Identifying microbes from environment water samples in a discovery-based learning module

Wenfa Ng*†

Department of Chemical and Biomolecular Engineering, National University of Singapore

*Corresponding author, Email: ngwenfa771@hotmail.com
†Present address: Novena, Singapore

Abstract

What is the microbe we are dealing with? Irrespective of cholera or anthrax, we want to know the disease causing microorganism as quickly as possible since prompt identification of the etiology organism would help control disease spread, and save lives through provision of appropriate care and medicine. But despite the promulgation of rapid microbe identification tools (such as those based on mass spectrometry), most undergraduate curricula continue to focus on culture and nucleic acid-based identification techniques since they are widely used for detecting and identifying microbes in clinical and environment samples. Mass spectrometry-based methods, however, have increasingly complemented traditional approaches in clinical and research laboratories - but they rarely feature in undergraduate curricula. Motivated by the desire to bridge the curriculum gap, I developed an inquiry-based laboratory exercise for introducing students to the operating principles and methodology of mass spectrometry enabled microbe identification. By requiring students to identify microbes in environment water samples (a real life problem with unknown answers), the exercise piqued the students’ interest in learning, while helping stir their curiosity in science through an interesting field activity where they put on a scientist’s hat in solving a mystery. This synopsis article summarizes a piece of published education research and expands on the discussion of concepts underlying matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) microbe identification. Specifically, the article discusses the relative advantages and disadvantages of the pattern recognition and proteome database search approaches for analyzing mass spectra data. In addition, the effect of different sample preparation protocols on identification accuracy is also discussed in detail. Finally, the pedagogy utility of field and inquiry-based education tools is also discussed in greater detail from a post-publication perspective. A full length synopsis of the work and a structured abstract can be found in the accompanying PDF file, while the original article, “Teaching Microbial Identification with Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS) and Bioinformatics Tools”, and supplementary material is published in the Journal of Microbiology and Biology Education, Vol. 14, No. 1, pp. 103-106, and is available as an open access article at http://www.asmscience.org/content/journal/jmbe/10.1128/jmbe.v14i1.494.
**Keywords:** ribosome proteins; biomarker; phylogeny; microbial ecology; taxonomy; education research; inquiry-based; mass spectrum fingerprinting; scientific method; microbe identification;

**Subject areas:** microbiology; education; environment sciences; ecology;

**Structured abstract**

**Background**

Rapid detection and identification of microorganisms is important, for example, in clinical diagnostics and quality control in food industry. Current methods for identifying microbes rely heavily on cell cultivation, or nucleic acid analysis (e.g., 16S rRNA sequencing). Besides long time to result of a few days, culture-based methods are also prone to false negatives and cultivation bias, as 99% of all known microorganisms have not been cultured under laboratory conditions. Although 16S rRNA analysis is culture independent (and thus, more expeditious than cell culture assays), bias associated with primer selection reduces identification accuracy. In contrast, matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) has emerged, over the past decade, as a rapid and relatively low cost microbe identification tool. Specifically, microbes are identified either by matching the mass spectrum of an unknown microorganism with those present in a reference database/library (i.e., the pattern recognition approach), or through a proteome database search method that relates individual mass peaks to biomarker proteins with species or strain specific signatures. Though commercialized, MALDI-TOF MS microbe typing has received relatively little coverage in undergraduate life science curricula compared to culture and 16S rRNA-based techniques.

**Pedagogy tool**

To help bridge the curriculum gap, a simple inquiry-based laboratory exercise for teaching microbe identification using a combined MALDI-TOF MS and proteome database search approach was developed. Specifically, students participated in a variety of activities - ranging from sample collection, cell cultivation, mass spectrometry analysis and bioinformatics annotation of mass spectra data - during identification of microorganisms from an environment water sample. By encouraging students to use deductive and inductive thinking skills in solving a real life problem with unknown answers, the education module helped ignite their inquiring minds, while teaching leading edge mass spectrometry-based microbe identification technique and concepts.

**Potential significance**

Collectively, besides introducing the concepts and operating principles of MALDI-TOF MS
Microbe typing to students, the activity oriented education module also helped students appreciate the connection between the scientific method (as explained in the textbook) and its application in real world problem solving. The latter helped demystify the work of scientists and connects scientific research to real life, as well as seeding a science oriented perspective in thinking about societal issues (many of which are influenced by science and technology).

**Synopsis**

Time is of the essence in identifying disease causing microbes (i.e., pathogens) since high growth rates of most microbes means that delays in detection and identification would severely hamper efforts aimed at containing their spread, or provide timely and appropriate treatment to infected patients.\(^1\)\(^-\)\(^3\) The latter is especially important given the increasing prevalence of multi (MDR), extremely (XDR), and totally (or pan) drug resistant (TDR) microbes.\(^4\) Despite the introduction of numerous techniques for rapid and accurate microbe identification,\(^1\)\(^,\)\(^5\)\(^-\)\(^13\) culture (chromogenic agar) and nucleic acid (16S rRNA)\(^14\) approaches remain standard techniques employed in most clinical laboratories around the world - due primarily to the robustness of the methods and availability of analytical equipment. However, the commercialization of mass spectrometry instruments (and accompanying bioinformatics software) capable of identifying microbes with high sensitivity and accuracy over the past decade, may usher in a paradigm shift in the field.\(^1\)\(^,\)\(^15\)\(^-\)\(^17\) This is demonstrated by increasing adoption of the technique in both research and clinical laboratories.\(^18\),\(^19\) Unfortunately, the situation is not mirrored on the education front where the focus remains on conventional identification techniques, with lack of coverage of emerging methods such as mass spectrometric profiling of biomarkers that identify specific species. Driven by the desire to help fill the curriculum gap, I developed an inquiry-based education activity (Figure 1) to introduce students to a leading edge mass spectrometry microbe identification technique (i.e., matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS), (Ng, 2013)\(^20\) where students gain practical and theoretical knowledge of the analytical instrument and associated bioinformatics tools, while employing the scientific method in solving a real world problem with unknown answers: what are the microbes present in an environment water sample?
Figure 1: Schematic diagram outlining the major steps of the inquiry focused education activity. From water sample collection and characterization of microbes by microscopy, to cell cultivation and mass spectrometry analysis, students had the opportunity of participating in all aspects of the discovery process in identifying unknown microbes from an environment water sample. (Adapted from Figure 1 in Ng (2013), JMBE, Vol. 14, No. 1, pp. 103-106)

Capable of ionizing large biomolecules without fragmention, and detecting the resulting singly charged molecular ions through a time of flight (TOF) detector, MALDI-TOF MS enables rapid (minutes per sample) identification of a variety of microbes such as bacteria, spores, mycobacteria, fungi, and viruses from various samples. Species level identification with high accuracy and sensitivity is possible with the technique. Strain and isolate level discrimination is also possible, but the accuracy is less than that at species and genus level. In general, accuracy of MALDI-TOF MS microbe identification progressively decreases from the genus to species and sub-species level. However, active research is underway to improve strain level identification; for example, by developing better sample preparation strategies, culture conditions, new matrixes for facilitating ionization of biomolecules, and data analysis algorithms capable of discerning small mass spectra differences of closely-related microbes. In the clinic, MALDI-TOF MS is able to deliver comparable performance at lower cost per sample in genus and species level identification compared to conventional biochemical, nucleic acid and culture-based methods.
Currently, MALDI-TOF MS microbe identification complements conventional techniques in the clinic and research lab. Specifically, MALDI-TOF MS is capable of identifying most (but not all isolates); thus, it could be used in earlier parts of an identification workflow and leave more challenging identification problems to confirmatory techniques, which are more laborious and time-consuming. In general, strains belonging to *Shigella*, pneumococci and viridian streptococci are difficult to be reliably identified by MALDI-TOF MS.

Compared to 16S rRNA sequencing, MALDI-TOF MS requires a culture step for generating sufficient cells to allow accurate analysis, but the total cell quantity required remains small. Thus, the high sensitivity (and low detection limit) of MALDI-TOF MS allows the direct identification, upon isolation of sufficient number of cells, of difficult to culture species such as those employing anaerobic metabolisms, are slow growing, or which require special growth conditions. In addition, microbes typically exist as communities comprising different species and strains; thus, any microbe identification technique must be able to identify individual species from a mixed population. To this end, various studies have demonstrated, under somewhat idealized conditions, the utility of MALDI-TOF MS in discriminating microbes from populations comprising multiple species and strains.

Sample preparation is crucial to achieving accurate and high sensitivity detection of target analytes in various analytical approaches – and depending on the analytical technique, may be the most complex and time-consuming aspect of the process. For intact cell (also known as whole cell) MALDI-TOF MS, sample preparation is easy and typically involves smearing a small amount of cell sample (for example, from an isolated colony on an agar plate) onto a MALDI target plate, mixing-in an organic matrix (for facilitating ionization of biomolecules), and placing the target in a mass spectrometer (for laser induced ionization). In addition, a protein extraction protocol can be included for enhancing identification accuracy. For example, higher quality mass spectra are obtained by treating cells with hydrolytic enzymes such as trypsin or lysozyme prior to MALDI-TOF MS analysis. Moreover, immunomagnetic and affinity-based separation of cell extracts also help improve mass spectra quality and analytical sensitivity. MALDI-TOF MS also allows multiple samples to be analyzed simultaneously; thus, increasing sample throughput while reducing analysis time. Finally, unlike methods based on polymerase chain reaction (PCR) which requires 16S rRNA gene information for generating requisite forward and reverse PCR primers, prior knowledge of the microbe’s identity is not required in MALDI-TOF MS microbe identification.
Sample preparation and culture conditions are known to affect MALDI-TOF MS’s ability at differentiating closely-related bacteria species and strains, but the issue does not affect the screening of microbes at the genus level. Moreover, delineating the effect of various culture and sample preparation conditions on identification accuracy has facilitated the development of specific protocols for discriminating different groups of closely-related microbes. This is especially important for narrowing down a microbe’s identity after an initial screen (using standard techniques) has revealed an isolate’s putative genus and preliminary species identity. Such tiered and complementary use of standard and case specific protocols for pinpointing microbe provenance would help improve the resolution (and confidence) of sub-species level identification.

MALDI-TOF MS determination of microbe identity depends on the existence of unique sets of biomarkers in each species, which upon mass spectrometric profiling, yields a species or strain specific mass spectrum fingerprint comprising mass peaks at specific mass/charge (m/z) ratios. Such biomarkers include ribosome proteins, unique metabolites (for example, toxins), or signaling molecules secreted by specific species and strains. Besides using ribosome proteins as biomarkers for species identification, MALDI-TOF MS is also capable of discriminating different strains of Shiga toxin secreting E. coli by profiling the toxin produced. Specifically, by combining top-down proteomics approaches with MALDI-TOF MS operated in the tandem MS/MS mode, sequence specific fragment ions indicative of subtypes of Shiga toxin are generated; thus allowing the differentiation of different strains of pathogenic E. coli based not on more conventional ribosome protein biomarkers but by the strain specific toxin secreted. Use of ribosome proteins would not be useful in this instance since they are of high degree of similarity across the different strains.

However, ribosome proteins are good biomarkers for tracking phylogeny across the tree of life given their high relative abundance (which allows ease of detection) and high degree of conservation across species. But small differences in protein sequence exist between various ribosome proteins from different species, which is useful for differentiating between microbes, and more importantly, allows the delineation of their evolutionary relationships. Thus, the remaining discussion will focus on ribosome protein biomarkers for MALDI-TOF MS microbe identification. Specifically, mutations in ribosome protein genes translate into changes in protein amino acid composition and sequence, and mass differences detectable by most commercial MALDI-TOF mass spectrometers, which forms the basis of a mass spectrum unique to each microbe species. With their important physiological role in mediating protein translation, and presence in almost all species across the three domains of life, ribosome proteins are ideally suited for serving as molecular time keepers of evolution, where extent of sequence difference is a good measure of evolutionary distance and relatedness. Thus, a given microbe can be identified by comparing the similarity of its MALDI-TOF mass spectrum with those of known microorganisms catalogued in a mass spectrum reference library, in an approach known as pattern recognition (or mass spectrum...
In a bioinformatics study, a variant of the pattern recognition approach is shown to be useful for discriminating between pairs of microbes at the species and strain level using as input, a single mass spectrum comprising mass peaks from both species. Individual sets of peaks can be discerned and grouped together by comparing the mixed species mass spectrum with those of microbes in a reference library. Collectively, the described method holds promise for differentiating various microbe species and strains present in complex mixed species consortia, if reference mass spectrum is available from each species to be identified or only one species with unknown mass spectrum is present in the set for identification.

Though conceptually feasible, the pattern recognition approach runs into difficulties associated with the high cost of the requisite reference library procured from the instrument maker or developed in-house. An alternative approach (i.e., proteome database search) is capable of circumventing the problem by tapping on publicly available genomic and proteomic information of diverse microbes for annotating mass peaks and thus microbe origins (Figure 2). Specifically, the method attempts to assign each peak in a mass spectrum to a specific biomolecule using information (such as molecule identities, sequences and molecular weight) present in genome and proteome databases.
Figure 2: Conceptual basis of proteome database search in identifying the provenance of mass peaks and microbes. Specifically, after the MALDI-TOF mass spectrum of a microorganism isolate is obtained, possible candidate biomarkers (e.g., ribosomal or ribosome proteins) are proposed and their nucleotide/ amino acid sequence retrieved from relevant sequence databases. With the amino acid sequence in hand, molecular weight of the biomarker can be obtained directly from the protein’s UniProt page, or calculated by summing the individual mass of amino acid residues constituting the protein. Since MALDI-TOF MS generates singly charged non-fragmented molecular ions, the mass/charge (m/z) ratio of the candidate biomarker equals its molecular mass. Finally, alignment of calculated m/z with a mass peak position (within a small margin of error since different instruments have differing mass resolution) associates a specific mass peak with a putative biomarker. Typically, positive microbe identification requires the annotation of at least a few mass peaks – with the goal of identifying as many of the biomarker peaks as possible.

Given the phylogenetic importance of ribosome proteins, annotation of ribosome protein peaks would significantly increase the chances and accuracy of identifying a microbe. Thus, a major focus of the proteome database search approach is the positive identification of peaks belonging to ribosome proteins. Although multiple biomolecules with closely similar molecular mass could potentially be assigned to a specific mass peak, greater cellular abundance of ribosome proteins meant that the probability of other biomolecules accounting for a given peak (especially those of high molecular weight) is low since ribosome proteins are typically of high molecular mass. As mentioned, utility of ribosome proteins as molecular chronometers arises from their high degree of conservation across species on various branches of the phylogenetic tree, where only small sequence differences in selected sections of ribosome protein genes (known as hypervariable regions) separate distinct species and strains. However, most commercial mass spectrometers have sufficient resolving power for discriminating between ribosome proteins that differ in just one amino acid. Hence, the proteome database search approach is capable of discriminating between closely-related species. Finally, although only about 2000 complete and ~2000 draft microbe genome are available in public databases, anticipated improvements in sequencing speed and declining cost (due to promulgation of new sequencing technologies) will likely provide a vastly expanded compendium of genome and proteome information of microbes in the near future; a resource that will undoubtedly facilitate the use of proteome database search approaches for MALDI-TOF MS microbe identification.

Compared to more conventional approaches, mass spectrometry enabled microbe identification techniques in general, and MALDI-TOF MS in particular, are generally not covered in undergraduate life sciences curricula. In addition, exercises designed for teaching bacteria identification do exist, but the focus is always on testing students’ ability at devising a sequence of biochemical and culture assays for identifying unknown microbes through progressive
elimination of candidate species and narrowing of search space, with a goal of using as few tests and in as short a time as possible. This is significantly different from the current case where bioinformatics tools are used in searching genomic and proteomic databases for annotating peaks in mass spectra. But, from the perspective of bridging the curriculum gap, what education tools would be most effective in introducing the necessary concepts and practical experience to students, while infusing an element of fun and igniting students’ curiosity? Accumulating evidence from education research indicates that inquiry-based exercises with practical experiential components (also known as active learning) are useful for increasing students’ motivation towards learning and enhancing content mastery. Specifically, acquisition of conceptual understanding and problem-solving skills often go beyond algorithmic exercises; for example, students need to feel a sense of challenge in working on an interesting problem, which would further motivate them to go the extra mile in thinking about various solution strategies. Finally, casting a forward glance, the next generation of students would likely judge the quality of teaching and the education they received by yardsticks different from those of earlier eras. These criteria include extent of critical thinking and active learning incorporated in course exercises (e.g., homework, laboratory classes, field work etc.), which further highlights the importance of active learning education tools in stimulating interest in the subject matter and motivating students to learn.

Drawing from teaching experiences documented in the pedagogy literature, I designed a learning module for guiding students in identifying microbes in environment water samples using a combined MALDI-TOF MS and proteome database search approach. By solving a research problem with unknown answers, students put into practice the scientific method they learnt (through flowcharts) in textbooks but seldom employ in solving problems. In addition, students were exposed to the complete suite of techniques and practical skills - ranging from collection of water samples and cell culture, to mass spectrometry analysis and bioinformatics search tools. Suitable either as a standalone laboratory module, or as a complement to a microbiology or bioinformatics class, the described activity helped, in a small way, to acquaint students with how scientists conduct field work – and thus, bring science closer to daily life. Specifically, students experienced the excitement inherent in exploring the natural world through the rarely used education tool of field work, while being equipped with technical and thinking skills useful either in industry or graduate school. More importantly, the practical experience also helps students realize that science is not an esoteric activity secluded in university labs – but rather, is a way of life where a logical sequence of questions guides investigation and solution of real world problems. By illuminating the close connection between science and the world around us, such inquiry-based education activity would also help increase students’ interest in science, and hopefully translates into higher retention of students in STEM (science, technology, engineering and mathematics) fields. With a strong interest in science and technology seeded in university, there is a higher chance that students will continue to stay engaged in science; for example, by participating in the wide variety of citizen science projects available. Even if students do not pursue a scientific career after graduation, or employed in science and technology jobs, the analytical, logical and critical thinking skill acquired through a science education would enable them to disentangle the
myriad interacting factors typical of problems in a modern economy, as well as assess the scientific evidence surrounding important societal issues such as climate change and genetically modified food. Modular in design, elements of the exercise can be easily tailored to suit the specific needs and objectives of individual cohorts and classes; for example, determining the set of microbes present on plants’ leaves would also achieve similar education outcomes as in determining microbe species in water samples. Finally, tips on implementing the exercise and some background information on MALDI-TOF MS are also available in the article and associated supplementary information.

To conclude, a laboratory exercise combining active inquiry by students during identification of microbes in water samples, with practical training in cell culture, mass spectrometer operation, and bioinformatics tools, was developed for teaching mass spectrometry enabled microbe identification. Imbued with a fun element for motivating students to learn, the activity put students through the paces in using deductive and inductive thinking for solving a real world problem with unknown answers, and in the process, helping them gain a better appreciation of how scientists work, while at the same time, igniting their inquiring minds and cultivating an interest in science and engineering.

The article and supplementary material is available at *Journal of Microbiology and Biology Education*, Vol. 14, No. 1, pp. 103-106, as an open access article, URL: http://www.asmscience.org/content/journal/jmbe/10.1128/jmbe.v14i1.494.

**References**

20. Ng, W. (2013), Teaching Microbial Identification with Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) and


New in this version

Language, sentence structure, readability, and explanation of concepts are improved in this version.

Conflicts of interest

This synopsis describes a published paper written by the author.

Author’s contribution

The author would like to discuss a piece of his published education research to a wider audience of scientists, students and public in an accessible way. Thus, an idea came to his mind to write a synopsis explaining the key idea of the article. Furthermore, he would also like to discuss his published research in the light of new developments in the field of mass spectrometry enabled microbe identification. He conceived the idea, read the literature and wrote the manuscript.
Funding

No funding was used in writing this synopsis article.