Supplementary Figure 8. Plasmid similarity and transmission traceability inference algorithm model. a) An example including five plasmids is provided. **b)** The identity between any two plasmids was evaluated from three dimensions. **c)** The "similarity" standard calculates the overall coverage of genes that are identical between two plasmids. The 'max_same' standard records the length of the longest identical gene between two plasmids, and the 'diff_num' standard shows the number of different sites between the longest identical genes. "Similarity" is utilized to construct a fully connected network where each node represents a plasmid and the line represents their similarity. **d)** A threshold is provided, and all lines below the threshold are removed to obtain the clustering results. **e&f)** A higher threshold leads to higher prediction accuracy. If there is no cyclic structure in the clustering results, the geographical transmission direction and time were inferred based on the temporal and geographical information (e.left). If there was a cyclic structure (e.right), their identity was determined based on "max_same" and "diff_num" (f).