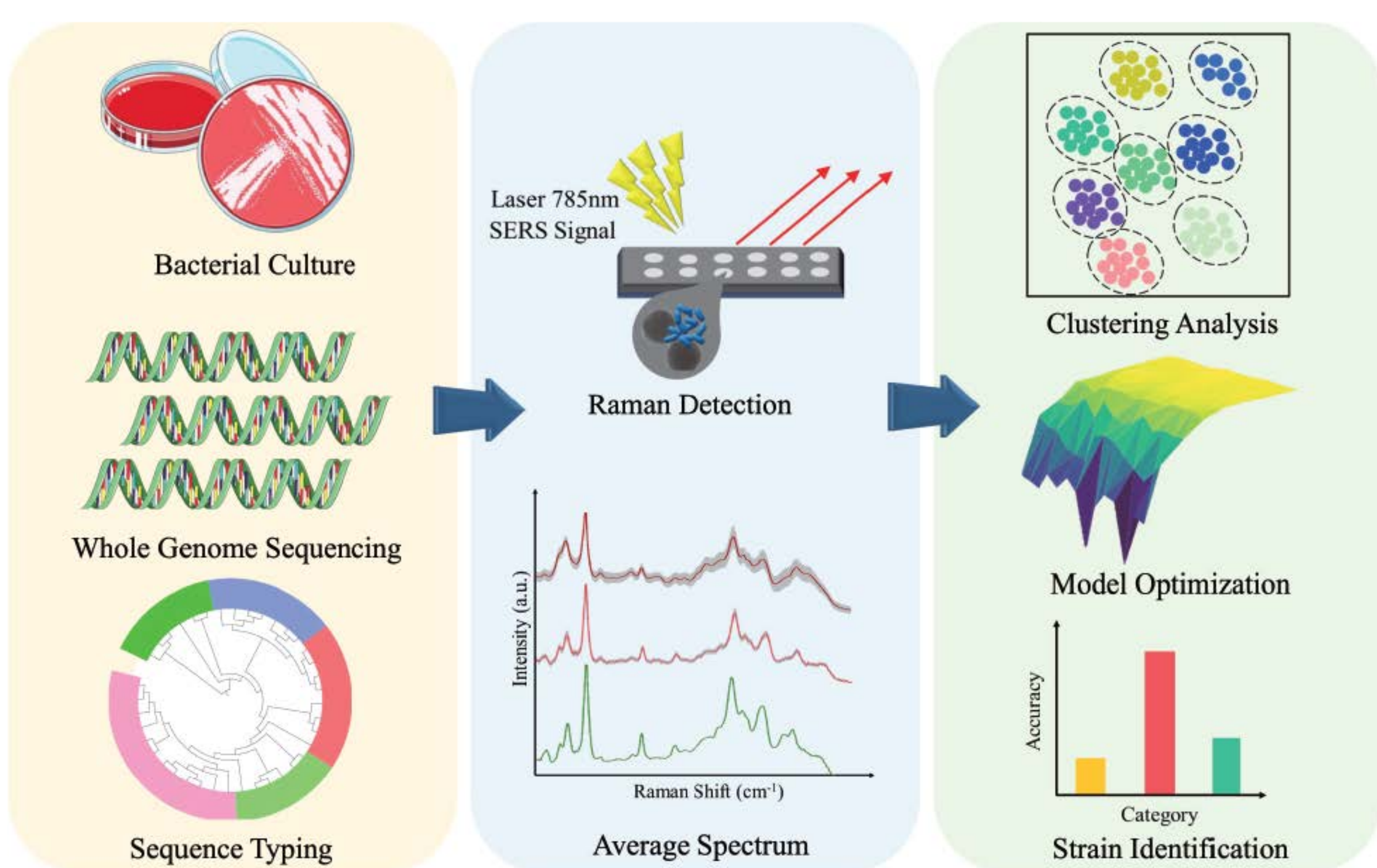
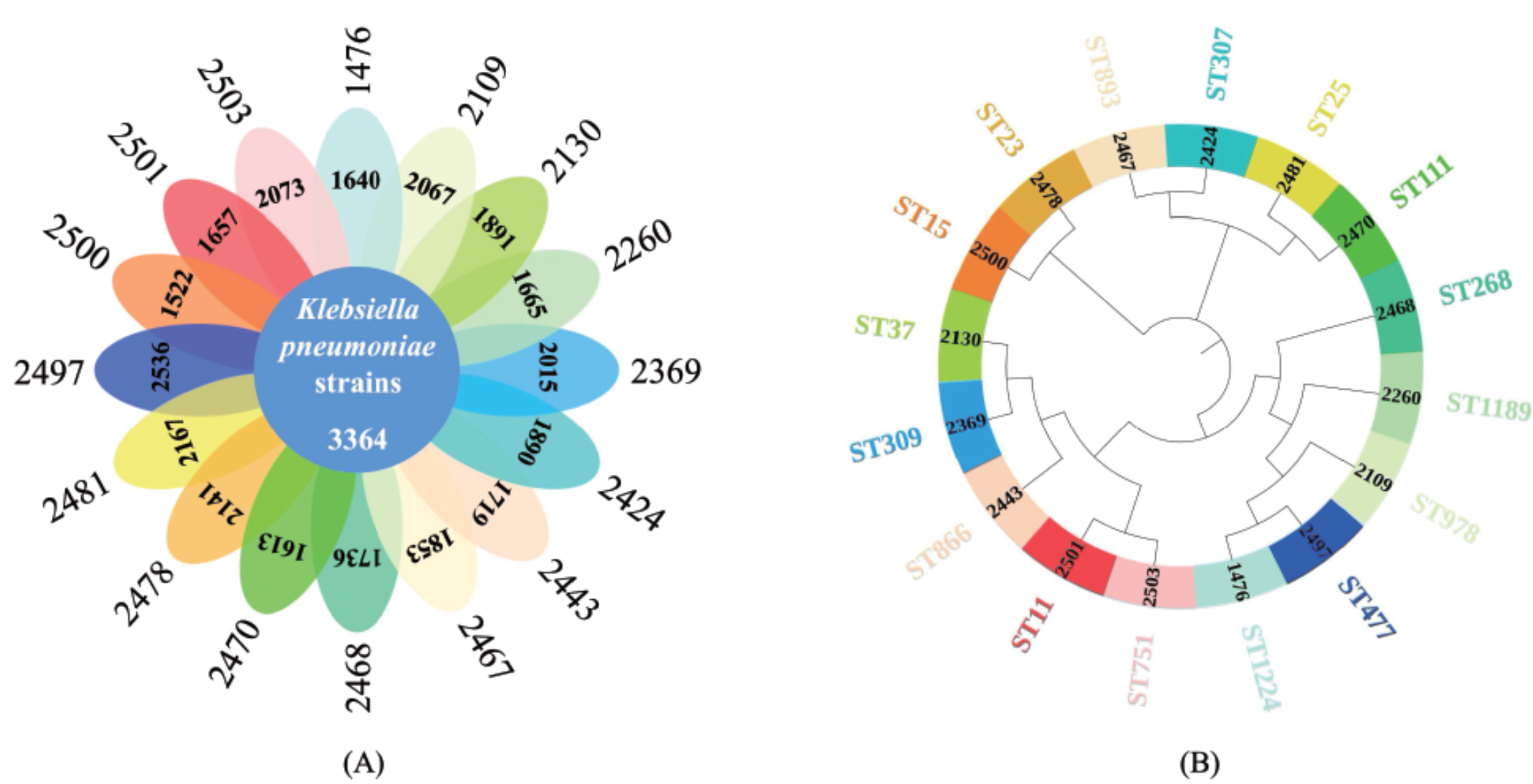


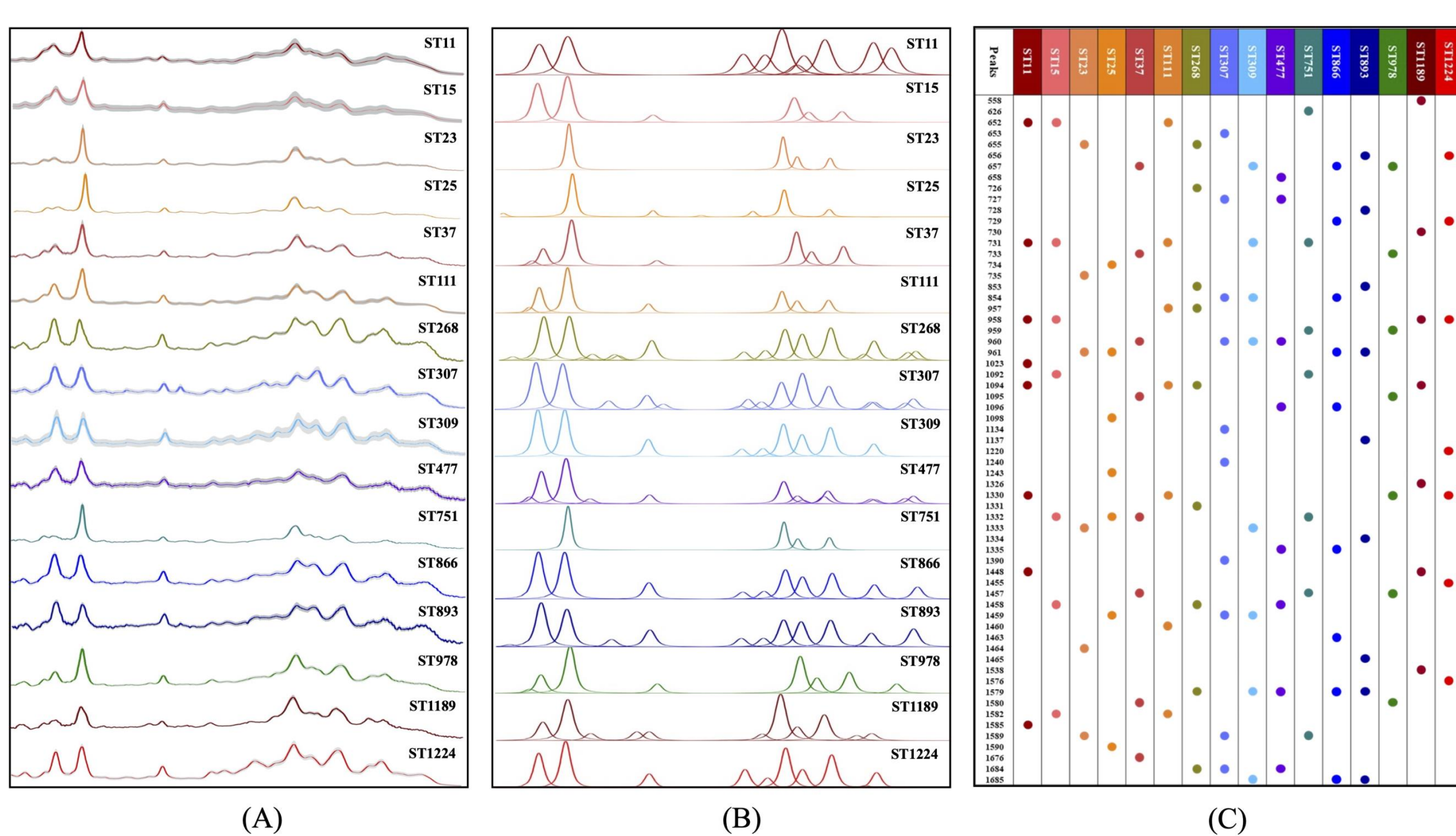
# Classification and prediction of *Klebsiella pneumoniae* strains with different MLST allelic profiles via SERS spectral analysis



In this study, we proposed a novel method based on the SERS-SVM model, which is able to accurately predict *K. pneumoniae*.



**Figure 1.** Core-genome and pan-genome analyses of *K. pneumoniae* strains with different STs. (A) Venn diagram of shared and unique CDSs among 16 *K. pneumoniae* strains. (B) Phylogenetic tree constructed via core genome analysis of 16 *K. pneumoniae* strains.



**Figure 2.** Core-genome and pan-genome analyses of *K. pneumoniae* strains with different STs. (A) Venn diagram of shared and unique CDSs. (B) Phylogenetic tree constructed via core genome analysis. (C) Dot plot of characteristic peaks for SERS spectra

We compared the performance of six supervised machine learning algorithms and explored their ability to identify ST types by analyzing SERS signal data from 16 *K. pneumoniae*. Seven machine learning evaluation metrics were used to evaluate different models as shown in **Table 1**.

Algorithm	ACC	Precision	Recall	F1	5-Fold CV	AUC	Time (s)
SVM	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0.50
RF	97.92%	97.92%	98.30%	97.88%	97.04%	98.08%	2.60
DT	96.53%	96.53%	96.16%	96.46%	94.98%	96.96%	0.01
XGB	93.75%	93.75%	94.16%	93.73%	94.30%	93.67%	21.40
QDA	76.39%	76.39%	76.96%	74.47%	76.81%	75.54%	0.09
AdaB	59.03%	59.03%	64.61%	52.56%	70.90%	59.00%	5.38

**Table 1.** Performance comparison of six supervised machine learning algorithms on the prediction of *K. pneumoniae* strains with distinct STs based on SERS spectral analysis.

In sum, our results show that SVM is an efficient and stable algorithm suitable for ST typing of *K. pneumoniae*, and has potential application for rapid tracing of the spread and control of *K. pneumoniae* in hospitals and communities.