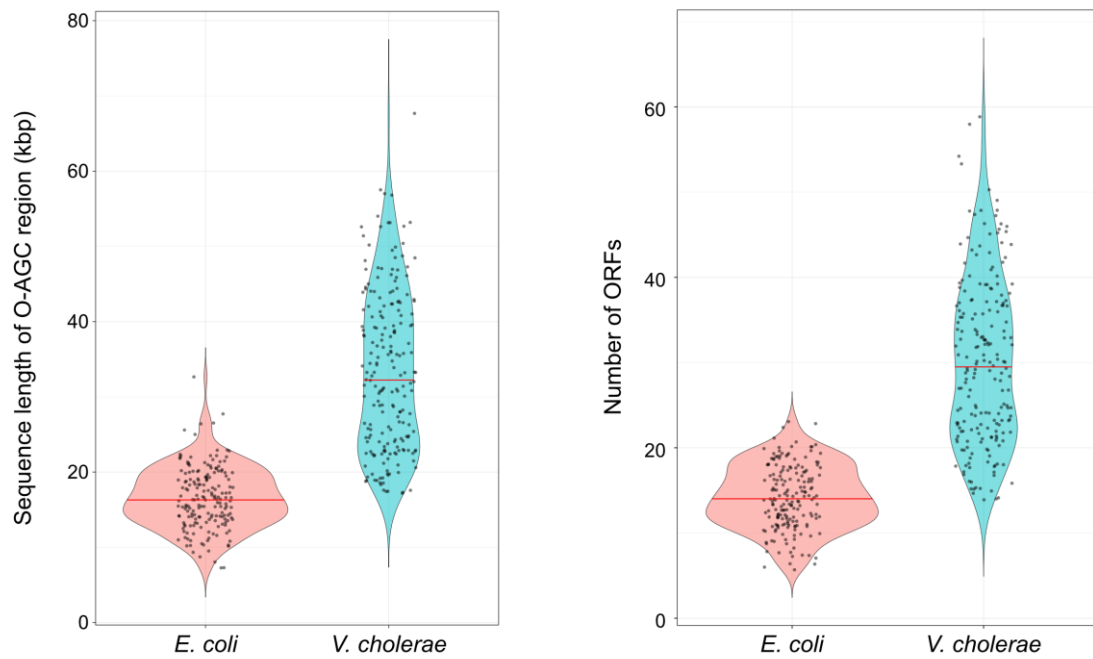


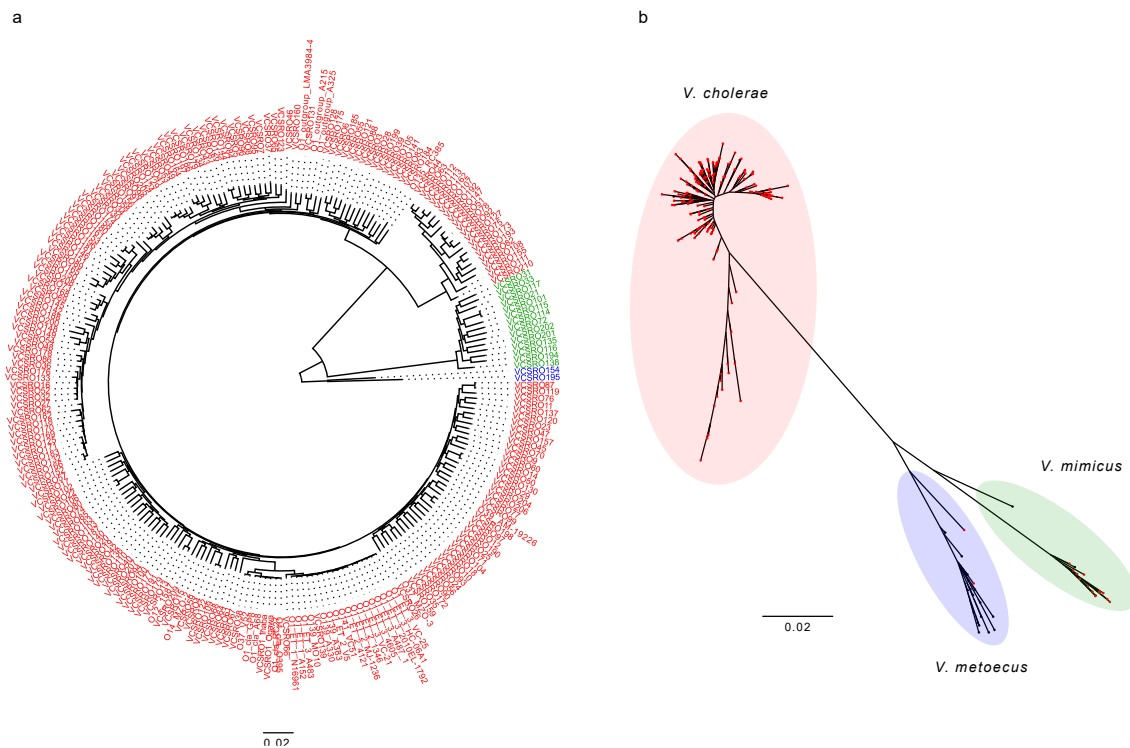
**Fig. S1. Genetic structure of O-antigen biosynthesis gene clusters of all 206 *V. cholerae* complex O-serogroup reference strains.**

Arrows indicate direction of transcription and relative size of genes and are color-coded according to their functions.



**Fig. S2. Comparison of O-AGC region lengths and the number of constitutive genes in *V. cholerae* and *E. coli*.**

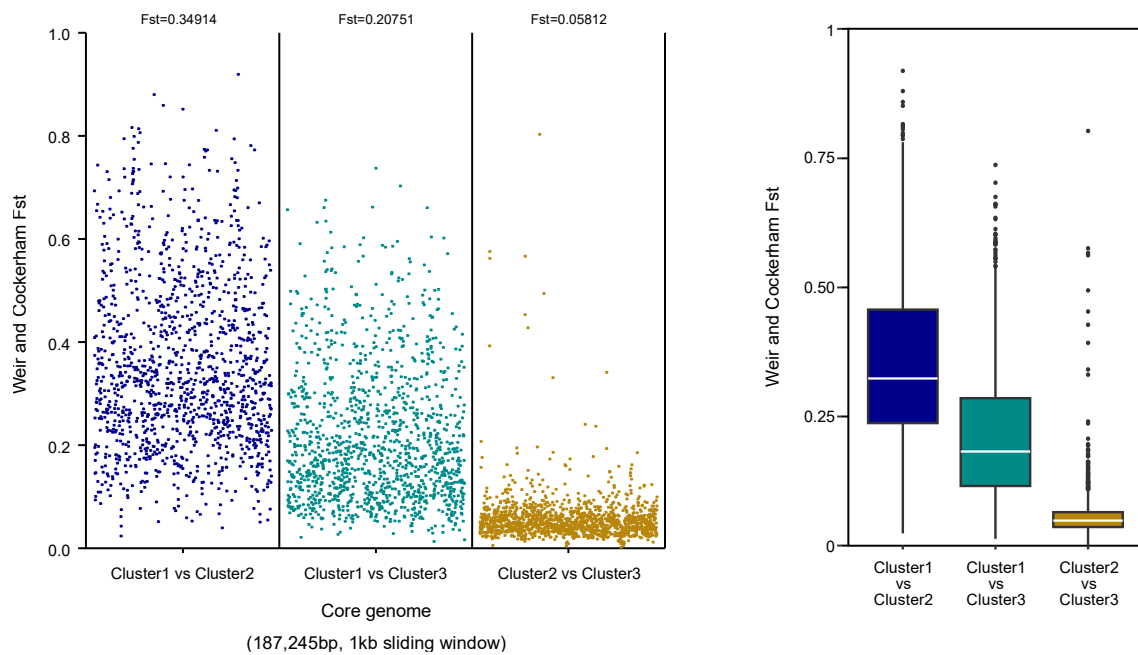
O-AGC region lengths and the number of constitutive genes in *V. cholerae* (n=206) and *E. coli* (n=184) were calculated, and the distribution of data was visualized by violin plot, using ggplot2 packages in R. Red line indicates the median value.



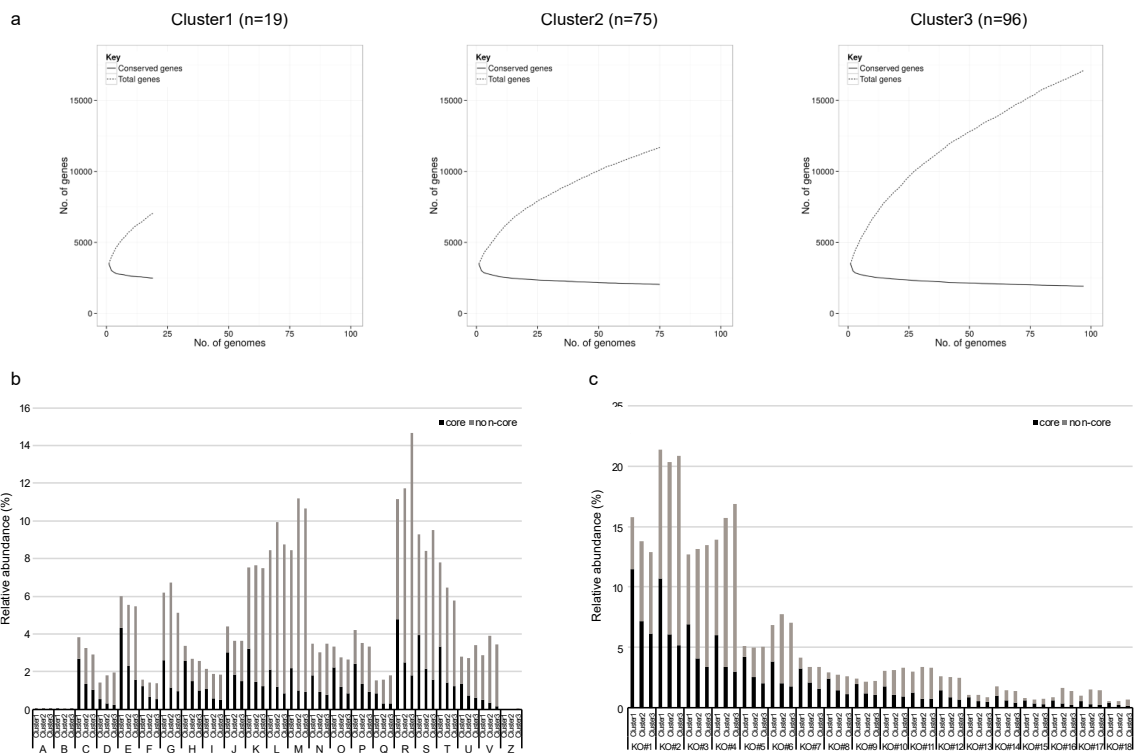
**Fig. S3. Phylogenetic relation of all 206 *V. cholerae* complex O-serogroup reference strains.**

(a), The 236 strains including 30 publicly available *V. cholerae* genome were further analyzed by kSNP3 with optimal k-mer size (k=19) and shown in the reconstructed maximum parsimony phylogenetic tree.

(b), Species classification of 206 *V. cholerae* complex O-serogroups reference strains (red dots) were confirmed by MLST analysis with six house-keeping genes (*adh*, *gyrB*, *mdh*, *pgi*, *recA*, and *rpoA*) using 51 publicly available sequence information (black dots; 30 *V. cholerae*, 2 *V. mimicus*, and 19 *V. metoecus*). Relationships between three *Vibrio* species (*V. cholerae*, *V. mimicus*, and *V. metoecus*) were shown in maximum-likelihood phylogenetic tree based on SNPs of above six genes. 51 publicly available sequence information were listed in Table S2.



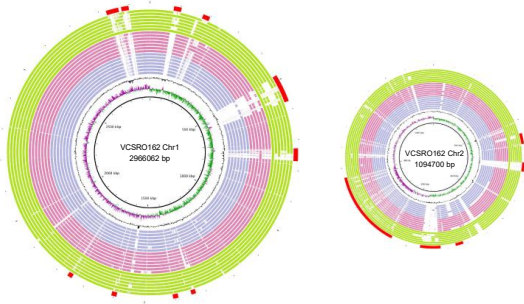
**Fig. S4. Genetic differentiation between three phylogenetic clusters in *V. cholerae* population.** The Weir and Cockerham's fixation index (Fst) value for SNPs on core genome sequences between two clusters was calculated by VCFtools (v0.1.17) with haploid mode and 1 kb sliding window (--haploid --fst-window-size 1000). Fst value between phylogenetic clusters was visualized by manhattan plot (left) and boxplot (right) using R with ggplot2 package.



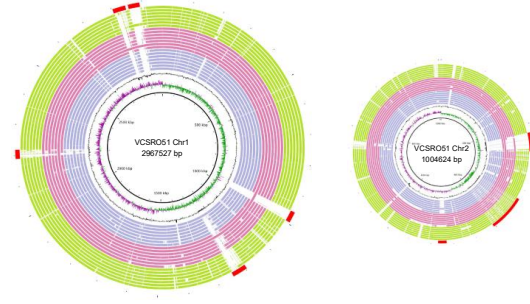
**Fig. S5. Pan-genome profile of three different *V. cholerae* clusters.**

(a) Pan-genome curve plotted in each *V. cholerae* cluster. (b) Distribution of COG functional categories in each *V. cholerae* cluster (left) and distribution of KEGG classifications in each *V. cholerae* cluster (right).

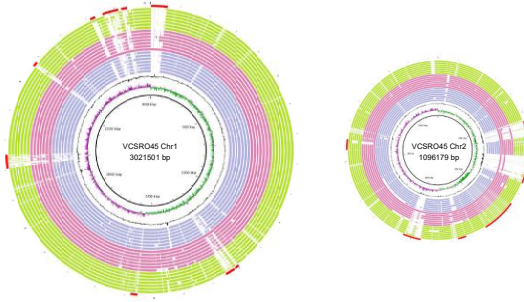
a



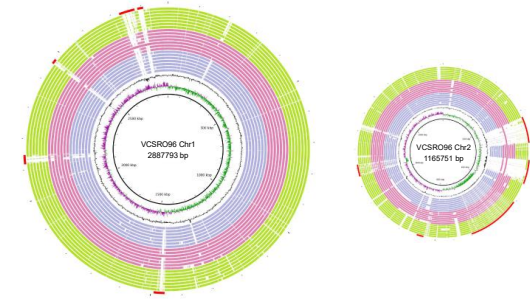
b



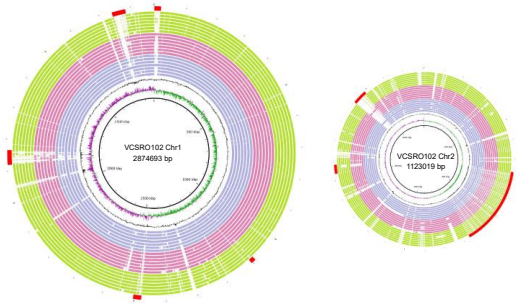
c



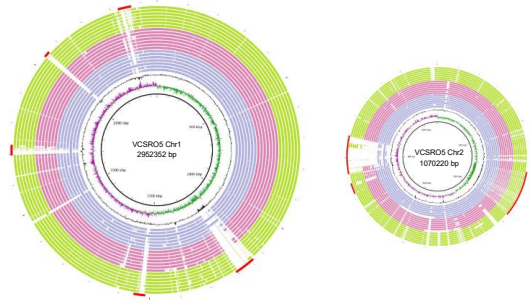
d



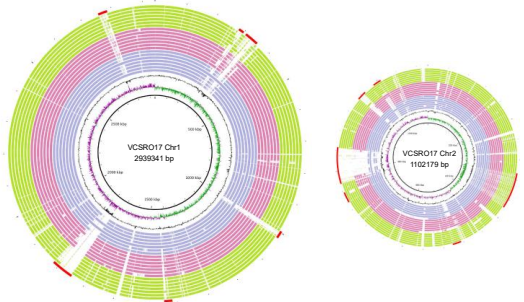
e



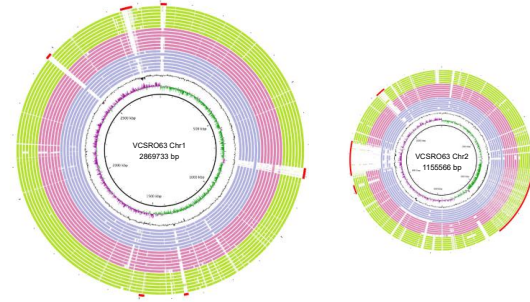
f



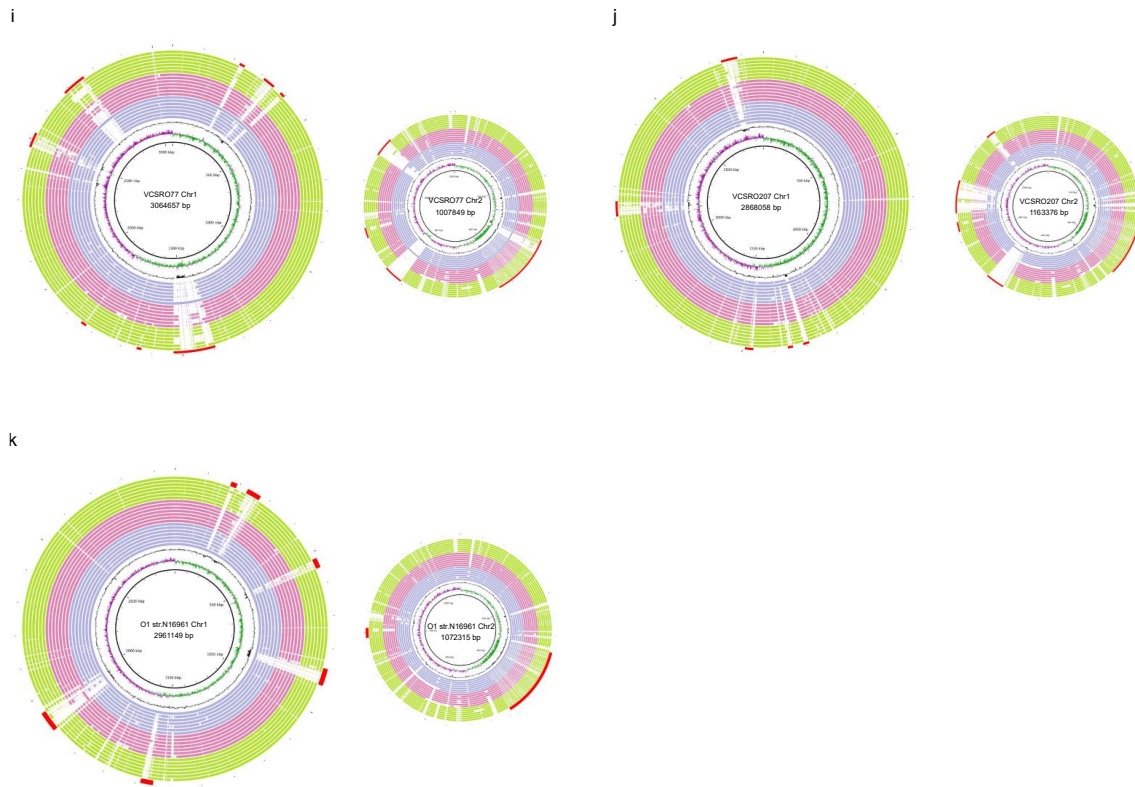
g



h



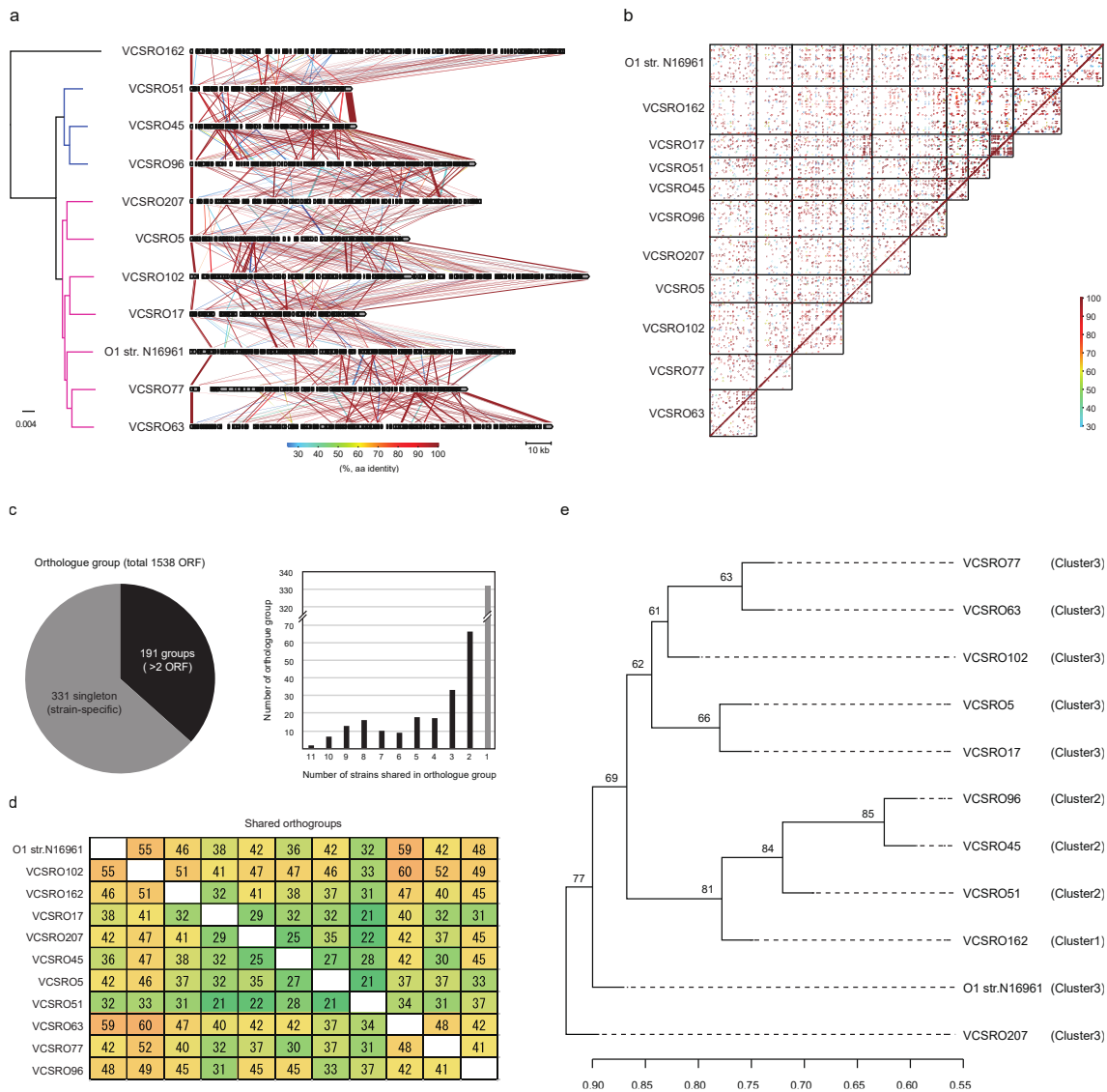




**Fig. S6. BRIG visualization of representative genome from each phylogenetic cluster compared with several *V. cholerae* genomes.**

The inner most circles represent the GC skew (purple/green) and GC contents (black) of each reference genome. BRIG (v0.95) shows the results of unfiltered BLASTn searches (threshold at 70% identity) against (a): O162, (b): O51, (c): O45, (d): O96, (e): O102, (f): O5, (g): O17, (h): O63, (i): O77, (j): O207, and (k): N16961 (O1). The colored rings were arranged from inner to outer colored circles as follows: N16961, O207, O5, O102, O17, O63, O77 from Cluster 3 (seven blue circles); O45, O51, O96, O186, O153, O72, O125 from Cluster 2 (seven pink circles); O162, O78, O192, O108, O193, O112, O210 from Cluster 1 (seven green circles). Red bar indicated region of GI listed in Table S5.





**Fig. S7. Comparative analysis of SI region on Chr2 from 11 *V. cholerae* complete genome.**

The GenomeMatcher (v2.30) was used for BLASTn analysis and visualization of the results. Total 1583 CDSs in SI regions from 11 *V. cholerae* complete genome were clustered by OrthoFinder (v2.3.7). **(a)**, Genetic structure with alignment profile of SI regions between strains. **(b)**, dotplot representation of DNA sequence homology of SI regions between strains. **(c)**, Frequency profiles of homology groups. **(d)**, one-to-one comparison between strains were represented by matrix with the number of shared homology groups. **(e)**, the cluster dendrogram analysis of 191 homology groups using R with pvclust package (method.hclust="average", method.dist="correlation") with multiscale bootstrap resampling method (500 iterations).