Supporting information

Lectin-Modified Bacterial Cellulose Nanocrystals Decorated with Au Nanoparticles for Selective Detection of Bacteria Using SERS Coupled with Machine Learning

Asifur Rahman,[a,b] Seju Kang,[a,b] Wei Wang[a,b], Qishen Huang,[a,b], Inyoung Kim[c], Peter J. Vikesland,[a,b*]

[a]Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia;

[b]Virginia Tech Institute of Critical Technology and Applied Science (ICTAS) Sustainable Nanotechnology Center (VTSuN), Blacksburg, Virginia;

[c]Department of Statistics, Virginia Tech, Blacksburg, Virginia;

***** *Corresponding author.* Tel.: +1; fax: +; E-mail address: pvikes@vt.edu (P. Vikesland).
Postal address: 415 Durham Hall, Virginia Tech, Blacksburg 24061, VA, USA

Keywords: Lectin, cellulose, classification, surface-enhanced Raman scattering (SERS), bacteria, support vector machine (SVM)

Table of Contents **Pages**

Experimental methods

Chemicals

Gold chloride trihydrate ($HAuCl_4 \cdot 3H_2O$), sodium citrate tribasic dihydrate ($Na_3Cit \cdot 2H_2O$), fructose, yeast extract, calcium carbonate (CaCO3, Reagent Plus), 2,2,6,6-tetramethylpiperidinyl-1-oxyl (TEMPO), concanavalin A (Con A) type VI, Phosphate buffered saline (PBS), 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC), N-hydroxysulfosuccinimide sodium salt (Sulfo-NHS), 2-(Nmorpholino) ethanesulfonic acid (MES) and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) powder were purchased from Sigma-Aldrich (St. Louis, MO). Sodium Hypochlorite Solution (5.65-6%/Laboratory), Sodium chlorite (80%), sodium hydroxide (NaOH, certified ACS) and hydrochloric acid (HCl, certified ACS plus) were purchased from Thermo Fisher Scientific (Waltham, MA, USA).

Bacterial Sample Preparation

For each bacterium, bacterial slurries were prepared by inoculating 1 ml of the bacterial glycerol stock in 200 mL of the growth media solution autoclaved at 120 °C for 20 mins after preparation. The culture conditions for each bacterium are listed in **Table S1.** After incubation time, 10 mL of the prepared bacterial slurries were diluted to 50 mL using 1x PBS solution and were centrifuged at 4200 rpm for 15 mins. After centrifugation, the supernatant was discarded and the pellets were re-suspended in 1x PBS. This process was repeated two more times with the final re-suspension in 1x PBS (1 mL). The OD_{600nm} was recorded for all samples (**Table S1**) and bacterial concentrations were obtained by serial dilutions and plate-counting results.

Au Nanoparticles Synthesis and Characterization

The AuNPs used in this study were synthesized using a seed-mediated growth method reported previously by Brown et al.[1] 1 mM $HAuCl_4 \cdot 3H_2O$ was mixed in DI water and heated until

it started boiling and 3.88 mM $Na_3Cit{\cdot}2H_2O$ was added as a reducing agent. The mixture was kept under magnetic stirring for 30 mins and then cooled and filtered through a 0.22 μm PTFE filter. In another round bottom flask, 100 mL solution of .254 mM $HAuCl_4{\cdot}3H_2O$ was heated to boil and then .85 mL of seed solution was added, along with 0.44 mL of 38.8 mM $Na_3Cit{\cdot}2H_2O$ as the reducing agent. The solution was refluxed using a condenser for 30 min and then cooled to room temperature. The as-synthesized AuNPs were characterized using a Cary 5000 UV-Vis-NIR Spectrophotometer, a JEOL 2100 Transmission Electron Microscopy (TEM) operating at 200 kV, and a Zetasizer Nano ZS Dynamic Light Scattering (DLS) instrument. The final product was stored at 4°C prior to use.

The as-synthesized AuNPs showed a characteristic localized surface plasmon resonance (LSPR) peak at ~542 nm (**Figure S1A**) . TEM images of the AuNPs showed spherical shaped structures (**Figure S1C**). The mean size of the AuNPs determined from TEM images was 49.69 $\pm$ 4.06 nm (**Figure S1B**). The electrophoretic mobility (EM) and zeta potential (ζ) of AuNPs suspended in 50 μM Na_3Cit solution was -2.7 (± 0.1, n=3) × 10^{-8} m^2/Vs (= -34.1 mV (± 0.8).

Synthesis of Bacterial Cellulose and BCNCs

One or two colonies of the cellulose producing bacteria *Gluconacetobacter xylinus* grown on nutrient agar were inoculated into 10 mL aliquots of ATCC medium 459 (10 g fructose, 1 g yeast extract and 2.5 g $CaCO_3$ in 200 mL of DI water) and then poured into Petri dishes (100 mm × 15 mm).[2] The Petri dishes were stored in an oven at 30 °C for five days until the hydrogel-like BC pellicles were fully grown. These pellicles were then removed, washed with tap water, and suspended in 0.1 M NaOH for three days to remove residual chemicals. The washed pellicles were stored in DI water at room temperature. To prepare the BCNC suspension, 10-15 mL of 37% (*w/w*) HCl was brought to 80 °C under continuous stirring and then a single BC pellicle was added. After

2 hr, the acid hydrolysis reaction was terminated, and 40 mL of DI water was added to the reaction vessel. The cloudy BCNC suspension was centrifuged at 2500 × g for 30 min and resuspended in DI water several times until the pH was between 6 and 7.[3]

Table S1. The scientific name and ATCC type strains, sample name used in this study, culture conditions and duration, optical density (OD) of the stock suspension and gram types for each of the 19 bacteria strains.

Serial	Bacteria name and ATCC type strain	Sample name	Culture conditions	OD of the stock	Gram positive /negative
1	*Acinetobacter baumannii* ATCC BAA-1789	*A. baumannii* 1789	Trypticase Soy Broth, 24 hours, 37 °C, aerobic	0.83	Gram negative
2	*Acinetobacter baumannii* ATCC BAA-2802	*A. baumannii* 2802	Trypticase Soy Broth, 24 hours, 37 °C, aerobic	0.81	Gram negative
3	*Bacillus subtilis* ATCC 6051	*B. subtilis*	Nutrient broth, 24 hours, 37 °C, aerobic	0.78	Gram positive
4	*Enterococcus faecalis* V583; ATCC 700802	*E. faecalis* V583	Brain Heart Infusion broth, 24 hours, 37 °C, aerobic	0.74	Gram negative
5	*Enterococcus faecium* ATCC 35667	*E. faecium* 35667	Brain Heart Infusion broth, 24 hours, 37 °C, aerobic	0.83	Gram positive
6	*Escherichia coli* ATCC 25922	*E. coli* 25922	Trypticase soy broth, 24 hours, 37 °C, aerobic	0.8	Gram negative
7	*Escherichia coli* ATCC 8739	*E. coli* 8739	Nutrient broth, 24 hours, 37 °C, aerobic	0.81	Gram negative
8	*Escherichia coli* K12; Strain SMG 123	*E. coli* K12	LB broth, 24 hours, 37 °C, aerobic	0.68	Gram negative
9	*Escherichia coli* MG 1655; ATCC 700926	*E. coli* MG 1655	M9 Minimal broth, 24 hours, 37 °C, aerobic	0.82	Gram negative
10	*Escherichia coli* SMS-3-5; ATCC BAA 1743	*E. coli* SMS-3-5	LB broth, 24 hours, 37 °C, aerobic	0.79	Gram negative
11	*Escherichia coli* DH5α ATCC 87399	*E. coli* DH5α	LB broth with 50 µg/mL ampicillin, 24 hours, 37 °C, aerobic	0.65	Gram negative
12	*Mycobacterium smegmatis* ATCC 14468	*M. smegmatis*	Middlebrook 7H9 Broth, 48 hours, 37 °C, aerobic	0.68	Gram positive
13	*Pseudomonas aeruginosa* ATCC BAA-2111	*P. aeruginosa* 2111	Trypticase Soy Broth, 24 hours, 37 °C, aerobic	0.8	Gram negative
14	*Pseudomonas aeruginosa* ATCC BAA-2795	*P. aeruginosa* 2795	Trypticase Soy Broth, 24 hours, 37 °C, aerobic	0.82	Gram negative
15	*Pseudomonas aeruginosa* ATCC 47085	*P. aeruginosa* 47085	LB broth, 24 hours, 30 °C, aerobic	0.76	Gram negative
16	*Pseudomonas syringae* pv. phaseolicola	*P. syringae*	LB broth, 48 hours, 25 °C, aerobic	0.793	Gram negative
17	*Staphylococcus aureus* ATCC 12600	*S. aureus* 12600	Nutrient broth, 24 hours, 37 °C, aerobic	0.79	Gram positive
18	*Staphylococcus aureus* ATCC BAA-1556	*S. aureus* 1556	Brain Heart Infusion broth, 24 hours, 37 °C, aerobic	0.77	Gram positive
19	*Shewanella oneidensis* MR-1	*S. oneidensis*	Trypticase Soy Broth, 24 hours, 30 °C, aerobic	0.76	Gram negative

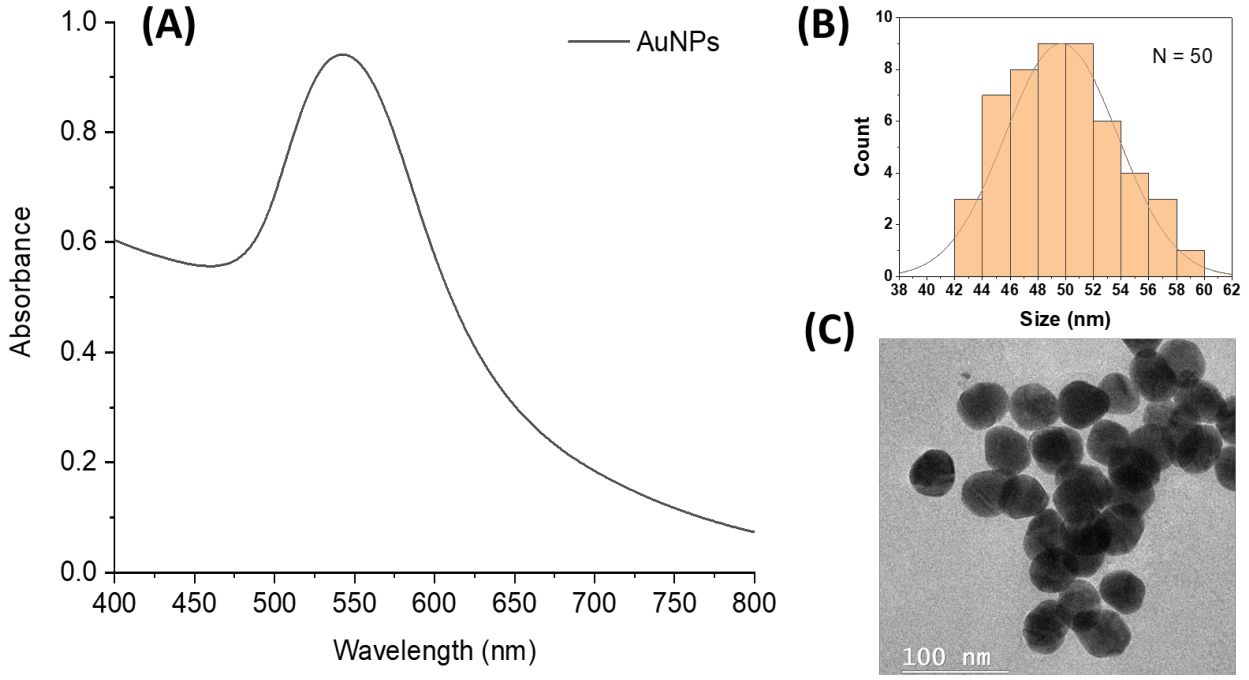


Figure S1. (A) UV-Vis absorbance spectrum of as-synthesized AuNPs with the peak at 542 nm. **(B)** Size distribution of AuNPs (N=50) measured by the ImageJ software. **(C)** TEM image showing the structure and size of AuNPs

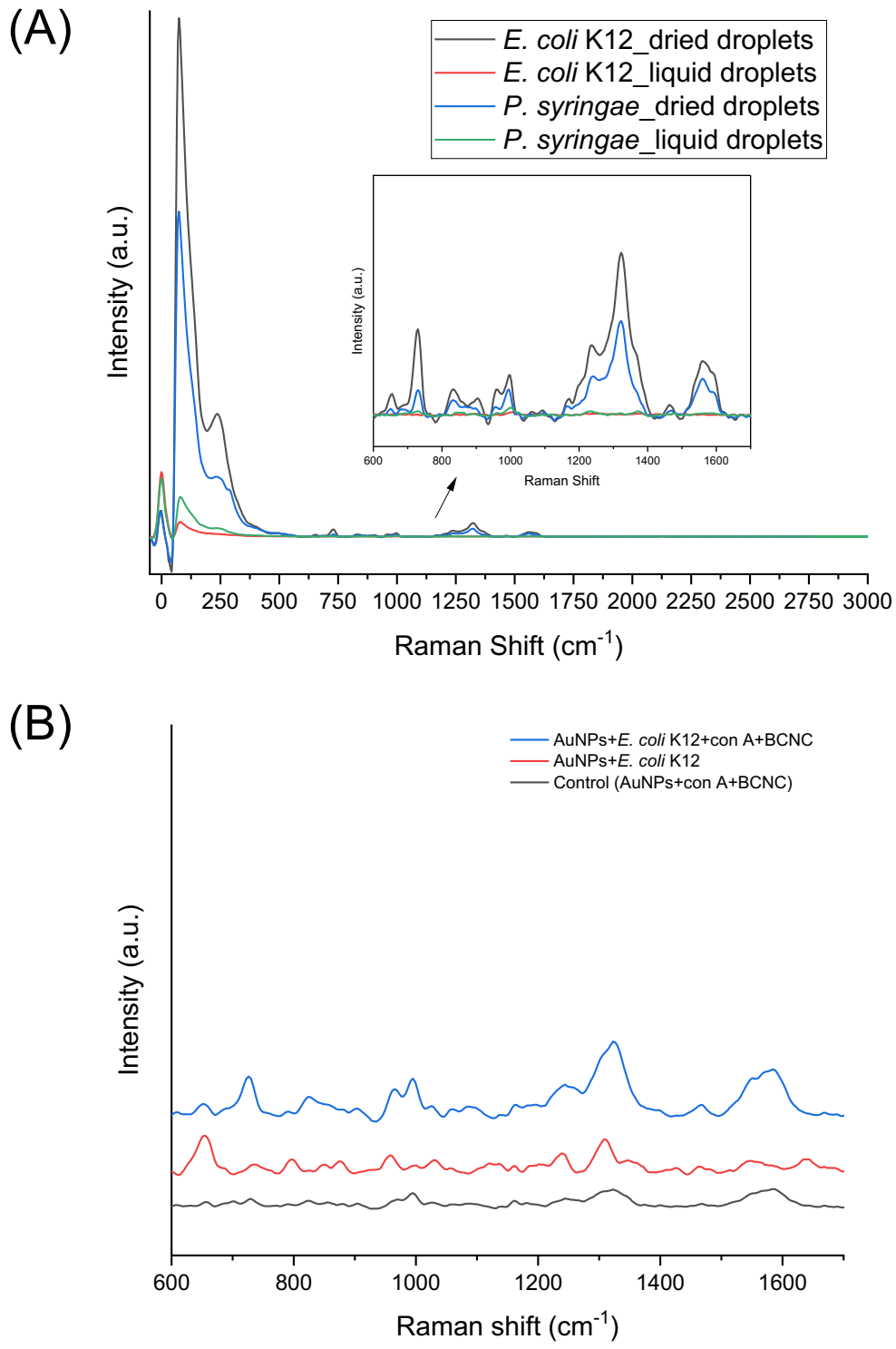


Figure S2. (A) Comparison of the SERS spectra of E. coli K12 and P. syringae under wet and dry conditions using AuNPs+*E. coli* K12/*P. syringae*+con A+BCNCs. Inset pictures show the spectra from 600 cm^{-1} and 1700 cm^{-1}. **(B)** Hot spot normalized SERS spectra of *E. coli* K12 (~10^8 CFU/ml) with AuNPs+con A @BCNCs, with AuNPs only, and the control with no bacteria present.

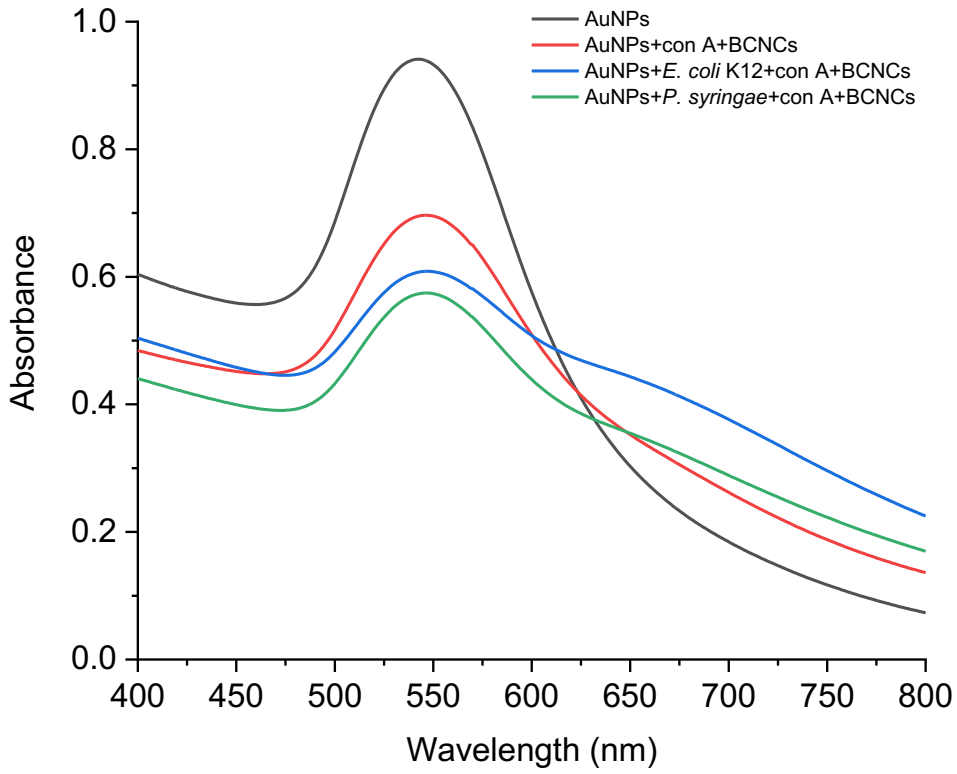


Figure S3. UV-Vis absorbance spectra of AuNPs, AuNPs conjugated to con A+BCNC (AuNPs+con A+BCNC), the spectra taken after conjugation of *E. coli* K12 and *P. syringae* bacteria to con A+BCNC, followed by conjugation with AuNPs (AuNPs+ *E. coli* K12+con A+BCNC and AuNPs+ *P. syringae* +con A+BCNC repectively)

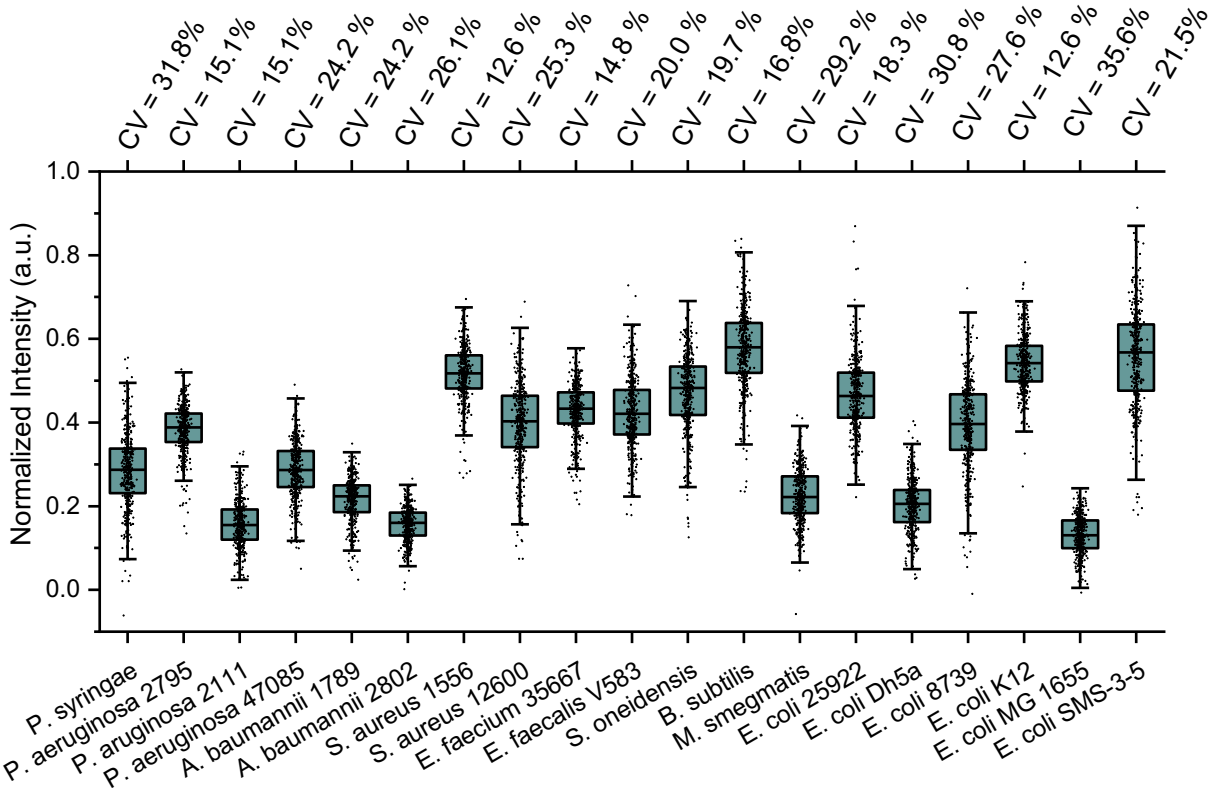


Figure S4. Boxplots with distribution of data points for the hot spot normalized peak intensity at 730 cm^{-1} for all 19 bacteria in this study. The upper X-axis shows the corresponding coefficients of variation (CV) of the normalized intensity at 730 cm^{-1} for each bacteria species.

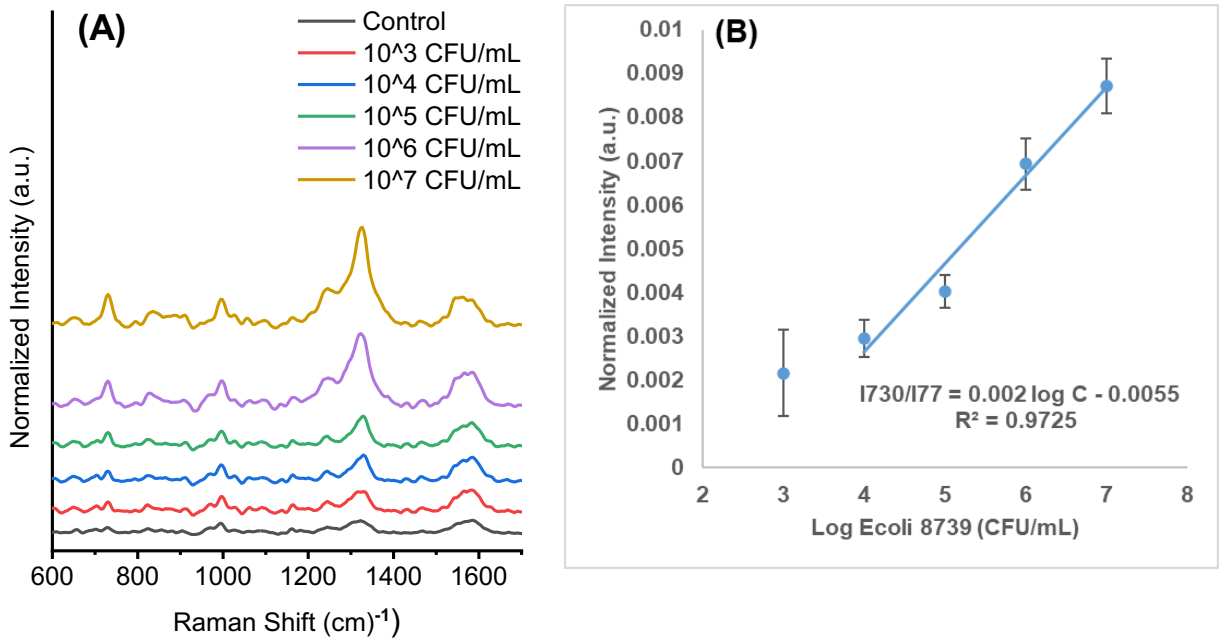


Figure S5. (A) SERS spectra taken from the aggregated AuNP + *E. coli* 8739 + con A+BCNC conjugates for various concentrations of *E. coli* 8739 (10^7-10^3 CFU/mL), along with a control with no *E. coli* 8739. **(B)** calibration curve for *E. coli* 8739 at a concentration range of 10^3−10^7 CFU/mL obtained by using SERS intensity at 730 cm^{-1}, normalized using the peak at 77 cm^{-1}.

Table S2. Previously reported the limit of detection (LOD) and linear concentration ranges for label-free detection of bacteria.

Nanomaterial used	Binding agent	LOD	Linear conc. range	Ref.
AgNPs	Aptamer	1.5 CFU/mL	10^1-10^7 CFU/mL	[4]
Ag colloids	-	1×10^5 CFU/mL	10^5-10^8 CFU/mL	[5]
Ag nanorods	Aptamer	10^8 CFU/mL	-	[6]
Au@Fe_3O_4+Au@Ag	PEI	10^3 CFU/mL	10^3-10^7 CFU/mL	[7]
AgNPs	-	2.5×10^2 CFU/mL	-	[8]
AuNPs+BCNCs	Lectin	1.5×10^3 CFU/mL	10^3-10^7 CFU/mL	This study

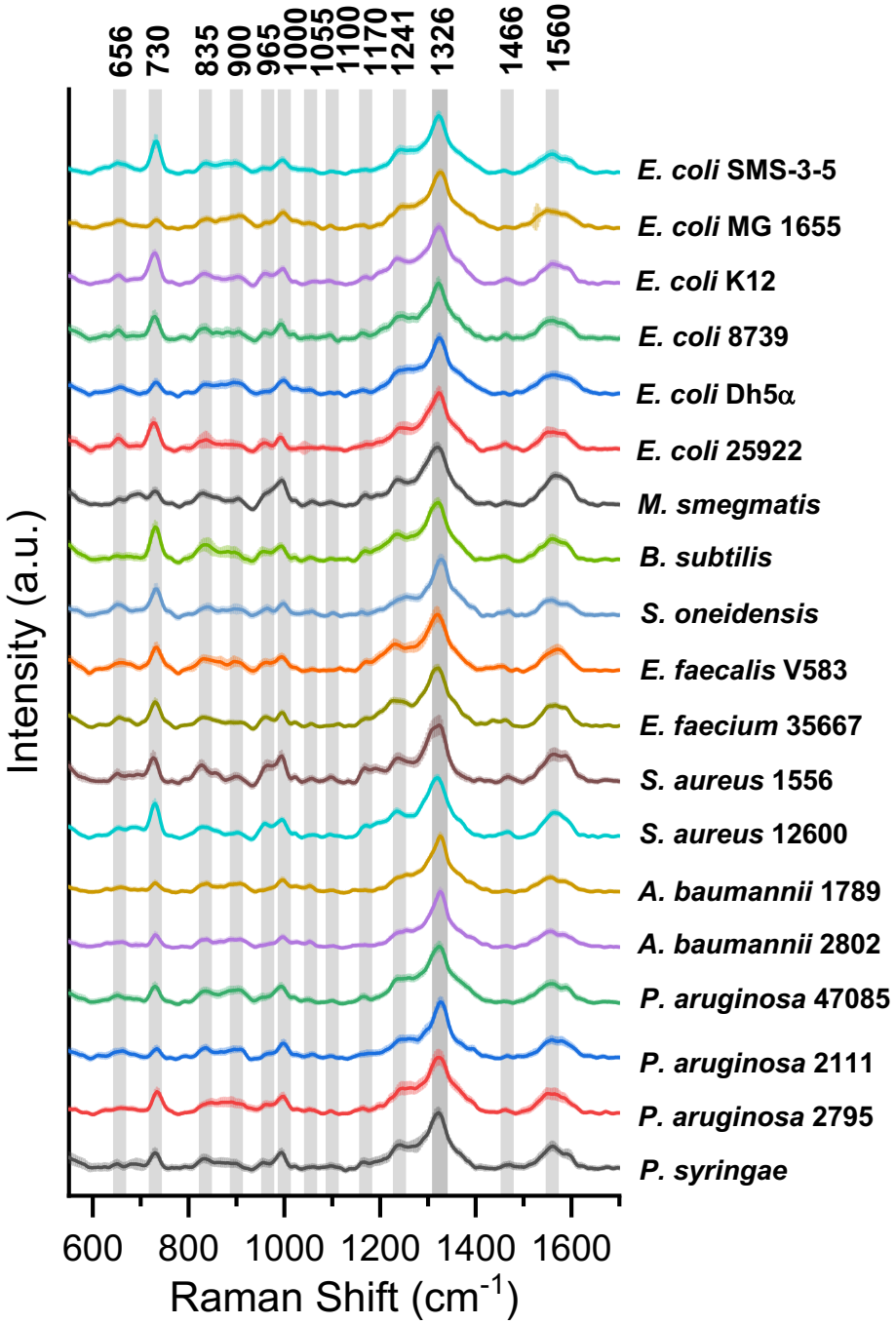


Figure S6. SERS spectra for the 19 bacteria strains used in this study, normalized using the peak at 1326 cm^{-1}. The SERS spectra presented for each bacterium is the average of 400 spectra collected using the Raman large area scan feature. Along with the average spectra, the standard deviation of each data point on the spectra is shown here. The 12 peak positions highlighted with gray bars were used as spectral features to develop the SVM model. The peak at 1326 highlighted with dark gray bar was not included in model development and was used to normalize the collected spectra.

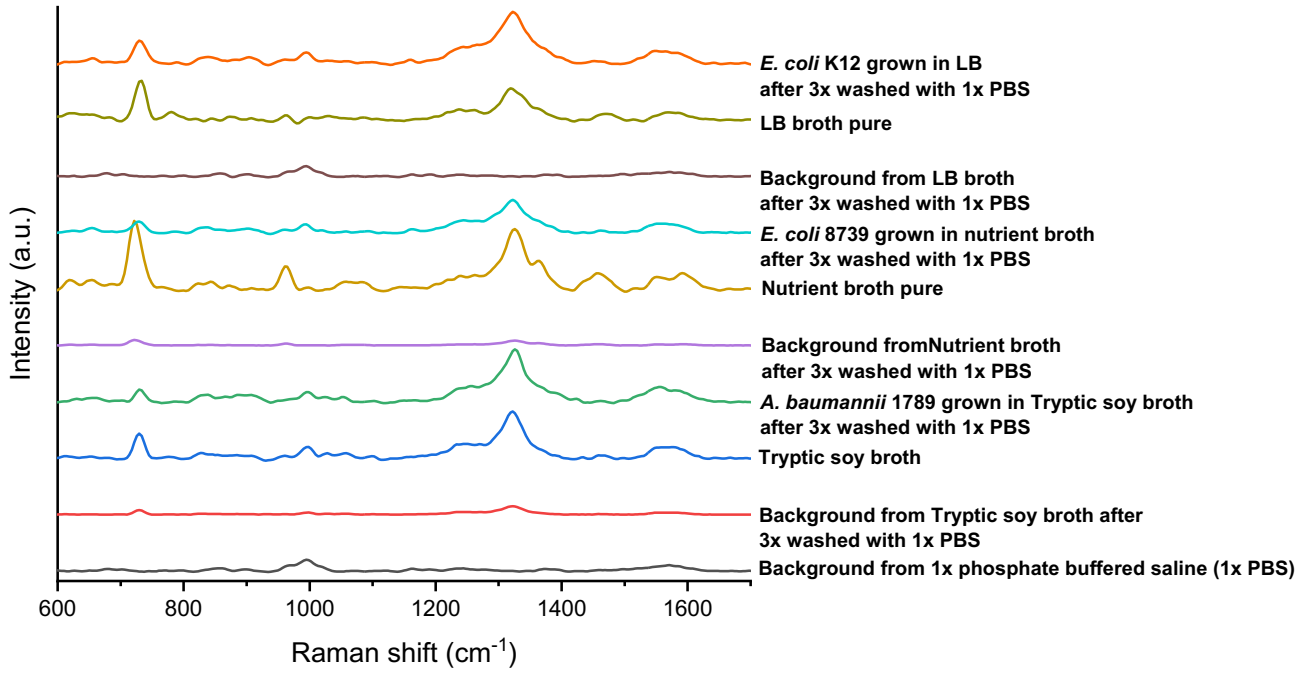


Figure S7. Comparison of the SERS spectra a representative bacterial strain grown in a certain media, pure media, and background SERS spectra from the growth media after 3x washing with 1x PBS for SERS detection of the bacterial strain. The background spectrum from 1x PBS is also provided here.

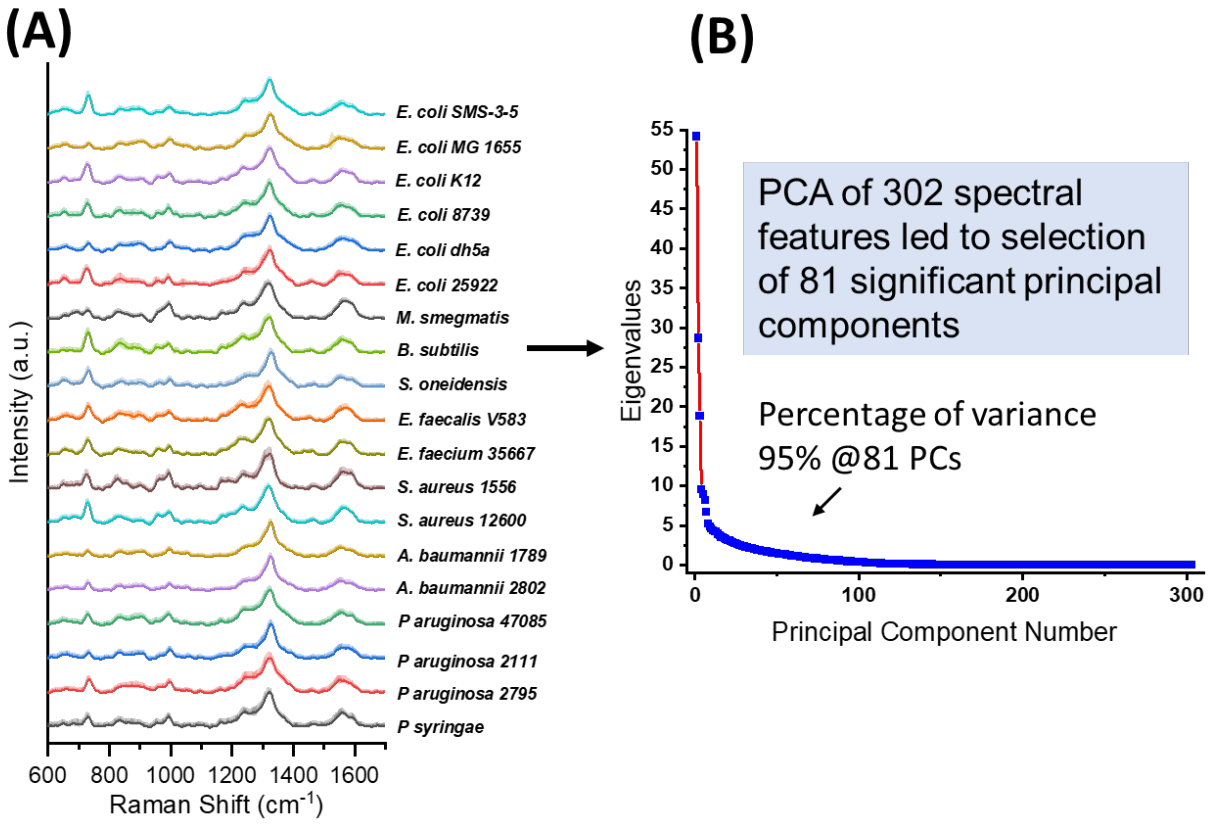


Figure S8. (A) SERS spectra for the 19 bacteria species used in this study, considering the whole spectra from 600 cm^{-1} to 1700 cm^{-1} for analysis. **(B)** The eigenvalues of the 302 features initially considered as principal components (PCs). From the 302 features, 81 PCs were selected by principal component analysis (PCA) with 95% variance.

Table S3. Comparison of the overall accuracy of different multiclass SVM classifiers, LDA and QDA

SVM classification		
Type of SVM classifier	**Overall accuracy (One vs one approach)**	**Overall accuracy (One vs all approach)**
Linear SVM	85.7%	73.3%
Quadratic SVM	87.4%	86.6%
Cubic SVM	86.0%	83.8%
Fine Gaussian SVM	31.2%	83.7%
Medium Gaussian SVM	**87.7%**	**87.0%**
Coarse Gaussian SVM	85.1%	76.8%
Discriminant analysis		
Type of classifier	**Overall accuracy**	
Linear Discriminant (LDA)	84.3%	
Quadratic Discriminant (QDA)	85.6%	

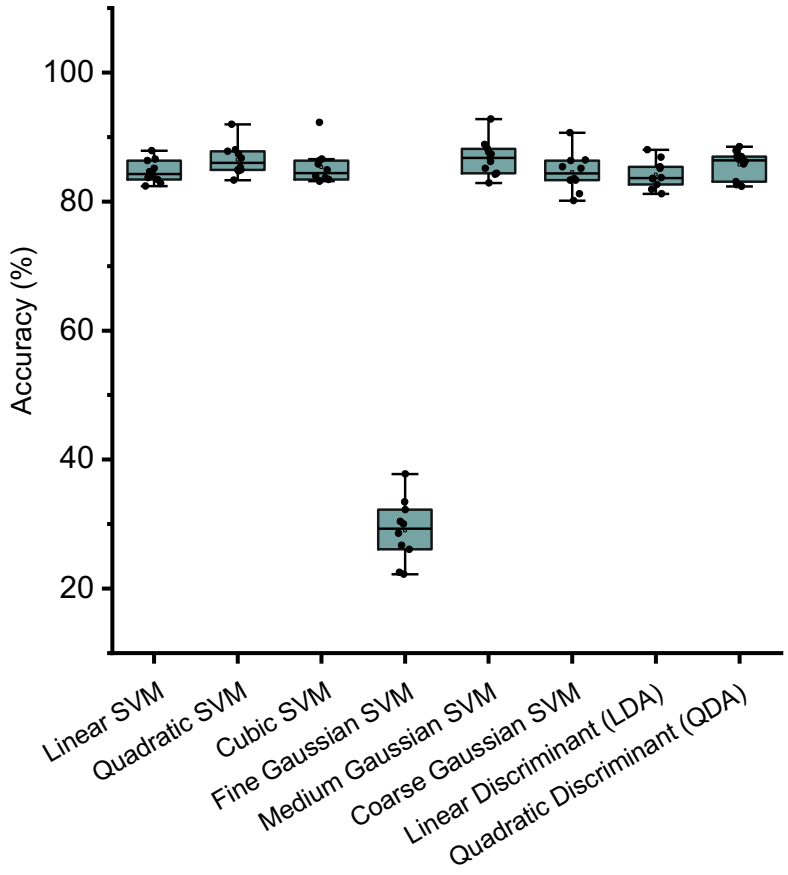


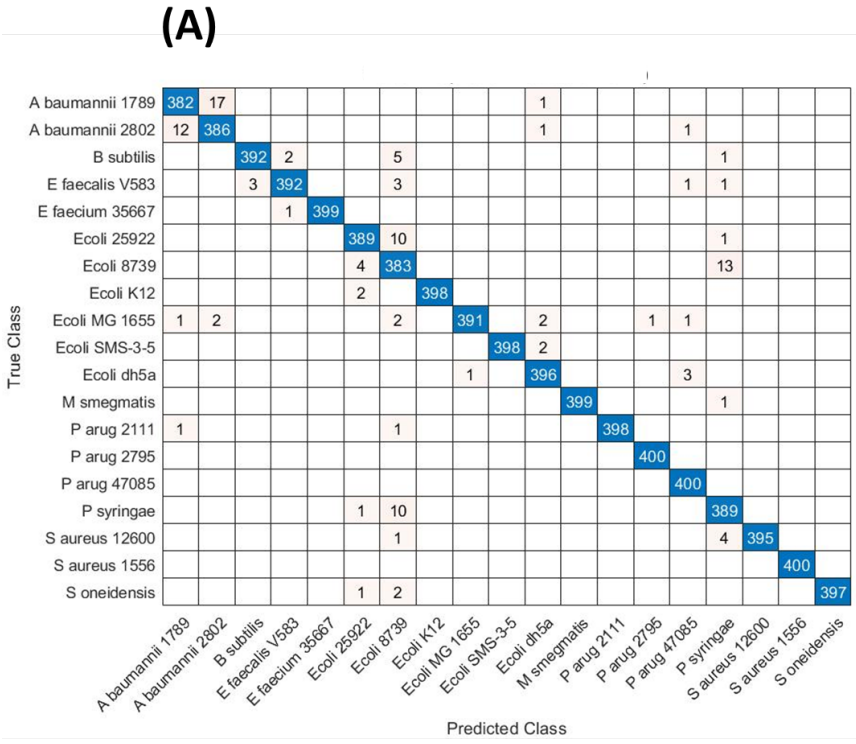
Figure S9. Comparison of the accuracy of the different trained multiclass SVM classifiers (linear, quadratic, cubic kernel functions and fine, medium, coarse Gaussian kernel functions) with one vs one voting strategy and PCA feature disabled. Discriminant analysis methods (LDA and QDA) are also included for comparison. Each boxplot has 10 points from 10-fold cross-validation results, upper and lower quartiles, and median.

Table S4. Summary statistics of the boxplot data on different trained multiclass SVM classifiers, along with discriminant analysis methods (LDA and QDA).

Type of classifier	Mean	Stdev	CV	Confidence level (95%)	Upper CI (95%)	Lower CI (95%)	Statistically different to all other types of classifiers? (Based on Wilcoxon ranked signed tests for (α = 0.05))
Linear SVM	84.67	1.70	2.01	1.05	85.73	83.62	No
Quadratic SVM	86.49	2.34	2.71	1.45	87.94	85.03	No
Cubic SVM	85.32	2.61	3.06	1.62	86.94	83.70	No
Fine Gaussian SVM	28.99	4.60	15.88	2.85	31.84	26.13	Yes
Medium Gaussian SVM	87.35	2.44	2.80	1.52	88.87	85.83	No
Coarse Gaussian SVM	84.54	2.83	3.35	1.75	86.29	82.79	No
Linear Discriminant (LDA)	83.98	1.74	2.07	1.08	85.06	82.90	No
Quadratic Discriminant (QDA)	85.69	2.11	2.47	1.31	87.00	84.38	No

Table S5. Summary statistics on SVM accuracy for the prediction of the E. coli K12 strain from five measurements using five different substrates.

1	2	3	4	5	6
Bacterial strain measured	Substrate no.	SVM accuracy (medium gaussian) (%)	Avg. ± Std. deviation (%)	Reported accuracy for *E. coli* K12 (%)	Difference = col.4 - col.5 (%)
E. coli K12	1	94	85.25 ± 6.62	83.50	1.75
	2	77.25			
	3	86.75			
	4	80.25			
	5	88			



Sample	Accuracy
A. baumannii 1789	
A. baumannii 2802	
B. subtilis	
E. faecalis V583	
E. faecium 35667	
E. coli 25922	
E. coli 8739	
E. coli K12	
E. coli MG 1655	
E. coli SMS-3-5	**98.5%**
E. coli DH5α	
M. smegmatis	
P. aeruginosa 2111	
P. aeruginosa 2795	
P. aeruginosa 47085	
P. syringae	
S. aureus 12600	
S. aureus 1556	
S. oneidensis	

Figure S10. (A) Confusion matrix for 19 bacteria strains. The rows and columns represent the true and predictive class from a total of 400 spectra for each strain of bacteria. The diagonal entries represent the accuracies for each class (i.e., number of spectra correctly predicted from a total of 400 spectra). Entries other than the diagonal ones represent misclassifications for a given strain of bacteria. **(B)** The overall accuracy is 98.5% when the whole spectra is considered to develop the model

References

1. Brown, K. R.; Walter, D. G.; Natan, M. J. Seeding of colloidal Au nanoparticle solutions. 2. Improved control of particle size and shape. *Chem. Mater.* **2000**, *12*, 306-313.
2. Wei, H.; Rodriguez, K.; Renneckar, S.; Leng, W.; Vikesland, P. J. Preparation and evaluation of nanocellulose–gold nanoparticle nanocomposites for SERS applications. *Analyst*. **2015**, *140*, 5640-5649.
3. Kang, S.; Rahman, A.; Boeding, E.; Vikesland, P. J. Synthesis and SERS application of gold and iron oxide functionalized bacterial cellulose nanocrystals (Au@Fe_3O_4@ BCNCs). *Analyst*. **2020**, *145*, 4358-4368.
4. Gao, W.; Li, B.; Yao, R.; Li, Z.; Wang, X.; Dong, X.; Qu, H.; Li, Q.; Li, N.; Chi, H. Intuitive label-free SERS detection of bacteria using aptamer-based in situ silver nanoparticles synthesis. *Anal. Chem.* **2017**, *89*, 9836-9842.
5. Yang, D.; Zhou, H.; Haisch, C.; Niessner, R.; Ying, Y., Reproducible *E. coli* detection based on label-free SERS and mapping. *Talanta*. **2016**, *146*, 457-463.
6. Chen, J.; Park, B.; Huang, Y.-w.; Zhao, Y.; Kwon, Y. Label-free SERS detection of *Salmonella* Typhimurium on DNA aptamer modified AgNR substrates. *J. Food Meas. Charact.* **2017**, *11*, 1773-1779.
7. Wang, C.; Wang, J.; Li, M.; Qu, X.; Zhang, K.; Rong, Z.; Xiao, R.; Wang, S. A rapid SERS method for label-free bacteria detection using polyethylenimine-modified Au-coated magnetic microspheres and Au@ Ag nanoparticles. *Analyst*. **2016**, *141*, 6226-6238.
8. Zhou, H.; Yang, D.; Ivleva, N. P.; Mircescu, N. E.; Niessner, R.; Haisch, C. SERS detection of bacteria in water by in situ coating with Ag nanoparticles. *Anal. Chem.* **2014**, *86*, 1525-1533.