Supplementary Information

Identification of Mycobacteria Based on Spectroscopic Analyses of Mycolic Acid Profiles

Omar E. Rivera-Betancourt[1], Russell Karls[2], Benjamin Grosse-Siestrup[2], Shelly Helms[2], Frederick Quinn[2], and Richard Dluhy[1*]

[1]Department of Chemistry, University of Georgia, Athens, GA 30602

[2]Department of Infectious Diseases, University of Georgia, Athens, GA 30602

*Author to whom correspondence should be addressed

Tel: +1 706 542 1950

Fax: +1 706 542 9454

dluhy@uga.edu

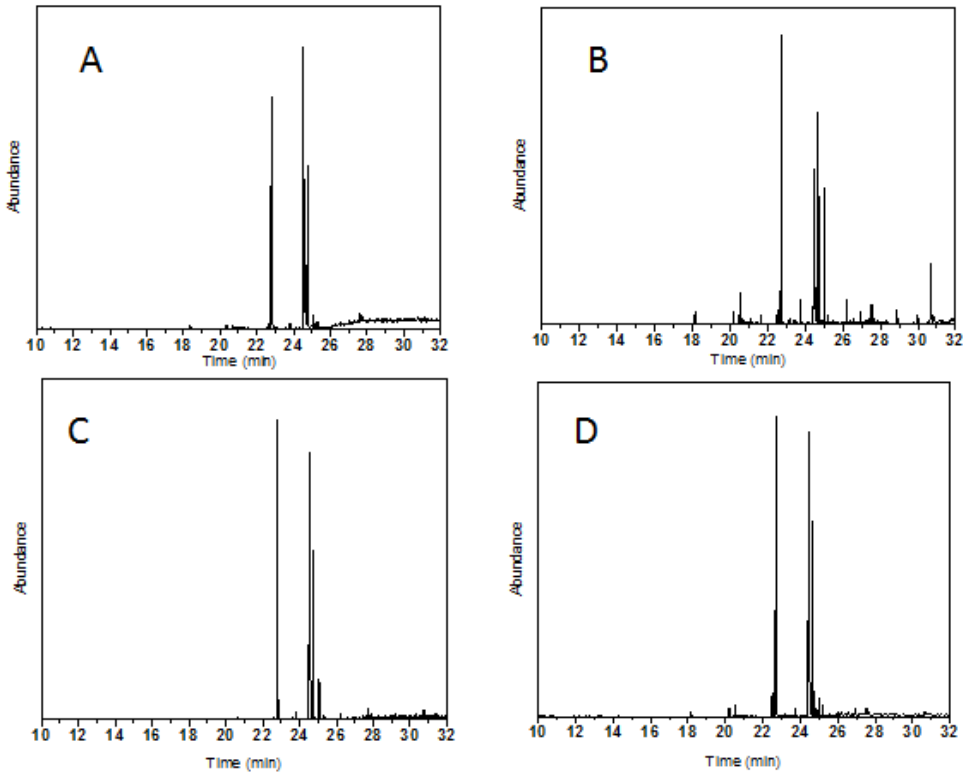

Figure S.1. Representative total ion chromatograms of three *Mycobacterium* species investigated by GC-MS. (A) *M. smegmatis* ; (B) *M. avium* ; (C) *M. tuberculosis* strain H37Rv ; (D) *M. tuberculosis* strain Erdman

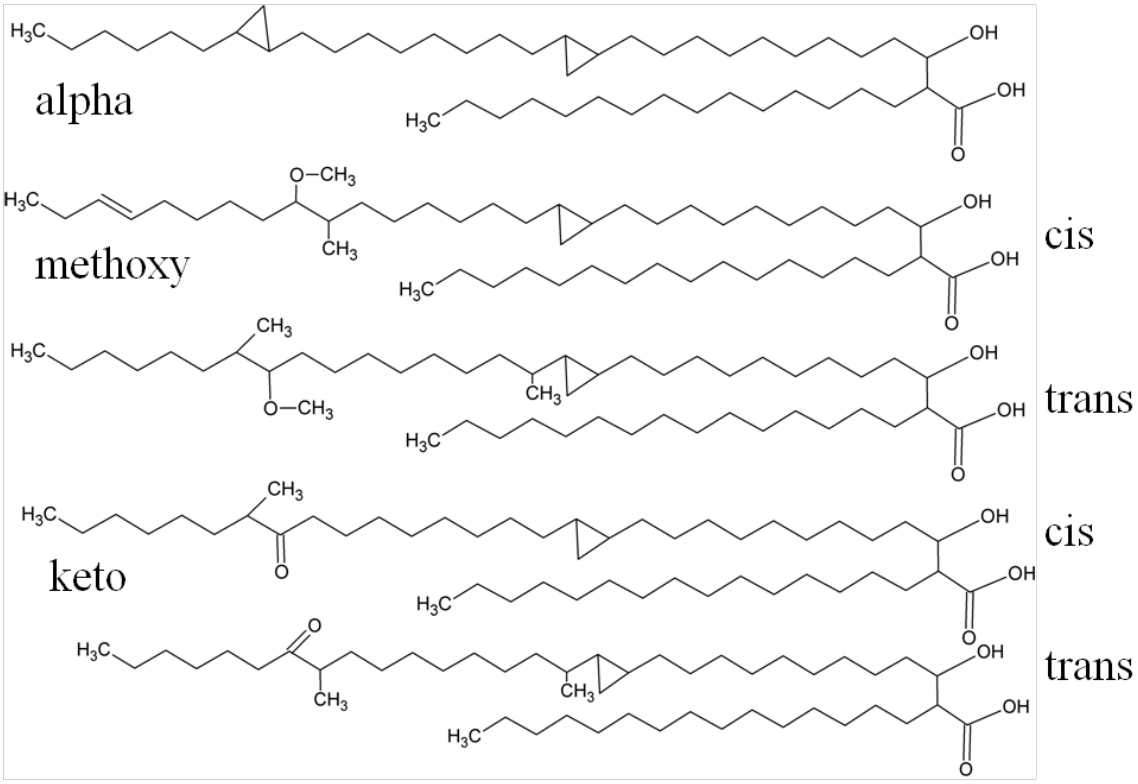

Figure S.2. Mycolic acid sub-structures commonly found in mycobacteria.

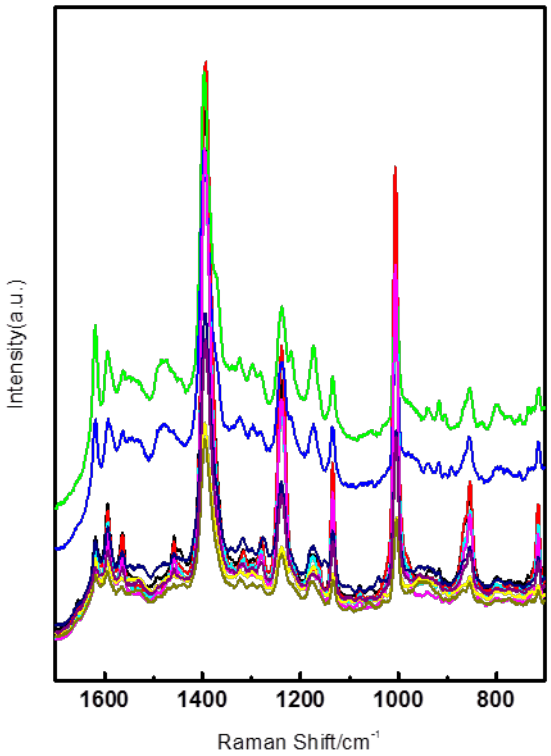

Figure S.3. Ten individual, unprocessed SERS spectra obtained for the *M. tuberculosis* Erdman strain. Spectra illustrate the reproducibility of the raw data used in this analysis.

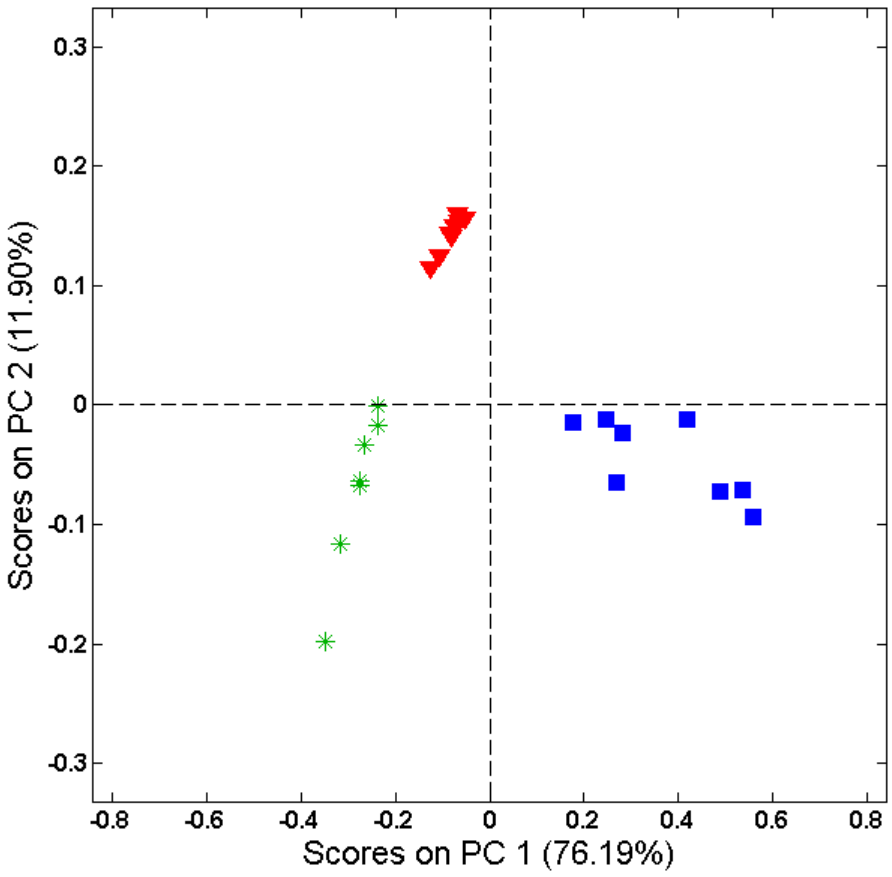

Figure S.4. PCA scores plot corresponding to the NTM species *M. smegmatis* (⓿), *M. avium* (⑤), *M. bovis* BCG (☒). The PC model was constructed from the SERS spectra of the corresponding species using the spectral range 1700-700 cm^{-1}.

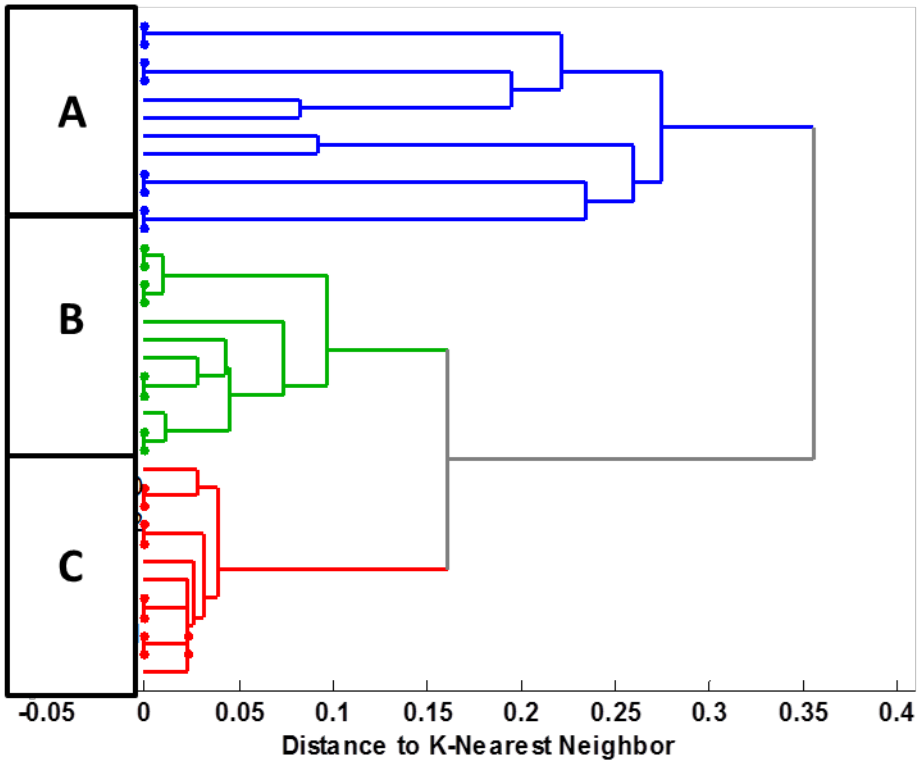

Figure S.5. A hierarchical cluster analysis dendrogram derived from the PC scores of the NTM species. The nodes group into three recognized clusters and are labeled according to the samples. (A) *M. avium* (blue) ; (B) *M. bovis* BCG (green) ; (C) *M. smegmatis* (red).

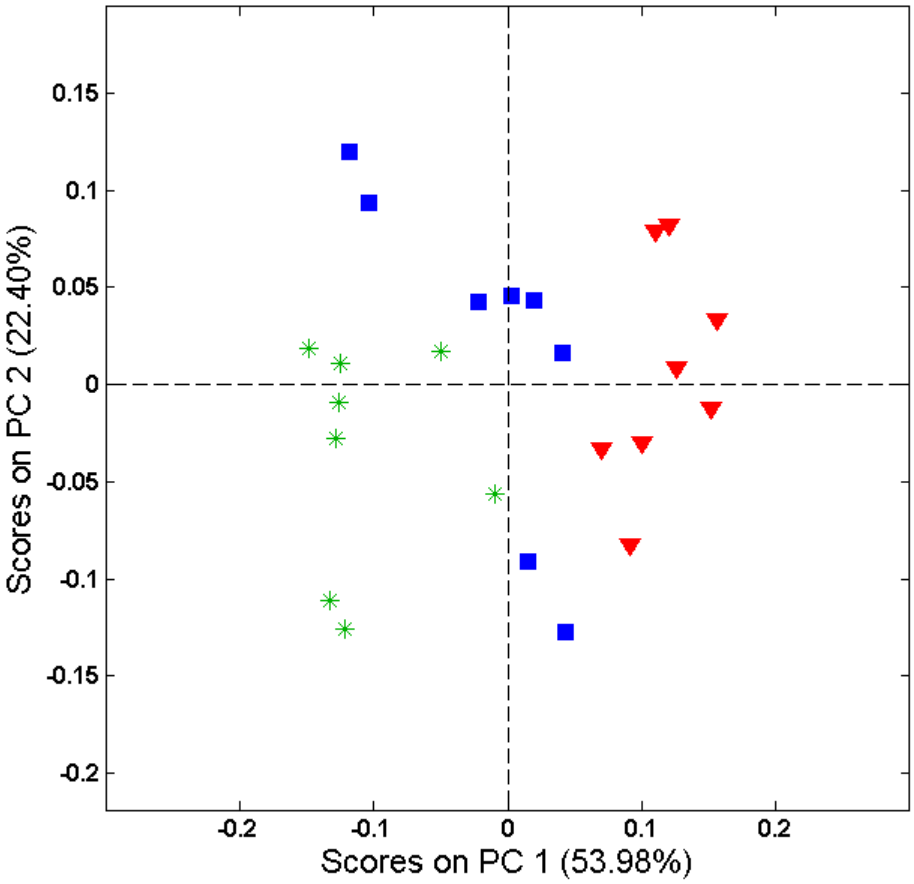

Figure S.6. PCA scores plot corresponding to *M. tuberculosis* Erdman strains. Wild type (⑤), EΔsigC mutant (⓿), and EΔCcomp complement (☒) strains. The PC model was constructed from the SERS spectra of the corresponding species using the spectral range 1700 - 700 cm^{-1}.

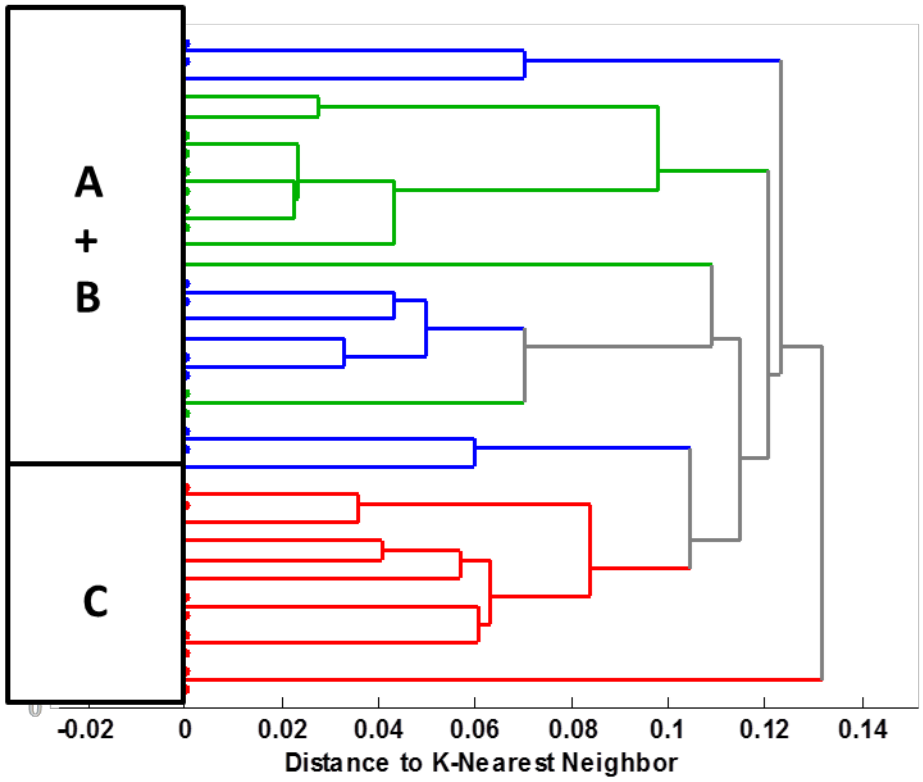

Figure S.7. A hierarchical cluster analysis dendrogram derived from the PC scores of the *M. tuberculosis* Erdman strains. The nodes group into two distinct clusters and are labeled according to the samples. (A + B) wild-type (blue) and *sigC* complement (green) ; (C) *sig C* mutant (red).

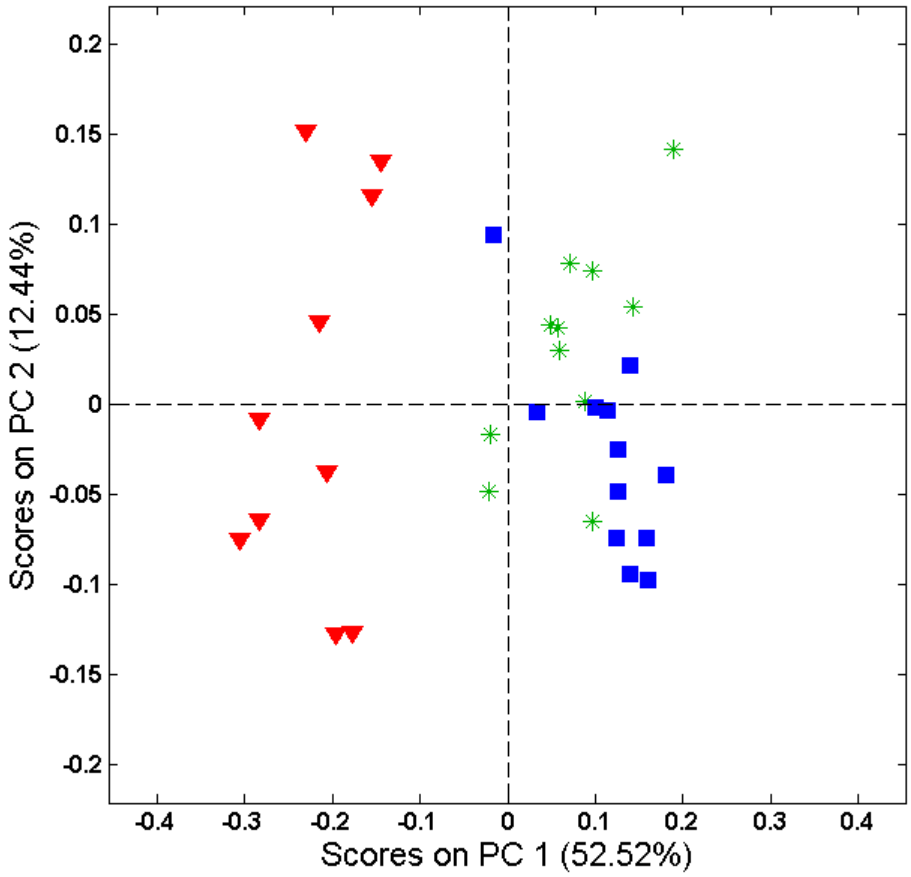

Figure S.8. PCA scores plot corresponding to *M. tuberculosis* H37Rv strains. Wild Type (⑤), RvΔsigC mutant (⓿), and RvΔCcomp complement (☒) strains. The PC model was constructed from the SERS spectra of the corresponding species using the spectral range 1700 - 700 cm^{-1}.

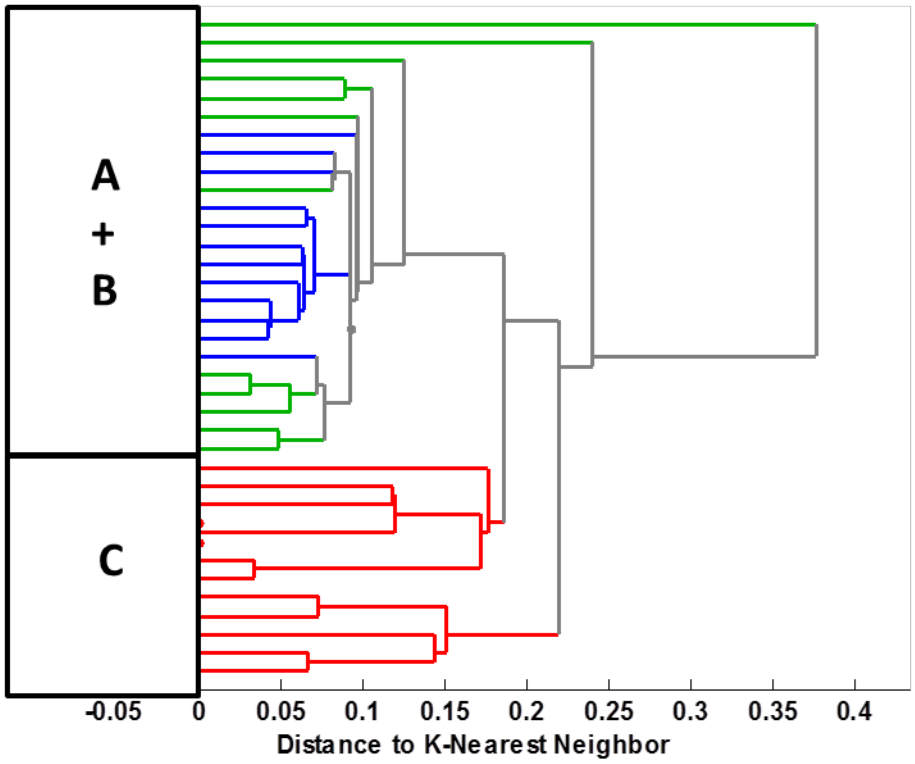

Figure S.9. A hierarchical cluster analysis dendrogram derived from the PC scores of the *M. tuberculosis* H37Rv strains. The nodes group into two distinct clusters and are labeled according to the samples. (A + B) parent (blue) and *sigC* complement (green) ; (C) *sigC* mutant (red).

Table S.1. Observed NMR chemical shift resonances appearing in the ^{1}H-NMR spectra of the non-tuberculous mycobacteria *M. smegmatis*, *M. avium* and *M. bovis* BCG.

Species	***M. smegmatis***	***M. bovis* BCG**	***M. avium***
Mycolic Acid Type Sub-Class	keto-mycolate k4	keto- & alpha k3, α1	keto-mycolate k3, k4
Functional Group	**Observed Resonances (ppm)**		
$(CH_2)_nCH_3$	0.85	0.85	0.85
$CH_2(CH_2)_nCH_2$	1.0 – 1.5	1.0 – 1.5	1.0 – 1.5
R_1=CH-R_2	2.45	2.45	2.45
CHOR		2.96	2.96
$CHOCH_3$		3.32	3.32
CO_2CH_3	3.7	3.7	3.7
trans CH=CH		5.39 , 5.35	5.39 , 5.35
cis CH=CH	5.3		
cis-cyclopropane	0.73	0.66 , -0.31	0.66
trans-cyclopropane			

Table S.2. Observed NMR chemical shift resonances appearing in the ^{1}H-NMR spectra of the *M. tuberculous* clinical Erdman strain.

Species	***Wild Type***	***EΔsigC***	***EΔCcomp***
Mycolic Acid Type Sub-Class	keto-mycolate	keto-mycolate	keto-mycolate
Functional Group	**Observed Resonances (ppm)**		
$(CH_2)_nCH_3$	0.85	0.85	0.85
$CH_2(CH_2)_nCH_2$	1.0 – 1.5	1.0 – 1.5	1.0 – 1.5
R_1=CH-R_2	2.45	2.45	2.45
CHOR		2.96	2.96
$CHOCH_3$		3.32	3.32
CO_2CH_3	3.7	3.7	3.7
trans CH=CH		5.39 , 5.35	
cis CH=CH	5.3		5.3
cis-cyclopropane	0.70 , -0.31	0.70 , -0.31	0.70 , -0.31
trans-cyclopropane	0.60 , 0.10	0.60 , 0.10	0.60 , 0.10

Table S.3. Observed NMR chemical shift resonances appearing in the ^{1}H-NMR spectra of the *M. tuberculous* laboratory-passaged H37Rv strains.

Species	**Wild Type**	**RvΔsigC**	**RvΔCcomp**
Mycolic Acid Type Sub-Class	keto-mycolate	keto-mycolate	keto-mycolate
Functional Group	**Observed Resonances (ppm)**		
$(CH_2)_nCH_3$	0.85	0.85	0.85
$CH_2(CH_2)_nCH_2$	1.0 – 1.5	1.0 – 1.5	1.0 – 1.5
R_1=CH-R_2	2.45	2.45	2.45
CHOR		2.96	2.96
$CHOCH_3$		3.32	3.32
CO_2CH_3	3.7	3.7	3.7
trans CH=CH		5.39 , 5.35	
cis CH=CH	5.3		5.3
cis-cyclopropane	0.70 , -0.36	0.70 , -0.36	0.70 , -0.36
trans-cyclopropane	0.6	0.6	0.6

Table S.4. Representative Raman bands appearing in the SERS spectra of the non-tuberculous mycobacteria *M. smegmatis*, *M. avium* and *M. bovis* BCG.

Raman Shift, cm^{-1}			Vibrational Mode Assignment
M.smegmatis	*M.avium*	*M. bovis* BCG	
1654	1612	1659	(C=C); C=O-β conjugated; v_s(C=O) carboxylic acid
1598	1594	1593	
		1500	v(C=C)
1442	1442	1441	δ(C-H_2) sci.; CH_3 antisym. bend
	1393		C-O-H bend 1° alcohol
	1350	1366	C-O-H bend; $(CH_2)_n$ in-phase twist
1299	1289	1298	C-O-H bend
		1265	cis dialkyl C-H sym. rock
1240	1242		CH-O epoxy ring breathing mode
1210		1210	C-O-H bend
1164	1170	1164	v_{as}(COC); δ(CH)
1129	1135	1135	C-C skel. str in alkane
	1101	1097	v_{as}(COC); C-C skeletal
1058		1065	v(CHR_2) C-C skel. str ; C-C skel. str in alkane; v_{as}(COC)
1028		1030	oop C-C-O stretch
1001	1003	1004	v(C-C-O) out-of-phase stretch of primary alcohol
	970	962	O-H-O wagging; trans dialkyl wag
	931		v(CHR_2) C-C skel. str where R≠CH_3; COO^- str carboxylic acid
893			v(C-C-O) in-phase stretch of primary alcohol, v_S(C-O-C); C-C skel. str in alkane; COO^- str carboxylic acid
841	855		v(C-C-O) in-phase stretch of primary alcohol; C-C skel. str in alkane; COO^- str carboxylic acid
		832	v(C-C-O) in-phase stretch of primary alcohol; v(CHR_2) C-C skel. str where R≠CH_3
773		778	CH_2 in-phase rock; v(CHR_2) -C-C skel. str where R≠CH_3
698	672	696	cis dialkyl C-H wag

Table S.5. Representative Raman bands appearing in the SERS spectra of the *M. tuberculosis* clinical Erdman strains.

Raman Shift, cm^{-1}			Vibrational Mode Assignment
Wild Type	EΔsigC	EΔCcomp	
1592	1586	1592	C=O, alkyl ketone; (C=C); C=O-β conjugated; v_s(C=O) carboxylic acid
1564	1564	1564	C=C
	1458	1492	CH_3 antisym. bend.
1458		1457	δ(C-H_2) sci.
1391	1387	1392	C-O-H bend
1319	1315	1314	C-O-H bend; $(CH_2)_n$ in-phase twist
1275	1280	1280	C-O-H bend
1241	1244	1241	C-O-H bend
1158	1157	1163	
1130		1130	C-C skel. str in alkane
1075	1076	1080	v_{as}(COC)
1008	1005	1008	v(C-C-O) out-of-phase stretch of primary alcohol
963			OHO wagging; trans dialkyl wag
	928		v(CHR_2) C-C skel. str where R≠CH_3;
	892		v(C-C-O) in-phase stretch of primary alcohol, v_S(C-O-C); C-C skel. str in alkane
857	856	852	v(C-C-O) in-phase stretch of primary alcohol; C-C skel. str in alkane
786			v(C-C-O) in-phase stretch of primary alcohol; v(CHR_2) C-C skel. str where R≠CH_3
	714	713	CH_2 in-phase rock; v(CHR_2) C-C skel. str where R≠CH_3
707			cis dialkyl C-H wag

Table S.6. Representative Raman bands appearing in the SERS spectra of the *M. tuberculosis* laboratory-passaged H37Rv strains.

Raman Shift, cm^{-1}			Vibrational Mode Assignment
Wild Type	RvΔsigC	RvΔCcomp	
1592	1586	1592	C=O, alkyl ketone; (C=C); C=O-β conjugated; v_s(C=O) carboxylic acid
1564	1564	1564	C=C
	1491	1497	CH_3 antisym. bend
1457	1460	1458	δ(C-H_2) sci.
1392	1387	1391	C-O-H bend
1314	1315	1319	C-O-H bend; -$(CH_2)_n$- in-phase twist
1280	1280	1275	C-O-H bend
1241	1244	1241	C-O-H bend
1163	1157	1158	
1130	1130	1130	C-C skel. str in Alkane
1080	1076	1075	v_{as}(COC)
1008	1005	1008	v(C-C-O) out-of-phase stretch of primary alcohol
924			v(CHR_2) C-C skel. str where R≠CH_3
		869	v(C-C-O) in-phase stretch of primary alcohol, v_S(C-O-C); C-C skel. str in alkane
852	856	857	v(C-C-O) in-phase stretch of primary alcohol; C-C skel. str in alkane
	816	786	v(C-C-O) in-phase stretch of primary alcohol; v(CHR_2) C-C skel. str where R≠CH_3
713	714		CH_2 in-phase rock; v(CHR_2) C-C skel. str where R≠CH_3
707	713	707	cis dialkyl -CH wag

Table S.7. Quantitative statistics calculated from the PLS-DA model developed from the SERS spectra[a] of the three NTM species *M. avium*, *M. bovis* BCG, and *M. smegmatis*.

Modeled Class[b]	*M. avium*	*M. bovis* BCG	*M. smegmatis*
Sensitivity (CV)[c]	1.000	1.000	1.000
Specificity (CV)	1.000	1.000	1.000
Class. Error (CV)[d]	0.000	0.000	0.000
RMSECV[e]	0.104	0.141	0.140

[a]Thirty-six spectra used, 12 for each NTM species. Before calculation, spectra were pre-processed using Savitzky-Golay 1st derivatives, vector normalized, and mean-centered.

[b]Five latent variables, accounting for 94.86% of the captured variance, was used in this regression model.

[c]CV, cross-validation based on Venetian blinds method with 6 splits

[d]Class. Error, classification error after cross-validation

[e]RMSECV, root-mean square error after cross-validation

Table S.8. Quantitative statistics calculated from the PLS-DA model developed from the SERS spectra[a] of the three *M. tuberculosis* Erdman strains.

Modeled Class[b]	Wild Type	EΔCsigC	EΔCcomp
Sensitivity (CV)[c]	1.000	1.000	1.000
Specificity (CV)	1.000	1.000	1.000
Class. Error (CV)[d]	0.000	0.000	0.000
RMSECV[e]	0.203	0.188	0.142

[a]Thirty-six total spectra used, 12 for each MTB Erdman strain. Before calculation, spectra were pre-processed using Savitzky-Golay 1st derivatives, vector normalization, and mean-centering.

[b]Two latent variables, accounting for 87.88% of the captured variance, was used in this regression model.

[c]CV, cross-validation based on Venetian blinds method with 6 splits

[d]Class. Error, classification error after cross-validation

[e]RMSECV, root-mean square error after cross-validation

Table S.9. Quantitative statistics calculated from the PLS-DA model developed from the SERS spectra[a] of the three *M. tuberculosis* H37Rv strains.

Modeled Class[b]	Wild Type	RvΔCsigC	RvΔCcomp
Sensitivity (CV)[c]	1.000	1.000	0.833
Specificity (CV)	1.000	0.917	0.792
Class. Error (CV)[d]	0.000	0.042	0.188
RMSECV[e]	0.154	0.351	0.352

[a]Thirty-six total spectra used, 12 for each MTB H37Rv strain. Before calculation, spectra were pre-processed using Savitzky-Golay 1st derivatives, vector normalized, and mean-centered.

[b]Three latent variables, accounting for 72.59% of the captured variance, was used in this regression model.

[c]CV, cross-validation based on Venetian blinds method with 6 splits

[d]Class. Error, classification error after cross-validation

[e]RMSECV, root-mean square error after cross-validation