Supporting Information for

Toward Spectral Library-Free Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Bacterial Identification

Ding Cheng[1], Liang Qiao[1,] * and Peter Horvatovich[2,] *

1. Department of Chemistry, Shanghai Stomatological Hospital, Fudan University, Shanghai 200000, China
2. Department of Pharmacy, University of Groningen, 9700 AD Groningen, The Netherlands

*: Correspondence should be addressed to Dr. Liang Qiao (liang_qiao@fudan.edu.cn, Tel: +86 21 31249161) or Prof. Peter Horvatovich (p.l.horvatovich@rug.nl, Tel: +31 50 363 3341)

Table of Contents

S1: Proteomic Database

Fasta File of proteins of 42 genera / 1276 species of microorganisms was downloaded from UniProt (http://www.uniprot.org/). After cleaning according to the rules as explained in the methods section, 83452 protein entries were left, and uploaded to the Github (https://github.com/dipcarbon/BacteriaMSLF).

The list of 42 genera:

Acinetobacter
Bacillus
Bacteroides
Bordetella
Brucella
Buchnera
Burkholderia
Campylobacter
Chlamydia
Chlorobium
Clostridium
Corynebacterium
Escherichia
Flavobacterium
Francisella
Geobacter
Haemophilus
Helicobacter
Klebsiella
Lactobacillus
Lactococcus
Lecanicillium
Legionella
Listeria
Mycobacterium
Mycoplasma
Neisseria
Pasteurella
Prochlorococcus
Pseudomonas
Rhizobium
Rhodopseudomonas
Rickettsia Salmonella
Shewanella
Shigella
Staphylococcus
Streptococcus
Streptomyces
Synechococcus

Vibrio
Xanthomonas
Yersinia

S2: Spectra Details

[C] indicates that the spectrum was selected for training/testing procedure.

1. Spectra obtained from Folkhalsomyndigheten

Mass spectra of 296 microorganisms strains were obtained from Spectra by the Public Health Agency of Sweden (Folkhälsomyndigheten):

Gordonia terrae NRRL B-16283T HKU
Gordonia lacunae NRRL B-24551T HKU
Gordonia hongkongensis HKU50T HKU
Gordonia hongkongensis HKU46 HKU
[C]Yersinia similis CCUG 52882 SMI
[C]Yersinia rohdei SMI Y81-12 SMI
[C]Yersinia rohdei SMI 6-09 SMI
[C]Yersinia rohdei SMI 4-11 SMI
[C]Yersinia rohdei SMI 4-09 SMI
[C]Yersinia rohdei SMI 10-12 SMI
[C]Yersinia pseudotuberculosis SMI 78-08 SMI
[C]Yersinia pseudotuberculosis SMI 46-09 SMI
[C]Yersinia pseudotuberculosis SMI 42-08 SMI
[C]Yersinia pseudotuberculosis SMI 33-12 SMI
[C]Yersinia pseudotuberculosis SMI 33-08 SMI
[C]Yersinia pseudotuberculosis SMI 26-10 SMI
[C]Yersinia pseudotuberculosis SMI 24-12 SMI
[C]Yersinia pseudotuberculosis SMI 18-09 SMI
[C]Yersinia pseudotuberculosis SMI 10-10 SMI
[C]Yersinia pestis NCTC 8775 SMI
[C]Yersinia pestis NCTC 570 SMI
[C]Yersinia pestis NCTC 2868 SMI
[C]Yersinia pestis NCTC 2028-4 SMI
[C]Yersinia pestis NCTC 144 SMI
[C]Yersinia pestis NCTC 10330 SMI
[C]Yersinia pestis NCTC 10329-3 SMI
[C]Yersinia pestis NCTC 10030 SMI
[C]Yersinia pestis KIM5 SMI
[C]Yersinia pekkanenii CCUG 60991 SMI
[C]Yersinia mollaretii SMI 65-09 SMI
[C]Yersinia mollaretii SMI 6-10 SMI
[C]Yersinia mollaretii SMI 30-12 SMI
[C]Yersinia mollaretii SMI 25-12 SMI
[C]Yersinia mollaretii CCUG 37901 SMI
[C]Yersinia massiliensis SMI 37-07 SMI
[C]Yersinia massiliensis SMI 32-11 SMI
[C]Yersinia massiliensis SMI 2-08 SMI

[C]Yersinia kristensenii SMI Y35-13 SMI
[C]Yersinia intermedia ATCC 29909 SMI
[C]Yersinia frederiksenii SMI Y71-12 SMI
[C]Yersinia frederiksenii SMI Y34-13 SMI
[C]Yersinia frederiksenii SMI Y3-12 SMI
[C]Yersinia frederiksenii SMI Y1-11 SMI
[C]Yersinia enterocolitica CCUG 8252 SMI
[C]Yersinia enterocolitica CCUG 8251 SMI
[C]Yersinia enterocolitica CCUG 8250 SMI
[C]Yersinia enterocolitica CCUG 8249 SMI
[C]Yersinia enterocolitica CCUG 8248 SMI
[C]Yersinia enterocolitica CCUG 8245 SMI
[C]Yersinia enterocolitica CCUG 8244 SMI
[C]Yersinia enterocolitica CCUG 8243 SMI
[C]Yersinia enterocolitica CCUG 8240 SMI
[C]Yersinia enterocolitica CCUG 8238 SMI
[C]Yersinia enterocolitica CCUG 8237 SMI
[C]Yersinia enterocolitica CCUG 8236 SMI
[C]Yersinia enterocolitica CCUG 8235 SMI
[C]Yersinia enterocolitica CCUG 8233 SMI
[C]Yersinia enterocolitica CCUG 8232 SMI
[C]Yersinia enterocolitica CCUG 8231 SMI
[C]Yersinia enterocolitica CCUG 4588 SMI
[C]Yersinia enterocolitica CCUG 37895 SMI
[C]Yersinia enterocolitica CCUG 33553 SMI
[C]Yersinia enterocolitica CCUG 31436 SMI
[C]Yersinia enterocolitica CCUG 29125 SMI
[C]Yersinia enterocolitica CCUG 12369 SMI
[C]Yersinia bercovieri SMI Y45-13 SMI
[C]Yersinia bercovieri SMI Y44-13 SMI
[C]Yersinia bercovieri SMI Y36-13 SMI
[C]Yersinia bercovieri SMI Y33-13 SMI
[C]Yersinia bercovieri SMI Y16-13 SMI
[C]Yersinia bercovieri SMI Y07-13 SMI
[C]Yersinia bercovieri SMI 5-12 SMI
[C]Yersinia bercovieri SMI 24-09 SMI
[C]Yersinia bercovieri SMI 23-12 SMI
[C]Yersinia bercovieri SMI 16-12 SMI
[C]Yersinia bercovieri SMI 14-12 SMI
[C]Yersinia bercovieri CCUG 37900 SMI
[C]Vibrio cholerae SMI V6-13 SMI
[C]Staphylococcus pseudintermedius AB 680 SVA
[C]Staphylococcus aureus ATCC 29213 SMI
[C]Pseudomonas aeruginosa ATCC 27853 SMI
Pasteurella pneumotropica LDJ6208-12
Pasteurella pneumotropica LDJ507-12

Parengyodontium album UAMH 9836 HKU
Parengyodontium album UAMH 8313 HKU
Parengyodontium album UAMH 4512 HKU
Parengyodontium album UAMH 1441 HKU
Parengyodontium album UAMH 11234 HKU
Parengyodontium album UAMH 10043 HKU
Parengyodontium album NRRL 2312 HKU
Parengyodontium album IHEM 4198 HKU
Parengyodontium album HKU48 HKU
Parengyodontium album CBS 836.71 HKU
Parengyodontium album CBS 570.71 HKU
Parengyodontium album CBS 504.83 HKU
Parengyodontium album CBS 368.72 HKU
Parengyodontium album CBS 121919 HKU
Legionella worsleiensis CCUG 44923 SMI
Legionella waltersii CCUG 44924 SMI
Legionella wadsworthii ATCC 33877 SMI
Legionella tucsonensis ATCC 49180 SMI
Legionella taurinensis SMI L005 SMI
Legionella taurinensis CCUG 44901 SMI
Legionella steigerwaltii CCUG 56441 SMI
Legionella spiritensis ATCC 35249 SMI
Legionella shakespearei CCUG 31228A SMI
Legionella sainthelensi SMI L025 SMI
Legionella rubrilucens SMI L026 SMI
Legionella rubrilucens SMI L023 SMI
Legionella rubrilucens SMI L022 SMI
Legionella rubrilucens SMI L021 SMI
Legionella rubrilucens SMI L014 SMI
Legionella rowbothamii CCUG 43886 SMI
Legionella quinlivanii ATCC 43830 SMI
Legionella quateirensis CCUG 44899 SMI
Legionella pneumophila ssp pneumophila OLDA ATCC 43109 SMI
Legionella pneumophila ssp pneumophila NCTC 11191 SMI
Legionella pneumophila ssp pneumophila ATCC 43703 SMI
Legionella pneumophila ssp pneumophila ATCC 43290 SMI
Legionella pneumophila ssp pneumophila ATCC 43283 SMI
Legionella pneumophila ssp pneumophila ATCC 43111
Legionella pneumophila ssp pneumophila ATCC 35096 SMI
Legionella pneumophila ssp pneumophila ATCC 33823 SMI
Legionella pneumophila ssp pneumophila ATCC 33215 SMI
Legionella pneumophila ssp pneumophila ATCC 33154 SMI
Legionella pneumophila ssp fraseri ATCC 33216 SMI
Legionella pneumophila ssp fraseri ATCC 33156 SMI
Legionella pneumophila HPA EQA1 SMI
Legionella pneumophila ATCC 33152 SMI

Legionella parisiensis ATCC 35299 SMI
Legionella oakridgensis SMI L013 SMI
Legionella oakridgensis SMI L012 SMI
Legionella oakridgensis ATCC 33761 SMI
Legionella norrlandica SMI LEGN SMI
Legionella norrlandica SMI 4329II SMI
Legionella norrlandica SMI 4327I SMI
Legionella nautarum CCUG 44900 SMI
Legionella micdadei ATCC 33218 SMI
Legionella maceachernii SMI 4270I SMI
Legionella longbeachae SMI L002 SMI
Legionella longbeachae SMI L001 SMI
Legionella longbeachae ATCC 33484 SMI
Legionella longbeachae ATCC 33462 SMI
Legionella londiniensis ATCC 700510 SMI
Legionella jordanis ATCC 33623 SMI
Legionella jordanis 4308I SMI
Legionella israelensis ATCC 43119 SMI
Legionella hackeliae ATCC 35250 SMI
Legionella gresilensis CCUG 45920 SMI
Legionella gratiana CCUG 44894 SMI
Legionella gormanii SMI L024 SMI
Legionella gormanii SMI L006 SMI
Legionella gormanii ATCC 33297 SMI
Legionella geestiana CCUG 44893 SMI
Legionella feeleii SMI L017 SMI
Legionella feeleii SMI L016 SMI
Legionella feeleii SMI L015 SMI
Legionella fallonii CCUG 43887 SMI
Legionella erythra ATCC 35303 SMI
Legionella dumoffii ATCC 33279 SMI
Legionella drozanskii CCUG 43885 SMI
Legionella dresdenensis CCUG 59858 SMI
Legionella cincinnatiensis ATCC 43753 SMI
Legionella cherrii ATCC35252 SMI
Legionella bozemanii SMI L028 SMI
Legionella bozemanii SMI L027 SMI
Legionella bozemanii SMI L020 SMI
Legionella bozemanii SMI L011 SMI
Legionella bozemanii SMI L010 SMI
Legionella bozemanii SMI L008 SMI
Legionella bozemanii SMI L004 SMI
Legionella bozemanii SMI L003 SMI
Legionella bozemanii SMI F41ENK-M SMI
Legionella bozemanii SMI F41ENK SMI
Legionella bozemanii ATCC 35545 SMI

Legionella bozemanii ATCC 33217 SMI
Legionella birminghamiensis ATCC 43702 SMI
Legionella anisa SMI L019 SMI
Legionella anisa ATCC 35292 SMI
Legionella adelaidensis CCUG 31231B SMI
Lecanicillium tenuipes CBS 658.80 HKU
Klebsiella pneumoniae KLIN K697 SMI
Klebsiella pneumoniae KLIN K695 SMI
Klebsiella pneumoniae KLIN K693 SMI
Klebsiella pneumoniae KLIN K691 SMI
Klebsiella pneumoniae KLIN 700 SMI
Klebsiella pneumoniae CCUG 58547 SMI
Klebsiella pneumoniae CCUG 58546 SMI
Klebsiella pneumoniae CCUG 58545 SMI
Klebsiella pneumoniae CCUG 56233 SMI
Klebsiella pneumoniae CCUG 45421 SMI
[C]Francisella tularensis ssp tularensis RKI Schu4 SMI
[C]Francisella tularensis ssp tularensis FOI FSC046 SMI
[C]Francisella tularensis ssp tularensis FOI FSC042 SMI
[C]Francisella tularensis ssp tularensis FOI 237 SMI
[C]Francisella tularensis ssp tularensis FOI 054 SMI
[C]Francisella tularensis ssp tularensis FOI 041 SMI
[C]Francisella tularensis ssp novicida RKI Ft27 SMI
[C]Francisella tularensis ssp novicida RKI Ft26 SMI
[C]Francisella tularensis ssp novicida QUANDHIP RG3-05 SMI
[C]Francisella tularensis ssp mediasiatica RKI Ft31 SMI
[C]Francisella tularensis ssp mediasiatica RKI Ft24 SMI
[C]Francisella tularensis ssp holarctica SMI R14-23 SMI
[C]Francisella tularensis ssp holarctica RKI Q2-6 SMI
[C]Francisella tularensis ssp holarctica RKI Ft101 6 SMI
[C]Francisella tularensis ssp holarctica RKI A146 1 SMI
[C]Francisella tularensis ssp holarctica RKI A104 15 SMI
[C]Francisella tularensis ssp holarctica RKI A101 4 SMI
[C]Francisella tularensis ssp holarctica FOI 152 SMI
[C]Francisella tularensis ssp holarctica FOI 151 SMI
[C]Francisella tularensis ssp holarctica FOI 133 SMI
Flavobacterium psychrophilum F510
Flavobacterium columnare F512
Engyodontium rectidentatum CBS 641.74 HKU
Engyodontium rectidentatum CBS 547.82 HKU
Engyodontium rectidentatum CBS 206.74 HKU
Engyodontium parvisporum IHEM 22910 HKU
[C]Burkholderia pseudomallei RKI Q2-1 SMI
[C]Burkholderia pseudomallei RKI Bp9 SMI
[C]Burkholderia pseudomallei RKI A173 SMI
[C]Burkholderia pseudomallei RKI A110 SMI

[C]Burkholderia pseudomallei RKI A106 SMI
[C]Burkholderia pseudomallei RKI 104 SMI
[C]Burkholderia mallei RKI Bm9 SMI
[C]Burkholderia mallei RKI A106 4 SMI
[C]Burkholderia mallei RKI A106 3 SMI
[C]Burkholderia mallei RKI A104 4 SMI
[C]Brucella suis NCTC 1720 SMI
[C]Brucella suis NCTC 10316 SMI
[C]Brucella suis NCTC 1030 SMI
[C]Brucella pinnipedialis NCTC 12890 SMI
[C]Brucella ovis NCTC 10512 SMI
[C]Brucella neotomae ATCC 23459 SMI
[C]Brucella microti CAPM 6434 SMI
[C]Brucella melitensis SMI R13-100 SMI
[C]Brucella melitensis SMI R12-209 SMI
[C]Brucella melitensis RKI A104-11 SMI
[C]Brucella melitensis NCTC 16M SMI
[C]Brucella melitensis NCTC 10503 SMI
[C]Brucella melitensis NCTC 10094 SMI
[C]Brucella inopinata CAPM 6436 SMI
[C]Brucella ceti NCTC 12891 SMI
[C]Brucella canis SVA 007450-2 SMI
[C]Brucella canis SVA 007041-6 SMI
[C]Brucella canis NCTC 10854 SMI
[C]Brucella canis Evira E20140122-166 SMI
[C]Brucella abortus NCTC 624 SMI
[C]Brucella abortus NCTC 544 SMI
[C]Bordetella pertussis SMI 22 SMI
[C]Bordetella pertussis SMI 17 SMI
[C]Bordetella parapertussis SMI 18 SMI
[C]Bordetella parapertussis SMI 16 SMI
[C]Bordetella bronchiseptica SMI 20 SMI
[C]Bordetella bronchiseptica SMI 19 SMI
[C]Bordetella avium SMI 21 SMI
[C]Bacillus weihenstephanensis DSM 11821 SMI
[C]Bacillus thuringiensis CIP 105674 SMI
[C]Bacillus thuringiensis BGSC 4Y1 SMI
[C]Bacillus thuringiensis BGSC 4BD1 SMI
[C]Bacillus thuringiensis BGSC 4AW1 SMI
[C]Bacillus thuringiensis BGSC 4AJ1 SMI
[C]Bacillus thuringiensis ATCC 35646 SMI
[C]Bacillus thuringiensis ATCC 10792 SMI
[C]Bacillus thuringiensis 4CC1 SMI
[C]Bacillus flexus SMI M14-03 SMI
[C]Bacillus cereus WSBC 10483 SMI
[C]Bacillus cereus WSBC 10286 SMI

[C]Bacillus cereus SJ1 SMI
[C]Bacillus cereus RIVM BC67 SMI
[C]Bacillus cereus NVH 1518-99 SMI
[C]Bacillus cereus NVH 0597-99 SMI
[C]Bacillus cereus NVH 0597-98 SMI
[C]Bacillus cereus MHI 1305 SMI
[C]Bacillus cereus MHI 13 SMI
[C]Bacillus cereus IH41385 SMI
[C]Bacillus cereus F837-76 SMI
[C]Bacillus cereus F4810-72 SMI
[C]Bacillus cereus F2085-98 SMI
[C]Bacillus cereus ATCC 4342 SMI
[C]Bacillus cereus ATCC 10987 SMI
[C]Bacillus cereus ATCC 10702 SMI
[C]Bacillus anthracis SVA11-BKT082753 SMI
[C]Bacillus anthracis SVA081216 SMI
[C]Bacillus anthracis SVA 13Bakt082753 SMI
[C]Bacillus anthracis SMI R82-04 SMI
[C]Bacillus anthracis NC08234 SMI
[C]Bacillus anthracis NC07753-03 SMI
[C]Bacillus anthracis NC07752-04 SMI
[C]Bacillus anthracis NC01328
[C]Bacillus anthracis ATCC 4229 SMI
[C]Bacillus anthracis 7702 SMI
[C]Bacillus anthracis 10340 SMI
[C]Bacillus anthracis 05444-02 SMI
[C]Bacillus anthracis 02620-03 SMI
[C]Bacillus anthracis 00109-03 SMI
Acinetobacter haemolyticus SMI A058 SMI

2. FoodBIMS

Mass spectra of strains from 11 different genera (*Escherichia, Listeria, Yersinia, Staphylococcus, Salmonella, Microccocus, Morganella, Pseudomonas, Proteus, Sarcina, Lactobacillus*) were obtained from FoodBIMS by Institute of Food Sciences, National Research Council of Italy:

Yersinia kristensenii NCTC 11471 / ATCC 33638
Yersinia intermedia NCTC 11469 / ATCC 29909
Yersinia frederiksenii NCTC 11470 / ATCC 33641
Yersinia enterocolitica subsp. Enterocolitica DSM 9499 / NCTC 11174 / ATCC 700823
Yersinia enterocolitica subsp
Staphylococcus aureus ATCC 6538 / DSM 799 / NCTC 10788
[C]Staphylococcus aureus ATCC 25923
[C]Staphylococcus aureus ATCC 14458
Sarcina flava
[C]Salmonella typhimurium ATCC 6994

Salmonella enteritidis
[C]Pseudomonas putida
[C]Pseudomonas fluorescens
Proteus vulgaris
Proteus mirabilis
Morganella morganii
Micrococcus spp
Listeria welshimeri
Listeria seeligeri
Listeria monocytogenes type strain ATCC 15313 / NCTC 10357
[C]Listeria monocytogenes ATCC 9525
Listeria monocytogenes ATCC 19114 / 5214
[C]Listeria ivanovii 750
Listeria innocua
Leuconostoc mesenteroides
[C]Lactococcus lactis
[C]Escherichia vulneris DSMZ 4564
[C]Escherichia fergusonii DSMZ 13698
[C]Escherichia coli
[C]Escherichia coli O157 H7 NCTC
[C]Escherichia coli DSMZ 9033
[C]Escherichia coli DSMZ 9031
[C]Escherichia coli DSMZ 9030
[C]Escherichia coli DSMZ 9029
[C]Escherichia coli DSMZ 9024
[C]Escherichia coli DSMZ 8711
[C]Escherichia coli DSMZ 8700
[C]Escherichia coli DSMZ 8696
[C]Escherichia coli DSMZ 8695
[C]Escherichia coli DSMZ 5802
[C]Escherichia coli DSMZ 301
[C]Escherichia coli DSMZ 30083
[C]Escherichia coli DSMZ 11752
[C]Escherichia coli DSMZ 10722

3. URMS spectra

Mass spectra of 90 species of microorganisms were obtained from URMS Database by Unité des Rickettsies, France:

Acidaminococcus massiliensis CSUR P2828
Vaginella massiliensis CSUR P2517
Timonella senegalensis
Streptococcus timonensis CSUR P2915
Senegalemassilia anaerobia
Ruminococcus phoceensis CSUR P2086

Romboutsia timonensis CSUR P326
Rasbobacterium massiliensis
Prevotella phoceensis CSUR P2259
Prevotella caccae CSUR P2931
Peptoniphilus vaginalis
Peptoniphilus timonensis
Peptoniphilus senegalensis
Peptoniphilus phoceensis CSUR P2183
Peptoniphilus pacaensis
Peptoniphilus obesi
Peptoniphilus grossensis
Peptoniphilus duodenumensis CSUR P2932
Peptoiphilus raoultii
Paenibacillus ihumii CSUR P1981
Olsenella phoceensis CSUR P2936
Olegusella massiliensis
Oceanobacillus massiliensis
Nesterenkonia massiliensis
Neglecta timonensis CSUR P2265
Negativicoccus massiliensis CSUR P2082
Murdochiella vaginalis
Micromassilia vaginalis CSUR P2986
Megasphaera massiliensis
Mediterranea massiliensis CSUR P2645
Massiliomicrobiota timonensis CSUR P2264
Massilioculturomica massiliensis CSUR P2935
Marseilla massiliensis CSUR P2745
Levasseurella massiliensis CSUR P3275
Lascolacoccus vaginalis
Lascolabacillus massiliensis
Lachnoclostridium touaregense CSUR P2415
Lachnoclostridium massiliosenegalense CSUR P299
Kallipyga massiliensis
Intestinimonas massiliensis CSUR P1930
Ihuprevotella massiliensis CSUR P2826
Hugonella massiliensis CSUR P2118
Haloferax massiliensis CSUR P194
Fusobacterium massiliensis CSUR P2749
Fournierella massiliensis CSUR P2014
Ezakiella massiliensis CSUR P2951
Enterococcus massiliensis CSUR P1928
Enterobacter massiliensis
Enorma massiliensis
Emergencia timonense CSUR P2260
Eisenbergiella massiliensis CSUR P2120
Duodena massiliensis CSUR P2968

Dielma fastidiosa
Corynebacterium timonense
Corynebacterium jeddahense
Corynebacterium fournierii CSUR P2948
Colonella massiliensis CSUR P2911
Collinsella vaginalis CSUR P2666
Clostridium senegalense
Clostridium saudii
Clostridium nigeriense CSUR P2414
Clostridium massiliodielmoense CSUR P2255
Clostridium jeddahense
Clostridium dakarense
Clostridium bouchedurhonense CSUR P2181
Christensenella timonensis CSUR P2437
Christensenella massiliensis CSUR P2438
Cellulomonas massiliensis
Caecumella massiliensis CSUR P2974
Butyricimonas phoceensis CSUR P2742
Brevibacillus massiliensis
Bretellus vaginalis CSUR P2746
Blautia timonensis CSUR P2398
Bartonella florenciae
Bariatricus massiliensis CSUR P2221
Bacteroides mediterraneense CSUR P2644
[C]Bacillus timonensis
[C]Bacillus massiliosenegalensis
[C]Bacillus massilionigeriensis CSUR P2366
[C]Bacillus massilioglaciei CSUR P2600P
Arcanobacterium urinomassiliense CSUR P3240
Anaerotruncus massiliensis CSUR P2007
Anaeromassilibacillus senegalense CSUR P1511
Anaerococcus mediterranensis CSUR P2765
Alistipes timonensis
Alistipes senegalensis
Alistipes obesi
Africanella massiliensis CSUR P2538
Actinomyces ihumii CSUR P2006
Actinomyces bouchedurhonensis CSUR P2825

4. EMbaRC Spectra

Mass spectra of 204 species of microorganisms were obtained from European Consortium of Microbial Resources Centres.

[C]Bacillus amyloliquefaciens ssp amyloliquefaciens DSM 7
[C]Bacillus anthracis A112
[C]Bacillus anthracis A119
[C]Bacillus anthracis A126
[C]Bacillus anthracis A128
[C]Bacillus anthracis A133
[C]Bacillus anthracis A38
[C]Bacillus anthracis A86
[C]Bacillus anthracis ATCC 4229
[C]Bacillus anthracis NC01328 SMI
[C]Bacillus cereusATCC 10987
[C]Bacillus cereus B10
[C]Bacillus cereus B108
[C]Bacillus cereus B146
[C]Bacillus cereus B16
[C]Bacillus cereus B181
[C]Bacillus cereus B185
[C]Bacillus cereus B192
[C]Bacillus cereus B194
[C]Bacillus cereus B198
[C]Bacillus cereus B238
[C]Bacillus cereus B245
[C]Bacillus cereus B252
[C]Bacillus cereus B253
[C]Bacillus cereus B256
[C]Bacillus cereus B257
[C]Bacillus cereus B260
[C]Bacillus cereus B265
[C]Bacillus cereus B271
[C]Bacillus cereus B272
[C]Bacillus cereus B274
[C]Bacillus cereus B277
[C]Bacillus cereus B289
[C]Bacillus cereus B294
[C]Bacillus cereus B296
[C]Bacillus cereus B297
[C]Bacillus cereus B298
[C]Bacillus cereus B299
[C]Bacillus cereus B300
[C]Bacillus cereus B304
[C]Bacillus cereus B305

[C]Bacillus cereus B306
[C]Bacillus cereus B309
[C]Bacillus cereus B329
[C]Bacillus cereus B33
[C]Bacillus cereus B332
[C]Bacillus cereus B69
[C]Bacillus cereus DSM 4490
[C]Bacillus cereus WIS-St. Nr. 283 ATCC 12826
[C]Bacillus cereus WIS-St. Nr. 29
[C]Bacillus cereus WSBC 10441
[C]Bacillus cereus biovar anthracis Affenstamm (B) SK1b
[C]Bacillus mycoides DSM2048
[C]Bacillus subtilis B130
[C]Bacillus subtilis B154
[C]Bacillus subtilis B156
[C]Bacillus subtilis DSM 10
[C]Bacillus subtilis DSM 347
[C]Bacillus subtilis WIS-St. Nr. 491
[C]Bacillus subtilis ssp. inaquosorum DSM 22148
[C]Escherichia coliNm I
[C]Escherichia coli 08 01585
[C]Escherichia coli ATCC 25922 DSM 1103
[C]Escherichia coli DSM 3871 (K12)
[C]Escherichia coli NCTC 104118
[C]Escherichia coli O157 H7 380 UD
[C]Escherichia coli O157 H7 381
[C]Escherichia coli O157 H7 643
[C]Escherichia coli O157 H7 719
[C]Escherichia coli O157 H7 ATCC 43894
[C]Escherichia coli O157 H7 ED 82
[C]Escherichia coli O157 H7 MI 380
[C]Escherichia coli RKI A139-1
[C]Escherichia coli RKI A139-2
Klebsiella oxytoca ATCC 13182
Klebsiella oxytoca CCUG 15788
Klebsiella oxytoca Oman 61
Klebsiella pneumoniae ssp ozeanae ATCC 25926
[C]Lactobacillus amylovorus CECT 4132T
[C]Lactobacillus amylovorus DSM 20531T
[C]Lactobacillus amylovorus LMG 9496T
[C]Lactobacillus brevis CECT 4121T
[C]Lactobacillus brevis CECT 4669
[C]Lactobacillus brevis CECT 5172
[C]Lactobacillus brevis CECT 5354
[C]Lactobacillus brevis CECT 5918
[C]Lactobacillus brevis DSM 20054T

[C]Lactobacillus brevis DSM 20556
[C]Lactobacillus brevis LMG 6906T
[C]Lactobacillus coryniformis subsp. coryniformis LMG 9196T
[C]Lactobacillus crispatus CIP 102990T
[C]Lactobacillus crispatus DSM 20584T
[C]Lactobacillus delbrueckii subsp. bulgaricus DSM 20080
[C]Lactobacillus delbrueckii subsp. delbrueckii DSM 20074T
[C]Lactobacillus delbrueckii subsp. lactis CIP 101028T
[C]Lactobacillus delbrueckii subsp. lactis CIP 53_61
[C]Lactobacillus delbrueckii subsp. lactis CIP 54_4
[C]Lactobacillus fermentum CECT 4007T
[C]Lactobacillus fermentum CECT 562
[C]Lactobacillus fermentum CIP 53_163
[C]Lactobacillus gallinarum DSM 10532T
[C]Lactobacillus gallinarum LMG 9435T
[C]Lactobacillus gasseri DSM 20077
[C]Lactobacillus iners DSM 13335T
[C]Lactobacillus paracasei LMG 9192T
[C]Lactobacillus paracasei subsp. paracasei CIP 103918T
[C]Lactobacillus paracasei subsp. paracasei LMG 13087T
[C]Lactobacillus paracasei subsp. paracasei LMG 9438
[C]Lactobacillus paracasei subsp. tolerans CECT 4175T
[C]Lactobacillus paracasei subsp. tolerans CIP 103024T
[C]Lactobacillus paracasei subsp. tolerans DSM 20258T
[C]Lactobacillus paracasei subsp. tolerans LMG 9191T
[C]Lactobacillus pentosus CIP 103156T
[C]Lactobacillus pentosus LMG 10755T
[C]Lactobacillus plantarum CIP 71_39
[C]Lactobacillus plantarum subsp. argentoratensis LMG 9205T
[C]Lactobacillus reuteri CIP 101887T
[C]Lactobacillus sanfranciscensis LMG 16002T
[C]Lactococcus garvieae DSM 20064
[C]Lactococcus garvieae DSM 20684T
[C]Lactococcus lactis subsp. lactis DSM 20250
Legionella longbeachae ATCC 33462 SMI
Legionella longbeachae SMI L001 SMI
[C]Listeria monocytogenes type strain ATCC 15313 NCTC 10357
[C]Pseudomonas aeruginosaA-269_02 B
[C]Pseudomonas aeruginosa CECT 4407
[C]Pseudomonas aeruginosa CIP 103282
[C]Pseudomonas aeruginosa CIP 103467
[C]Pseudomonas aeruginosa CIP 103837
[C]Pseudomonas aeruginosa CIP 104116
[C]Pseudomonas aeruginosa CIP 104467
[C]Pseudomonas aeruginosa CIP 105094
[C]Pseudomonas aeruginosa CIP 59_35

[C]Pseudomonas aeruginosa CIP 59_43
[C]Pseudomonas aeruginosa DSM 1117
[C]Pseudomonas aeruginosa DSM 1128
[C]Pseudomonas aeruginosa DSM 1253
[C]Pseudomonas aeruginosa DSM 1299
[C]Pseudomonas aeruginosa DSM 3227
[C]Pseudomonas aeruginosa DSM 46358
[C]Pseudomonas aeruginosa DSM 50069
[C]Pseudomonas aeruginosa DSM 50071T
[C]Pseudomonas aeruginosa DSM 939
[C]Pseudomonas aeruginosa HST 50426
[C]Pseudomonas brassicacearum subsp. brassicacearum DSM 13227T
[C]Pseudomonas chlororaphis CECT 4470T
[C]Pseudomonas chlororaphis subsp. aureofaciens DSM 6698T
[C]Pseudomonas chlororaphis subsp. chlororaphis DSM 50083T
[C]Pseudomonas fluorescens CECT 378T
[C]Pseudomonas fluorescens DSM 50090T
[C]Pseudomonas fluorescens DSM 50091
[C]Pseudomonas fluorescens DSM 50108
[C]Pseudomonas meliae DSM 6759T
[C]Pseudomonas mendocina DSM 50017T
[C]Pseudomonas psychrotolerans DSM 15758T
[C]Pseudomonas putida CECT 324T
[C]Pseudomonas putida DSM 084
[C]Pseudomonas putida DSM 12735
[C]Pseudomonas putida DSM 291T
[C]Pseudomonas putida DSM 3263
[C]Pseudomonas putida DSM 46608
[C]Pseudomonas putida DSM 50026
[C]Pseudomonas putida DSM 7314
[C]Pseudomonas putida II
[C]Pseudomonas stutzeri CECT 930T
[C]Pseudomonas stutzeri DSM 13592
[C]Pseudomonas stutzeri DSM 13627
[C]Pseudomonas stutzeri DSM 50227
[C]Pseudomonas stutzeri DSM 5190T
[C]Pseudomonas stutzeri DSM 6082
[C]Pseudomonas syringae DSM 10604T
[C]Pseudomonas syringae DSM 50312
[C]Pseudomonas syringae DSM 50315
[C]Salmonella enterica serovar minnesota SF 1111
[C]Salmonella typhimurium T104 08 01627
[C]Vibrio alginolyticus CECT 521T
[C]Vibrio breoganii CECT 7222T
[C]Vibrio cholerae A171-1
[C]Vibrio cholerae A222

[C]Vibrio cholerae A225 O1 Toxin positiv
[C]Vibrio cholerae A226-1
[C]Vibrio cholerae A227
[C]Vibrio cholerae A240
[C]Vibrio cholerae A312
[C]Vibrio cholerae A313
[C]Vibrio cholerae A314
[C]Vibrio cholerae A325
[C]Vibrio cholerae A327
[C]Vibrio cholerae A339
[C]Vibrio cholerae A351
[C]Vibrio cholerae A361
[C]Vibrio cholerae A371 O1 Toxin negativ
[C]Vibrio cholerae A378
[C]Vibrio cholerae Inaba 1573
[C]Vibrio cholerae NIH35A3
[C]Vibrio harveyi CECT 525T
[C]Vibrio mimicus CECT 4218T
[C]Yersinia enterocolitica DSM 4780
[C]Yersinia enterocolitica O15 31089
[C]Yersinia enterocolitica O3 29211
[C]Yersinia enterocolitica O50,51 1
[C]Yersinia enterocolitica O9 7192
[C]Yersinia frederiksenii NCTC 11470 ATCC 33641
[C]Yersinia pestis 10329

S3: Optimization of the number of the most informative genes for bacterial identification (Figure S1)

We have assessed the effect of using 5, 10 and 15 most informative genes for the library-free MALDI-TOF MS bacterial identification. The most informative genes were obtained individually with the double cross-validation procedure. As shown in **Figure S1**, identification accuracy was clearly lower using 5 genes. The identification accuracies were similar using 10 or 15 most informative genes, while the computation time was much longer when 15 most informative genes were considered. Therefore, the 10 most informative genes were used in this study for library-free MALDI-TOF MS bacterial identification.

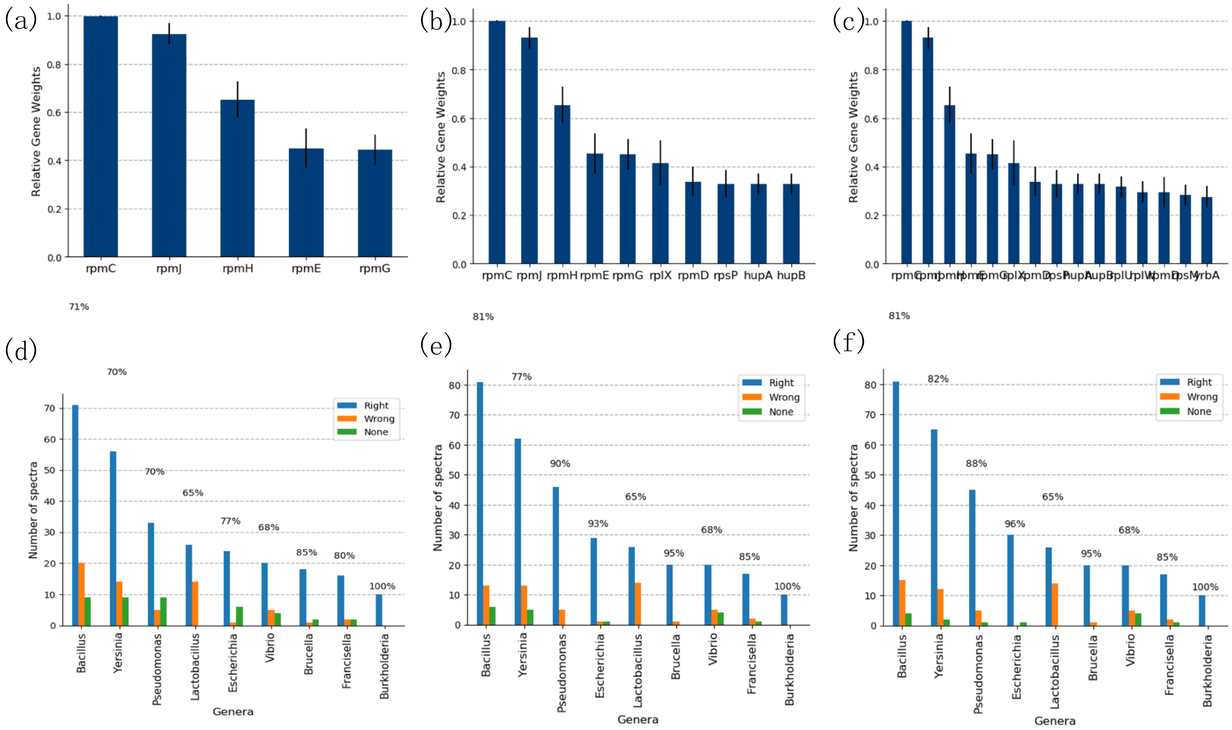


Figure S1. The 5 (a), 10 (b) and 15 (c) most informative genes obtained with the double cross-validation procedure. Identification accuracy of mass spectra from different genera using the 5 (d), 10 (e) and 15 (f) most informative genes.

S4: Example of identification

For clarity, we present here an example mass spectrum identification obtained from *Bacillus anthracis* with our algorithms.

Usually, the gene weights table (***W***) is obtained from double cross-validation training. But in this example, we use the gene weights presented in **Figure 3**:

rpmc	0.9993
rpmj	0.9311
rpmh	0.6529
rpme	0.4537
rpmg	0.4510
rplx	0.4146
rpmd	0.3375
rpsp	0.3289
hupa	0.3284
hupb	0.3271

Calculation details of spectrum identification:

Step 1:

Get the molecular weight comparison matrix (***M***), where the columns correspond to different genes and the rows correspond to different species. The numbers in the cells correspond to the molecular weight of the proteins encoded by the genes in the species obtained from UniProt.
The comparison matrix (***M***, 1276 rows × 10 columns) is:

entry	rpmc	rpmj	rpmh	rpme	rpmg	rplx	rpmd	rpsp	hupa	hupb
0	7435	4265	5175	8300	6090	11096	6610	9200	0	9089
1	7434	4265	5175	8316	6090	11303	6638	9342	0	9308
2	7449	4265	5175	8261	6104	11227	6628	11324	0	0
3	7409	4265	5175	8314	6104	11166	6613	9326	0	0
4	7395	4265	5175	8569	6090	11203	6613	9326	0	0
5	7380	0	5175	8429	6104	11147	6626	9340	0	0
6	7405	4251	5175	8371	6090	11219	6642	9300	0	0
7	7435	4265	5175	8245	6090	11181	6628	9200	0	9322
8	7395	4265	5175	8287	6090	11243	6611	9312	0	0
9	7337	4265	5175	8273	6090	11187	6610	9340	0	0
10	7449	4265	5175	8314	6090	11123	6610	9326	0	0
11	7379	4265	5175	8314	6104	11156	6627	9326	0	9250
12	7449	4265	5175	8330	6104	11227	6628	9356	0	0
13	7449	4265	5175	8289	6104	11173	6642	9342	0	9294
14	7435	4265	5175	8357	6104	11202	6626	9326	0	0
15	7395	4265	5175	8401	6104	11146	6610	9326	0	9222
16	7449	4265	5175	8344	6104	11201	6642	9326	0	9320

17	7435	4265	5175	8328	6090	11181	6642	9200	0	0
18	7395	4265	5175	8244	6090	11158	6595	9340	0	9105
19	7380	4265	5175	8219	6090	11381	6654	9396	0	0
20	7435	4265	5175	8328	6090	11167	6642	9170	0	0
21	7435	4265	5175	8245	6104	10000	6628	9200	0	0
22	7449	4265	5175	8316	6104	11175	6628	9372	0	0
23	7435	4265	5175	8328	6090	11181	6642	9200	0	9322
24	7419	4265	5175	8328	6090	11287	6626	9356	0	0
25	7421	4265	5175	8371	6090	11248	6654	9311	0	0
26	7452	4265	5175	8300	6090	11128	6568	9352	0	0
27	7449	4265	5175	8344	6090	11167	6642	9200	0	0
28	7421	4265	5175	8344	6090	11303	6638	9372	0	0
29	7395	4265	5175	8399	6104	11233	6642	9326	0	0
...	...	...	...	...	...	...	...	...	...	...
1246	7282	4755	5310	9268	6406	11284	7189	9650	0	0
1247	7237	4872	5310	9364	6406	11226	7207	9546	0	9328
1248	7249	4872	5310	9264	6388	11240	7117	9657	0	0
1249	7249	4872	5310	10491	6406	11240	6316	0	0	0
1250	7135	4872	5148	9445	6353	11170	7248	9177	0	0
1251	7237	4872	5310	9363	6406	11226	7207	9546	0	0
1252	7237	4872	5310	8907	6392	11173	7180	9534	0	0
1253	7237	4872	5310	9364	6406	11240	7207	9546	0	0
1254	0	4872	5310	9304	0	0	0	0	0	0
1255	7207	4842	5367	9378	6229	11128	7198	9662	0	0
1256	7207	4867	5367	9389	6260	11156	7162	9572	0	9377
1257	7237	4872	5310	9304	6406	11240	7207	9602	0	0
1258	7237	4872	5310	9304	6406	11226	7207	9647	0	0
1259	7261	0	5440	7769	6358	11328	6546	9039	9593	0
1260	7261	0	5412	7838	6358	11344	6546	9068	9635	0
1261	7261	4349	5412	7767	6358	11344	6546	9054	9548	9238
1262	0	5420	0	0	0	0	0	0	0	0
1263	7261	4349	5412	7767	6358	11314	6546	9038	9560	9238
1264	7261	0	5412	7767	6358	11314	6546	9038	9607	0
1265	7261	0	5412	7781	6358	11314	6546	9068	9635	0
1266	7261	0	5412	7838	6358	11344	6546	9054	9635	0
1267	7261	5464	5412	7838	6358	11344	6546	9054	9635	0
1268	7261	0	5426	7781	6358	11344	6546	9068	9637	0
1269	7274	0	5426	7827	6372	11374	6546	9010	9591	0
1270	5733	4349	5426	7782	6358	11347	6546	9096	9661	9268
1271	7274	4349	5426	7782	6358	11347	6546	9096	9661	9268
1272	7261	0	5426	7767	6358	11344	6546	9038	9635	0
1273	7246	4349	5426	7781	6358	11344	6546	9068	0	0
1274	7274	0	5426	7782	6358	11347	6546	9096	9661	0
1275	7274	0	5426	7782	6358	11347	6546	9096	9661	0

Step 2:

Get the difference matrix (***C***) using the comparison matrix (***M***) and input mass spectrum peaks (**P**) with the formula (2) in the manuscript.
The difference matrix (***C*,** 1276 rows × 10 columns) is:

entry	rpmc	rpmj	rpmh	rpme	rpmg	rplx	rpmd	rpsp	hupa	hupb
0	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.0	0.00	0.84	0.00	0.00	0.60	0.00	0.00	0.00	0.00
3	0.0	0.00	0.84	0.00	0.00	0.00	0.58	0.74	0.00	0.00
4	0.0	0.00	0.84	0.00	0.00	0.00	0.58	0.74	0.00	0.00
5	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.88
8	0.0	0.00	0.84	0.00	0.00	0.40	0.43	0.00	0.00	0.00
9	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.74	0.00	0.00
11	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.74	0.00	0.00
12	0.0	0.00	0.84	0.00	0.00	0.60	0.00	0.00	0.00	0.00
13	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.53
14	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.74	0.00	0.00
15	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.74	0.00	0.00
16	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.74	0.00	0.71
17	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
23	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.88
24	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	0.0	0.00	0.84	0.00	0.00	0.00	0.55	0.00	0.00	0.00
27	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	0.0	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	0.0	0.00	0.84	0.00	0.00	0.98	0.00	0.74	0.00	0.00
...	...	...	...	...	...	...	...	...	...	...
1253	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.00
1254	0.0	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.00	0.60
1255	0.0	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00
1256	0.0	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00
1257	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1258	0.0	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.00	0.00
1259	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

1260	0.0	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00
1261	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1262	0.0	0.00	0.00	0.00	0.47	0.00	0.00	0.00	0.00	0.00
1263	0.0	0.00	0.00	0.00	0.46	0.00	0.53	0.00	0.00	0.00
1264	0.0	0.00	0.00	0.00	0.00	0.53	0.00	0.38	0.00	0.00
1265	0.0	0.00	0.00	0.00	0.00	0.55	0.00	0.76	0.00	0.00
1266	0.0	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00
1267	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00
1268	0.0	0.00	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.00
1269	0.0	0.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1270	0.0	0.00	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.00
1271	0.0	0.00	0.00	0.54	0.00	0.00	0.00	0.00	0.64	0.00
1272	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00
1273	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00
1274	0.0	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00
1275	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.00

Step 3:

Multiply the ***C*** matrix by the normalized gene weight table (***W***) to obtain the scores for the list of eligible species. The entry with the top score is selected as identification of the MALDI spectra:

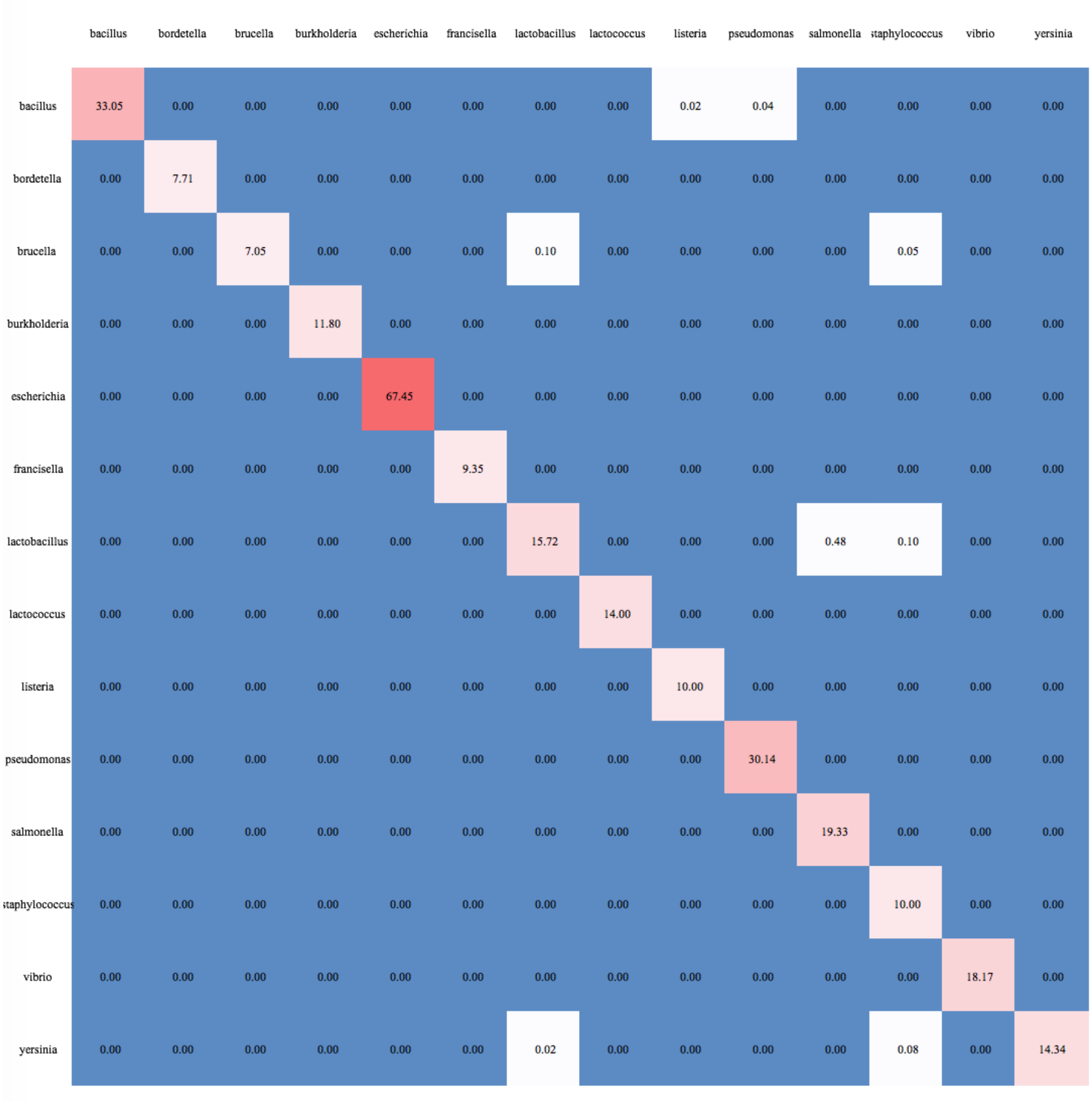
entry	genus	species	score
39	**bacillus**	**anthracis**	**2.462446**
110	**bacillus**	**thuringiensis**	**2.462446**
111	**bacillus**	**toyonensis**	**2.462446**
77	**bacillus**	**manliponensis**	**2.462446**
63	**bacillus**	**gaemokensis**	**2.462446**
56	**bacillus**	**cytotoxicus**	**2.462446**
45	bacillus	bombysepticus	1.816004
82	bacillus	mycoides	1.503078
93	bacillus	pseudomycoides	1.503078
118	bacillus	weihenstephanensis	1.503078
50	bacillus	cereus	1.503078
37	bacillus	aminovorans	1.119903
1264	yersinia	intermedia	0.966819
1263	yersinia	frederiksenii	0.966819
1261	yersinia	enterocolitica	0.966819
639	lactococcus	piscium	0.959368
672	legionella	sainthelensi	0.909401
55	bacillus	coahuilensis	0.856637
57	bacillus	decisifrondis	0.856637
43	bacillus	badius	0.856637
44	bacillus	bataviensis	0.856637
52	bacillus	circulans	0.856637

58	bacillus	encimensis	0.856637
61	bacillus	firmus	0.856637
71	bacillus	infantis	0.856637
81	bacillus	methanolicus	0.856637
116	bacillus	vireti	0.856637
84	bacillus	nealsonii	0.856637
85	bacillus	niacini	0.856637
87	bacillus	oceanisediminis	0.856637
...	...	...	...
502	lactobacillus	composti	0.000000
503	lactobacillus	concavus	0.000000
504	lactobacillus	coryniformis	0.000000
506	lactobacillus	crustorum	0.000000
507	lactobacillus	curieae	0.000000
508	lactobacillus	curvatus	0.000000
510	lactobacillus	dextrinicus	0.000000
511	lactobacillus	diolivorans	0.000000
512	lactobacillus	equi	0.000000
514	lactobacillus	equigenerosi	0.000000
515	lactobacillus	fabifermentans	0.000000
516	lactobacillus	farciminis	0.000000
517	lactobacillus	farraginis	0.000000
521	lactobacillus	fructivorans	0.000000
537	lactobacillus	helsingborgensis	0.000000
522	lactobacillus	frumenti	0.000000
523	lactobacillus	fuchuensis	0.000000
524	lactobacillus	futsaii	0.000000
525	lactobacillus	gallinarum	0.000000
526	lactobacillus	gasseri	0.000000
527	lactobacillus	gastricus	0.000000
529	lactobacillus	gigeriorum	0.000000
530	lactobacillus	ginsenosidimutans	0.000000
531	lactobacillus	graminis	0.000000
532	lactobacillus	hammesii	0.000000
533	lactobacillus	hamsteri	0.000000
534	lactobacillus	harbinensis	0.000000
535	lactobacillus	hayakitensis	0.000000
536	lactobacillus	heilongjiangensis	0.000000
1275	yersinia	wautersii	0.000000

The result of identification at genus level is *Bacillus*. At species level, it can not be distinguished between *Bacillus cytotoxicus, Bacillus gaemokensis, Bacillus manliponensis, Bacillus toyonensis, Bacillus thuringiensis,* and *Bacillus anthracis*. The original label of the MALDI spectra was *Bacillus anthracis*, which species is listed in the group.

S5: Chance of random matches (Figure S2)

403 MALDI-TOF spectra corresponding to 14 genera and protein sequences of the genera in UniProt were used to test the effect of random peak matching. The averaged numbers of matched peaks were calculated for mass spectra from each individual genus (*y*-axis in **Figure S2**) when using the protein sequences of a specific genus (*x*-axis in **Figure S2**) for peak labelling. As shown on **Figure S2**, the chance for random matches is minimal. Much more peaks in MALDI-TOF spectra can be labeled when using the protein sequences of the corresponding genus than using protein sequences from any other genera.



	bacillus	bordetella	brucella	burkholderia	escherichia	francisella	lactobacillus	lactococcus	listeria	pseudomonas	salmonella	staphylococcus	vibrio	yersinia
bacillus	33.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.00	0.00	0.00	0.00
bordetella	0.00	7.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
brucella	0.00	0.00	7.05	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.05	0.00	0.00
burkholderia	0.00	0.00	0.00	11.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
escherichia	0.00	0.00	0.00	0.00	67.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
francisella	0.00	0.00	0.00	0.00	0.00	9.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
lactobacillus	0.00	0.00	0.00	0.00	0.00	0.00	15.72	0.00	0.00	0.00	0.48	0.10	0.00	0.00
lactococcus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	14.00	0.00	0.00	0.00	0.00	0.00	0.00
listeria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00	0.00
pseudomonas	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	30.14	0.00	0.00	0.00	0.00
salmonella	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	19.33	0.00	0.00	0.00
staphylococcus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.00	0.00	0.00
vibrio	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.17	0.00
yersinia	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.08	0.00	14.34

Figure S2. Values in the cell show the averaged numbers of matched peaks for mass spectra from each individual genus (y-axis) when using the protein sequences of a specific genus (x-axis) for peak labelling.

S6: Bacterial identification by considering only ribosomal proteins (Figure S3)

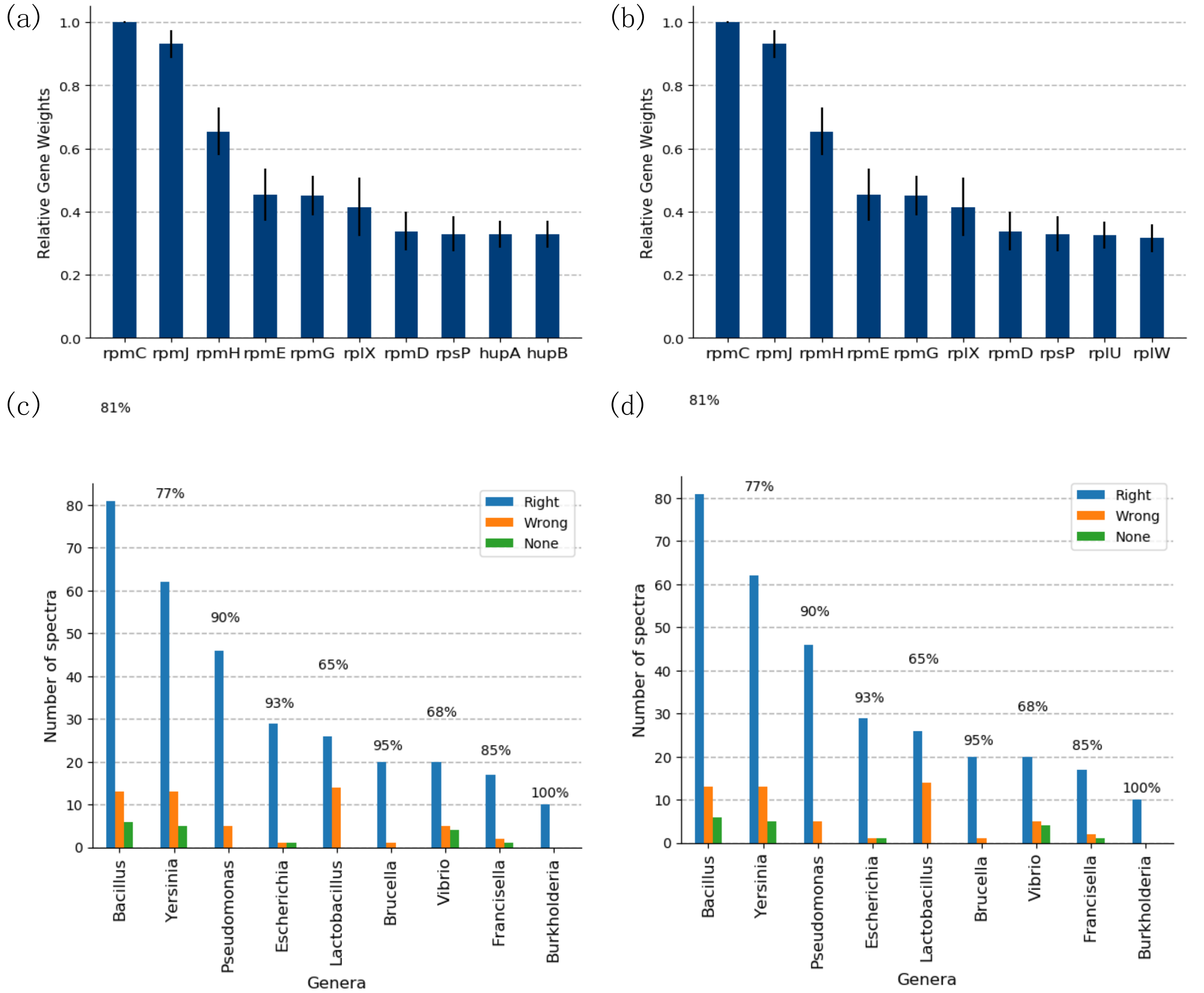


Figure S3. (a) The 10 most informative genes obtained with the double cross-validation procedure using all the protein sequences in a genus; (b) The 10 most informative genes obtained with the double cross-validation procedure using only ribosomal protein sequences in a genus; (c) Identification accuracy for mass spectra from different genera using the 10 most informative genes from (a); (d) Identification accuracy for mass spectra from different genera using the 10 most informative genes from (b).

S7: Variation in molecular weights of the proteins encoded by the most informative genes (Figure S4a and b)

Variation in molecular weights of the proteins encoded by the following 9 genes in different genera (the results for gene rpmC is shown in **Figure 5** in the manuscript):

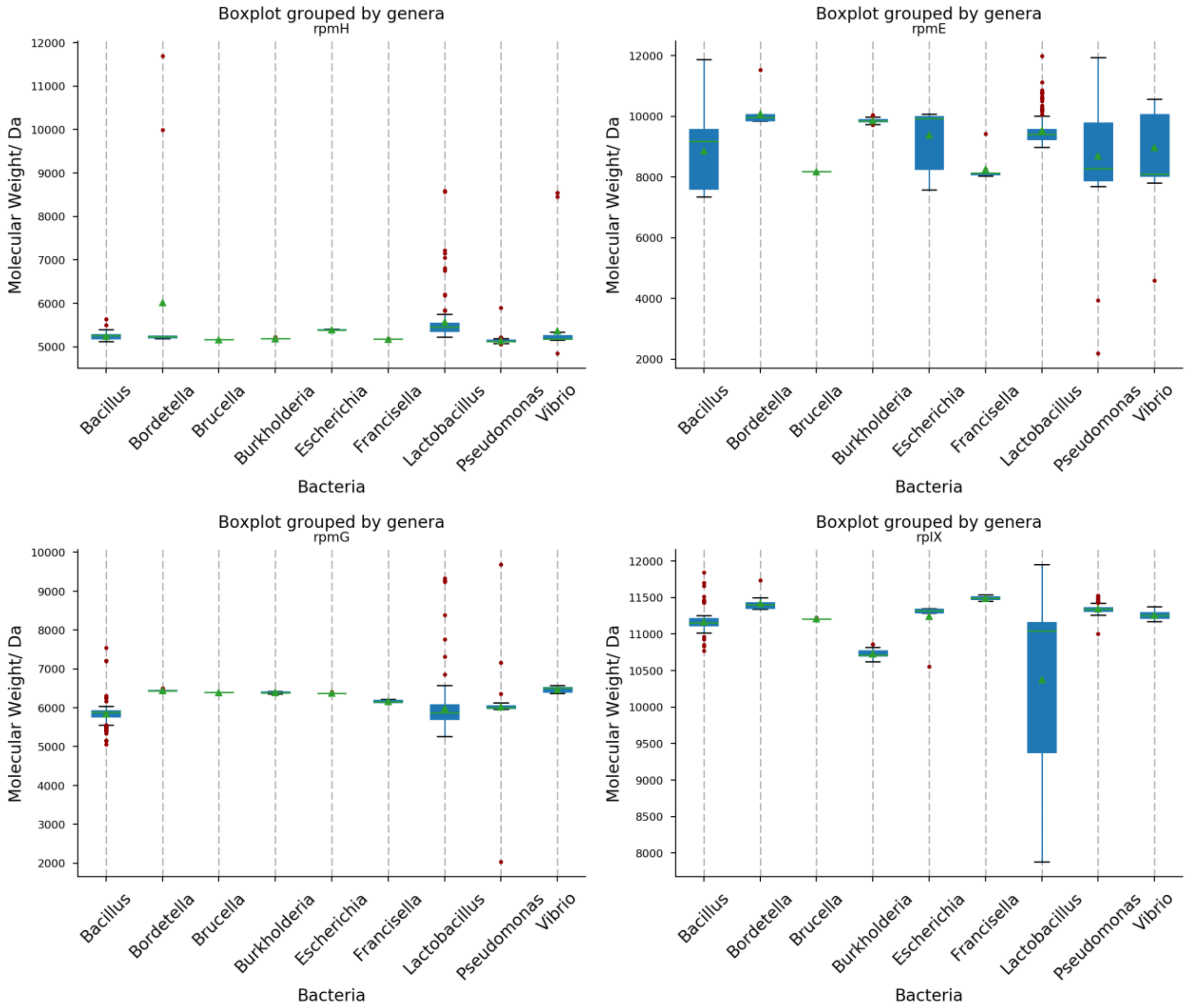


Figure S4a. Variation of the molecular weights of the proteins encoded by rpmH, rpmE, rpmG, rplX in different genera using the protein sequences database from UniProt. Red dots shows outliers, the green triangle indicates mean value, the green line shows the median value and the blue box shows data values between inter quartile range (25%-75%).

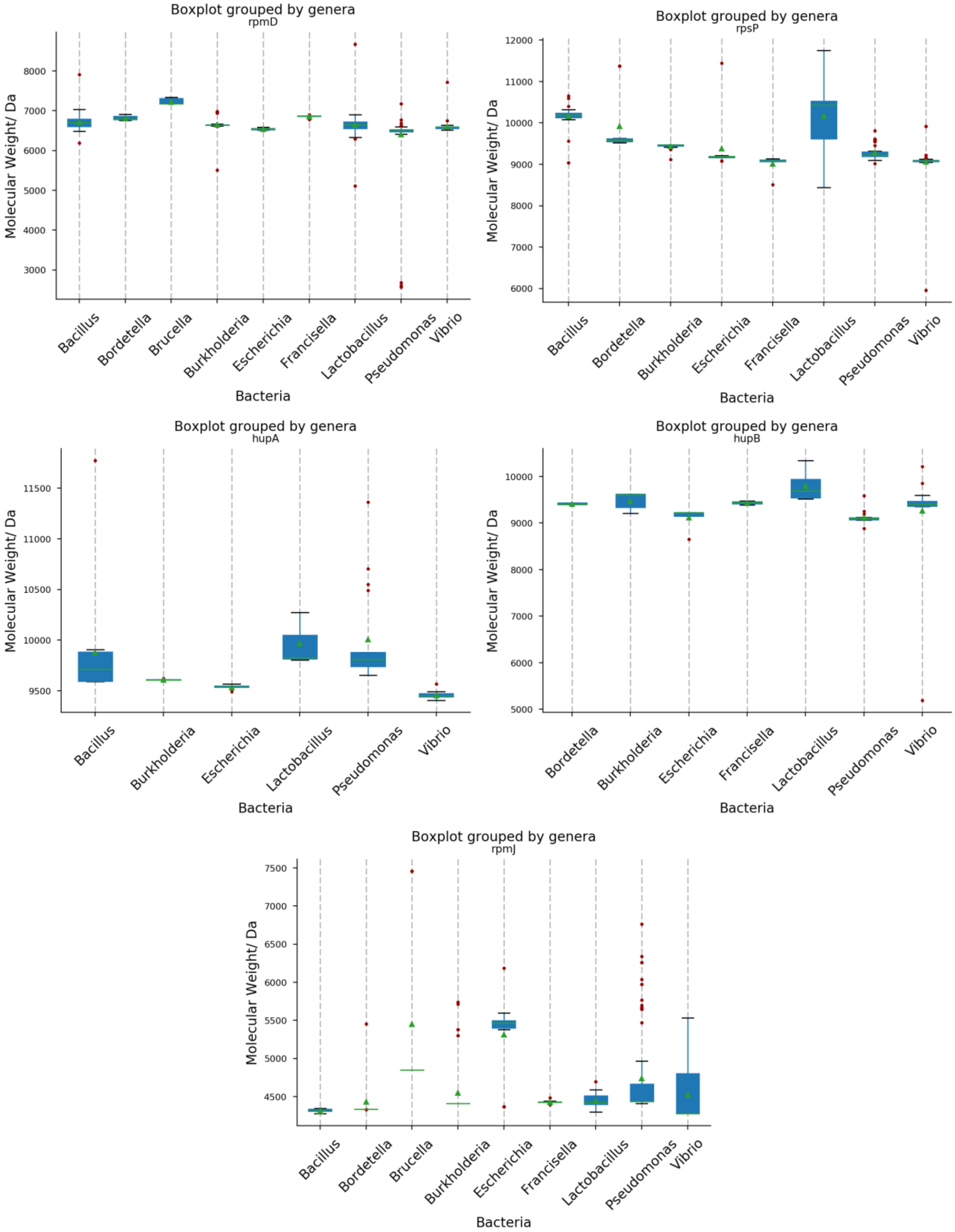


Figure S4b. Variation of the molecular weights of the proteins encoded by rpmD, rpsP, hupA, hupB, rpmJ in different genera using the protein sequences database from UniProt. Red dots shows outliers, the green triangle indicates mean value, the green line shows the median value and the blue box shows data values between inter quartile range (25%-75%).