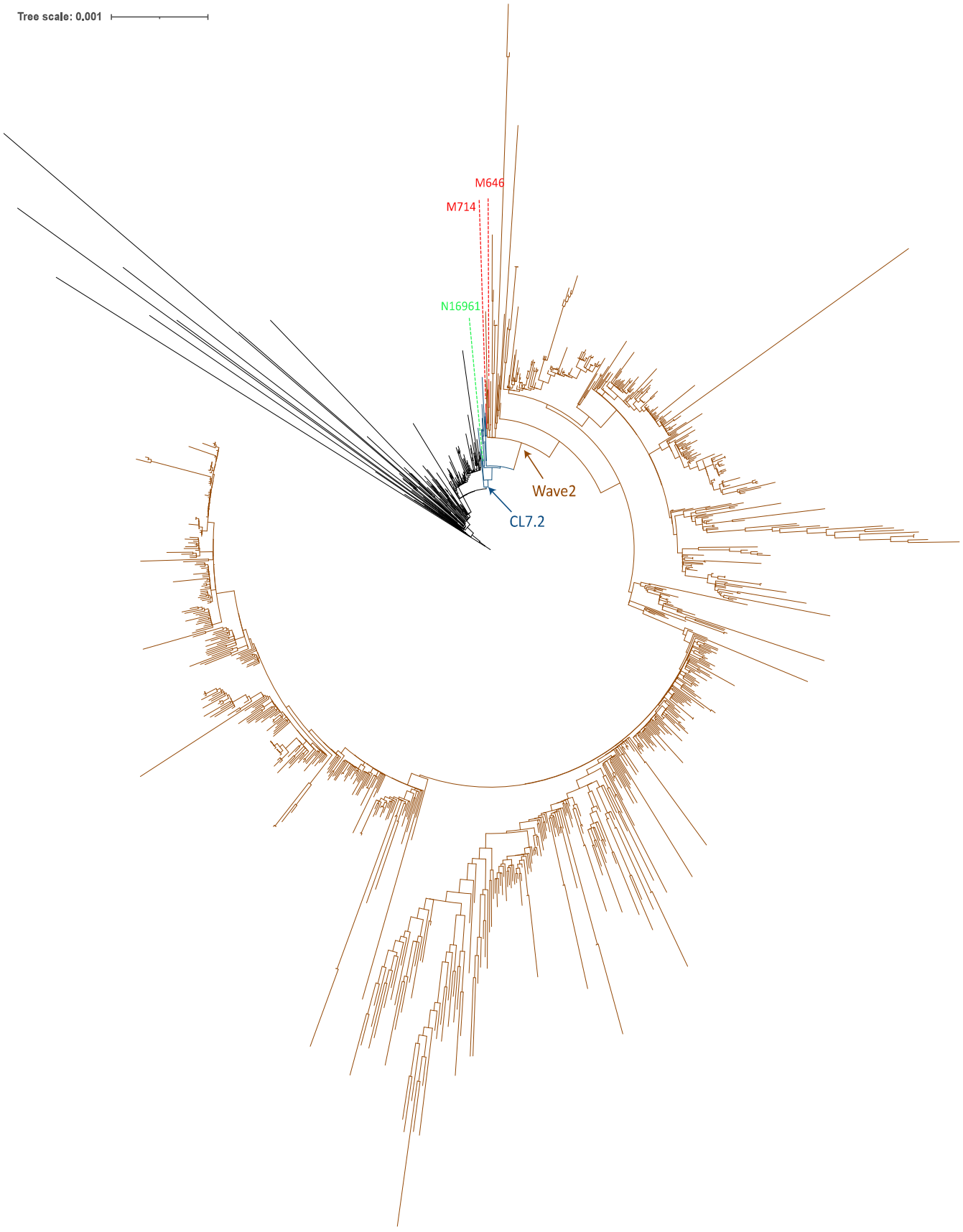
Supplementary Fig S2. Locations and annotations of SNPs on each branch

Branch ID are marked in red on top of each branch. In the brackets, the numbers delimited with hyphens represent total SNPs, total CDS, SNPs on functional genes and SNPs on intergenic regions successively. Blue stars represent nonsingleton SNPs on single gene.

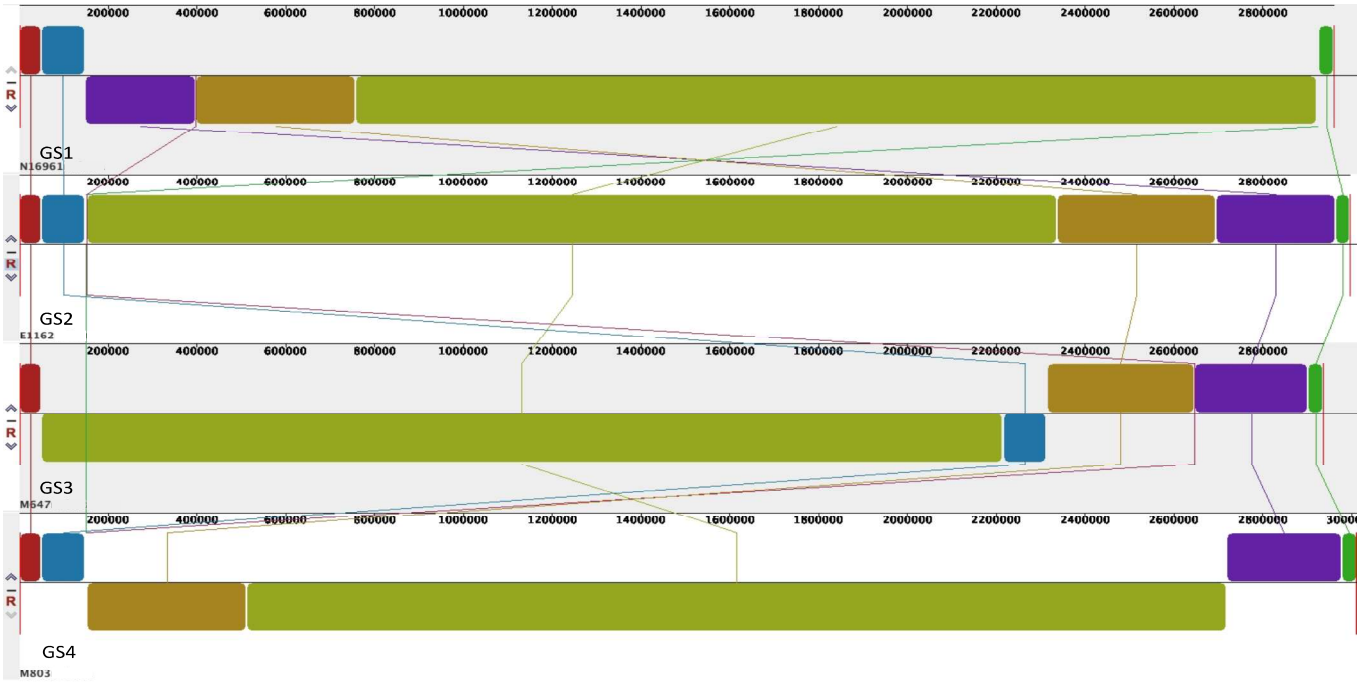


Supplementary Fig S3. Phylogenetic tree of complete genomes compared with 152 Illumina sequenced genomes.

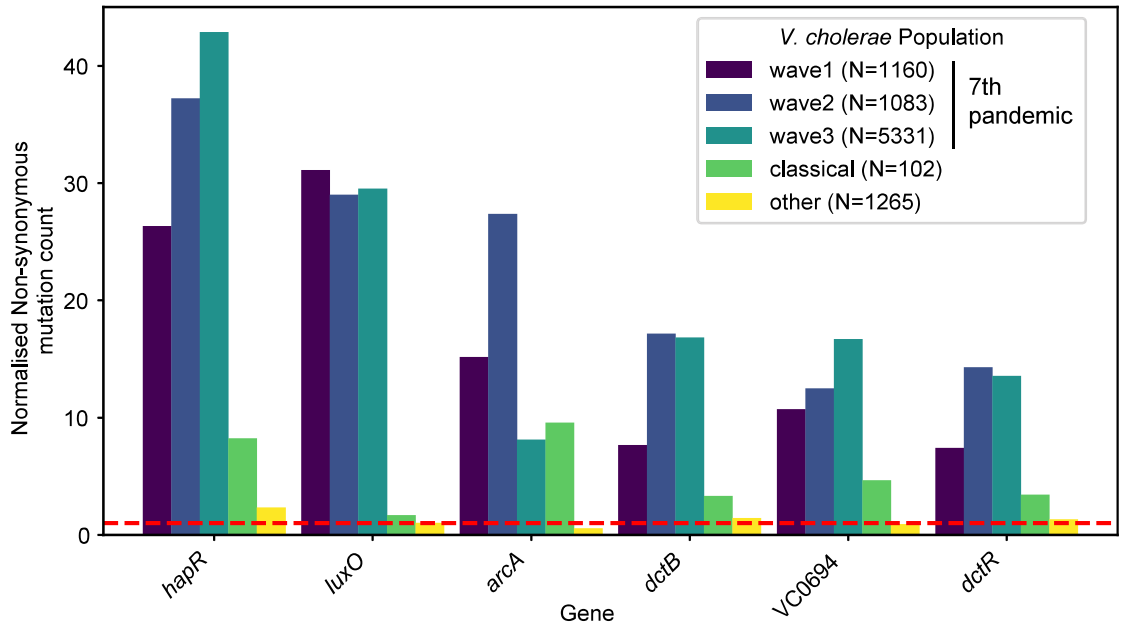
The Metadata and clusters are labeled in different colors same as Figure 1. Stage 1 and stage 2 were demarcated with blue and red dash squares. Cluster 1 (CL1) to cluster 7 (CL7) are demarcated with different color branches and labeled at each node. The red stars represent the isolates carrying IncA/C plasmids.



Supplementary Fig S4. Phylogenetic tree of complete genomes compared with wave 1 and wave 2 genomes.



Supplementary Fig S5. Representative chromosome 1 genome structures
Complete chromosomes 1 were rearranged by starting from the origin of replication (*dnaA*). Four representative genome structures (GS1, GS2, GS3 and GS4 represented by N16961, E1162, M647 and M803 respectively) were aligned by progressiveMauve. The distribution of these structures amongst the 29 complete genomes in the study are shown in Figure 1.The coloured blocks represent the homologous segments. Blocks that are shifted downward represent the inverted segments.



Supplementary Fig 6. Elevated Non-synonymous mutation rate in signal transduction genes across the 7th pandemic, classical and non 7th pandemic and non classical isolates (marked as other).

Non-synonymous mutation count was normalised in two ways, first the counts were adjusted to be per Kb to account for different gene lengths, second the mutation count for each gene in each population was divided by the average count for all genes in that population. Therefore a score of 1(marked with red dashed line) represents an average number of non-synonymous mutations for a gene in each population. Only genes that had a normalised count of more than 5 in all three 7th pandemic waves were included.