Extended Data Fig. 1 | Genome quality assessment flowchart (A) and characteristics of the high-quality genomes (B).
Extended Data Fig. 2 | Pairwise SNP distance distribution between all the 3,642 isolates.
Extended Data Fig.3 | Temporal dynamics of the number of isolates from patients. Years (2002–2005) with <50 patient isolates were merged. PCG, pathogenic clonal group.
Extended Data Fig.4 | Temporal distribution of the number of patient isolates from pathogenic clonal group (PCG)-others. Point sizes are scaled with the number of patient isolates in a PCG.
Extended Data Fig. 5 | Geographical distributions of patient isolates in 23 pathogenic clonal groups (PCGs). The numbers in the heatmap indicate the number of patients.
Extended Data Fig.6 | Pairwise SNP distance distribution (bar, left Y-axis) between Foodborne Disease Outbreak Surveillance (FDOS)-outbreak isolates of different pathogenic clonal group (PCGS), PCGs, and the proportion of clustered isolates under different SNP cutoffs (lines and points, right Y-axis). Blue and red indicate all the SNPs and non-recombined SNPs, respectively.

PCG3
(n-FDOS-outbreaks = 97, n-isolates = 489)

PCG189
(n-FDOS-outbreaks = 18, n-isolates = 81)

PCG-others
(n-FDOS-outbreaks = 11, n-isolates = 45)

Pairwise SNP distance between isolates

No. of paired isolates

Fraction of clustered isolates (%)

All SNPs
Non-recombined SNPs
Extended Data Fig.7 | Pairwise SNP distance distribution between 34 isolates from four outbreaks of external independent datasets.
Extended Data Fig.8 | Number and size distribution of Foodborne Disease Outbreak Surveillance (FDOS) outbreaks, P-clusters, and Ob-clusters under different SNP cutoffs (3, 6, and 10 SNPs) and time intervals (1 week and 1 month).
Extended Data Fig. 9 | Size distribution of Ob-clusters detected/not detected by the Foodborne Disease Outbreak Surveillance (FDOS).
Extended Data Fig.10 | Inferred source district distribution (posterior probability >0.7) of cross-district Ob-clusters before and after subsampling.

![Bar chart showing proportion of inferred source districts before and after subsampling. The chart compares the proportions of different districts before and after subsampling.]
Extended Data Fig.11 | SNP distance distribution over different geographical distance to hotspot district between isolates from Ob-clusters sourced from the hotspot. The bold black horizontal line indicates the mean SNP distance.