Reconciling Microbial Systematics

Chapter 10

MALDI TOF MS and currently related proteomic technologies in reconciling bacterial systematics

Haroun N. Shah[1], Ajit J. Shah[1], Omar Belgacem[2], Malcolm Ward[1], Itaru Dekio[3], Lyna Selami[2], Louise Duncan[4], Kenneth Bruce[4], Zhen Xu[5], Hermine V. Mkrtchyan[6,7], Rory Cave[7], Laila M.N. Shah[8] and Saheer E. Gharbia[9].

[1]Department of Natural Sciences, Middlesex University, London, NW4 4BT

[2]Ascend Diagnostics Limited, CityLabs 1.0, Nelson Street, Manchester, M13 9NQ

[3]Department of Biochemistry and Integrative Medical Science, School of Medicine, Keio University Japan

[4]King's College London, Molecular Microbiology Research Laboratory, Pharmaceutical Science Research Division, Franklin-Wilkins Building, 150 Stamford Street, London, SE1 9NH

[5]Tianjin Key Laboratory of Environment, Nutrition and Public Health, Department of Toxicology and Sanitary Chemistry, School of Public Health, Tianjin Medical University, Qixiang Road No. 22, Tianjin, China, 300070

[6]School of Biomedical Sciences, University of West London, St Mary's Rd, London, UK, W5 5RF

[7]School of Health, Sport and Biosciences, University of East London, Water Lane, London, UK. E1 4NS.

[8]King's College London, Faculty of Natural and Mathematical Sciences, Department of Chemistry, Franklin-Wilkins Building, 150 Stamford Street, London, SE1 9NH

[9]Public Health England, Department of Gastrointestinal Pathogens, 61 Colindale Avenue, London, NW9 5EQ

10.1 Introduction

The introduction of the Gram stain over a century ago into Clinical Microbiology has been profound and enduring (Gram, 1884). The test, based on the cell's colour reaction to the stain, enabled the dichotomous division of the microbiological kingdom and stimulated microbiologists to seek new methods to further characterise potential pathogens. Today, even in the current era of whole genome sequencing (WGS), many traditional diagnostic laboratories such as those long-established in Public Health England, UK are designated as enteric 'Gram- negative' or 'Gram positive' food pathogens to describe their function. As new physiological tests and subsequently chemotaxonomic methods were introduced into microbiology, various editions of Bergey's Manual of Determinative Bacteriology (from 1923 to 1974), assigned chapters/volumes of different taxa in accord with their reaction to the Gram stain. There was a brief period when peptidoglycan patterns, based mainly upon the characteristic dibasic amino acids lysine or diaminopimelic acid were considered as a new, well defined dichotomous key for the microbial kingdom (Schleifer and Kandler,1972). However, while this was retained as a key character for the description of bacterial species, various anomalies within some genera such as the *Bacteroides* or *Fusobacterium* prevented this from being universally adopted.

Microbiology changed fundamentally with the introduction by Sneath (1957) of numerical phenetic analysis which strongly advocated a transition from single weighted tests to multiple based characters to delineate nomenclatural groups. The arrival of DNA/DNA reassociation and subsequently16S rRNA re-enforced a move towards a phylogenetic-based structure and added rigor to the definition of a species. Thus, as microbiology progressed from an arbitrary identification system to one based upon systematic classification principles, new characters were reported and microbial cellular components were analysed for the first time to maintain a polyphasic approach (see Goodfellow and Board, 1980; Goodfellow and Minnikin, 1985, Minnikin and Goodfellow,1980). During this process, mass spectrometry was used for the first time for microbial systematics. Both polar and particularly non-polar lipids revealed enormous complexity and diversity and exhibited good congruence with established taxonomic methods. Very multifaceted, hitherto, poorly defined groups such as the "Acid-Fast" bacteria that encompassed several ill-defined taxa such as *Mycobacterium, Nocardia, Actinomadura, Corynebacterium, Rhodococccus* and others began to show resolution. Such data provided the impetus to drastically transform microbiology from a 'determinative' to a 'systematic' based approach and for the first time Bergey's Manuel reflected this change to *Bergey's Manual of Systematic Bacteriology* (1984) and microbial classification was no longer pivotally structured around the Gram stain (Shah and Gharbia, 2011).

10.2 Proteins in microbial systematics

The use of proteins profiles in microbial systematics has long been regarded as highly compatible with phylogenetic approaches since proteins were envisioned as a direct product of the genome. Numerous studies were reported that showed excellent correlation between protein profiles and methods such as DNA/DNA reassociation (see review Jackman, 1985). The immense complexity of the bacterial proteome was already established through studies of intermediary metabolism and microbial biochemistry. However, although protein sequences were being reported as early as 1967 (Ambler, 1967) and its immense value in microbial systematics demonstrated (Ambler, 1985), existing technologies were considerably laborious, costly and outside the scope of bacterial systematics. Microbiologists therefore

turned towards electrophoretic-based platforms to provide insight into the expressed proteome of bacterial species. Three basic approaches were pursued viz. multilocus enzyme electrophoresis (MLEE, Selander *et. al.,* 1986), peptide/protein profiles using sodium dodecyl sulphate-polyacrylamide gel electrophoresis ([SDS-PAGE, Laemmli, 1970) and isoelectric focusing protein profiling (IEF-PP) (eg. Shah *et al.,* 1982). When analyses commenced, optimisation of methods to release cellular proteins were based largely upon the Gram stain but this was due to differences in complexity of the cell envelope rather than taxonomic rationale. Early results revealed enormous implications for bacterial systematics and soon methods such as multilocus enzyme electrophoresis (MLEE), based upon the electrophoretic mobility of specific enzymes (see eg. Selander *et. al.*, 1986) became established as a tool for microbial systematics. This subsequently translated into its high-resolution counterpart, Multilocus Sequence Typing (MLST), based upon comparative DNA sequencing of the selected gene. MLST is used widely today for systematics, evolution and epidemiology (see eg. High *et al.,* 2015). However, both SDS-PAGE and IEF-PP continued to have a significant impact in microbial identification and phylogeny. In an early study, strains designated *Bacteroides oralis* were shown to exhibit such profound heterogeneity using IEF-PP that it led to the proposal of three new species (Shah and Collins 1981, Shah *et al.,* 1982). Results corroborated with other chemotaxonomic methods and subsequently by comparative 16S rRNA sequencing (Shah and Collins, 1981). SDS-PAGE is robust, inexpensive, simple and rapid and is still used in its original form today by some laboratories (see eg. Berber, 2004).

10.3 Arrival of MALDI-TOF MS in Microbiology

A major drawback of the above electrophoretic methods is that they rely solely on pattern-matching algorithms and provide no information on the protein biomarkers that were used to differentiate taxa. In theory new forms of mass spectrometry (MS) such as MALDI-TOF MS and Tandem MS/MS should have provided the analytical tools to develop these methods, analogous to that achieved for transforming MLEE to MLST. However, systematic analysis using whole cell proteins profiles are only now gaining momentum as new forms of mass spectrometry become more accessible. Current MS enable identification of post-translational modification of proteins that cannot be deduced from WGS and is likely to gain widespread application both is cell biology and systematics as these techniques become automated, accessible, miniaturised, portable and less costly.

The platform that paved the way for these developments is Matrix-Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry (MALDI-TOF MS). Earlier forms of mass spectrometry that were used in microbiology such as electron impact or pyrolysis mass spectrometry had upper detection limits of *ca.*1,500 daltons and were therefore restricted to small molecules such as lipids and metabolites (Goodfellow and Minnikin, 1985). Proteins are orders of magnitude greater and initial success was reported in the analysis of small proteins using Fast Atom Bombardment Mass Spectrometry in the mid-1990s (see eg Drucker, 1997) but this soon gave way to the more efficient and versatile platform, MALDI-TOF MS which prominent microbiologists have described as 'transformational'.

Pioneered by the late Franz Hillenkamp's group in the mid-1980s (Karas, Bachmann and Hillenkamp, 1985; Karas *et al.,* 1987), MALDI-TOF MS soon became established as a method for analysis of high molecular weight compounds, particularly proteins (Karas and Hillenkamp, 1988;Tanaka *et. al.,*1988). The work of Cain, Lubman and Weber (1994) drew attention to its first application in microbiology through their publication, *"Differentiation of bacteria using protein profiles from matrix assisted laser desorption/ionization time-of-flight mass spectrometry"*. The response by the Mass Spectrometry company, Kratos Analytical,

was to launch the first 'bench-top' MS instrument that was ground breaking for this early period. Almost simultaneously, three papers reported its use 1996, (Claydon *et. al.,*1996; Holland *et. al.,*1996 and Krishnamurthy and Ross, 1996) and provided further evidence of its potential. The response by microbiologists was inconsequential perhaps due to the work being reported in non-microbiological journals and possibly more significant, the previous failure of MS methods such as Pyrolysis MS (see eg Shute *et. al.,*1985) to gain acceptance as a tool for bacterial systematics.

10.4 Establishing MALDI-TOF MS in clinical Microbiology

In mid-1997, Public Health England (then Public Health Laboratory Service and later Health Protection Agency) established a new laboratory (Molecular Identification Services Unit, *MISU*) to identify novel, emerging and atypical pathogens that were increasingly being reported in the UK. This provided an opportune moment to access the potential of new and emerging technologies and the various MALDI-TOF MS platforms were among those introduced with almost immediate impact for *MISU*. The laboratory was in the midst of assembling a database of MALDI-TOF MS that included most staphylococcal species. A sample labelled "atypical *Staphylococcus aureus"* was received by *MISU* for identification. Its MALDI-TOF MS spectrum was shown to be incompatible with any staphylococci and the isolate was identified by 16S rRNA as *Exiguobacterium aurantiacum.* The inclusion of spectral data into the database facilitated the subsequent rapid identification of 18 blood cultures using the Kratos Linear MALDI-TOF MS and allowed notification to be sent to hospitals to alert them of impending threat. The clinical application of MALDI-TOF MS increased substantially and early reviews reflected its increasing optimism (see eg Belén *et al.,* 2019; Singhal *et al.,* 2015). It was already hailed as "the quantum leap" by Greub (2010) and later regarded as the gold standard for clinical microbiology (Schubert and Kostrzewa (2017). However, the clinical applications of MALDI-TOF MS, which was accredited and widely used in Europe, only gained widespread success in North America when quality assurance procedures were subsequently established (Bourassa, 2018) and FDA approval received.

From 1998, 21 consecutive international conferences were held at PHE and subsequently with Middlesex University, London to advance the application of these technologies in microbiology (see Shah and Gharbia, 2017). Among the early MALDI-TOF MS methods was the novel application of DNA resequencing that was developed by Sequenom GmbH Hamburg. The system designated MassARRAY single-nucleotide polymorphism (SNP) typing platform employed a MALDI-TOF MS coupled with single-base extension PCR for high-throughput multiplex SNP detection (see review, Honisch, Chen and Hillenkamp, 2010). Very complex traditional serotyping methods such as the Kauffmann-White for the rapid typing of Salmonella were shown to be amenable for transfer to the Sequenom MALDI-MS platform (Bishop, Arnold, and Gharbia, 2010; Syrmis *et. al.,* 2011). However, despite its huge potential for microbial analysis, the technology never gained traction in Microbiology although it continued to find applications in clinical diagnostics and is currently used for detection of cancer biomarkers (see eg Gloss *et. al.*, 2012). Subsequently, PCR based products were analysed using tandem MS/MS which was excellent for species identification and typing but perhaps because of its colossal cost, the system, designated PCR-ESI-MS, was not acquired by laboratories and soon became obsolete (see review, Emonet et. al., 2010).

Protein-based target molecules were viewed with more optimism by microbiologists and different approaches were explored. While the standard MALDI instrument developed by Kratos Analytical Ltd employed a stainless target plate for sample analysis, the Ciphergen Biosystems instrument utilised a chemically active surface (ProteinChip Array) to selectively

capture various classes of proteins (Figure 10.1). Thus, the same sample, using sinapinic acid as its matrix solution, could be analysed by using various ProteinChips to obtain more in-depth coverage of the proteome (Shah *et al.,* 2010).

[**Insert Figure 10.1**. MALDI-TOF MS profiles obtained for the same sample using three different ProteinChip arrays (NP1, SAX2 and WCX1) to reveal the complexity of a sample. To enable rapid visual comparison of spectra, the software converted the mass ions into a Gel view images; the greater the mass intensity, the darker the band. (Adapted from Shah *et al.,* 2010, Changing concepts in the characterisation of microbes and the influence of mass spectrometry. In: Shah H.N. and Gharbia, S.E. (eds). '*Mass Spectrometry for Microbial Proteomics'* Wiley, Chichester, pp. 3- 34].

Datasets obtained from SELDI-TOF MS analysis were complex and the accompanying software utilised heatmaps to visualise differences in mass ion intensities (Shah *et. al.,* 2010). However, raw data could also be extracted and analysed separately. Artificial neural networks were generally employed to assess the diversity of species (Encheva *et. al.,* 2005; Schmid *et. al.,* 2005; Lancashire *et. al.* 2005, Hamilton *et. al.,* 2010, Chiu, 2014). Large proteins of >150,000 daltons were reported while mass spectra were so vast that PHE considered this as more a comprehensive and robust method for typing bacterial species (Shah *et. al.,* 2005, Encheva *et. al.,* 2006). The launch of Ciphergen Biosystems next generation ProteinChip System, Series 4000 in 2004 enabled rapid biomarker discovery and development of predictive, high throughput SELDI-based analysis. Unfortunately, laboratories with the first generation MALDI -TOF MS instruments on which considerable background was undertaken, were left unsupported. In 2006, the company was acquired by Bio-Rad and there was an abrupt loss of interest for the technology in microbiology.

It was evident that for any of these technologies to have a future, a comprehensive microbial database of mass spectral profiles of species was required. Early protocols for analysis of whole cells by MALDI-TOF MS again resorted to the use of different matrix solutions based upon the Gram stain. In general, laboratories utilised alpha-cyano-4-hydroxycinnamic acid for Gram-negative cells and 5-chloro-2-mercaptobenzothiazole for Gram positive cells. Having optimised a method based on available data, considerable in-house laboratory work was undertaken to initiate this process (Shah *et. al.,*2000, 2002). With the arrival of a dedicated new upright linear MALDI-TOF MS instrument by Micromass (later acquired by Waters Inc), four years was spent developing the first mass spectral database (Key *et. al.,* 2004). Proof of concept was demonstrated through a direct trial at The Royal London Hospital, University of London's diagnostic laboratory and was highly successful. More than 600 isolates were analysed in parallel with traditional tests undertaken in the hospital's laboratory and yielded an 80% concordance (Rajakaruna *et. al.,* 2009).

During this period, infections due to *Clostridium difficile* in the UK was at its peak and MALDI-TOF MS was again used to access its potential value. Initial results were spurious and led to a re-examination of the basic protocol. A change from 5-chloro-2-mercaptobenzothiazole to 2,5-dihydroxy benzoic acid in acetonitrile: ethanol: water (1:1:1) with 0.3 % TFA (see Shah *et. al.,*2010) led to a dramatic improvement in results. Similar results were also obtained using α-cyano-4-hydroxycinnamic acid. Furthermore, re-optimisation of protocols led to the use of a preliminary extraction with formic acid and the use of α-cyano-4-hydroxycinnamic acid for both Gram positive and Gram-negative cells. This was demonstrated in several laboratories by a group who met periodically to optimise methods (Kallow *et. al.,* 2010). It was soon evident that MALDI-TOF MS spectra gave similar results to its corresponding ribosomal RNA preparation of various species and provided unequivocal proof that these are the molecules that gave rise to the specific mass

ions that comprised the spectrum of a given species. This removed the need to standardise the growth conditions since the mass spectra were not adversely affected by the physiological conditions of the cell. A universal protocol was now established for use with all microbial species and interest in the technique gained momentum. Commercial databases were now available through Bruker Daltonik GmbH (Bremen, Germany) and AnagnosTec (Potsdam, Germany, subsequently acquired by bioMerieux).

In 2011, the UK was in the midst of preparations for staging of the London 2012 Olympics. Part of these preparations was a responsibility by Public Health England (PHE) to implement technologies that enabled rapid identification of pathogens in the event of an infectious disease outbreak. A proposal by the Molecular Identification Services Unit, PHE led to the placement of six MALDI-TOF MS instruments initially and a further 10 instruments in major cities in the UK. It represented the largest network of instruments in one organisation globally and endorsement of the technology after 13 years of consolidated work. It continues to function to the present and retained parallel analysis using comparative 16S rRNA sequence analysis.

10.5 MALDI-TOF MS in the non-clinical laboratory and its role is searching for new diversity.

Because MALDI-TOF MS databases were assembled using mainly clinically relevant strains for species identification, it was assumed initially that application of the technology for identification of isolates from other sites such as the soil, marine, freshwater, agricultural, poultry, and industrial waste, bioremediation in landfill sites, inhospitable sites, manufacturing etc would require new databases with correspondingly relevant species. The method began attracting interest a decade ago and several specialist groups utilised an existing database in tandem with 16S rRNA sequencing to extended applications of the technology. Species that could not be detected by MALDI-TOF MS but identified by 16S rRNA were added to the database for studying microbial communities of hitherto poorly studied sites. Such an approach in which MALDI-TOF MS and 16S rRNA are used in parallel and the latter used to add new diversity to a pre-existing mass spectral database has successful moved the technology well beyond the clinical laboratory where it had its roots. For example, MALDI-TOF MS and the 16S rRNA were used in tandem for characterising the cultivable bacterial communities from polluted soils and water following copper mining. The results revealed that MALDI-TOF MS analysis was reliable and could be used as a rapid tool for identifying copper-resistant bacteria (Avanzi *et.al.,* 2017). As anticipated some results were equivocal but accuracy was improved by enhancing the reference database. In a very interesting application, Timperio *et. al.,* (2017) studied Arctic bacteria isolated from the White Sea, Russia in parallel with 16S rRNA. Agreement between both methods was 100% at the genus level but decreased to 48% at the species level but subsequently remedied by the inclusion of the missing reference spectra to the database. Notable examples were strains of *Exiguobacterium oxidotolerans* and *Pseudomonas costantinii* that were misidentified initially using the MALDI BioTyper due to the absence of reference spectra in the database. Interestingly, the concordance for *Pseudomonas* species was low (29%), confirming the problematic taxonomy of this genus and highlighted areas where the use of MALDI-TOF MS may be used in systematics. In vastly different study, the geo-microbiology of a 300 m stretch of a rivulet with very hard water (>120 mg Ca^{2+}/L) in a northern-Germany karst hardwater creek, the Westerhofer Bach was investigated using 16S rRNA and MALDI-TOF MS as a phenotypic method. Some 35 genera were identified which included a predominance of Flavobacterium, Pseudomonas and Stenotrophomonas and some 60 novel phylospecies. The study provided deep insight into the bacterial community of the geosphere and biosphere

of a non-marine environment (Cousin *et al.,* 2008) and demonstrates the versatility of the technique.

Habitat selection and human transmission of *P. aeruginosa* is currently being investigated. Strains from environments across London are being sampled and analysed by MALDI-TOF MS following presumptive identification on agar plates. MALDI-TOF MS in parallel with 16S rRNA revealed that several phenotypically closely related isolates that produce the characteristic fluorescent and green pigments on agar may have been incorrectly misidentified *P. aeruginosa,* instead of *Pseudomonas citronellolis* (Louise Duncan, current PhD in progress, unpublished). *P. citronellolis* is a documented soil microbe that exhibits a strong biotic relationship with pine trees or basil plant host (Seubert, 1960; Remus-Emsermann *et. al.,* 2016). Interestingly, *P. citronellolis* was reported for the first time in a case of human infection (Williams, 2019) and demonstrates the value of MALDI-TOF MS in studying microbial community habitats (Duncan *et. al*., 2019). Investigation into these microbial communities reveal that phenotypical colour formation is not an exclusive property of *P. aeruginosa* (Pirnay *et. al*., 2009; Batrich *et. al*., 2019). Therefore, MALDI-TOF MS proved to be a highly sensitive and accurate tool for identifying diverse species of *Pseudomonas*.

The application of MALDI-TOF MS to bacterial identification is key to both diagnostic laboratories and industry. Food and beverage manufacturers are disposing of their traditional microbiology laboratories methods and turning toward the routine usage of MALDI-TOF MS through various stages of production (see egs, Pavlovic *et al.,* 2013; Santos, Hildenbrand and Schug, 2016). It facilitates rapid communication of important quality control results thus preventing the contaminated products from leaving the manufacturing environment to distributors and the subsequent consumer (see Bourassa, 2018)

These studies, together with the huge volume of reports on the successful use of MALDI-TOF MS in clinical laboratories indicate that where the taxonomy of species are unambiguous and the associated reference spectra are in the database employed, the use of MALDI-TOF MS for microbial species identification appears likely to be continued for the foreseeable future. The exception to date is Mycobacterium spp. where the rigid cell wall contains very complex long-chained fatty acid of up to 80 carbon atoms and additional extraction methods are required. Furthermore, many species are so poorly resolved that a separate database is generally created for this genus and lower threshold scores are used to identify some species (see egs, Kim and Kim, 2017, Kim *et. al.,* 2020, Pranada *et. al.,* 2017, Rose *et. al.,*2017).

10.6 MALDI-TOF MS in subspecies identification, typing and screening for genetic variants. Implication for systematics.

Among certain taxa where subspecies level identification is apparent and better defined, MALDI-TOF MS has been used to delineate such diversity. Among members of the genus *Lactobacillus,* several species and subspecies are recognised and commonly used as probiotics. Most species of *Lactobacillus* may be identified using MALDI-TOF MS and several subspecies for example, *Lactobacillus delbrueckii* subspecies *delbrueckii, Lactobacillus delbrueckii* subspecies *lactis, Lactobacillus delbrueckii* subspecies *indicus*, which are favoured by various probiotic companies and needs authentication and verification, may also be delineated by MALDI-TOF MS. Many companies such as Natren Inc, Westlake Village, California, USA, previously used long-chained fatty acids to detect minor strain variation to characterise starter cultures and ascertain strain stability during manufacturing. Today, this has been transformed by the use of MALDI-TOF MS for rapid and routine monitoring. Public Health England has worked closely with Natren Inc to assess the potential

of MALDI-TOF MS to identify probiotic *Lactobacillus* spp and *Bifidobacterium* spp. in parallel with 16S rRNA and, in some species, whole genome sequencing and found excellent correlation (see eg Figure 10.2).

[**Insert Figure 10.2**. The separation of *Lactobacillus* species and subspecies using Bruker Biotyper (software version 3.0). The results highlight the capacity of MALDI-TOF MS to delineate some taxa to the subspecies level. These include: *Lactobacillus delbrueckii* subspecies *delbrueckii, Lactobacillus delbrueckii* subspecies *lactis* and *Lactobacillus delbrueckii* subspecies *indicus* which are frequently used by various probiotic companies].

Numerous reports highlight the use of MALDI-MS in the detection of genetic variants and the typing of a vast range of species such as *Staphylococcus aureus*, *Pseudomonas* spp, *Clostridium difficile*, *Bacillus* spp, *Listeria monocytogenes, Salmonella* spp, *Streptococcus* spp, *Escherichia coli*, *Klebsiella pneumoniae*, *Acinetobacter baumannii* and others (see Biswas, Gouriet and Rolain, 2016). We have investigated the capacity of MALDI-TOF MS to subtype the above taxa against a recognised DNA-based typing system and have consistently failed to substantiate such correlations. It is sometimes assumed that because MALDI-TOF MS spectra comprise mainly highly conserved ribosomal protein structures, dendrograms would mirror 16S rRNA trees. However, concordance between both is uncommon except where taxa have very high gene sequence similarities (>98.2 %) (Schumann and Maier, 2014). However, we could not demonstrate this for strains of *P. aeruginosa* that were >99% similarity in 16S rRNA and bifurcations were not shared in dendrograms. Figure 10.3 shows an example of the discordance among isolates of *Pseudomonas aeruginosa* using MALD-TOF MS against the DNA-based typing system (VNTR's) that is used routinely at Public Health England. Similarly, Schumann and Maier (2014) reported the dendrogram generated on the basis of MALDI-TOF MS did not correspond with the topology of the tree derived from their 16S rRNA gene sequences for the type strains of type species of genera of the family Microbacteriaceae or genus *Arthrobacter.*

[**Insert Figure 10.3.** Dendrograms showing phenotypic similarities and relationships of 53 *Pseudomonas aeruginosa* strains from cystic fibrosis (CF) and non-CF sites and the same strains analyzed genetically using VNTR (variable number tandem repeat). Congruence between the phylotypes was negligible and indicate that one method cannot supplant another. (Adapted from Olkun, Shah and Shah, 2017, 'Elucidating the intraspecies proteotypes of *Pseudomonas aeruginosa* from cystic fibrosis. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 579 – 592].

Each method is based on different target molecules and was expected to differ as these results demonstrate. If the long-term goal of such studies is to use MALDI-TOF MS to replace a recognised DNA-based typing system, our studies have shown that this is not achievable. However, if the objective is to use MALDI-TOF MS exclusively to establish a 'mass spectral typing' system for use within a specific laboratory, indications are that this is achievable. Figure 10.4 shows samples from a study to map the transmission of *Staphylococcus aureus* and related species in the community. A total of 411 isolates were recovered from the general public and various environmental sites in London. The MALDI-TOF MS data was analysed using the software BioNumerics (Applied Maths; Vranckx, De Bruyne and Pot, 2017) and evidence of clustering within the same sites were apparent with the human hand being the reservoir (Xu *et al.,* 2017;). Nineteen species of staphylococci were

identified, most of which were coagulase negative. BioNumerics revealed hierarchical interrelationships of nine major clusters comprising *S. hominis, S. haemolyticus, S. epidermidis, S. pasteuri, S. warneri, S. aureus, S. saprophyticus, S. capitis* and *S. simiae* (Figure 10.4).

[Insert. Figure 10.4 (Left) Unrooted cluster analysis of staphylococcus species in the community based upon MALDI-TOF MS spectral profiles. While each species was distinctively delineated, the data enabled the intraspecies diversity to be clearly discerned (Xu *et. al,* 2017. Adapted from "Subtyping of *Staphylococcus* spp. based upon MALDI-TOF MS data analysis. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 563 - 378.

[Insert. Figure 10.4 (Right) Three-dimensional scatter plot *Staphylococcus aureus* isolates using a supervised method such as linear discriminant analysis (LDA, Bionumerics) to show relationships among isolates from specific sites which may be useful for identifying unique traits during transmission (Adapted from Vranckx, K., De Bruyne, K. and Pot, B. (2017) "Analysis of MALDI-TOF MS spectra using the BioNumerics software". In: Shah, H.N. and Gharbia, S.E. (eds). MALDI-TOF and Tandem MS for Clinical Microbiology. Wiley, Chichester, UK, pp 539-562.]

It is our view that with the current level of resolution of the technology, MALDI-TOF MS is excellent for species and in some instances subspecies-level identification but does not provide a means to type isolates in a manner similar to DNA-based methods. Table 10.1 highlights some of obstacles to develop a universal typing platform analogous to the databased developed for species identification. However, the value MALDI-TOF MS as a tool for microbial systematics has been grossly undervalued and reports in systemic journals remain sparse despite an early report by Schumann and Maier (2014) that provides detailed methods for adapting MALDI-TOF MS towards systematics applications.

[Insert Table 10.1 Barriers to using MALDI-TOF MS exclusively as a universal typing tool]

10.7 MALDI-TOF MS in microbial systematics; a case study involving *Cutibacterium acnes.*

The widespread use of comparative 16S rRNA sequence analysis enabled the transition of a determinative microbial classification system to a phylogenetic structure (see above, 10.1). Systematic subcommittees provide essential information on the proposal of new species and continues to ensure that a polyphasic description is still the desired approach (see Chapter 2). However, several species are still proposed solely on the basis of 16S rRNA where a level >98.8% has been detected and where there is a paucity of reliable characters to facilitate its identification. While whole genome sequencing (WGS) represents the most comprehensive data available to describe a new taxon, acquired genetic elements may skew the description of the new entity. Consequently, *in silico* DNA-DNA hybridisation or ANI values are now used to circumscribe new diversity (see eg. Dekio *et al.,*2015). For many new fastidious species, especially non-fermentative taxa, characteristic phenotypic characters may be extremely difficult to discern. MALDI-TOF MS has the potential to provide unique characters. The current case report (below), based on *Propionibacterium acnes,* helped in its

reclassification to *Cutibacterium acnes* and provides an example of how these techniques may be used to reconcile microbial systematics.

10.7.1 Brief biology of *Cutibacterium acnes.*

C. acnes (previously, *Propionibacterium acnes*) is a facultative anaerobic bacterium and one of the most abundant resident species of the human skin, with a count of 10^5-10^6 cells/cm^2 on the facial skin (Dekio *et al.,* 2007). Although it is considered beneficial to humans, it plays a fundamental role in the development of inflammatory acne, a disease that gave the species its specific epithet. Moreover, it also inhabits the eye, oral cavity, large intestine, and genito-urinary tract of humans, and causes opportunistic infections in the cornea, prostate cancer, surgically-treated bone, and blood-borne illnesses.

Differences in cell wall composition led to subdivision of the species into types I and II (Johnson and Cummins 1972), and later a group with a unique filamentous morphology, designated type III (McDowell *et al.,* 2008). These groups were found to be serologically unique and using the housekeeping genes, *recA* and *tly* to be phylogenetically distinct (McDowell *et al.* 2005, 2008). In attempt to resolve the taxonomic substructure of the group, in-depth analysis using genomics and proteomics against known morphological and biological features were undertaken in our laboratory (Dekio *et al.* 2015) and those of McDowell *et al.* (2016). To clarify the role of *C. acnes* in acne, studies have focussed on potential virulence determinants which finally led to the conclusion that type I was responsible for inflammatory acne (Lomholt & Kilian, 2010).

10.7.2 MALDI-TOF MS delineates three proteotypes

Finding characters that consistently discriminate bacterial strains below the species level is often fraught with difficulties. In an attempt to delineate these subtypes, we collected MALDI-TOF MS spectra of *C. acnes* strains of a dozen MLST types belonging to the types I, II, and III. As expected, the majority of peaks were common among the three types, but peaks at 6950-7200 m/z were found to be highly indicative of each type and may be used as biomarkers (Figure 10.5), (Nagy *et al.,* 2013, Dekio *et al.,* 2015). These biomarkers are clearly detected by MALDI-TOF MS and have been confirmed at four different institutes, Keio University, Japan, Japan Collection of Microorganisms, Japan, Public Health England, UK and Middlesex University, U.K. with acceptable shifts in m/z values.

[Insert Figure 10.5. Unique biomarker mass ions in the MALDI-TOF MS spectrum of *Cutibacterium acnes* subspecies (Dekio *et al*. 2015).

These results were corroborated using another form of MALDI-TOF MS, SELDI-TOF MS which analysed a far greater depth of the proteome (see above, Sect 10.4). Furthermore, the more precise analysis with the wider mass range of mass ions provided by SELDI-TOF MS was found to be useful in functional investigations. *C. acnes* isolates showed significantly higher expression of a protein with a mass range *ca.*15-17kD in type I (Figure 10.6). This expression disappears almost completely when cells were cultured under aerobic conditions (Dekio *et al*. 2013).

[Insert Figure 10.6. Unique biomarker mass ions between 10-20 kD segment of the SELDI-TOF MS spectra of anaerobically-cultured *Cutibacterium acnes* isolates

We further investigated this peak following preliminary separation by 1D SDS-PAGE, protein extraction and tandem LC/MS/MS using a ThermoFisher Orbitrap. Among several proteins contained in this gel-band, CAMP factor, an infection-related protein, was identified. This protein is hypothesised to be a key protein related to development of inflammatory acne (Dekio *et al.,* 2013).

10.7.3 Correlation of proteotypes with whole genome sequencing

Since the taxonomic basis of the prototypes of *Cutibacterium acnes* was established using primarily MALDI-TOF MS and its extended form, SELDI-TOF MS, a key consideration was to investigate the rigor of these subtypes by their correlation with genome sequencing. While they showed 16S rRNA gene similarity of > 99.3%, the genome similarity of each group, calculated by *in silico* DNA-DNA hybridization, were in the range 78-72%, slightly higher than that accepted for a species cut off of 70% (Dekio *et al.,* 2015). This resulted in the confirmation of the three subspecies *viz*. *C. acnes* subspecies *acnes*, *C. acnes* subspecies *defendens*, and *C. acnes* subspecies *elongatum*, for types I, II, and III, respectively (Dekio *et al.,* 2019, McDowell *et al.* 2016). Unlike many taxa where there is a paucity of reliable characters for subspecies-level identification, these results demonstrate unequivocally the huge potential of MALDI-TOF MS and proteomics in microbial systematics.

[**Insert** Figure 10.7. Genome phylogram for type strains of *Cutibacterium* species and subspecies that confirmed the initial data obtained by MALDI-TOF and SELDI-TOF MS analyses and emphasise the value of the latter techniques in preliminary screening of isolates from diverse habitats].

10.8 MALDI-TOF MS and the future interest of MS companies

10.8. 1 The current and long-term microbiological usage of MALDI-TOF MS is entirely dependent on the interest and development of leading MS companies. Around the globe there are numerous MS companies and several manufacture MALDI-TOF MS instruments, yet despite the success of the technology and its transformative impact in clinical microbiology, only two companies, Bruker Biotyper (Bruker Daltonics, Germany) and bioMérieux (in partnership with Shimadzu, Japan; VITEK ® MS System, France) dominated the global market. However, this is likely to change because more companies are showing an interest to enter the field. The arrival of ASTA's MALDI-TOF MS (Tinkerbell LT, ASTA, Korea) and recent developments at Ascend Diagnostics Ltd, Manchester are examples. The latter's new MALDI-TOF MS platform designated Lexi ^{TM} is a state of the art MALDI benchtop Mass Spectrometer that allows rapid identification of microorganisms. Its novel electronic and mechanical modules are embedded in a compact and robust design generating the smallest footprint on the market at present. The unique features and versatility of this instrument and other platforms that enable users to have more flexibility is likely to stimulate further interest across the breadth of microbiology.

The most significant barrier to more MS companies entering clinical microbiology is the monumental task of developing an accredited microbial database for use with a new instrument. ASTA has successfully achieved this and demonstrated the reliability of their MicroIDSys system (ASTA, Korea) by carrying out parallel studies using the Bruker Biotyper (Bruker Daltonics, Germany) on over 5,000 clinical isolates. Identical results with high confidence scores (≥ 2.0 for Bruker Biotyper), and (≥140 for ASTA MicroIDSys) were obtained for 86.1% of isolates with 99.2% (4,267 strains) showing good scores in both systems. The authors concluded that the ASTA MicroIDSys had the capacity to reliably

identify clinically important microorganisms (Lee *et. al.,* (2017). In further developments, the ASTA MicroIDSys was challenged with 370 clinical anaerobic isolates and attained 91.6% success at the species level. Among these were many poorly described species and many non-fermentative taxa that are normally difficult to speciate (Kim *et al.* 2020).

The MS companies SAI (Scientific Analysis Instruments, Manchester, UK), Waters Corporation, Elstree, UK (formerly, Micromass) and Shimadzu Corporation, Japan have been manufacturing MALDI-TOF MS instruments at the onset of these developments and have shown a long-term interest in promoting microbiological applications. Two other MS companies, AB Sciex and Thermo Fisher Scientific supplies tandem MS instruments that are likely to feature significantly in the future development of microbial proteomics and diagnostic applications in microbiology.

10.8.2 The initial microbiological development and acceptance of the MALDI-TOF MS was mainly in Europe. A decade later when accreditation was achieved, the technology entered the USA market and met considerable approval by some of the leading laboratories (see eg Patel, 2013, Angeletti, 2017). This was subsequently mirrored in many developing countries largely because after the outlay for an instrument, the accuracy, robustness, low running cost and simplicity of the method markedly surpasses traditional methods.

[Insert Table 10.2. Summary of the current advantages of MALDI-TOF MS]

10.8.3 Limitations of MALDI-TOF MS as currently used.

Use of mass spectral pattern matching profiles of mass ions to delineate bacterial species has enabled microbiologists to demonstrate unequivocal proof of concept. These biomarker mass ions are largely derived from the high abundant ribosomal proteins of the cell and their stability and reproducibility has been pivotal in the development of a universal method that has gained widespread appeal at the special level. Its application to the subspecies level using mass spectral profiles is equivocal (see above 10.6). Indications are that several of the minor peaks in the mass spectrum may be lipids/lipoproteins/glycolipids and may provide biomarkers for subspecies level identification as shown for yeast and fungi (Stübiger *et. al.,* 2016). These molecules have been extensively studied in bacteria using older forms of MS (see Minnikin and Goodfellow, 1980) but to date the use of MALDI-TOF MS to aid species/subspecies identification has not been demonstrated.

DNA-based methods are inherently more precise and, an elegant method developed by Hiroto Tamura, takes advantage of the exactitude of DNA and combines it with the ease and speed of MALDI-TOF MS (Tamura, 2017). The method expands the use of the ribosomal proteins used in current MALDI-TOF MS analysis but differs in that it decodes the base sequence for the S10-spc-alpha operon which codes for approximately half of the ribosomal proteins. Designated the S-10-GERMS method, it allows rigorous typing of bacterial isolates. Compared to conventional gene-based typing methods which requires DNA sequencing, Tamura's method is simpler and rapid and has been adopted by Shimadzu Corporation, Japan (see Microorganism Identification System).

10.9 Retaining the interest of Mass Spectrometry companies.

While there is enormous interest by microbiologist to retain the use of MALDI-TOF MS for the foreseeable future**,** interest by the MS companies will depend on financial returns. Several MALDI TOF MS companies have contributed significantly to the development of

applications of the technology but subsequently discontinued their interest due to budgetary constraints or a change in the direction of the company's strategy as noted above for Micromass M@LDI -MS (Waters Corporation), Sequenom MassArray system or Ciphergen Biosystems SELDI-TOF MS. In general MS companies need to work with microbiologists to assemble appropriate databases; the exception is Bruker Daltonik GmbH, who established their own microbiology facility at Bremen, Germany. The Shimadzu Corporation developed a database for use in the Far East while in Europe and America, they utilised the expanded accredited database of bioMerieux who drives the application of the method globally.

Several studies have shown near congruent results through parallel analyses of the same samples using the Bruker Biotyper and bioMerieux VITEK MS (see eg. Lévesque, *et. al.,* 2015). Each sample needed to be run according to the methodology and analysed using the search engine of each platform and may have accounted for minor differences in output. The ASTA (MicroIDSys system, ASTA, Korea) is unique in that they are the only non-specialist MS company to have successfully built their own MALDI-TOF MS instrument for microbiological applications. Their design features include a target plate that may be fitted to a Bruker's Microflex LT/SH or Autoflex LRF MALDI TOF MS. The 384 well target plate therefore allows for the first time MALDI-TOF MS analysis of the same sample on two entirely different platforms. The ability to be able to combine the use of two different instruments and databases adds new depth and confidence to the type of analyses that may now be undertaken. This was demonstrated in our laboratory using several staphylococcal species analysed on both instruments; Bruker's Autoflex LRF and ASTA's Tinkerbell MALDI TOF MS, with considerable ease and confidence (see Figure 10.8)

[Insert. Figure 10.8 (Left). MALDI-TOF MS analysis of **s**taphylococcal species using ASTA's Tinkerbell Linear MS and Bruker's Autoflex LRF MALDI TOF MS. An example using *Staphylococcus sciuri* showing the spectrum obtained on ASTA's Tinkerbell Linear MS (red) which was superimposed on the Bruker's Autoflex LRF MALDI TOF MS spectrum (light blue). The results show unequivocally the correspondence of the major mass ions of this species.

[Insert. Figure 10.8 (Right). 16S rRNA identification of 10 atypical strains of *Staphylococcus cohnii.* Both instruments revealed low identification scores, but most samples were correctly identified. The identification scores of seven strains were too low to provide an identification while strain 319 was incorrectly identified as *Mycobacterium lylae* using ASTA's Tinkerbell MS. Once added to the ASTA database, this and other strains were correctly identified].

The key to successful identification of an isolate is the comprehensive databases held by these companies which are continually expanding and released periodically. While valuable for users this remains a significant financial burden for more modest laboratories, particularly those in developing countries where paradoxically, more diversity may be encountered. Ideally, an online database of MALDI-TOF MS profiles analogous to DNA sequencing databases would remove the constrains currently imposed by a company-based database and add value to a more diverse and pragmatic solution for future MALDI-TOF MS platforms.

10.10 Potential to identify the biomarker peaks in a MALDI-TOF MS Spectrum; Towards MALDI-TOF MS global database.

The development and clinical applications of MALDI-TOF MS began in Europe in the late 1990s, gained confidence in the USA a decade ago and in now rapidly being implemented in in many developing countries. To date attempts have not been made to identify the key mass ions in the MALDI-TOF MS spectra that give each species its unique signature. However, as MALDI-TOF MS is used more for systematics it will be necessary to identify signature mass ions that may be added to the description of a new species as described above (Section 10.7) for the subspecies of *Cutibacterium acnes.* Being able to identify the signature mass ions of species will add immensely to the development of an open access global database in the future and help develop the next generation of MALDI-TOF MS applications. To undertake such a task, it will be necessary to utilise several high-resolution forms of MS which MALDI-TOF MS is unable to deduce.

10.11 High-resolution forms of MS that may be used to deduce peptide/protein taxon-specific signatures.

Although available MALDI-TOF MS databases encompass mass ions within the range 500 – 20,000 daltons, most bacterial species signature biomarkers are within the 500 -12,000 daltons and remain uncharacterised. MALDI-TOF MS is efficient at ionising proteins in this mass range. However, to characterise these proteins and species-specific proteins that are much larger in the range 50,000-100,000 daltons they need to be isolated using techniques such as gel electrophoresis or high-performance liquid chromatography and subsequently digested with a protease before MALDI-TOF MS-MS analysis and database searching. This can result in a better level of taxonomic identification. Furthermore, peptides that are highly abundant and detected reproducibly may be identified and used as biomarkers (Fox, *et. al.,* 2011). Alternatively, proteins can be extracted from an organism and analysed using liquid chromatography with high resolution accurate mass spectrometry using either top down or bottom up proteomic approach (Wynne, *et. al.,* 2010). Apart from simpler sample preparation, compared to the bottom-up approach, the top-down workflow provides additional information such as post-translational modifications and identification of proteoforms. With rapid improvements in ion optics and vacuum technology and fragmentation techniques, the upper limit for top down is towards >100kDa. Groups that have used bottom-up approaches for proteotyping have created their own database (Jabbour, *et. al.,* 2010). Currently there are no databases for processing data from either tryptic digests or intact proteins. Such a database could lead not only to better identification, especially for closely related species, but also to identification of specific protein markers. An area where bottom-up workflow has been used extensively is in comparative proteomics. Studies in this area include comparison of protein profiles of methicillin-susceptible and resistant *Staphylococcus aureus* (Xu *et. al.,* 2020), antibacterial mechanism of antibiotics (Ma, *et.al.,* 2017), resistant mechanisms (Chen, *et. al.,* 2019) and effect of culture conditions on organisms such as *P. aeruginosa* (Duncan *et. al.,* 2019). A bottom-up workflow has also been used for discovery of biomarkers that can be used for detection and identification of bacteria (Charretier, *et. al.,* 2015). The application of high-resolution mass spectrometry approaches for identification and studying the bacterial proteome would become even more widespread if databases were available for both top-down and bottom-up workflows.

10.12 From Linear MALDI-TOF MS to Tandem LC-MS/MS; unravelling the proteome of microbial species and future implications for bacterial systematics.

To date, MALDI-TOF MS and tandem MS/MS analyses have operated independently, consequently the ability to rapidly link bacterial phylogenetics with potential environmental function has not been explored. To bridge this gap, Clark *et. al.,* (2018) designed a novel MALDI-TOF MS data acquisition and bioinformatics pipeline (IDBac) to integrate data from both intact protein and specialised metabolite spectra directly from bacterial cells grown on agar. This technique organised bacteria into highly similar phylogenetic groups and allowed for comparison of metabolic differences of hundreds of isolates in just a few hours.

The literature detailing the proteome of bacterial species is colossal and continues to expand appreciably. Nearly all investigations pertain to aspects of the physiological flux of the cell; from the effect of environmental stimuli, antibiotic resistance to studying detailed insight into pathogenic mechanisms of infectious agents (see eg Chamot-Rooke *et. al.,* 2011, Gault et al., 2017, Soufi and Soufi, 2016, Chilton *et. al.*, 2014, 2017). Although protein profiling methods gained a strong presence in microbial systematics (see Section 10.2), MALDI-TOF MS which utilises protein signatures dominates taxonomic and species identification outputs. The capability of high-resolution tandem MS methods to identify novel characters and translate genomic traits into phenotype that may be used in microbial systematics remains poorly studied. We embarked on peptide/protein analysis to decipher species and subspecies markers for predicting microbial behaviour and devising identification markers in the late 1990s using initially MALDI and SELDI-TOF MS (see Section 10.4) and a few years later explored the value of LC-MS/MS for higher resolution and direct biomarkers sequencing (see review, Shah and Gharbia, 2017). With acquisition of the first generation of Thermo Fisher's LTQ Orbitrap MS, new rapid extraction methods and bioinformatic approaches applicable across the microbial kingdom were established (Lancashire *et. al.,* 2005, Schmid *et. al.,* 2005) Initially, these were confined to high risk pathogens that are genetically indistinguishable from non-pathogenic taxa such as separating *Bacillus anthracis* from *B. cereus, Shigella dysenteriae* from *S. boydii, Clostridium botulinum* from *Cl. sporogenes.* Similarly, *Burkholderia cepacian* and its relationship to *Pseudomonas aeruginosa* in the cystic lung and*, Mycobacterium tuberculosis,* a member of an almost identical complex of species and subspecies known as the TB complex that presents huge challenges for the clinical laboratory and systematics.

A coherent bioinformatic pipeline incorporating comparative genome sequence and open reading frame prediction and translation was developed (Al-Shahib *et. al.,* 2010, Misra *et al.,* 2012, 2015, 2017). This enabled the discovery of novel peptide biomarkers characterising isolates of each species tested using peptide sequences purified from GelC or nono-LC-MS/MS. Such approaches integrating peptide discovery against comparative genomics (eg Karlsson *et. al.,*2017, 2018) have the capacity to effectively utilise bacterial proteomes for both species, subspecies and strain level typing.

10.13. Case Study: Use of Tandem LC-MS/MS during major disease outbreak of pathogenic *E. coli*; taxonomic implications

During outbreak infections, samples received by a diagnostic laboratory often comprise a diverse range of isolates of the putative pathogen in which the boundaries between subspecies are blurred or the recognition of pathogenic/non-pathogenic strains are indiscrete due to exchange of genetic elements. The expression of a virulence factor/s such as toxins, which are nearly always protein in nature, make proteomics a powerful and sensitive tool to

simultaneously identify isolates and map the pathogenic factors within strains. Ideally this should be considered against a background of *in silico* analysis of the relevant genome to provide a sound basis to screen for genetic variants to corroborate the pathogenic potential of the isolate in terms of its proteome. The *E. coli* 0104:H4 outbreak in the summer of 2011 in Europe is a poignant example of the role of modern systematics in unravelling the pathogenic mechanisms of a well-defined taxon and controlling the outbreak.

10.13.1 Nature of the Outbreak

On 1st May 2011, an *E. coli* outbreak began in Germany with a few patients presenting with bloody diarrhoea. Within days, the aggressiveness of the strain was realised as the outbreak extended to eight other European countries. With over 4000 reported cases and 50 deaths, the status of the outbreak was soon elevated to one of the "deadliest *E. coli* outbreaks". The outbreak strain was positively identified on 25th May by Karch and colleagues at the University of Münster and the Robert Koch Institute, Germany using serotyping and PCR assays. Multilocus sequence typing confirmed that the outbreak was caused by a single clone, and that it was a rare serotype viz O104:H4. This serotype is normally associated with enteroaggregative *E. coli* (EAEC) that are known to cause persistent diarrhoea, but not haemorrhaging or HUS. In June, patients exhibiting symptoms of the outbreak strain arrived in the UK. Public Health England (then Health Protection Agency) initiated plans to undertake the sequencing of its first in house bacterial genome and proteome sequencing to aid investigation of this outbreak strain. WGS confirmed the mosaic nature of the strain's chromosome which was consistent with reports from other isolates of this outbreak. Genome sequence of multiple strains identified virulence factors that may account for the higher incidence of HUS and the unique traits of the strain. Two independent groups completed about 80% sequencing of the outbreak isolate's 5.2 million base pair genome and two large plasmids using short-read DNA sequencers (Mellmann *et. al.,* 2011, Editorial, 2011) which partly explained the strain's pathogenicity, acquisition of new genetic elements and evolutionary origin (Shah and Gharbia, 2012).

10.13.2 Proteomics and systematics in a high containment laboratory

Cell culture, mechanical disruption of cells, centrifugation and sample preparations were carried in Class 111 facilities at PHE. Proteomic analysis was performed on five strains from patients with confirmed *E. coli* 0104 infections; three clinical isolates from patients infected during the "German outbreak" and two other comparator strains that were from PHE's archives and previously characterized as EAEC and EHEC respectively.

Two parallel approaches were used to reduce complexity of the protein mixture prior to mass spectral analysis. Initially, lysates were separated by (SDS-PAGE), gel slices digested with trypsin and peptides analysed using nano-LC-MS/MS. Secondly, the entire cell lysate was digested directly with trypsin and injected on to two LC-MS/MS systems (Thermo Scientific LTQ Orbitrap and Thermo Scientific LTQ Orbitrap Velos, each with a front-end Ultimate 3000 Dionex nano/capillary liquid chromatography system, Thermo Fisher Scientific) that provided ultra-high resolution and accurate mass for differentiating closely related peptides.

The recorded peptide MS/MS spectra were matched to both protein and *in silico* genome-translated databases to identify sequenced proteins. The peptides were then subjected to further analysis using bioinformatic pipelines to characterise unique signatures. In total, 2,500 proteins were identified from the *E. coli* outbreak isolates which included 68 peptide signatures that were not shared by EAEC or EHEC or other Enterobacteriaceae. Species-level peptide signatures including those for the AggR transcription factor, haemolysin protein, Aaf fimbriae protein, and Iha adhesion protein were detected. Furthermore, genus characteristic

peptides were also mapped, suggesting that using proteome mining, an unknown isolate can be identified at genus, species and strain level from a single analysis.

10.14. Towards a combined MALDI-TOF MS/Tandem MS/MS Platform.

Over the last two decades we have employed a parallel approach to maximising the application of proteomics to characterise bacterial species from a vast range of environments. The manner in which proteomic data may be used in bacterial taxonomy was reported previously (Nomura, 2015, Shah, 2014). While MALDI-TOF MS has gained formidable recognition in microbiological laboratories, the huge potential available through in-depth analysis of the proteome by new advanced high-resolution MS-based technologies is grossly underestimated (Xu *et. al.,* 2020). The latter may be due to the absence of a high throughout, low cost. microscale instrument that is comparable with a MALD-TOF MS platform which therefore restricts the technology largely to the confines of the research laboratory. The methods described above (eg. Section 10.13) utilises a bottom-up approach in which cells need to be efficiently disrupted, lysates prepared and trypsinised prior to LC-MS/MS. This introduces a higher level of technical expertise and lengthy sample preparation steps which perhaps has also discouraged MALDI-TOF MS users.

More recently, Thermo Fisher Scientific usurped the well-established LTQ Orbitrap MS with the Q Exactive™ Hybrid Quadrupole- Orbitrap™ Mass Spectrometer. This highly advanced instrument, with the facilities to now undertake Top-down MS proteome analysis paves the way for a more automated, high throughput platform that is comparable with MALDI-TOF MS. Early work done at PHE and reported at ECCMID (Shah *et. al.,* 2015) showed that even for species that are genetically very closely related such as *E. coli* and *Shigella sonnei* that are inseparable even using WGS, Top-down MS analysis revealed differential biomarkers (see Figure 10.9). The next generation of MS to supersede MALDI-TOF MS is likely to be an instrument such as a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer that can analyse the bacterial proteome using a Top-down approach that is as manageable as processing samples using a MALDI-TOF MS instrument (Gault *et. al.,* 2017; Armengaud, 2017). However, the data derived from such analyses that take advantage of vast amount of data derived from the entire proteome will propel microbiology into a new era of systematics and biology.

[Insert -Figure 10.9. Use of Top-Down Proteomics to delineate genetically closely related species. In an early attempt to demonstrate the high resolution of this approach, proteins unique to both species were evident but were differentially expressed (Shah *et. al.,* 2015)].

References:

Al-Shahib, A., Misra, R., Ahmod, N., Fang, M., Shah, H.N. and Gharbia, S.E. (2010) Coherent pipeline for biomarker discovery using mass spectrometry and bioinformatics. *BMC Bioinformatics*, 11:437

Ambler, R.P. and Brown, L.H. (1967) The amino acid sequence of *Pseudomonas fluorescens* azurin. *Biochemistry Journal,* 104, 784 – 825.

Ambler RP (1985) Protein sequencing and taxonomy. In: Jones D, Goodfellow M, Priest FG (eds) Computer-aided bacterial systematics. Academic Press, London, pp 307–335

Armengaud, J. (2017) Proteogenomics of non-model microorganisms. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 529-538

Angeletti, S. (2017) Matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) in clinical microbiology. *Journal Microbiological Methods,* 138, 20-29.

Avanzi, I.R., Gracioso, L.H., Baltazar, M.D.P.G., Karolski, B., Perpetuo, E.A., do Nascimento, C.A.O., (2017) Rapid bacteria identification from environmental mining samples using MALDI-TOF MS analysis. *Environmental Science and Pollution Research,* 24, 3717–3726.

Batrich, M., Maskeri, L., Schubert, R., Ho, B., Kohout, M., Abdeljaber, M., Abuhasna, A., Kholoki, M., Psihogios, P., Tahir, R., Sawhney, S., Siddiqui, S., Xoubi, E., Cooper, A., Hatzopoulos, T. and Putonti, C. (2019) *Pseudomonas* diversity within urban freshwaters. *Frontiers in Microbiology,* 10:195. doi: 10.3389/fmicb.2019.00195.

Belén, R-S., Emilia, C., Coste, A. T. and Gilbert, G. (2019) Review of the impact of MALDI-TOF MS in public health and hospital hygiene, 2018. *Eurosurveillance*.24, https://doi.org/10.2807/1560-7917.ES.2019.24.4.1800193.

Berber, I. (2004) Characterization of Bacillus species by numerical analysis of their SDS-PAGE protein profiles. *Journal of Cell and Molecular Biology*, 3, 33-37.

Bergey's Manual of Determinative Bacteriology. 1st Edition. 1923. Genus *Bacteroides* Castellani and Chalmers 1919. (edited by D.H. Bergey, F.C. Harrison, R.S. Breed, B.W. Hammer and F.M. Huntoon). Williams and Wilkins Company. Baltimore. pp. 255- 264.

Bishop, C, Arnold, C. and Gharbia, S.E. (2010) Transfer of a traditional serotyping system (Kauffmann–White) onto a MALDI-TOF-MS platform for the rapid typing of Salmonella isolates. In: Shah, H.N. and Gharbia, S.E. (eds). *Mass Spectrometry for Microbial Proteomics*. Wiley, Chichester, UK, pp 463-496.

Biswas, S., Gouriet, F. and Rolain, J-M (2016) Molecular Typing of Bacteria/Fungi Using MALDI-TOF MS. In: Kostrzewa, M. and Schubert, S. (eds). *MALDI-TOF Mass Spectrometry in Microbiology.* Caister Academic Press, Munich, Germany, pp.79-92

Bourassa, L. (2018) MALDI TOF MS Quality Control in Clinical. Microbiology. *Clinical Laboratory News.* https://www.aacc.org/publications/cln/articles/2018/janfeb/maldi-tof-ms-quality-control-in-clinical-microbiology

Cain, T.C., Lubman, D.M. and Weber Jr. W.J. (1994) Differentiation of bacteria using protein profiles from matrix assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 8, 1026-1030.

Chamot-Rooke, J., Mikaty, G., Malosse, C., Soyer, M., Dumont, A., Gault, J., Imhaus, A-F., Martin, P., Trellet, M., Clary, G., Chafey, P., Camoin, L., Nilges, M., Nassif, X. and Duménil, G. (2011) Posttranslational modification of pili upon cell contact triggers N. meningitidis dissemination. *Science*, 331, 778-782.

Charretier, Y., Dauwalder, O., Franceschi, C., Degout-Charmette, E., Zambardi, G., Cecchini, T., Bardet, C., Lacoux, X., Dufour, P., Veron, L., Rostaing, H., Lanet, V., Fortin, T., Beaulieu, C., Perrot, N., Dechaume, D., Pons, S., Girard, V., Salvador, A., Durand, G., Mallard, F., Theretz, A., Broyer, P., Chatellier, S., Gervasi, G., Van Nuenen, M., Roitsch, C. A.,, Van Belkum, A., Lemoine, J., Vandenesch, F. and Charrier, J.P. (2015) Rapid bacterial identification, resistance, virulence and type profiling using selected reaction monitoring mass spectrometry. *Scientific Rep*orts. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4563557/

Chen, C.Y., Clark, C.G., Langner, S., Boyd, D.A., Bharat, A., McCorrister, S.J., McArthur, A.G., Graham, M.R., Westmacott, G.R., Van Domselaar, G. (2019) Detection of antimicrobial resistance using proteomics and the comprehensive antibiotic resistance database: a case study. *Proteomics Clinical Appl*ications. https://doi.org/10.1002/prca.201800182

Chiu, T-C. (2014) Recent Advances in Bacteria Identification by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Using Nanomaterials as Affinity Probes. *International Journal of Molecular Sciences.* 15, 7266-7280.

Chilton, C.H., Gharbia, S.E., Fang, M., Misra, R., Poxton, I.R., Borriello, S.P. and Shah, H. N. (2014) Comparative Proteomic Analysis of *Clostridium difficile* Isolates of Varying Virulence. *Journal of Medical Microbiology,* 63, 489 - 503.

Chilton, C.H., Gharbia, S.E., Misra, R.V., Fang, M., Poxton, I.R., Borriello, S.P. and Shah, H.N. (2017) Mapping of the Proteogenome of *Clostridium difficile* Isolates of Varying Virulence. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 379 – 398.

Clark, C.M., Costa, M.S., Sanchez, L.M. and Murphy, B.T. (2018) Coupling MALDI-TOF mass spectrometry protein and specialized metabolite analyses to rapidly discriminate bacterial function. *PNAS*,115, 4981- 4914.

Claydon, M.A., Davey, S. N., Edwards-Jones, V., and Gordon, D.B. (1996) The rapid identification of intact microorganisms using mass spectrometry. *Nature Biotechnology*, 14,1584-1586.

Cousin, S., Brambilla, E., Yang, J. and Stackebrandt, E. (2008) Culturable aerobic bacteria from the upstream region of a karst water rivulet. *International Microbiology*, 11, 91-100.

Dekio, I., Sakamoto, M., Hayashi, H., Amagai, M., Suematsu, M. and Benno, Y. (2007) Characterization of skin microbiota in patients with atopic dermatitis and normal subjects using 16S rRNA gene-based comprehensive analysis. *Journal of Medical Microbiology*, 56,1675-1683.

Dekio, I., Rajendram, D., Morita, E., Gharbia, S.E. and Shah, H.N. (2012) Genetic diversity of *Propionibacterium acnes* strains isolated from human skin in Japan and comparison with their distribution in Europe. *Journal of Medical Microbiology*, 61, 622 - 630.

Dekio, I., Culak, R., Fang, M., Ball, G., Gharbia, S., Shah, H. N. (2013) Correlation between phylogroups and intracellular proteomes of *Propionibacterium acnes* and differences in the protein expression profiles between anaerobically and aerobically grown cells. BioMed Res Int, article 151797, doi: 10.1155/2013/151797.

Dekio, I., Culak, R., Misra, R., Gaulton, T., Fang, M., Sakamoto, M., Oshima, K., Hattori, M., Klenk, H-P., Rajendram, D., Gharbia, S.E., Shah, H.N. (2015) Dissecting the taxonomic heterogeneity within *Propionibacterium acnes*: proposal for *Propionibacterium acnes* subsp. *acnes* subsp. nov. and *Propionibacterium acnes* subsp. *elongatum* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology.* 65, 4776-4787.

Dekio, I., McDowell, A., Sakamoto, M., Tomida, S., Ohkuma M. (2019) Proposal of new combination, *Cutibacterium acnes* subsp. *elongatum* comb. nov., and emended descriptions of genus *Cutibacterium*, *Cutibacterium acnes* subsp. *acnes*, and *Cutibacterium acnes* subsp. *defendens*. *International Journal of Systematic and Evolutionary Microbiology.* 69, 1087–1092.

Duncan, L.K., Kenneth, B., Shah, A.J., Selami, L., Belgacem, O., Ward, M., Shah, H. N. (December 2019). Isolation of putative strains of *P. aeruginosa* and verification using Mass Spectrometry. Society of Chemical Industry December 2019 Conference, *Rapid conformation using MALDI-TOF III*. SCI, 14/15 Belgrave Square London SW1X 8PS

Drucker, D.B. 1997. Use of Fast Atom Bombardment Mass Spectrometry for analysis of polar lipids of anaerobes. In *Anaerobic Pathogens.* (edited by A.R. Eley and K.W. Bennett. Sheffield Academic Press, Sheffield, England, pp. 331-341.

Editorial. Outbreak Genomics. (2011) Whole-genome sequencing and crowdsourced analyses proved a powerful adjunct to traditional typing in the recent *Escherichia coli* outbreak. *Nature Biotechnology*. 29, 769.

Emonet, S., Shah, H.N., Cherkaoui, A. and J. Schrenzel, J. (2010) Application and use of various mass spectrometry methods in clinical microbiology. *Clinical Microbiology and Infection*. 16, 1604-1613.

Encheva V, Wait, R., Gharbia, S.E., Begum, S., Shah, H.N. (2005) Proteome analysis of serovars Typhimurium and Pullorum of *Salmonella enterica* subspecies. *BMC Microbiology*, 5,42 -52.

Encheva, V., Wait, R., Gharbia, S.E., Begum, S., Shah, H.N. (2006) Comparison of extraction procedures for proteome analysis of *Streptococcus pneumoniae* and a basic reference map. *Proteomics,* 6, 3306-3317.

Fox, K., Fox, A., Rose, J. and Walla, M. (2011) Speciation of coagulase negative staphylococci, isolated from indoor air, using SDS page gel bands of expressed proteins followed by MALDI-TOF MS and MALDI TOF-TOF MS-MS analysis of tryptic peptides. *Journal of Microbiological Methods,* 84, 243–250.

Gloss, B.S., Patterson, K.I., Barton, C.A., Gonzalez, M., Scurry, J.P., Hacker, N.F., Sutherland, R.L., O'Brien, P.M., Clark, S.J. (2012) Integrative genome-wide expression and promoter DNA methylation profiling identifies a potential novel panel of ovarian cancer epigenetic biomarkers. *Cancer Letters.* 318, 76–85.

Gault, J., Vorontsov, E., Dupré, M., Calvaresi, V., Duchateau, M., Lima, D.B., Malosse, C. and Chamot-Rooke, J. (2017) Top-Down Proteomics in the Study of Microbial Pathogenicity. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 493-504.

Goodfellow, M. and Board, 1980 *Microbiological classification and Identification.* Society for Applied Bacteriology, Academic Press, London, New York, Toronto, Sydney, San Francisco.

Goodfellow, M and Minnikin, D. M (1985) Chemical methods in bacterial systematics. Society for Applied Bacteriology, Technical Series 20, Academic Press, London, Orlando, San Diego, New York, Toronto, Sydney, Tokyo.

Gram, H.C. (1884) "Über die isolierte Färbung der Schizomyceten in Schnitt- und Trockenpräparaten". *Fortschritte der Medizin,* 2, 185-189.

Greub, G. (2010) MALDI-TOF mass spectrometry: the quantum leap. *Clinical Microbiology and Infection*. 16, 1603

Hamilton, S., Levin, M., Knoll, J.S. and Langford, P.R. (2010) Characterisation of microorganisms by pattern matching of mass spectral profiles and biomarker approaches requiring minimal sample preparation. In: Shah H.N. and Gharbia, S.E. (eds). *Mass Spectrometry for Microbial Proteomics.* Wiley, Chichester, pp. 223- 254.

High, N.J, Fan, F., and Schwartzman, J.D. (2015) *Haemophilus influenzae.* In: Molecular Medical Microbiology 2nd Ed. Chapter. 97. Vol. 3. pp. 1709 – 1728.

Holdeman, L.V. and W.E.C. Moore. 1974. Gram-negative anaerobic rods; Bacteroidaceae Primbram 1933, 10. In: *Bergey's Manual of Determinative Bacteriology.* 8th Edition. (edited by R.E. Buchanan and N.E. Gibbons). The Williams and Wilkins Company. pp. 384-404.

Holdeman, L.V., R. W. Kelley and W.E.C. Moore. 1984. Bacteroidaceae Primbram 1933, 10. In: *Bergey's Manual of Systematic Bacteriology* . 1st edition. (edited by N.R. Krieg and J.G. Holt). Williams and Wilkins, Baltimore/London. pp. 602- 631

Holland, R.D., Wilkes, J.G., Rafii, F., and Lay, J.O. (1996) Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 10, 1227-1232

Honisch, C., Chen, Y and Hillenkamp, F. (2010) DNA Resequencing by MALDI-TOF Mass Spectrometry and its application to traditional microbiological problems. In: Shah, H.N. and Gharbia, S.E. (eds). *Mass Spectrometry for Microbial Proteomics*. Wiley, Chichester, UK, pp 443-462.

Jackman, P.J.H. (1985) Bacterial taxonomy based on electrophoretic whole-cell protein patterns. In Goodfellow M. and Minnikin D E (eds). *Chemical Methods in Bacterial Systematics.* Academic Press, London and New York, pp. 115 - 129.

Jabbour, R.E., Deshpande, S.V., Wade, M.M., Stanford, M.F., Wick, C.H., Zulich, A.W., Evan W. Skowronski, E.W. and Snyder, A.P. (2010) Double-Blind Characterization of Non-Genome-Sequenced Bacteria by Mass Spectrometry-Based Proteomics. *Applied and Environmental Microbiol*ogy, 76, 3637–3644.

Jarman, K. H. and Wahl, k. l. (2006) Development of spectral pattern-matching approaches to Matrix Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry for bacterial identification. In: Wilkins, C. L. and Lay, J.O. (ed.) *Identification of Microorganisms by Mass Spectrometry*. John Wiley & Sons, Inc. Hoboken, New Jersey, pp 153 - 160

Johnson, J. L. and Cummins C. S. (1972) Cell wall composition and deoxyribonucleic acid similarities among the anaerobic coryneforms, classical propionibacteria, and strains of *Arachnia propionica*. *Journal of Bacteriology*, 109, 1047-1066.

Kallow, W., Erhard, M., Shah, H.N., Raptakis, E. and Welker, M. (2010) MALDI-TOF MS for microbial identification: years of experimental development to an established protocol. In: Shah H.N. and Gharbia, S.E. (eds). *Mass Spectrometry for Microbial Proteomics,* Wiley, Chichester, pp. 255- 276.

Karas, M., Bachmann, D., Hillenkamp, F. (1985) Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry*, 57, 2935 - 2939.

Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F. (1987) Matrix-Assisted Ultraviolet Laser Desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes*, 78, 53–68.

Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*. 60, 2299- 2301.

Karlsson, R., Gonzales-Siles, L., Boulund, F., Lindgren, Å., Svensson-Stadler, L., Karlsson, A., Kristiansson, E., Moore, E.R.B. (2017) Proteotyping: Tandem Mass Spectrometry Shotgun Proteomic Characterization and Typing of Pathogenic Microorganisms. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 419-450

Karlsson, R., Gonzales-Siles, L., Gomila, M., Busquets, A., Salvà-Serra, F., Jaén-Luchoro, D., Jakobsson, H.E., Karlsson, A., Boulund, F., Kristiansson, E. and Moore, E.R.B. (2018) Proteotyping bacteria: Characterization, differentiation and identification of pneumococcus and other species within the Mitis Group of the genus *Streptococcus* by tandem mass spectrometry proteomics. PLOS One. https://doi.org/10.1371/journal.pone.0208804

Keys, C.J., Dare, D.J., Sutton, H., Wells, G., Lunt, M., McKenna, T., McDowall, M. and Shah, H.N. (2004) Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infection, Genetics and Evolution*: 4, 221-242.

Kim, Y. and Kim, J-S (2017) ASTA'S MicroID System and its MycoMp database for mycobacterium. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 110-12

Kim, D., Ji, S., Kim, J.R., Kim, M., Byun, J-H., Yum, J.H., Yong, D., Lee, K. (2020) Performance evaluation of a new matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, ASTA MicroIDSys system, in bacterial identification against clinical isolates of anaerobic bacteria. *Anaerobe.* 61 (February 2020, in press).

Krishnamurthy, T., Ross, P.L. (1996) Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 10, 883-888.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227, 680–685.

Lancashire, L., Schmid, O., Shah, H.N. and Ball, G. (2005) Classification of bacterial species from proteomic data using combinatorial approaches incorporating artificial neural networks, cluster analysis and principal components analysis. *Bioinformatics:* 21: 2191 - 2199.

Lee, Y., Sung, J.Y., Kim, H., Yong, D., Lee, K. (2017) Comparison of a new matrix-assisted laser desorption/ionization time-of-flight mass spectrometry platform, ASTA MicroIDSys, with Bruker Biotyper for species identification. *Annals of Laboratory Medicine* 37, 531-535.

Lévesque, S., Dufresne, P.J., Soualhine, H., Domingo, M-C, Bekal, S., Lefebvre, B. and Tremblay, C. (2015) A Side by Side Comparison of Bruker Biotyper and VITEK MS: Utility of MALDI-TOF MS Technology for Microorganism Identification in a Public Health Reference Laboratory. *PLoS One;* 10(12): doi: 10.1371/journal.pone.0144878

Lomholt, H. B. and Kilian, M. (2010) Population genetic analysis of *Propionibacterium acnes* identifies a subpopulation and epidemic clones associated with acne. PLoS ONE 5, e12277.

Ma, W., Zhang, D., Li, G., Liu, J., He, G., Zhang, P., Yang, L., Zhu, H., Xu, N. and Liang, S. (2017) Antibacterial mechanism of daptomycin antibiotic against Staphylococcus aureus based on a quantitative bacterial proteome analysis. *Journal of Proteomics,*150, 242-251.

Mahé, P., Arsac, M, Perrot, N., Charles, M-H., Broyer, P., Hyman, J., Walsh, J., Chatellier, S., Girard, V., van Belkum A. and Veyrieras, J-B. (2017) Identification of species in mixed microbial populations using MALDI-TOF MS. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 167 - 186.

McDowell, A., Valanne, A., Ramage, G., Tunney, M.M., Glenn, J.V., McLorinan, Bhatia, A., Maisonneuve, J.F., Lodes, M., Persing, D.H. and Patrick, S. (2005) *Propionibacterium acnes* Types I and II Represent Phylogenetically Distinct Groups. *Journal of Clinical Microbiology*, 43, 326–334.

McDowell, A., Perry, A. L., Lambert, P. A. and Patrick, S. (2008) A new phylogenetic group of *Propionibacterium acnes*. *Journal of Medical Microbiology* 57, 218–224.

McDowell, A., Barnard, E., Liu, J., Li, H. and Patrick S. (2016) Proposal to reclassify *Propionibacterium acnes* type I as *Propionibacterium acnes* subsp. *acnes* subsp. nov. and *Propionibacterium acnes* type II as *Propionibacterium acnes* subsp. *defendens* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 66, 5358-5365.

Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., McLaughlin, S.F., Henkhaus, J.K., Leopold, B., Bielaszewska, M., Prager, R., Brzoska, P.M., Moore, R.L., Guenther, S., Rothberg, J.M. and Karch H. (2011) Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by rapid next generation sequencing technology. PLoS ONE. 6 (7), e22751.

Minnikin, D. E and Goodfellow, M. (1980) Lipid composition in the classification and identification of Acid-Fast bacteria. In: Goodfellow, M. and Board, R. G. (eds.). *Microbiological Classification and Identification.* Academic Press. London, pp. 189 - 256.

Misra, R.V., Ahmod, N.Z., Parker, R., Fang, M., Shah, H.N. and Gharbia, S.E. (2012) Developing an integrated proteo-genomic approach for the characterisation of biomarkers for the identification of *Bacillus anthraci*s. *Journal of Microbiological Methods*, 88, 237- 247.

Misra, R.V., Gaulton, T., Ahmod, N., Fang, M., Hornshaw, M.M., Ho, J., Gharbia, S.E. and Shah, H.N.(2017) Tandem Mass Spectrometry Analysis as an Approach to Delineate

Genetically Related Taxa. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 313 - 378.

Misra, R.V., Gaulton, T., Fang, M., Culak, R.A., Hornshaw, M. M., Ho, J., Gharbia, S.E. and Shah, H.N. (2015) Tandem Mass Spectrometry Analysis as an Approach to Delineate Genetically Related Taxa, with Specific Implication for Differentiating *Escherichia coli* from amongst the Complex Enterobacteriaceae Family. *Journal of Proteomics & Enzymology,* 4, 1 -11.

Nagy, E,. Urbán, E., Becker, S., Kostrzewa, M., Vörös, A., Vörös, A., Hunyadkürti, J. and Nagy I. (2013) MALDI-TOF MS fingerprinting facilitates rapid discrimination of phylotypes I, II and III of *Propionibacterium acnes*. *Anaerobe* 20, 20–26.

Nomura, F., 2015. Proteome-based bacterial identification using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS): A revolutionary shift in clinical diagnostic microbiology. *Biochimica Biophysica Acta* 1854, 528–537. 2014.10.022

Olkun, A., Shah, A. and Shah, H.N. (2017) Elucidating the Intra-Species Proteotypes of *Pseudomonas aeruginosa* from Cystic Fibrosis. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 579 - 592.

Pavlovic, M., Huber, I., Konrad R. and Busch U. (2013) Application of MALDI-TOF MS for the identification of food borne bacteria. *The Open Microbiology Journal,* 7, 135–141.

Patel, R. (2013) MALDI-TOF Mass Spectrometry: Transformative proteomics for clinical microbiology. *Clinical Chemistry*, 59, 340-342.

Pirnay, J.-P., Matthijs, S., Colak, H., Chablain, P., Bilocq, F., Van Eldere, J., De Vos, D., Zizi, M., Triest, L. and Cornelis, P. (2005) Global *Pseudomonas aeruginosa* biodiversity as reflected in a Belgian river. *Environmental Microbiology*, 7: 969-980.

Pranada, A. B., Witt, E., Bienia, M., Kostrzewa, M., and Timke, M. (2017) Accurate differentiation of *Mycobacterium chimaera* from *Mycobacterium intracellulare* by MALDI-TOF MS analysis. *Journal Medical Microbiology,* 66, 670–677.

Rajakaruna L., Hallas, G., Diane Dare, D., Sutton, H., Encheva, V., Renata Culak, R., Innes, I., Ball, G., Sefton, A.M., Eydmann, M., Kearns, A.M and Shah, H.N. (2009) High Throughput Identification of Clinical Isolates of *Staphylococcus aureus* using MALDI-TOF-MS of Intact Cells. *Infection, Genetics and Evolution*, 9. 507 -513.

Remus-Emsermann, M., Schmid, M., Gekenidis, M. T., Pelludat, C., Frey, J. E., Ahrens, C. H. and Drissner, D. (2016) Complete genome sequence of *Pseudomonas citronellolis* P3B5, a candidate for microbial phyllo-remediation of hydrocarbon-contaminated sites. *Standards in genomic sciences*, *11*, 75. doi:10.1186/s40793-016-0190-6

Rose, G., Culak, R., Chambers, T., Gharbia, S.E. and Shah, H.N. (2017) The challenges of identifying mycobacterium to the species level using MALDI-TOF MS. In: Shah, H.N. and

Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 93-121.

Santos, I.C., Hildenbrand, Z.L. and Schug, K.A. (2016) Applications of MALDI-TOF MS in environmental microbiology. *The Analyst* 141, 2827–2837.

Schleifer, K. H. and Kandler, O. (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriological Reviews*, 36, 407- 477.

Schmid, O., Ball, G., Lancashire, L., Culak, R. and Shah, H.N. (2005) New approaches to identification of bacterial pathogens by surface enhanced laser desorption/ionisation time of flight mass spectrometry in concert with artificial neural networks, with special reference to *Neisseria gonorrhoeae*. *Journal of Medical Microbiology* 54**,** 1205-11.

Schubert S. and Kostrzewa, M. (2017) MALDI-TOF MS in the Microbiology Laboratory: *Current Issues in Molecular Biology*. doi: 10.21775/cimb.023.017.

Schumann, P. and Maier, T. (2014) MALDI-TOF Mass Spectrometry Applied to Classification and Identification of Bacteria.In: Goodfellow, M, Sutcliffe I. and Chun, J. (eds.) *Methods in Microbiology*, *New approaches to prokaryotic systematics.* Vol. 41. Elsevier, Oxford Chapter 13, 275 - 306.

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N. and Whittam, T.S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology*. 51, 873 - 884.

Seubert, W. (1960). Degradation of isoprenoid compounds by microorganisms. *Pseudomonas citronellolis* nov. sp. Isolation and characterization of an isoprenoid-degrading bacterium. *Journal of Bacteriology*, 79, 426-434

Shah, H.N. (May, 2014) "*New Methods bring new Information; What is hidden behind a bacterial species name*". European Congress of Clinical Microbiology & Infectious Diseases (ECCMID), Congress, Barcelona.

Shah, H.N., Chilton, C., Rajakaruna, L., Gaulton, T., Hallas, G., Atanassov, H., Khoder, G., Rakowska, P.D., Cerasoli, E., Gharbia, S.E. (2010). Changing concepts in the characterisation of microbes and the influence of mass spectrometry. In: Shah H.N. and Gharbia, S.E. (eds). '*Mass Spectrometry for Microbial Proteomics'* Wiley, Chichester, pp. 3- 34.

Shah H N, Collins M D. (1981). *Bacteroides buccalis* sp nov *Bacteroides denticola* sp nov and *Bacteroides pentosaceus* sp nov, new species of the genus *Bacteroides* from the oral cavity. *Zentralblatt fur Bakteriologie Mikrobiologie und Hygiene I. Abteilung Originale C*. 2, 235-241.

Shah, H.N., Encheva, V., Schmid, O., Nasir, P., Culak, R.A., Ines, I., Chattaway, M.A., Keys, C.J., Jacinto, R.C., Molenaar, L., Ayenza, R.S., Hallas, G., Hookey, J.V. and Rajendram. D. (2005) Surface Enhanced Laser Desorption/Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS): A Potentially Powerful Tool for Rapid Characterisation of Microorganisms. In: M.J. Miller, M.J. (ed). *Encyclopedia of Rapid Microbiological Methods.*Vol.3. DHI Publishing, LLC, River Grove, IL, USA. pp. 57 - 96.

Shah, H.N. and Gharbia, S.E. (2011) A century of systematics of the genus *Bacteroides,* from *a* single genus up to the 1980s to an explosion of assemblages and the dawn of MALDI-TOF-Mass Spectrometry. *The Bulletin of BISMiS*, 2, 87-106

Shah, H.N., and Gharbia, S.E. (2012) Using nano-LC-MS/MS to investigate the toxicity of outbreak *E. coli* 0104:h4 strain. Culture. **33**- ISSN 0965-0989.

Shah, H.N. and Gharbia, S.E. (2017) A paradigm shift from research to front-line microbial diagnostics in MALDI-TOF and LC-MS/MS: a laboratory's vision and relentless resolve to help develop and implement this new technology amidst formidable obstacles. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 1 – 38.

Shah, H. N., Keys, C., Gharbia, S.E., Ralphson, K., Trundle, F., Brookhouse, I. and Claydon M. (2000) The application of MALDI-TOF Mass Spectrometry to profile the surface of intact bacterial cells. *Microbial Ecology in Health and Disease*, 12: 241-246

Shah, H.N., Keys, C.J., Schmid, O. and Gharbia, S.E. (2002) Matrix-Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry and Proteomics; a New Era in Anaerobic Microbiology. *Clinical Infectious Diseases*, 35, 58-64.

Shah, H. N., Stephenson, J., Cardasis, H., Neil, J. Yip, P. Ravela, S. Ritamo, I. Damsbo, M. Grist, R. Freeke, J. Stielow, B. Gvozdyak, O. Dukik, K. de Hoog, S. Damoc, E. Valmu, L. Cherkassky, A. Fang, M., Gaulton, T. Misra. R. and Gharbia, S. E. (April, 2015). "A Global Diagnostic Approach for Microbial Identification: Accurate characterization of difficult to differentiate pathogens using Top-down proteomics". European Congress of Clinical Microbiology & Infectious Diseases (ECCMID) Congress, Copenhagen.

Shah H N, van Steenbergen T J M, Hardie J M, de Graaff J. DNA base composition, DNA-DNA reassociation and isoelectric focusing of proteins of strains designated *Bacteroides oralis.* (1982) *FEMS Microbiology Letters* 13, 125-130.

Shute, L.A., Berkeley, R.C.W., Norris, J.R. and Gutteridge, C.S. (1985) Pyrrolysis mass spectrometry in bacterial systematics. In: Goodfellow, M. and Minnikin D E. (eds.) Chemical Methods in Bacterial Systematics. London and New York: Academic Press, pp 95-114.

Singhal, N., Kumar, M., Kanaujia, P. K. and Virdi, J. S. (2015) MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2015.00791

Sneath, P. H. A. (1957) The Application of Computers to Taxonomy. *Journal of General Microbiology*, 17, 201-26.

Soufi, Y. and Soufi B. (2016) Mass Spectrometry-Based Bacterial Proteomics: Focus on Dermatologic Microbial Pathogens. *Frontiers in Microbiology* https://doi.org/10.3389/fmicb.2016.00181

Stübiger, C.G., Wuczkowski, M., Mancera, L., Lopandic, K., Sterflinger, K. and Belgacem, O. (2016) Characterization of yeasts and filamentous fungi using MALDI Lipid Phenotyping. *Journal of Microbiological Methods*. 130, 27- 37.

Syrmis, M.W., Moser, R.J., Whiley, D.M., Vaska, V., Coombs, G. W., Nissen, M. D., Sloots, T.P. and Nimmo, G. R. (2011) Comparison of a multiplexed MassARRAY system with real-time allele-specific PCR technology for genotyping of methicillin-resistant *Staphylococcus aureus*. *Clinical Microbiology and Infection*, 17, 1804-1810.

Tamura, H. (2017) MALDI-TOF MS Based on Ribosomal Protein Coding in S10-spc-alpha Operons for Proteotyping. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 269 - 310.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. and Matsuo, T. (1988) Protein and Polymer Analyses up to *m/z* 100 000 by Laser Ionization Time-of flight Mass Spectrometry. *Rapid Communications in Mass Spectrometry*, **2**, 151–3.

Timperio AM, Gorrasi S, Zolla L, Fenice M (2017) Evaluation of MALDI-TOF mass spectrometry and MALDI BioTyper in comparison to 16S rDNA sequencing for the identification of bacteria isolated from Arctic sea water. *PLoS One* 12: e0181860

Vranckx, K., De Bruyne, K. and Pot, B. (2017) Analysis of MALDI-TOF MS Spectra using the BioNumerics Software. In: Shah, H.N. and Gharbia, S.E. (eds). MALDI-TOF and Tandem MS for Clinical Microbiology. Wiley, Chichester, UK, pp 539-562

Veloo, A.C.M., Jean-Pierre, H., Justesen, U.S., Morris, T., Urban, E., Wybo, I., Shah, H.N., Friedrich, A.W., ENRIA workgroup, Nagy, E., Kostrzewa, M. (2017) A multi-center ring trial for the identification of anaerobic bacteria using MALDI-TOF MS. *Anaerobe*. 48, 94 - 97.

Williams, G. (2019). First report of infection with *Pseudomonas citronellolis*: a case of urosepsis. *New microbes and new infections*, *30*, 100531. doi:10.1016/j.nmni.2019.100531

Wynne, C., Edwards, N. J. and Fenselau, C. (2010) Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using cap LC-MS/MS on an Orbitrap. *Proteomics*, 10, 3631-43.

Xu, Z., Olkun, A., Vranckx, K., Mkrtchyan, H.V., Shah, A.J, Pot, B., Cutler, R.C. and H.N. (2017) Subtyping of *Staphylococcus* spp. Based upon MALDI-TOF MS Data Analysis. In: Shah, H.N. and Gharbia, S.E. (eds). *MALDI-TOF and Tandem MS for Clinical Microbiology*. Wiley, Chichester, UK, pp 563 - 378.

Xu, Z., Chen, J., Vougas, K., Shah, A.J., Shah, H.N., Misra, R. and Mkrtchyan, H.V. (2020) Comparative proteomic profiling of methicillin-susceptible and resistant *Staphylococcus aureus*. *Proteomics* (in press).

Yang, Y., Lin, Y. and Qiao, L (2018) Direct MALDI-TOF MS identification of bacterial mixtures. *Analytical Chemistry*, 17, 10400-10408