Conserved mass peaks in MALDI-TOF mass spectra of bacterial species at the genus and species levels

Wenfa Ng

Unaffiliated researcher, Singapore, Email: ngwenfa771@hotmail.com

Abstract

Microbes are identified based on their distinguishing characteristics such as gene sequence or metabolic profile. Nucleic acid approaches such as 16S rRNA gene sequencing provide the gold standard method for microbial identification in the contemporary era. However, mass spectrometry-based microbial identification is gaining credence through ease of use, speed, and reliability. Specifically, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has been used in identifying bacteria, fungus, molds and archaea to the species level with high accuracy. The approach relies on the existence of unique mass spectrum fingerprint for individual microbial species. By comparing the mass spectrum of an unknown microbe with that catalogued in a reference database of known microorganisms, microbes could be identified through mass spectrum fingerprinting. However, the approach lacks fundamental biological basis given the relative difficulty in assigning specific protein to particular mass peak in the profiled mass spectrum, which hampers a deeper understanding of the mass spectrum obtained. This study seeks to examine the existence of conserved mass peaks in MALDI-TOF mass spectra of bacteria at the species and genus levels using open access data from SpectraBank. Results revealed that conserved mass peaks existed for all bacterial species examined. Large number of conserved mass peaks such as that of *Escherichia coli* and *Morganella morganii* suggested more closely-related strains of a species though functional annotation of the mass peaks is required to provide a deeper understanding of the mechanisms underlying the conservation of specific proteins. On the other hand, strains of *Staphylococcus aureus* and *Pseudomonas putida* had the least number of conserved mass peaks. Presence of conserved mass peaks in many genus provided further evidence that MALDI-TOF MS microbial identification had a biological basis in identification of microbial species to the genus level. In addition, it also highlighted that a subset of proteins could define the taxonomical boundary between the species and genus level. Finally, existence of only one conserved mass peak in *Bacillus* genus corroborated the difficulty of discriminating *Bacillus* species based on MALDI-TOF mass spectra. Similarly, no conserved mass peak at the genus level could be found for the *Staphylococcus* genus. Overall, existence of conserved mass peaks of bacteria at the species and genus levels provided evidence of a firm biological basis in the mass spectrum fingerprinting approach of MALDI-TOF MS microbial identification. This could help identify specific species in mass spectrum of single or multiple microbial species. Further functional annotation of the conserved mass peaks could illuminate in greater detail the biological mysteries of why certain proteins are conserved in specific genus and species.

Keywords: microbial identification, MALDI-TOF MS, genus, species, strain, bacteria, conserved mass peaks, mass spectrum fingerprinting, biomarkers, pattern recognition,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Subject areas: biotechnology, microbiology, biochemistry, bioinformatics, computational biology,

Highlights

1) Conserved mass peaks were identified in MALDI-TOF mass spectra of different strains of the same bacterial species.

2) Large number of conserved mass peaks highlights that different strains are closely-related at the proteome level.

3) Conserved mass peaks were also identified at the genus level for many genus, but not *Bacillus* and *Staphylococcus*.

4) Presence of conserved mass peaks at the species and genus levels highlights the deep biological basis in MALDI-TOF MS microbial identification where highly conserved proteins could serve as biomarkers.

5) Overall, this study revealed the biological basis underlying the mass spectrum fingerprinting approach to microbial identification in MALDI-TOF MS where pattern recognition of sets of conserved mass peaks provides the basis for identification.

Introduction

Identification of an entity relies on definitive distinguishing characteristics of the entity to be found and characterized. Over the decades, microbiologists have progressively used different distinguishing characteristics for understanding differences between different microorganisms as well as classifying them into different taxonomy groups. While cell shape and colour were used in classifying microorganisms in earlier periods of the field, the lack of distinguishing features in microbes of similar colour motivated microbiologists in seeking better biomarkers for identifying microorganisms. To this end, biochemical assays were introduced which to a certain extent helped ameliorated the lack of methods for classifying microbes. However, similar metabolic characteristics by different microbes and possible influence of growth state on type of metabolism utilized remained barriers to definitive classification of different microbes based on metabolic traits. Hence, modern taxonomic classification takes on a molecular approach where biomarkers sought are biomolecules endowed with the evolutionary history of different species and strains. For example, exploration of the utility of 16S rRNA as a biomarker gene for classifying the large variety of microbial species laid the foundation of molecular taxonomy in enabling the classification of species without need for prior cultivation. Such culture-independent approaches provide a useful tool for microbiologists interested in probing the microbial dark matter where microbial cells could not be coaxed into growth on solid agar medium.

The need for rapid, cost-effective and robust methods for microbial identification especially in the clinical laboratory provided the impetus towards continued refinement of mass spectrometry based methods for identifying microorganisms after culture on solid medium.[1] Typically implemented with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) as readout of species-specific mass spectra of microorganisms,[2]

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

[3] [4] mass spectrometry-based microbial identification currently lack a firm basis in the origins of mass peaks profiled from individual microbial specimen. Specifically, MALDI-TOF MS relies on pattern recognition algorithms in vendor provided software for providing identification based on comparing mass spectrum of unknown microbes with those curated in a vendor provided reference database of mass spectra. The alternative method of searching for biomarker proteins in proteome database for annotating mass peaks,[5] while conceptually useful, nevertheless run into problems of the time and effort needed to calculate the molecular masses of all proteins in the proteome of many microbes. Additionally, given the storage of existing proteomic information of microorganisms in different proteomic databases, the collection and searching of all proteomic information of microbes in a centralized database would be an enabling tool for the proteome database search approach to microbial identification.

Thus, the reality of current proteome-based methods for annotating mass peaks in MALDI-TOF mass spectra of microbes is that it lags behind that of mass spectrum fingerprinting, where mass spectrum of an unknown microbe would be compared against that of known microorganisms for identifying distinguishing characteristics (i.e., mass peaks) useful for positive identification.[6] [7] Specifically, proteome-based methods for mass peak annotation in MALDI-TOF MS microbial identification feeds into a broader problem afflicting mass spectrometry-based proteomics: how to assign proteins to profiled mass peaks? The solution to the problem is non-trivial and awaits the building of curated proteome databases supported by software and libraries necessary for efficient protein search and mass calculation.

While mass spectrum fingerprinting is the main method by which MALDI-TOF MS helps identify microorganisms, questions remain especially in putting the methodology on a firmer conceptual and mechanistic basis.[8] Specifically, origins of the mass peaks are ignored in mass spectrum fingerprinting, where the set of profiled mass peaks offer distinguishing characteristics useful for classifying unknown microbes based on a reference database of known microorganisms. But, herein lies the problem, how does one identify an unknown microbe which has not been catalogued in a reference database? The solution would be 16S rRNA gene sequencing and comparison with a phylogenetic tree. An alternative approach could be the search of all proteomic information of microorganisms for annotating the mass peaks profiled in the mass spectrum of the unknown microbe. Such a method would require enormous amount of time and effort in collecting all proteomic information of microbes into a single database for search purposes, and is not guaranteed to succeed.

On the other hand, understanding the basis for the presentation of unique mass peaks of different species could provide important knowledge for developing MALDI-TOF MS microbial identification into a clinically important tool comparable to 16S rRNA gene sequencing which has a firm theoretical basis. One important problem besetting the field remains the inability to fully annotate all mass peaks profiled in a mass spectrum of a bacterial species. Another issue is the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

possible presence of conserved mass peaks in the mass spectra of strains belonging to a species. Tackling the second question, the goal of this study was to understand if there are conserved mass peaks in the MALDI-TOF MS mass spectra of different bacteria at the species and genus levels. Understanding this would provide a firm theoretical basis for developing the mass spectrum fingerprinting approach into a method grounded with the profiling of species-specific biomolecules that lend itself into identification of specific microbes based on mass spectrometry acquisition of unique set of mass peaks. Using mass spectra information of bacterial species catalogued in an open-access database, SpectraBank, this study aimed to identify conserved mass peaks in mass spectra of different strains and species of bacteria profiled in the database.

Materials and Methods

Mass spectra of different bacterial species were downloaded from SpectraBank (http://www.usc.es/gl/investigacion/grupos/lhica/spectrabank/Database.html). Analysis was conducted to identify conserved mass peaks of bacterial species in mass spectra of different strains of the same species. Similarly, efforts were also made to identify conserved mass peaks in different species of bacteria belonging to the same genus. Conserved mass peaks are defined by a difference in *m/z* of less than 10. Mass peaks (*m/z*) of bacterial strains and species found to be conserved were averaged to yield the final *m/z* of the conserved mass peaks.

Results and Discussion

Table 1: Conserved mass peaks *(m/z)* of *Bacillus subtilis*

2182
2745
3046
3342
3725
3859
3884
4305
4572
4944
5004
5031
6506
6599
6676
7715

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Table 1 shows the conserved mass peaks in mass spectra of different *Bacillus subtilis* strains catalogued in the SpectraBank database. Specifically, conserved mass peaks are 2182, 2745, 3046, 3342, 3725, 3859, 3884, 4305, 4572, 4944, 5004, 5031, 6506, 6599, 6676, and 7715 *m/z*. The large number of conserved mass peaks for *B. subtilis* implied that different strains of the species are closely-related at the proteome level.

Table 2: Conserved mass peaks *(m/z)* of *Bacillus thuringiensis*

2168
3090
3118
3652
3708
3746
4333
4551
4993
5474

Conserved mass peaks of another *Bacillus* species, *Bacillus thuringiensis,* is shown in Table 2. Specifically, conserved mass peaks are 2168, 3090, 3118, 3652, 3708, 3746, 4333, 4551, 4993, and 5474 *m/z*.

Table 3: Conserved mass peaks *(m/z)* of *Carnobacterium maltaromaticum*

2174
2902
3238
3350
3436
4347
5804
6347
6475
6872

Table 3 shows the conserved mass peaks of *Carnobacterium maltaromaticum*. The mass peaks are 2174, 2902, 3238, 3350, 3436, 4347, 5804, 6347, 6475, and 6872 *m/z*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Table 4: Conserved mass peaks *(m/z)* of *Escherichia coli*

2184
2692
2836
3129
3159
3580
3638
3674
3936
4186
4365
4769
4778
5097
5151
5381
6255
6316
6411
7158
7274
7869
8370
8995
9226
9543

Table 4 shows the conserved mass peaks of *Escherichia coli*. Specifically, large number of mass peaks were conserved in strains of this species and thus highlighted high level of relatedness of each strain at the proteome level. Conserved mass peaks are 2184, 2692, 2836, 3129, 3159, 3580, 3638, 3674, 3936, 4186, 4365, 4769, 4778, 5097, 5151, 5381, 6255, 6316, 6411, 7158, 7274, 7869, 8370, 8995, 9226, and 9543 *m/z*.

Table 5: Conserved mass peaks *(m/z)* of *Proteus vulgaris*

2243
2750

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

2826
3138
3554
3637
4185
4484
4738
4770
4802
5131
6274
7274
9477

Table 5 shows the conserved mass peaks of strains of *Proteus vulgaris*. Specifically, the conserved mass peaks are 2243, 2750, 2826, 3138, 3554, 3637, 4185, 4484, 4738, 4770, 4802, 5131, 6274, 7274, and 9477 *m/z*.

Table 6: Conserved mass peaks *(m/z)* of *Pseudomonas fluorescens*

2218
2534
3041
3310
3586
4128
4433
4980
5066
6080
6393
7172

Table 6 shows the conserved mass peaks of *Pseudomonas fluorescens*. Specifically, the mass peaks are 2218, 2534, 3041, 3310, 3586, 4128, 4433, 4980, 5066, 6080, 6393, and 7172 *m/z*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Table 7: Conserved mass peaks *(m/z)* of *Pseudomonas fragi*
2218
2534
3023
3306
3594
4128
4433
4946
5066
6044
6610
7186
8254

Table 7 shows the conserved mass peaks of *Pseudomonas fragi*. Specifically, the conserved mass peaks are 2218, 2534, 3023, 3306, 3594, 4128, 4433, 4946, 5066, 6044, 6610, 7186, and 8254 *m/z*.

Table 8: Conserved mass peaks *(m/z)* of *Pseudomonas putida*
5137
7171
8237

Table 8 shows the conserved mass peaks of *Pseudomonas putida*. Specifically, existence of only three conserved mass peaks (5137, 7171, and 8237 *m/z*) revealed that strains of the species were not closely-related at the proteome level.

Table 9: Conserved mass peaks *(m/z)* of *Pseudomonas syringae*
2219
2564
3587
3785
4128

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

4434
4832
5124
5673
5978
7172
9109

Table 9 shows the conserved mass peaks of *Pseudomonas syringae*. Specifically, the conserved mass peaks are 2219, 2564, 3587, 3785, 4128, 4434, 4832, 5124, 5673, 5978, 7172, and 9109 *m/z*.

Table 10: Conserved mass peaks *(m/z)* of *Serratia marcescens*

2691
2826
3962
4185
4349
4606
4768
5359
6116
6226
7924
9537

Table 10 shows the conserved mass peaks of *Serratia marcescens*. Specifically, the conserved mass peaks are 2691, 2826, 3962, 4185, 4349, 4606, 4768, 5359, 6116, 6226, 7924, and 9537 *m/z*.

Table 11: Conserved mass peaks *(m/z)* of *Serratia proteamaculans*

2176
2698
2825
3948
4183
4347
4781

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

5393
6236
6410
7892
9557

Table 11 shows the conserved mass peaks of *Serratia proteamaculans*. Specifically, conserved mass peaks are 2176, 2698, 2825, 3948, 4183, 4347, 4781, 5393, 6236, 6410, 7892, and 9557 *m/z*.

Table 12: Conserved mass peaks *(m/z)* of *Staphylococcus aureus*

3444
4304
5031
6887

Table 12 shows the conserved mass peaks of *Staphylococcus aureus*. Relatively few number of conserved mass peaks of *S. aureus* revealed that strains of the species are not closely-related at the proteome level. Specifically, the conserved mass peaks are 3444, 4304, 5031, and 6887 *m/z*.

Table 13: Conserved mass peaks *(m/z)* of *Stenotrophomonas maltophilia*

2631
2779
4242
5266
9583

Table 13 shows the conserved mass peaks of *Stenotrophomonas maltophilia*. Specifically, the conserved mass peaks of the species are 2631, 2779, 4242, 5266, and 9583 *m/z*. Relatively few number of conserved mass peaks in mass spectra of strains of the species revealed that strains were likely not closely-related at the proteome level.

Table 14: Conserved mass peaks *(m/z)* of *Bacillus cereus*

2589
4333

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

4994
5443
5548

Table 14 shows the conserved mass peaks of *Bacillus cereus*. Specifically, the conserved mass peaks of the species are 2589, 4333, 4994, 5443, and 5548 *m/z*. Relatively few number of conserved mass peaks revealed that strains of the species are not closely-related at the proteome level.

Table 15: Conserved mass peaks *(m/z)* of *Bacillus licheniformis*

2057
3022
3042
3253
3290
5893
9900

Table 15 shows the conserved mass peaks of *Bacillus licheniformis*. Specifically, the conserved mass peaks are 2057, 3022, 3042, 3252, 3290, 5893, and 9900 *m/z*.

Table 16: Conserved mass peaks *(m/z)* of *Bacillus megaterium*

2478
3047
3075
3131
3553
4304
4610
4794
5205
5830
6259
6579
6743
7728
9348

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Table 16 shows the conserved mass peaks of *Bacillus megaterium*. Specifically, the conserved mass peaks are 2478, 3047, 3075, 3131, 3553, 4304, 4610, 4794, 5205, 5830, 6259, 6579, 7728, 9348 *m/z*. Presence of large number of conserved mass peaks suggested that strains of the species, *B. megaterium*, are closely-related at the proteome level.

Table 17: Conserved mass peaks *(m/z)* of *Bacillus pumilus*

3019
3045
3761
4303
4586
5297
6617
7238
7724
9820

Table 17 shows the conserved mass peaks of *Bacillus pumilus*. Specifically, the conserved mass peaks are 3019, 3045, 3761, 4303, 4586, 5297, 6617, 7238, 7724, and 9820 *m/z*.

Table 18: Conserved mass peaks *(m/z)* of *Morganella morganii*

2187
2691
3109
3156
3177
3193
3241
3590
3637
3879
4336
4372
4454
4641

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

4733
5137
5380
6216
6480
7178
7272
8670
9279
9461

Table 18 shows the conserved mass peaks of *Morganella morganii*. Specifically, conserved mass peaks are 2187, 2691, 3109, 3156, 3177, 3193, 3241, 3590, 3637, 3879, 4336, 4372, 4454, 4641, 4733, 5137, 5380, 6216, 6480, 7178, 7272, 8670, 9279, and 9461 *m/z*. Large numbers of conserved mass peaks in *M. morganii* revealed that strains of the species are closely-related at the proteome level.

Overall, sets of conserved mass peaks for specific species could find use in bioinformatic approaches for identifying specific species from MALDI-TOF mass spectrum of mixtures of different microbial species. In addition, the same set of conserved mass peaks could also augment the mass spectrum fingerprinting approach in identifying specific bacterial species.

Table 19: Conserved mass peaks (*m/z*) in *Bacillus* genus

3046

Analysis of conserved mass peaks between species of the genus *Bacillus* revealed the existence of only one conserved mass peak at 3046 *m/z*. Lack of conserved mass peaks in the genus corroborate observations of difficulty of discriminating different *Bacillus* species via MALDI-TOF MS microbial identification.[9] Given the presence of only one conserved mass peak in the genus, *Bacillus*, it can be approximated that conserved mass peaks do not exist in the genus.

Table 20: Conserved mass peaks (*m/z*) in *Carnobacterium* genus

2173
3236
3350

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

4346

Table 20 shows the conserved mass peaks for the genus, *Carnobacterium*. The conserved mass peaks are 2173, 3236, 3350, and 4346 *m/z*.

Table 21: Conserved mass peaks (*m/z*) in *Clostridium* genus

2152
3660
3858
4304
5491

Table 21 shows the conserved mass peaks for the genus, *Clostridium*. Conserved mass peaks are 2152, 3660, 3858, 4304, and 5491 *m/z*.

Table 22: Conserved mass peaks (*m/z*) in *Enterobacter* genus

2183
2692
4185
4364
5383

Table 22 shows the conserved mass peaks for the genus, *Enterobacter*. Conserved mass peaks are 2183, 2692, 4185, 4364, and 5383 *m/z*.

Table 23: Conserved mass peaks (*m/z*) in *Klebsiella* genus

2182
3139
3579
3622
4185
4363
4736
4773
6382

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

7242

Table 23 shows the conserved mass peaks for *Klebsiella* genus. The conserved mass peaks are 2182, 3139, 3579, 3622, 4185, 4363, 4736, 4773, 6382, and 7242 *m/z*. Large numbers of conserved mass peaks revealed that species in the genus, *Klebsiella*, are closely-related at the proteome level.

Table 24: Conserved mass peaks (*m/z*) in *Listeria* genus

2163
3005
3195
3431
3509
3702
4323
4518
4696
4877
5120
5173
5598
6007
6362
6388
6860
7015
7402
9389
9751

Table 24 shows the conserved mass peaks of the genus, *Listeria*. Large numbers of conserved mass peaks indicated that different *Listeria* species are likely to be closely-related at the proteome level. The conserved mass peaks are 2163, 3005, 3195, 3431, 3509, 3702, 4323, 4518, 4696, 4877, 5120, 5173, 5598, 6007, 6362, 6388, 6860, 7015, 7402, 9389, and 9751 *m/z*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Table 25: Conserved mass peaks (*m/z*) in *Photobacterium* genus

2140
2982
3131
3140
3582
4184
4278
5159
5687
6261
6583
7165
8367

Table 25 shows the conserved mass peaks of *Photobacterium* genus. As the genus has large number of conserved mass peaks between different species of the genus, it is likely that different species of *Photobacterium* are closely-related at the proteome level. The conserved mass peaks are 2140, 2982, 3131, 3140, 3582, 4184, 4278, 5159, 5687, 6261, 6583, 7165 and 8367 *m/z*.

Table 26: Conserved mass peaks (*m/z*) in *Proteus* genus

2240
2749
2826
3132
3556
3637
4185
5131
7273

Table 26 shows the conserved mass peaks of the genus, *Proteus*. The conserved mass peaks are 2240, 2749, 2826, 3132, 3556, 3637, 4185, 5131, and 7273 *m/z*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Table 27: Conserved mass peaks (*m/z*) in *Providencia* genus

2219
2735
3119
3163
3554
3623
4185
4435
4778
5467
6230
7107
7244

Table 27 shows the conserved mass peaks for the genus, *Providencia*. Large number of conserved mass peaks revealed that species of *Providencia* are closely-related at the proteome level. The conserved mass peaks are 2219, 2735, 3119, 3163, 3554, 3623, 4185, 4435, 4778, 5467, 6230, 7107, and 7244 *m/z*.

Table 28: Conserved mass peaks (*m/z*) in *Shewanella* genus

2133
3302
3580
4108
4264
4490
4721
5026
8214

Table 28 shows the conserved mass peaks for the genus, *Shewanella*. Conserved mass peaks of the genus are 2133, 3302, 3580, 4108, 4264, 4490, 4721, 5026, and 8214 *m/z*.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

Table 29: Conserved mass peaks (*m/z*) in *Vibrio* genus

3598
4179
4278
4534
5151
6167

Table 29 shows the conserved mass peaks of the genus, *Vibrio*. The conserved mass peaks are 3598, 4179, 4278, 4534, 5151, and 6167 *m/z*.

Table 30: Conserved mass peaks (*m/z*) in *Pseudomonas* genus

2218
3589
4128
4433

Table 30 shows the conserved mass peaks for the genus, *Pseudomonas*. Specifically, *Pseudomonas* genus exhibited conservation in mass peaks at 2218, 3589, 4128, and 4433 *m/z*.

Table 31: Conserved mass peaks (*m/z*) in *Serratia* genus

2696
2826
4185
4349

Table 31 shows the conserved mass peaks for the genus, *Serratia*. Specifically, the mass peaks are 2696, 2826, 4185, and 4349 *m/z*. Thus, conserved mass peaks at the genus level are present in many genus of bacteria except the genus of *Staphylococcus* and *Bacillus*. Conservation in mass peaks at the genus level implied a deep biological basis underlaid the mass spectrum fingerprinting approach to species identification by MALDI-TOF MS. In particular, specific proteins are conserved at the genus level that helped define and chronicle the evolutionary history of the genus. Identifying the specific proteins responsible for the conserved mass peaks would help further our understanding of how evolutionary forces selected specific proteins as highly conserved biomolecules that chronicle the evolutionary trajectory taken by the genus. Given that

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

most of our understanding of how evolution shape biological organization and complexity was derived by examining the relatedness of conserved biomolecules at the species level, understanding how evolutionary processes works at the genus level could provide a lens into possible differences in evolutionary processes at the genus and species levels. More importantly, different sets of conserved mass peaks exist for different genus. This highlighted that each genus likely had specific highly conserved proteins that defined the characteristics of species in the genus.

Conclusions

Conserved mass peaks were found for MALDI-TOF mass spectra of different bacterial species that likely suggested highly conserved proteins that defined specific characteristics in metabolism, cell signalling and physiology for individual species. The number of conserved mass peaks differed between bacterial species, which suggested that different species had different levels of relatedness in proteome of strains of the same species. In summary, *Escherichia coli* and *Morganella morganii* exhibited the most number of conserved mass peaks, while *Pseudomonas putida* and *Staphylococcus aureus* had the least. Thus, strains of *E. coli* and *M. morganii* were likely more closely-related at the proteome level compared to those of *P. putida* and *S. aureus*. Although conserved mass peaks suggested highly conserved proteins that defined distinguishing characteristics for the species, closer examination of conserved mass peaks across species in a genus revealed the existence of conserved mass peaks in many genus such as *Carnobacterium, Enterobacter, Clostridium, Klebsiella, Listeria, Photobacterium, Proteus, Providencia, Pseudomonas, Shewanella,* and *Serratia*. The genus with only one conserved mass peak at the genus level was *Bacillus*, which corroborated experimental observations of difficulty in distinguishing different *Bacillus* species using MALDI-TOF mass spectrometry profiling of biomolecules. On the other hand, no conserved mass peaks could be found for the genus, *Staphylococcus*. Finally, the genus, *Listeria*, had the most number of conserved mass peaks, which suggested that species in the genus are closely-related at the proteome level. Differing number of conserved mass peaks in different genus highlighted different levels of relatedness amongst species belonging to different genus.

Overall, the analysis results suggested that MALDI-TOF mass spectrometry microbial identification could profile biomolecules of biological significance that defined molecular evolution at the genus and species level. More importantly, determination of conserved mass peaks at the genus and species level helped laid a firmer foundation for the mass spectrum fingerprinting approach to microbial identification. Specifically, uncovering a set of conserved biomolecules from a species helped provide a biological basis for subsequent identification by comparison of mass spectrum from unknown microbe with those of known microorganisms catalogued in a reference database. Future functional annotation of the conserved mass peaks would reveal how specific metabolic or cell signalling processes are important to definition of biological genus and species. In addition, questions of how evolutionary forces selected specific proteins and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

biomolecules as anchors for certain genus and species as well as how natural selection chose to retain similar variants of the proteins are also important. On the other hand, conserved mass peaks for specific species could also find practical use in helping identify specific species from MALDI-TOF mass spectrum of mixture of different microbial species. They could also complement mass spectrum fingerprinting in identifying bacterial species from mass spectrum of a single species. In essence, conserved mass peaks at the species and genus level could serve as biomarker peaks.

Supplementary materials

Raw data of comparison of mass peaks from mass spectra of bacterial strains is appended as an Excel file.

References

1. Cheng, K., Chui, H., Domish, L., Hernandez, D. & Wang, G. Recent development of mass spectrometry and proteomics applications in identification and typing of bacteria. *PROTEOMICS – Clin. Appl.* **10,** 346–357 (2016).
2. Ge, M.-C. *et al.* Routine identification of microorganisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Success rate, economic analysis, and clinical outcome. *J. Microbiol. Immunol. Infect.* **50,** 662–668 (2017).
3. Biswas, S. & Rolain, J.-M. Use of MALDI-TOF mass spectrometry for identification of bacteria that are difficult to culture. *J. Microbiol. Methods* **92,** 14–24 (2013).
4. Mass spectrometry tools for the classification and identification of bacteria | Nature Reviews Microbiology. Available at: https://www.nature.com/articles/nrmicro2243. (Accessed: 1st January 2018)
5. Fagerquist, C. K. *et al.* Web-Based Software for Rapid Top-Down Proteomic Identification of Protein Biomarkers, with Implications for Bacterial Identification. *Appl. Environ. Microbiol.* **75,** 4341–4353 (2009).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3524v2 | CC BY 4.0 Open Access | rec: 12 May 2018, publ: 12 May 2018

6. Randell, P. It's a MALDI but it's a goodie: MALDI-TOF mass spectrometry for microbial identification. *Thorax* **69,** 776–778 (2014).
7. Ng, W. Graphical illustration for explaining mass spectrum fingerprinting in microbial identification. (PeerJ Inc., 2017). doi:10.7287/peerj.preprints.580v3
8. Fagerquist, C. K. Unlocking the proteomic information encoded in MALDI-TOF-MS data used for microbial identification and characterization. *Expert Rev. Proteomics* **14,** 97–107 (2017).
9. Starostin, K. V. *et al.* Identification of *Bacillus* strains by MALDI TOF MS using geometric approach. *Sci. Rep.* **5,** 16989 (2015).

New in this version

Conserved mass peaks of bacterial species, *Bacillus cereus, Bacillus licheniformis, Bacillus megaterium, Bacillus pumilus,* and *Morganella morganii* were added. Analysis was expanded to investigate the conservation of mass peaks at the genus level. Conservation of mass peaks was found to exist in many genus except *Bacillus* and *Staphylococcus*.

Conflicts of interest

The author declares no conflicts of interest.

Author's contribution

The author hypothesized that there could be conserved mass peaks in mass spectra of different bacterial strains and species.. He analysed the peak lists of mass spectra of bacterial strains and species deposited in SpectraBank, and wrote the manuscript.

Funding

No funding was used in this work.