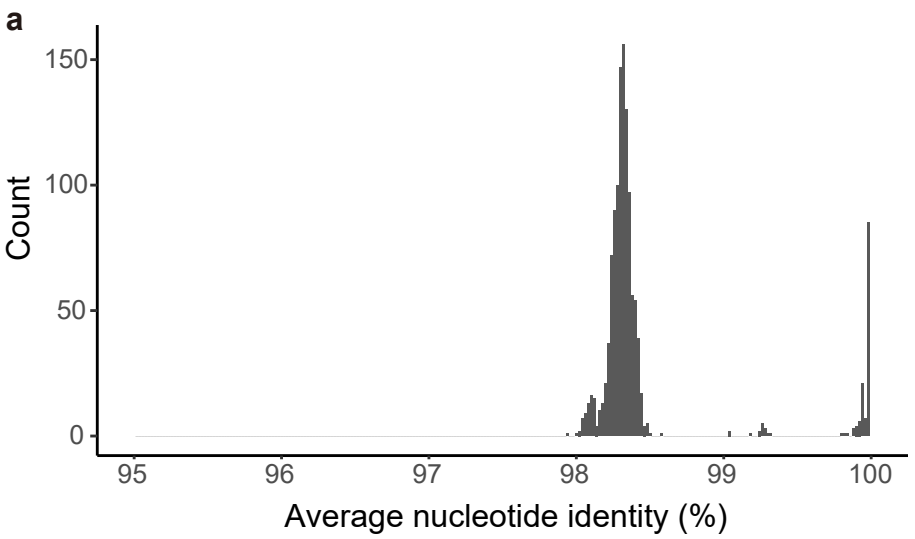
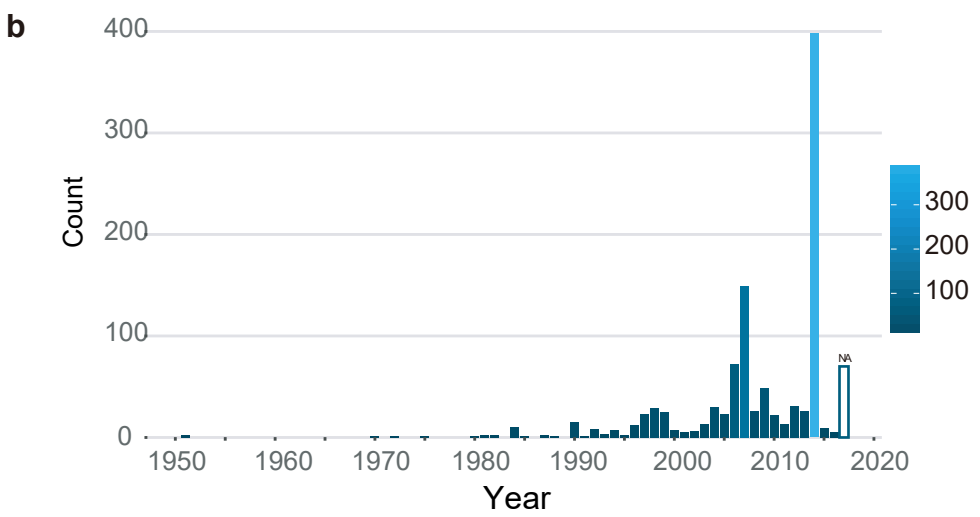
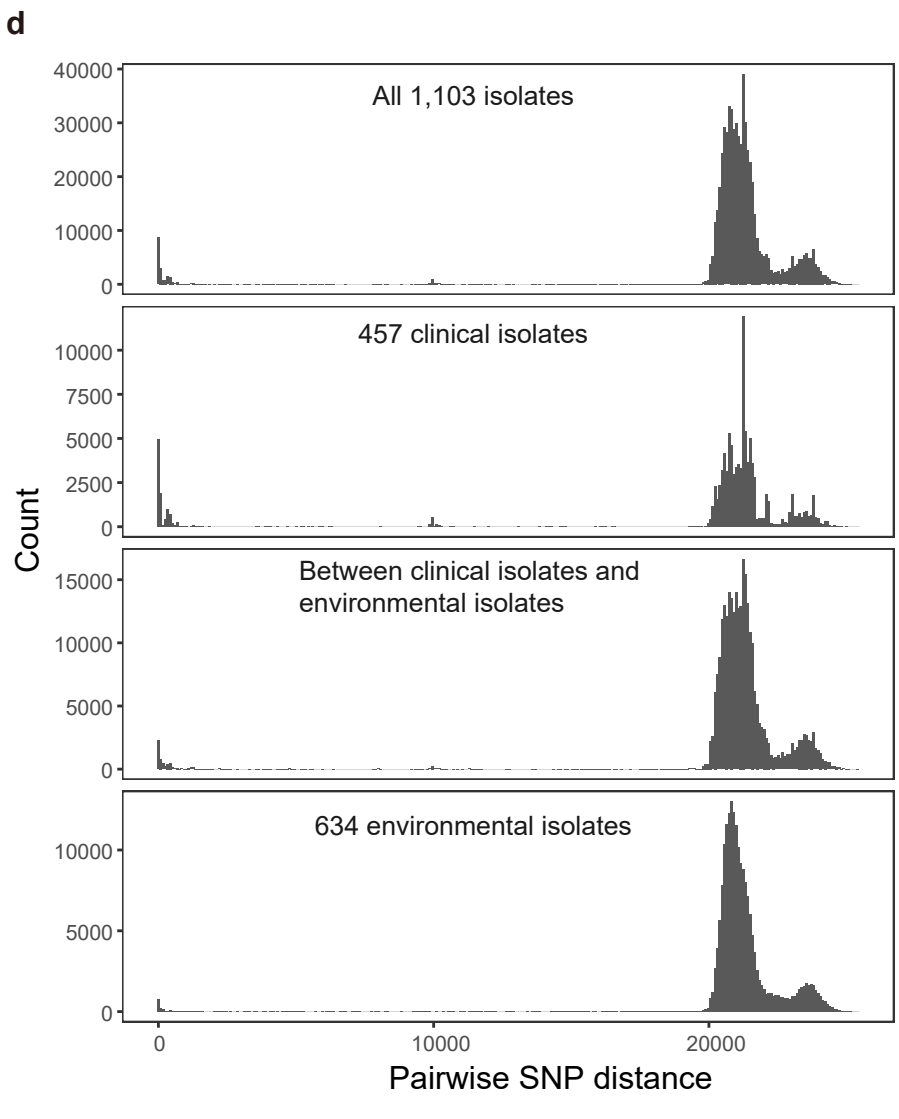
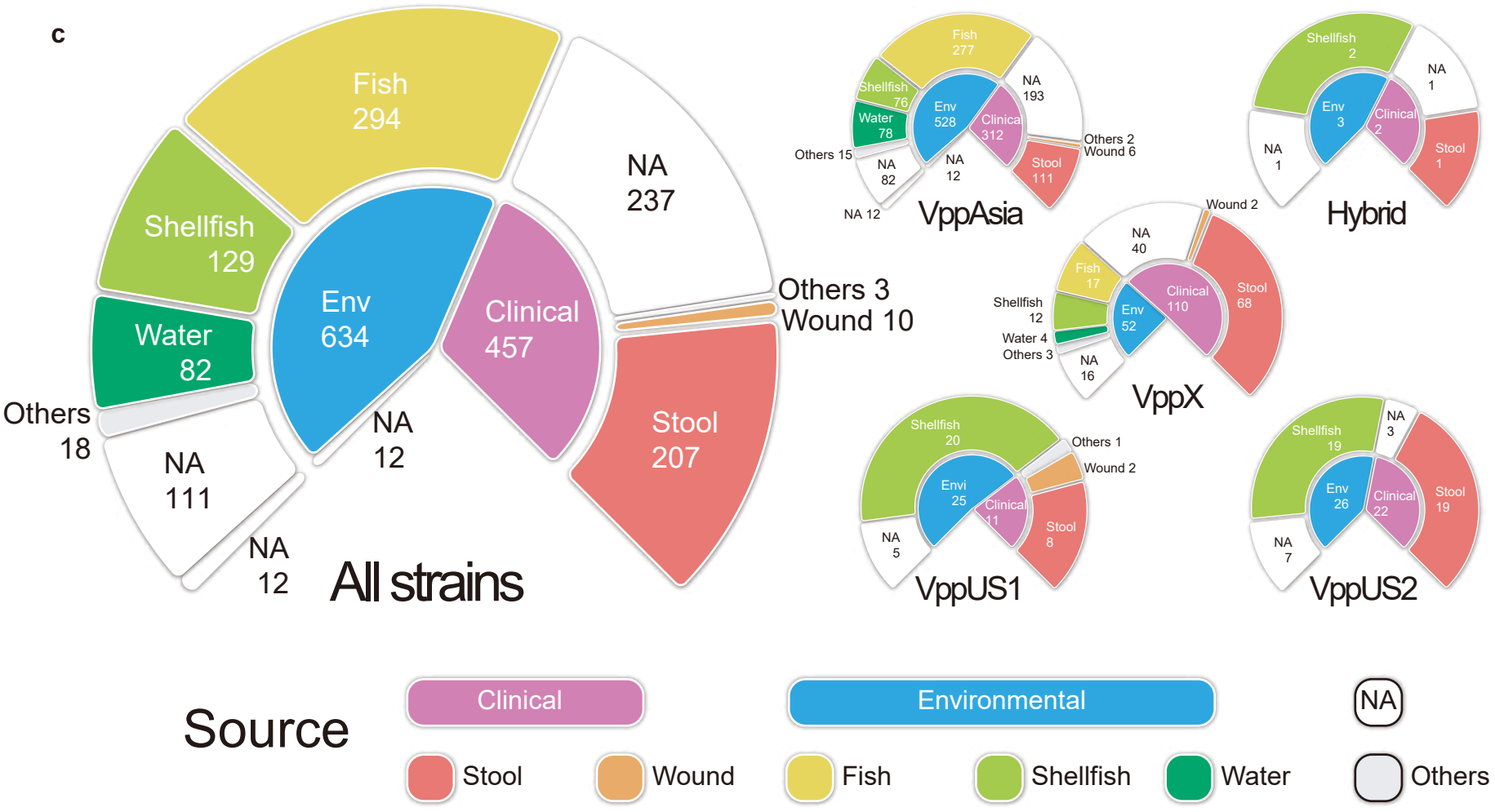
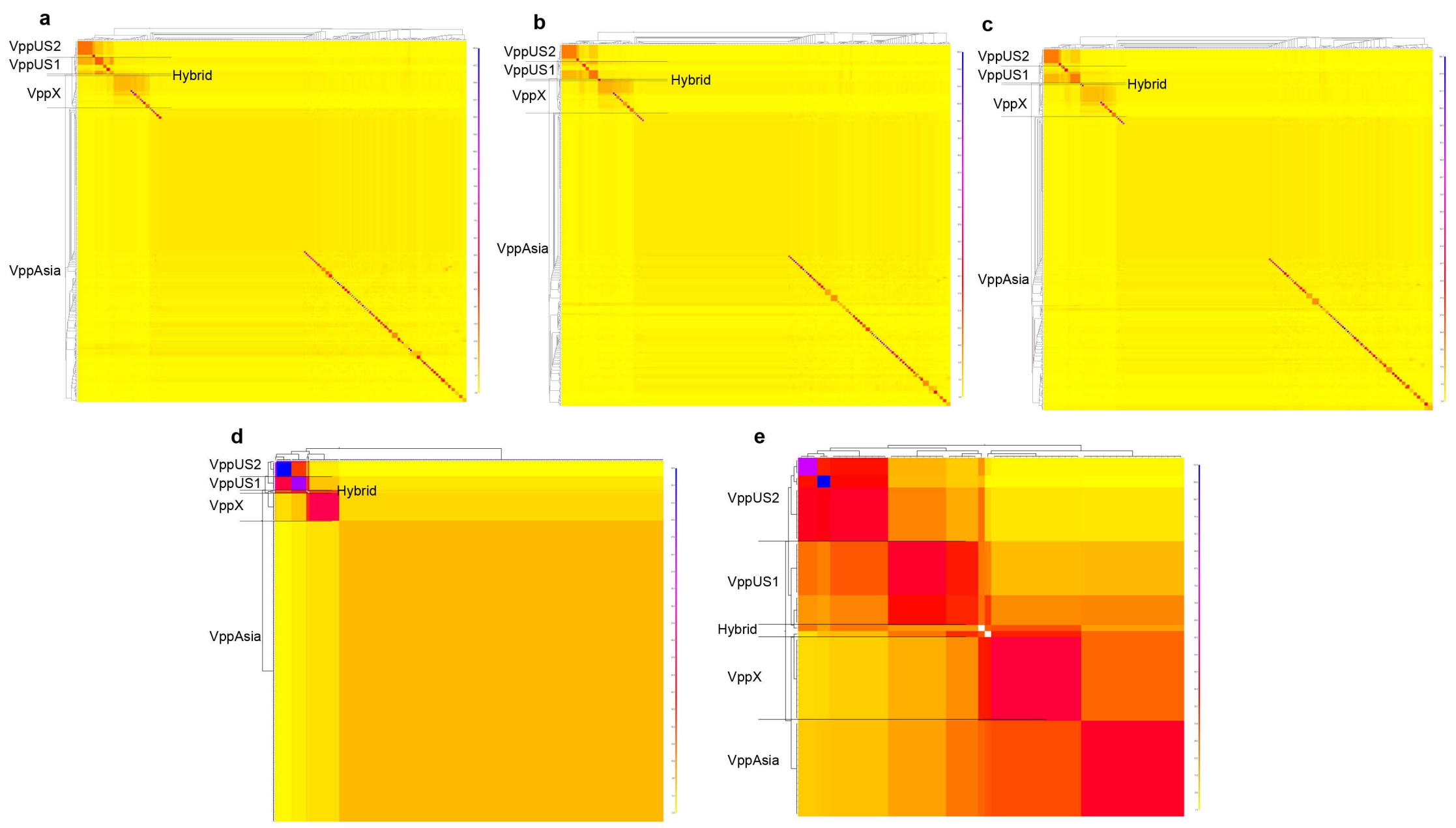
Supplementary Figure 1. Average nucleotide identity (a), isolation time (b) and source (c) of 1,103 *V. parahaemolyticus* strains and pairwise SNP distance distribution in different types of isolates (d). RIMD 2210633 was used as the reference genome in the calculation of average nucleotide identity in (a).





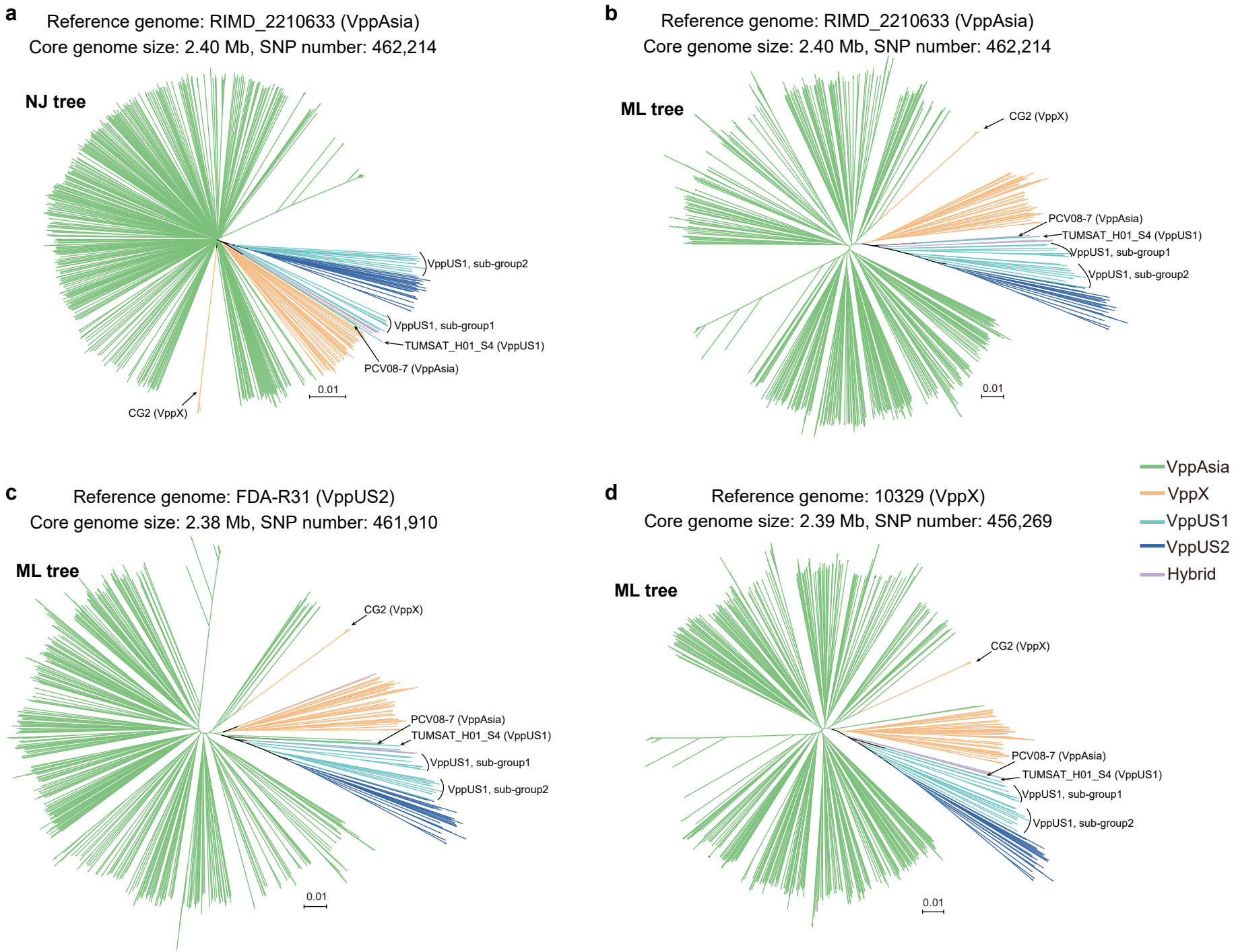




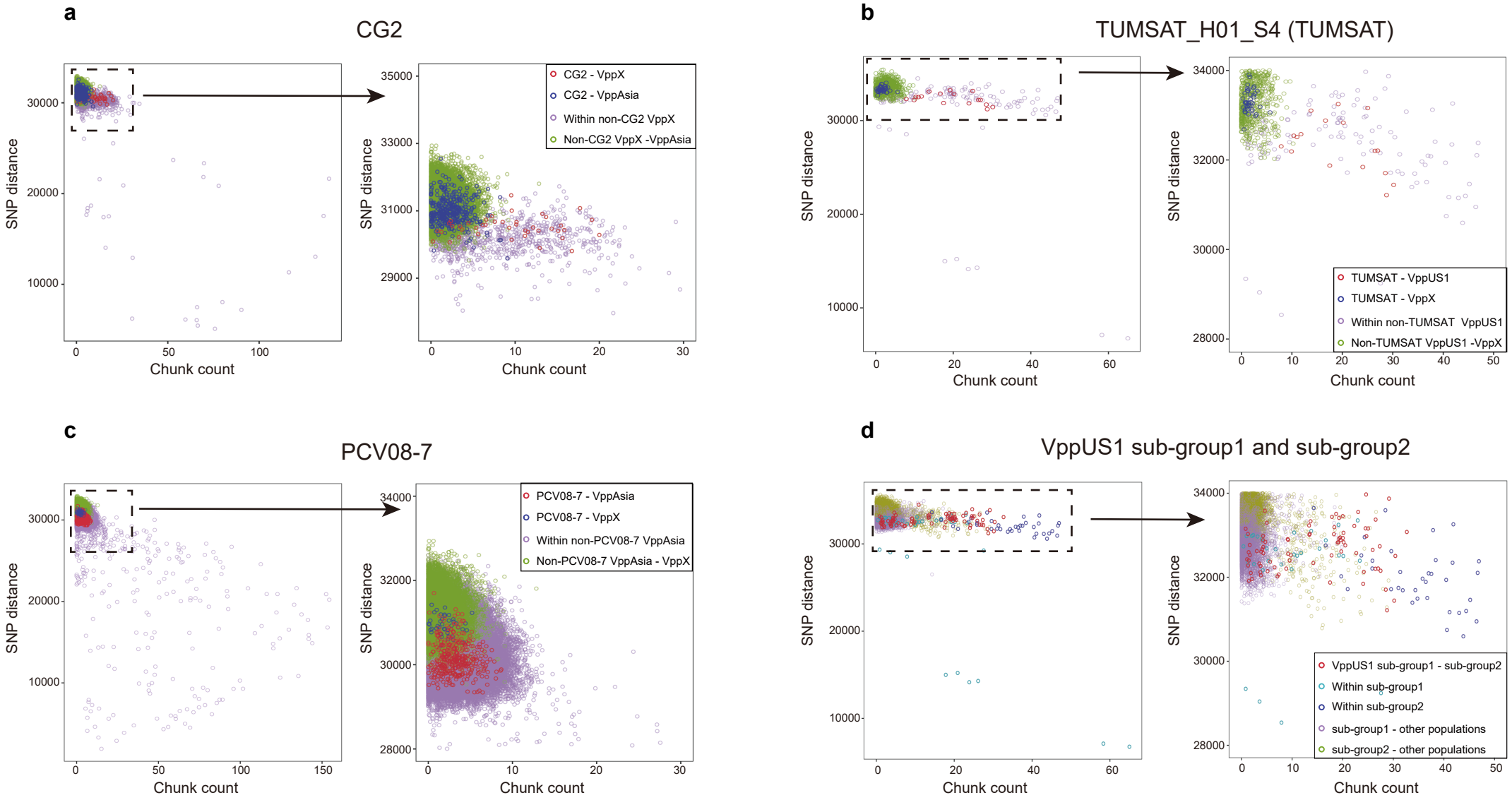


Supplementary Figure 2. The *V. parahaemolyticus* populations defined by fineSTRUCTURE analysis. (a-c) Coancestry matrixes calculated based on three independent sets of random selected 469 non-redundant strains. (d) Coancestry matrix of 260 strains after 6 more iterations of fineSTRUCTURE analysis to remove clonal signals. (e) Coancestry matrix of 60 strains with balanced sampling number from each population. The color of each cell indicates the expected chunks numbers imported from a donor (column) to a recipient (row). The boundaries between different populations are marked with lines.

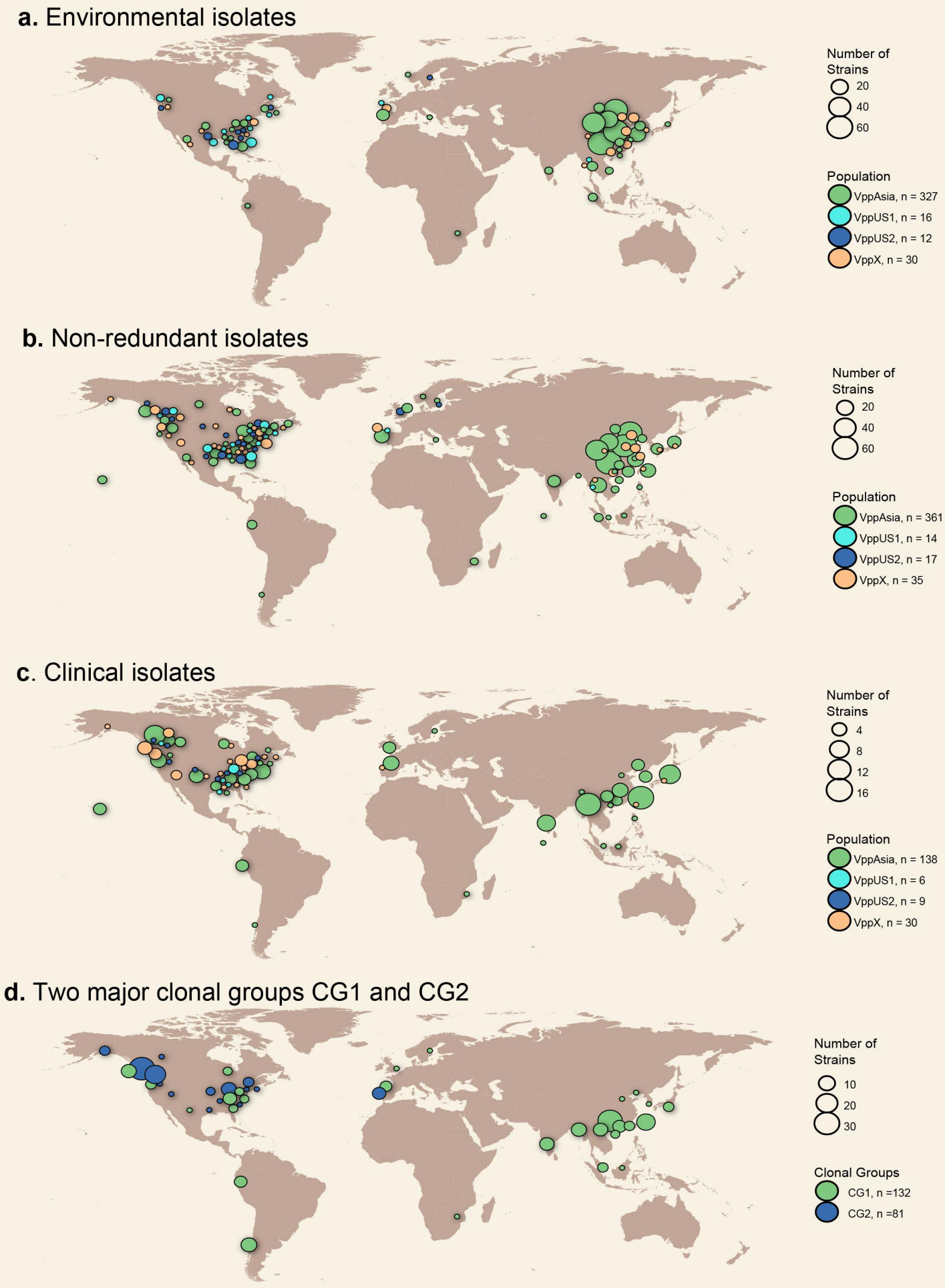
Supplementary Figure 3. Neighbor-Joining (NJ, a) and Maximum-Likelihood (ML, b-d) trees of 1,103 *V. parahaemolyticus* isolates. Branch colors indicate populations defined using fineSTRUCTURE. The inconsistent grouping results (CG2, PCV08-7, TUMSAT_H01_S4 and VppUS1) between phylogenetic tree and fineSTRUCTURE were indicated by arrows and arcs.

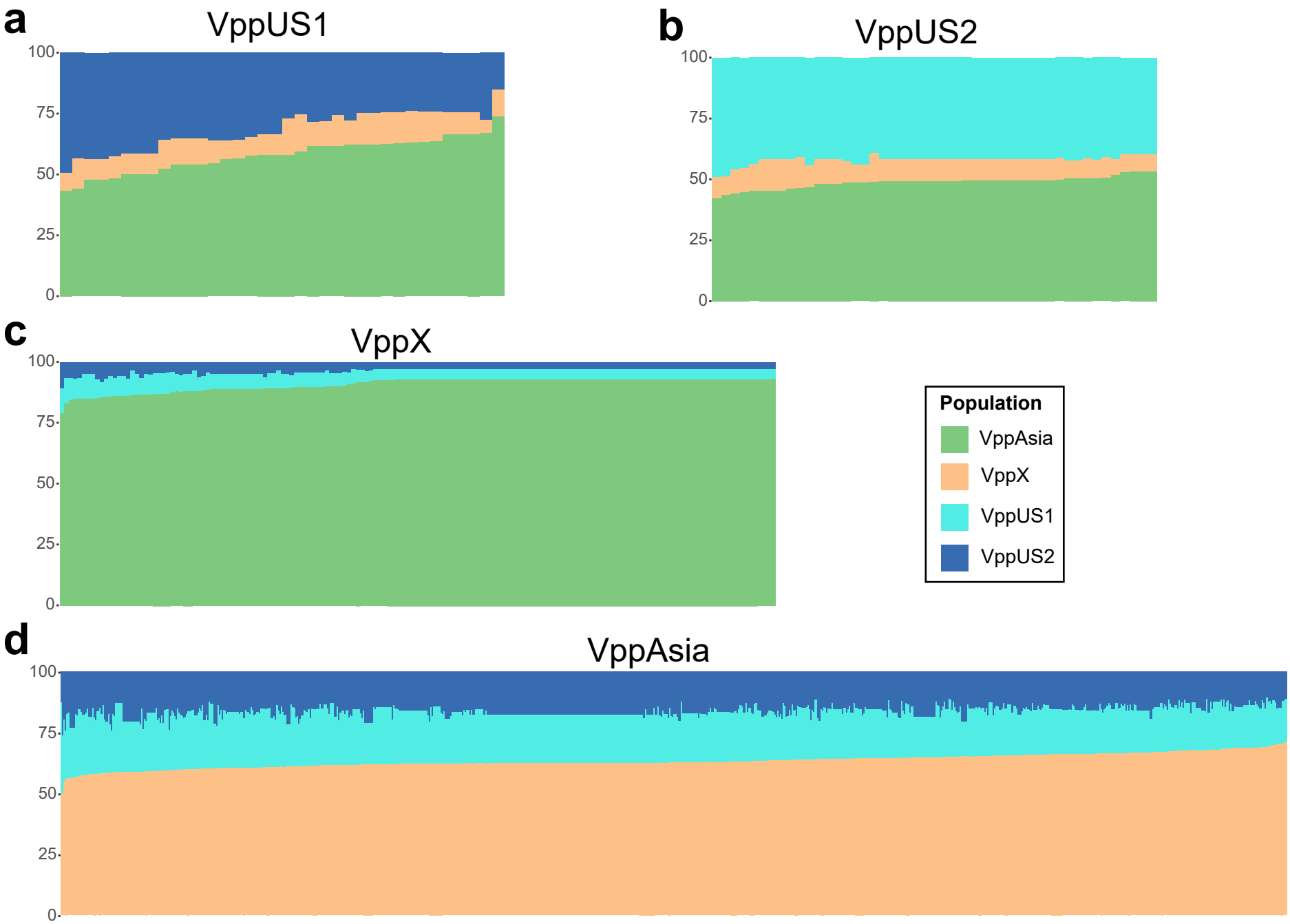


Supplementary Figure 4. The distribution of pairwise SNP distance and chunk count among isolates with inconsistent grouping results between phylogenetic method and fineSTRUCTURE. The colors of circles indicate different categories of genetic relationship as shown by legend in each panel. For each panel, the plot at the right side was the zoomed view of the dotted square at the left plot.



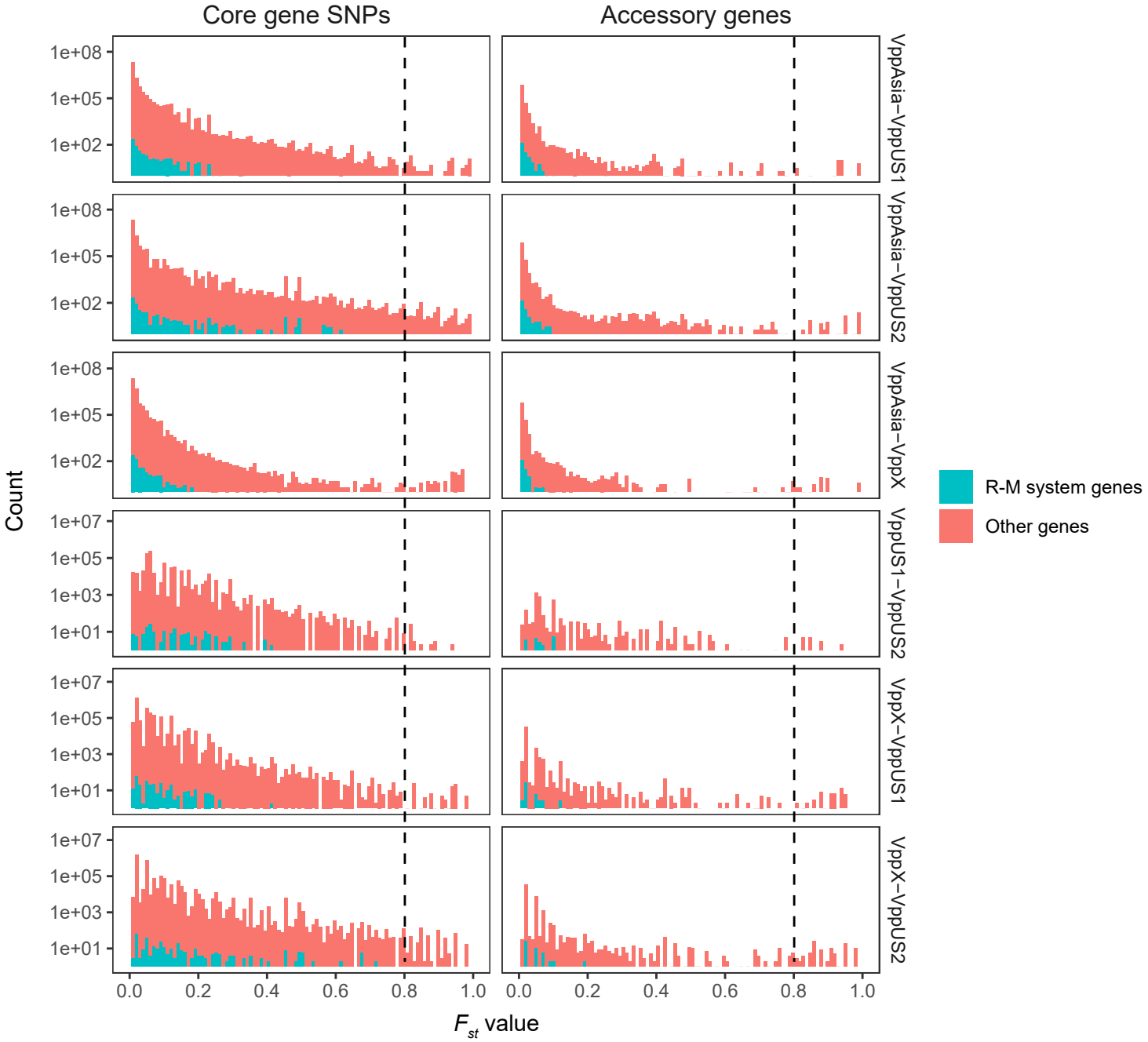
Supplementary Figure 5. Geographical distribution of environmental isolates (a), non-redundant isolates (b), clinical isolates (c) and two major clonal groups (d). Colors in circle indicate populations and are as in Figure 2. Each circle indicates the population composition of a city/country, with radius in proportion to the sample size. Clonal group strains within the same city/country were counted as one strain in panel a-c. Only strains with information of isolation location are included.



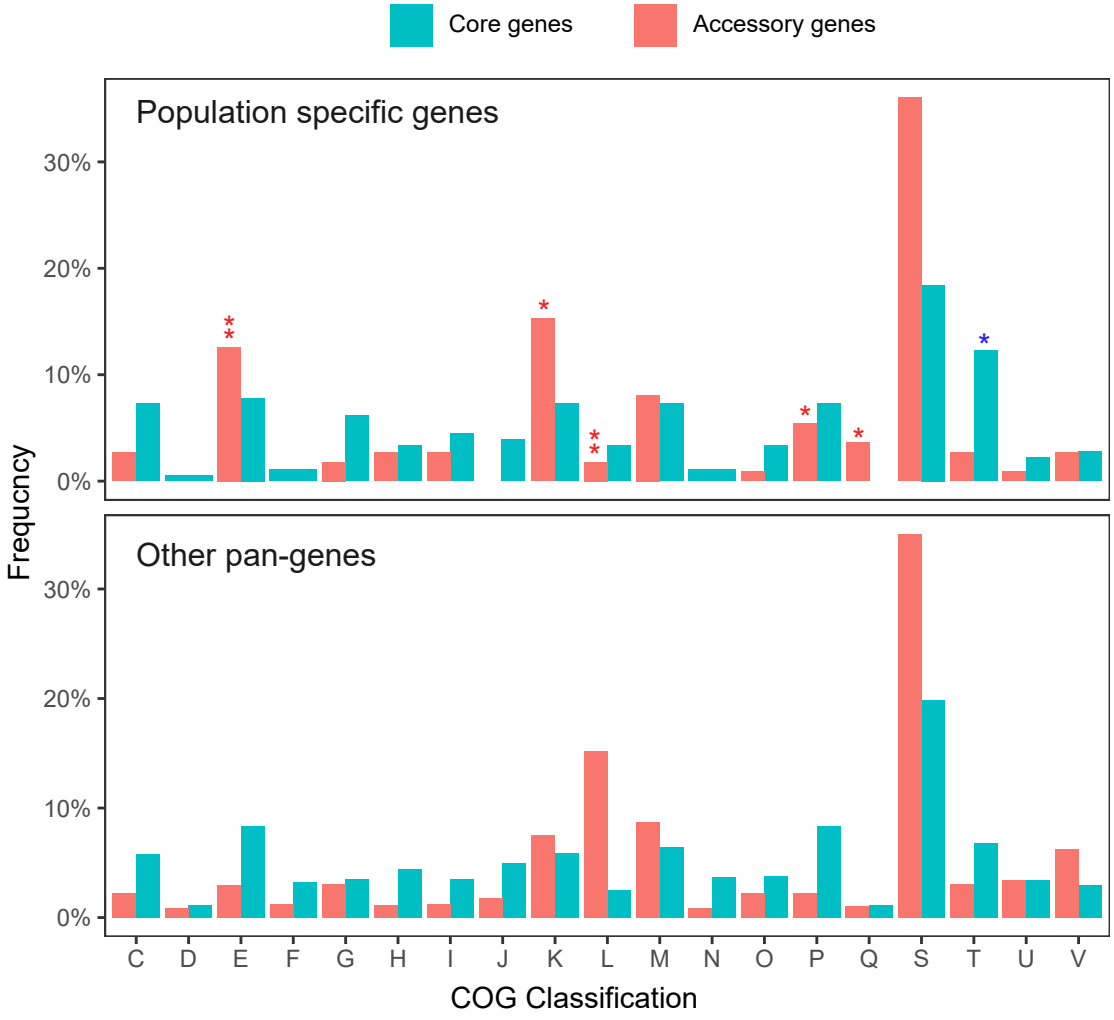


Supplementary Figure 6. The genome admixture of four *V. parahaemolyticus* populations according to chromosome painting. Chromosome painting results for all VppUS1 (a), VppUS2 (b), VppX (c) and VppAsia (d) strains inferred by using other populations as donor. Each vertical bar represents one isolate and the proportion indicates the contribution of each population.

Supplementary Figure 7. F_{st} values of core gene SNPs and accessory genes between different *V. parahaemolyticus* populations. Blue color indicates the genes encoding restriction modification system (R-M system) and red indicates other genes. The Y axis is in log-scale.Vertical dotted lines indicate the threshold (0.8) of F_{st} for defining population specific variations.



Supplementary Figure 8. COG classification of the population specific genes (F_{st} value >0.8, top) and other pan-genes (bottom). Blue color indicates core genes and red indicates accessory genes. Arrows indicate the COG classifications that showed significantly difference between population specific genes and other pan-genes. Asterisks indicate the Fisher's exact test *P* value, $^{**}P < 0.01$, $0.01 < {}^{*}P < 0.05$.



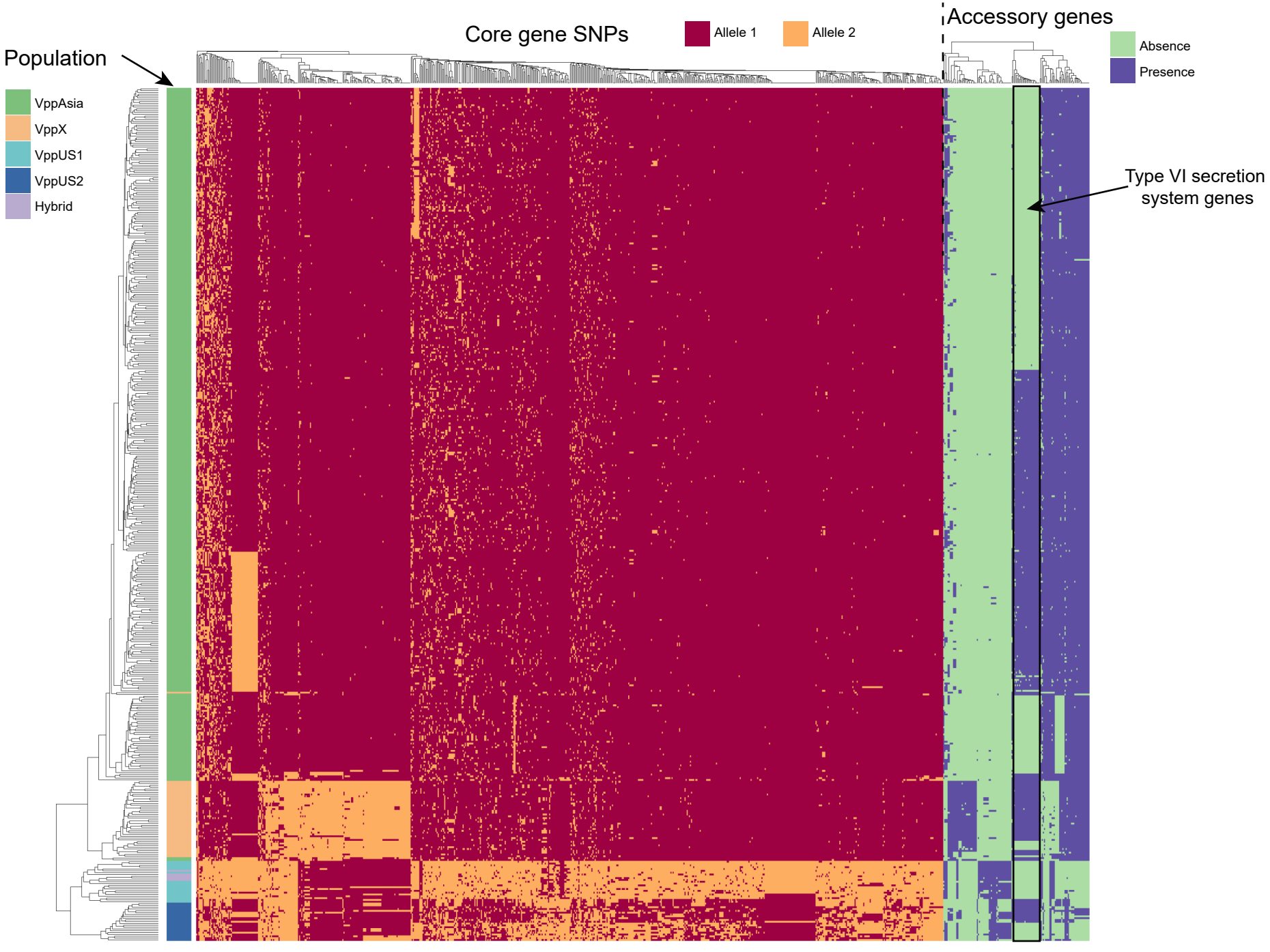
CELLULAR PROCESSES AND SIGNALING
[D] Cell cycle control, cell division, chromosome partitioning
[M] Cell wall/membrane/envelope biogenesis
[N] Cell motility
[O] Post-translational modification, protein turnover, and chaperones
[T] Signal transduction mechanisms ↗
[U] Intracellular trafficking, secretion, and vesicular transport
[V] Defense mechanisms
[W] Extracellular structures
[Y] Nuclear structure
[Z] Cytoskeleton

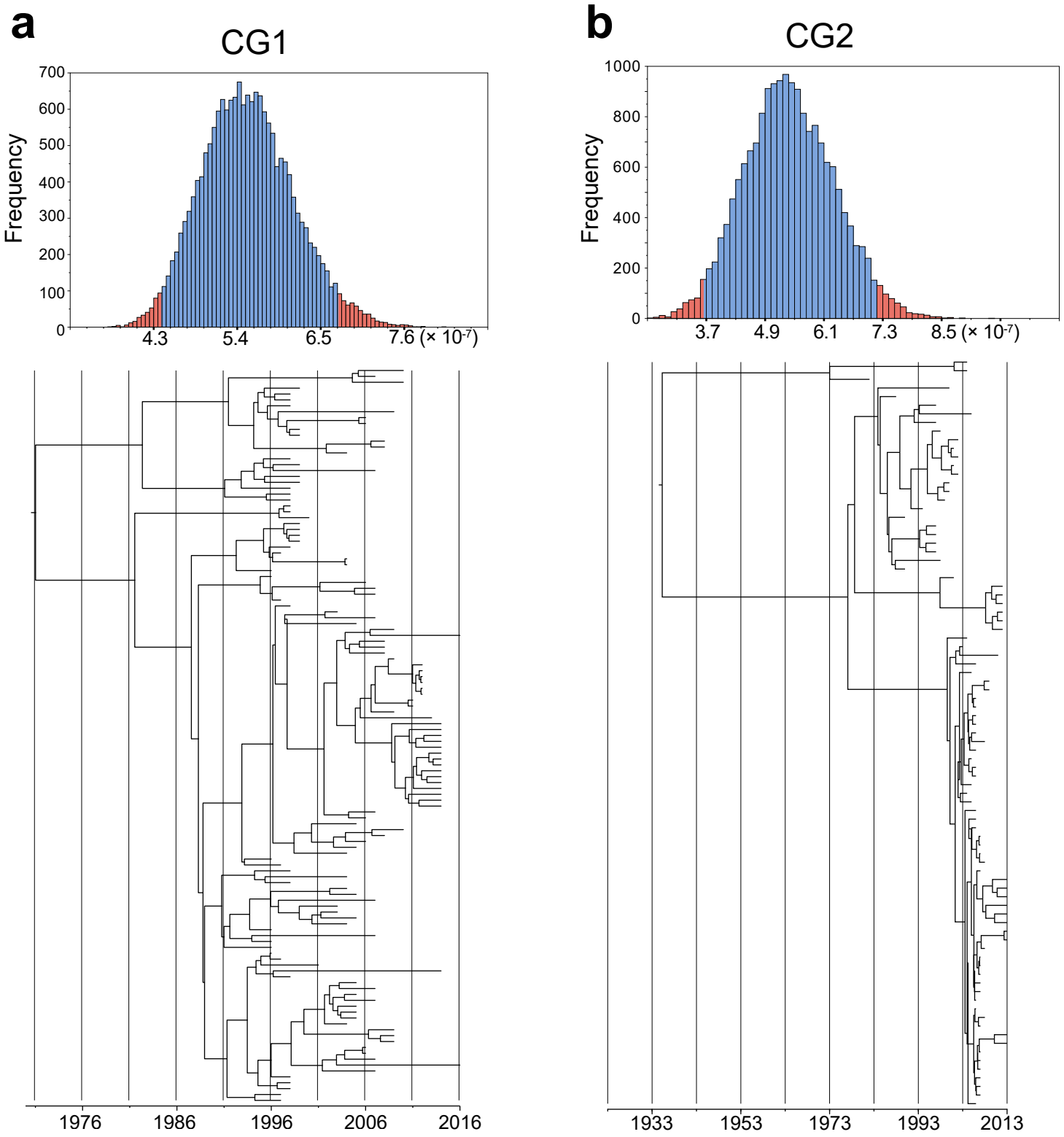
INFORMATION STORAGE AND PROCESSING
[A] RNA processing and modification
[B] Chromatin structure and dynamics
[J] Translation, ribosomal structure and biogenesis
[K] Transcription ↗
[L] Replication, recombination and repair ↘

METABOLISM
[C] Energy production and conversion
[E] Amino acid transport and metabolism ↗
[F] Nucleotide transport and metabolism
[G] Carbohydrate transport and metabolism
[H] Coenzyme transport and metabolism
[I] Lipid transport and metabolism
[P] Inorganic ion transport and metabolism ↗
[Q] Secondary metabolites biosynthesis, transport, and catabolism ↗

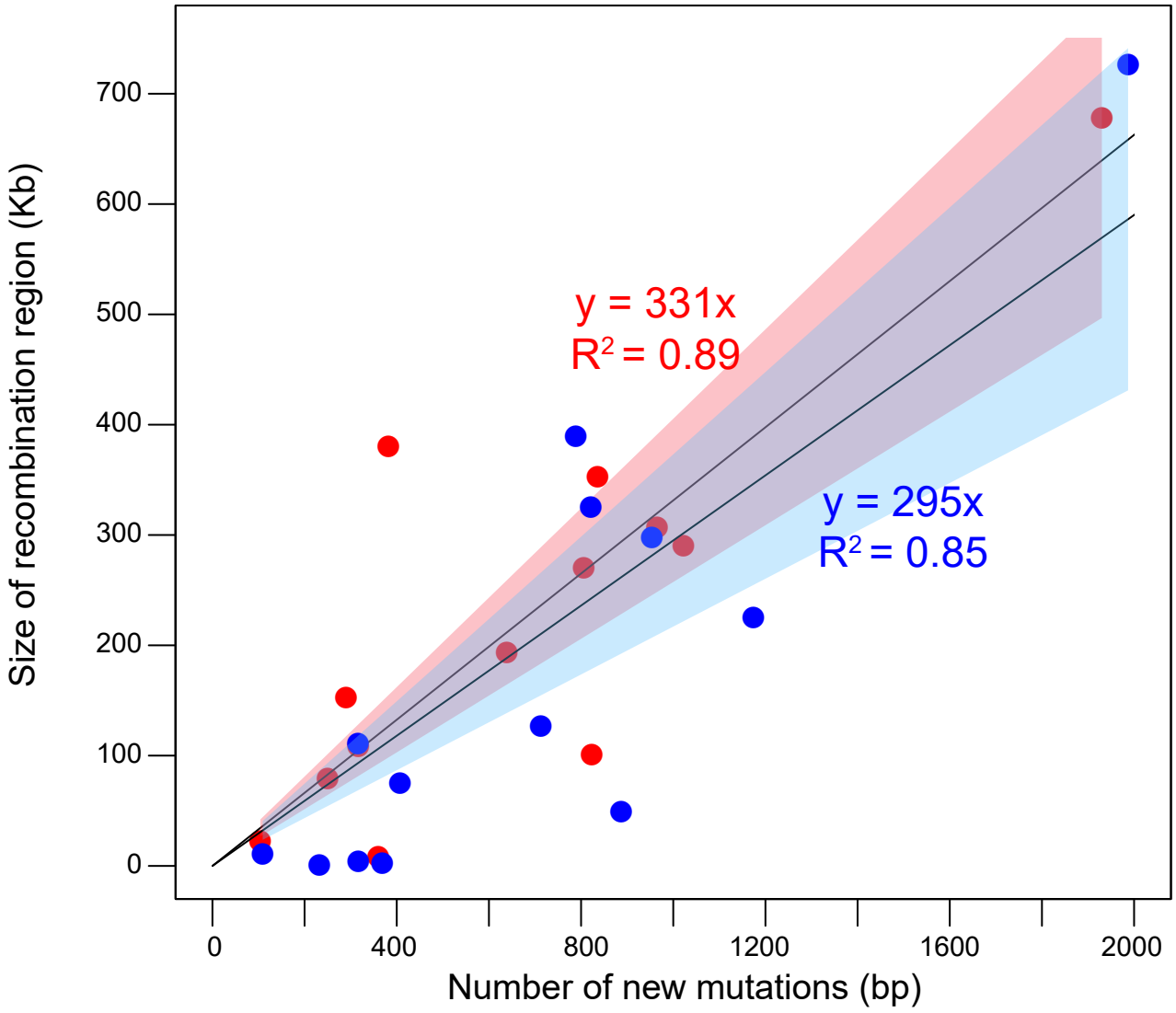
POORLY CHARACTERIZED
[R] General function prediction only
[S] Function unknown

Supplementary Figure 9. Hierarchical clustering of 1,103 *V. parahaemolyticus* strains (rows) based on population specific SNPs and genes (columns, F_{st} value > 0.8).The bar at the left of heatmap indicates the populations of *V. parahaemolyticus*. In the heatmap, the color rose and orange indicate alleles of SNPs, and blue/green are presence/absence of accessory genes. The distribution of type VI secretion system genes was highlighted with a black rectangle.

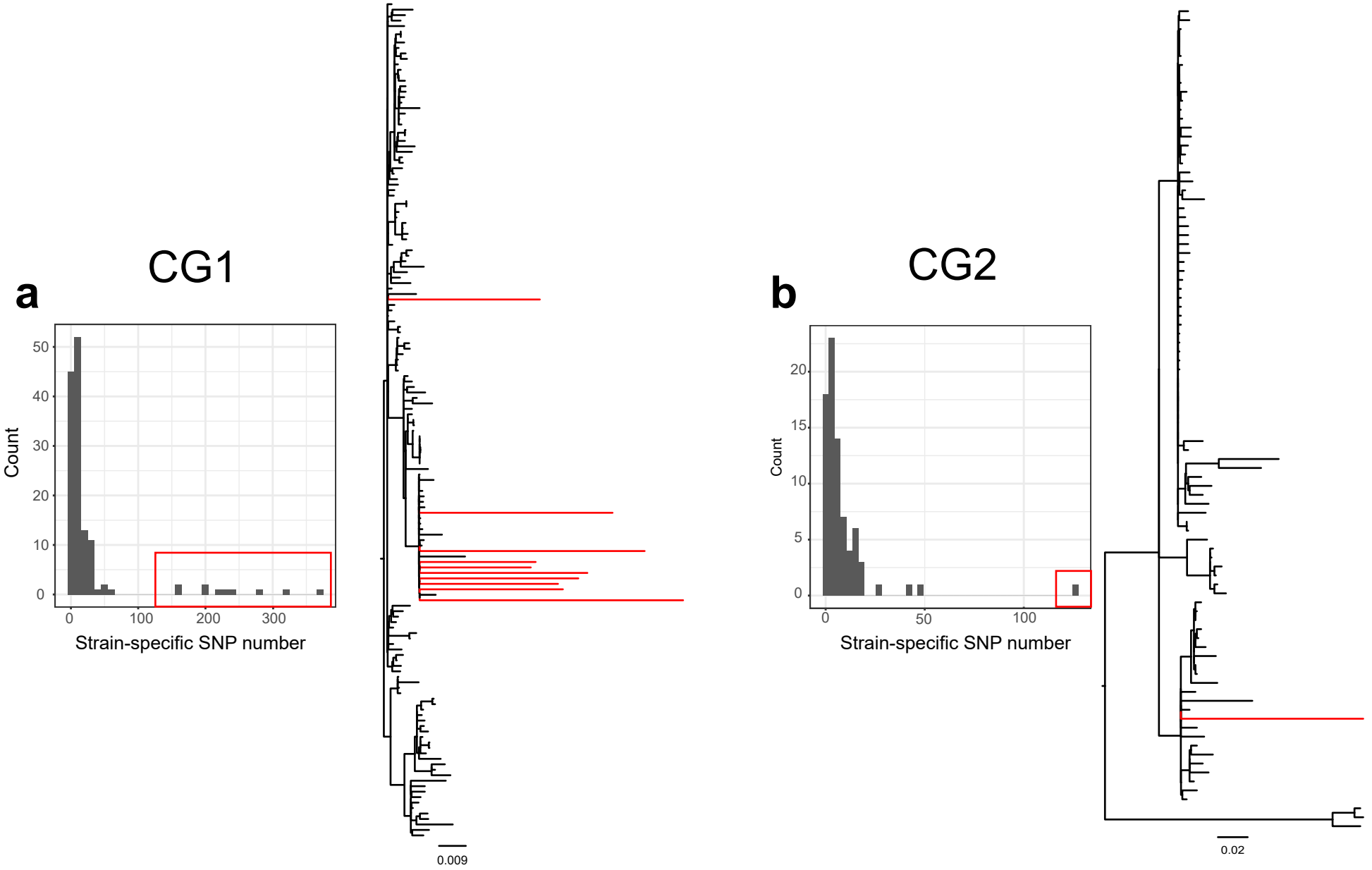




Supplementary Figure 10. Clock rates of two major clonal groups CG1 (a) and CG2 (b) inferred by BEAST. Top: the posterior probability density distribution of clock rate. Bottom: maximum clade credibility tree.



Supplementary Figure 11. Correlation between total size of recombination regions and amount of new mutation sites within 13 clonal groups. The red color indicates results inferred by our in-house pipeline, blue color indicates results inferred by ClonalFrameML. Each dot indicated one clonal lineage. The linear regression was constrained to go through the origin and the light red and blue shading indicates the 95% confidence interval (CI) of the slope, [257-405] for red and [217-373] for blue.



Supplementary Figure 12. *V. parahaemolyticus* strains contained unusual high number of strain-specific variations. The NJ trees of CG1 (a) and CG2 (b) were built based on SNPs after removing recombination sites. Both the long branches in NJ trees and outliers in the histograms indicated some *V. parahaemolyticus* strains contained unusual high number of strain-specific SNPs. Strains marked in red in histograms/NJ trees are excluded from BEAST analysis.