Accepted 23 September, 2013

Bacterial species identification getting easier

Phumudzo Tshikhudo, Ronald Nnzeru*, Khayalethu Ntushelo and Fhatuwani Mudau

University of South Africa, Department of Agriculture and Animal Health, Private Bag X6, Florida, 1710, South Africa.

*Corresponding author. E-mail: ronaldn@nda.agric.za/ ronaldn@daff.gov.za.

Abstract

The traditional methods of bacterial identification are based on observation of either the morphology of single cells or colony characteristics. However, the adoption of newer and automated methods offers advantage in terms of rapid and reliable identification of bacterial species. The review provides a comprehensive appreciation of new and improved technologies such fatty acid profiling, sequence analysis of the 16S rRNA gene, desorption/ionization time-of-flight (MALDI-TOF), metabolic finger profiling using BIOLOG, ribotyping, together with the computational tools employed for querying the databases that are associated with these identification tools and high-throughput genomic sequencing in bacterial identification. It is evident that with the increase in the adoption of new technologies bacterial identification is becoming easier.

Key words: Bacteria, Biolog, computational tools, fatty acids, Gram staining, identification, metagenomics, morphology, MALDI-TOF MS, RiboPrinter, 16S rRNA gene.

Abbreviations: MALDI-TOF MS, Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; **TFA**, trifluoroacetic acid; **SEM**, scanning electron microscopy; ISH, *in situ* hybridization; **GC,** gas chromatography; **CFA,** cellular fatty acid; **FAME,** fatty acid methyl ester analysis; **PCR,** polymerase chain reaction; **RFLPs,** restriction fragment length polymorphisms; **HTS,** high-throughput sequencing.

INTRODUCTION

Bacteria are primarily grouped according to their morphological characteristics (shape, presence or absence of flagella, and arrangement of flagella), substrate utilisation and Gram staining. Another important trait is their pattern of growth on solid media as different species can produce very diverse colony structures (Christopher and Bruno, 2003). The traditional methods that employ observation of either the morphology of single cells or colony characteristics remain reliable parameters for bacterial species identification. However, these traditional techniques have some disadvantages.

Firstly, they are time-consuming and laborious. Secondly, variability of culture due to different environmental conditions may lead to ambiguous results. Thirdly, a pure culture is required to undertake identification, making the identification of fastidious and unculturable bacteria difficult and sometimes impossible. To evade these problems, newer and automated methods which rapidly and reliably identify bacteria have been adopted by many laboratories worldwide. At least one of these methods, namely analysis of the 16S rRNA gene, does not require a pure culture. Combining these automated systems with the traditional methods provides workers with a higher level of confidence for bacterial identification. This review serves as a comprehensive appreciation of these new technologies. The methods we discuss are fatty acid profiling, sequence analysis of the 16S rRNA gene, protein profiling using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF), metabolic finger profiling using BIOLOG, and ribotyping, together with the computational tools employed for querying the databases that are associated with these identification tools. We further discuss the role of high-throughput genomic sequencing in bacterial identification. Unfortunately, laboratories in poor countries cannot afford some of these new systems. With increased access to these technologies, workers in many laboratories will find the identification of bacterial species easier.

THE MORPHOLOGICAL IDENTIFICATION OF BACTERIA

As it has always been the desire of humankind to understand the environment, the classification and identification of organisms has always been among the priorities of the early scientists. Unlike zoologists and botanists who have a plethora of morphological traits with which to identify animals and plants, the morphological characters for identifying bacteria are few and limiting. This not only provided a challenge, but also an opportunity for creativity. Gram staining was a result of the creative insight of Hans Christian Joachim Gram (1850-1938) to classify bacteria based on the structural properties of their cell walls. It was based on Gram staining that bacteria could be differentially classified as either Gram positive or Gram negative, a convenient identification and classification tool that remains useful today. Although there are few morphological traits, and little variation in those traits, identification based on morphology still has significant taxonomic value. When identifying bacteria, much attention is paid to how they grow on the media in order to identify their cultural characteristics, since different species can produce very different colonies (Christopher and Bruno, 2003). Each colony has characteristics that may be unique to it and this may be useful in the preliminary identification of a bacterial species. Colonies with a markedly different appearance can be assumed to be either a mixed culture or a result of the influence of the environment on a

bacterial culture which normally produces known colony characteristics or a newly discovered species.

The features of the colonies on solid agar media include their shape (circular, irregular or rhizoid), size (the diameter of the colony: small, medium, large), elevation (the side view of a colony: elevated, convex, concave, umbonate/umbilicate), surface (how the surface of the colony appears: smooth, wavy, rough, granular, papillate or glistening), margin/border (the edge of a colony: entire, undulate, crenated, fimbriate or curled), colour (pigmentation: yellow, green among others), structure/opacity (opaque, translucent or transparent), degree of growth (scanty, moderate or profuse) and nature (discrete or confluent, filiform, spreading or rhizoid). Cell shape has also been used in the description and classification of bacterial species (Cabeen and Jacobs-Wagner, 2005). The most common shapes of bacteria are cocci (round in shape), bacilli (rod-shaped) and spirilli (spiral-shaped) (Cambray, 2006).

Observations of bacterial morphologies are done by light microscopy, which is aided by the use of stains (Bergmans et al., 2005). Dutch microbiologist Antonie van Leeuwenhoek (1632-1723) was the first person to observe bacteria under a microscope. Without staining, bacteria are colourless, transparent and not clearly visible and the stain serves to distinguish cellular structure for a more detailed study. The Gram stain is a differential stain with which to categorise bacteria as either Gram positive or Gram negative. Observing bacterial morphologies and the Gram reaction usually constitutes the first stage of identification. Specialised staining for flagella reveals that bacteria either have or do not have flagella and the arrangement of the flagella differs between bacterial species. This serves as a good and reliable morphological feature for identifying and classifying bacterial species.

Light microscopy was traditionally used for identifying colonies of bacteria and morphologies of individual bacteria. The limitation of the light microscope was its often insufficient resolution to project bacterial images for clarity of identification. Scanning electron microscopy (SEM) coupled with high-resolution back-scattered electron imaging is one of the techniques used to detect and identify morphological features of bacteria (Davis and Brlansky, 1991). SEM has been widely used in identifying bacterial morphology by characterizing their surface structure and measure cell attachment and morphological changes (Kenzata and Tani, 2012). A combination of morphological identification with SEM and *in situ* hybridization (ISH) techniques (SEM-ISH) clarified the better understanding of the spatial distribution of target cells on various materials. This method has been developed in order to obtain the phylogenetic and morphological information about bacterial species to be identified using *in situ* hybridization with rRNA-targeted oligonucleotide probes (Kenzata and Tani, 2012).

These morphological identification techniques were improved in order to better identify poorly described, rarely isolated, or phenotypically irregular strains. An improved method was brought up for the bacterial cell characterization based on their different characteristics by segmenting digital bacterial cell images and extracting geometric shape features for cell morphology. The classification techniques, namely, 3σ and K-NN classifiers are used to identify the bacterial cells based on their morphological characteristics (Hiremath et al., 2013).

In addition to microscopy, several other tools for bacterial identification are useful to confirm identities based on morphology, thereby increasing the level of confidence of identity. Among these tools is the analysis of fatty acid profiles which will be discussed.

FATTY ACID ANALYSIS

Fatty acids are organic compounds commonly found in living organisms. They are abundant in the phospholipid bilayer of bacterial membranes. Their diverse chemical and physical properties determine the variety of their biochemical functions. This diversity, which is found in unique combinations in various bacterial species, makes fatty acid profiling a useful identification tool.

The fatty acid profiles of bacteria have been used extensively for the identification of bacterial species (Purcaro et al., 2010). Fatty acid profiles are determined using gas chromatography (GC), which distinguishes bacteria based on their physical properties (Núñez-Cardona, 2012).

Reagents to cleave the fatty acids are required for saponification (45 g sodium hydroxide, 150 ml methanol and 150 ml distilled water), methylation (325 ml certified 6.0 N hydrochloric acid and 275 ml methyl alcohol), extraction (200 ml hexane and 200 ml methyl tert-butyl ether) and sample clean-up (10.8 g sodium hydroxide dissolved in 900 ml distilled water). Information on the fatty acid composition of purple and green photosynthetic sulphur bacteria includes fatty acid nomenclature, the distribution of fatty acids in prokaryotic cells, and published information on the fatty acids of photosynthetic purple and green sulphur bacteria (Núñez-Cardona, 2012). This information also describes a standardised gas chromatography technique for the fatty acid analysis of these photosynthetic bacteria using a known collection and wild strains.

The cellular fatty acid analysis for bacterial identification is based on the specific fatty acid composition of the cell wall. The fatty acids are extracted from cultured samples and are separated using gas chromatography. A computer generated, unique profile pattern of the extracted fatty acids is compared through pattern recognition programs, to the existing microbial databases. These databases include fatty acid profiles coupled with an assigned statistical probability values indicating the confidence level of the match. This has become very common in biotechnology.

The fatty acid analysis for bacterial identification using gas-chromatography became simpler with the available computer-controlled chromatography and data analysis (Welch, 1991). The fatty acid analysis method uses electronic signal from the gas chromatographic detector and pass it to the computer where the integration of peaks is performed (Sasser, 2011). The whole cellular fatty acid methyl esters content is a stable tool of bacterial profile in identification because the analysis is rapid, cheap, simple to perform and highly automated (Geacomini et at., 2000). In addition, bacterial identification can be done at or below the species level.

Adams et al. (2004) determined the composition of the cellular fatty acid (CFA) of *Bacillus thuringiensis* var. *kurstaki* using the MIDI Sherlock microbial identification system on a Hewlett-Packard 5890 gas chromatograph. This study revealed the capability to detect the strain variation in the bacterial species *B. thuringiensis* var. *kurstaki* and to clearly differentiate strain variants on the basis of qualitative and quantitative differences in hydrolysable whole CFA compositions in the preparations examined. Since this technology was used to resolve strain differences within a species, we can easily assume that the differentiation of species is done more accurately when fatty acid profiling is used. Kloepper et al. (1991) isolated and identified bacteria from the geocarposphere, rhizosphere, and root-free soil of field-grown peanut at three sample dates, using the analysis of fatty acid methyl-esters to determine if qualitative differences exist between the bacterial microflora of these zones. The dominant genera across all three samples were *Flavobacterium* for pods, *Pseudomonas* for roots, and *Bacillus* for root-free soil. Heyrman et al. (1999) isolated 428 bacterial strains, of which 385 were characterised by fatty acid methyl ester analysis (FAME). The majority (94%) of the isolates comprised Gram-positive bacteria and the main clusters were identified as *Bacillus* sp., *Paenibacillus* sp., *Micrococcus* sp., *Arthrobacter* sp. and *Staphylococcus* sp. Other clusters contained nocardioform actinomycetes and Gram-negative bacteria, respectively. A cluster of the latter contained extreme halotolerant bacteria isolated in Herberstein (Heyrman et al., 1999). At present, no bacterial identification method is guaranteed to provide absolute identity to all presently known bacterial species and therefore a number of methods are employed for a single identification procedure. Another method that is widely used for bacterial identification is sequence analysis of the 16S rRNA gene.

SEQUENCE ANALYSIS OF THE 16S rRNA GENE

Ribosomal RNA genes are a critical part of the protein synthesis machinery. They are omnipresent and therefore classification based on the analysis of ribosomal RNA genes does not leave out any of the known bacteria. For this reason, analysis of ribosomal RNA genes is a suitable tool for bacterial species identification and taxonomic categorisation. Moreover, ribosomal RNA genes are

conserved but have sufficient variation to distinguish between taxa (Woese, 1987). In prokaryotes, ribosomal RNA genes occur in copies of three or four in a single genome (Fogel et al., 1999). The 16S rRNA gene has become a reliable tool for identifying and classifying bacteria. Over time, the 16S rRNA gene has shown functional consistency with a relatively good clocklike behaviour (Chanama, 1999) and its length of approximately 1,500 bp is sufficient for bioinformatic analysis (Janda and Abbott, 2007).

Analysis of the 16S rRNA gene requires that this gene be amplified by polymerase chain reaction (PCR) and the resultant PCR product sequenced. The gene sequence can then be matched with previously obtained sequences obtainable from various DNA databases. This method has been so widely adopted that DNA sequence databases are flooded with sequences of the 16S rRNA gene. Almost all new sequences deposited for query have matches and any 16S rRNA gene copy which does not match any known bacterial species is believed to be new (Chanama, 1999). In certain instances there is no requirement for pure colony amplification of the 16S rRNA gene, which makes this method suitable for studies of fastidious and unculturable bacteria and a good tool for the metagenomic analysis of environmental samples. Petrosino et al. (2009) defined metagenomics as "culture-independent studies of the collective set of genomes of mixed microbial communities, (which may) be applied to the exploration of all microbial genomes in consortia that reside in environmental niches, in plants or in animal hosts".

With the advent of metagenomic analyses of gross DNA samples, analysis of the 16S rRNA gene is proving its worth. In 16S rRNA-based metagenomics, gene sequencing has been widely used for probing the species structure of various bacteria in the environment (Shah et al., 2010). The 16S rRNA gene sequence is used to detect bacterial species in natural specimens and to establish phylogenetic relationships between them (Eren et al., 2011). This is made possible by the fact that all bacterial species contain the 16S rRNA gene, which has highly conserved regions on which to design universal primers, as well as hypervariable regions that are useful in distinguishing species.

The 16S rRNA gene has hypervariable regions which are an indication of divergence over evolutionary time. The 16S rRNA genes of bacteria possess nine hypervariable regions (V1 - V9) that display considerable sequence diversity in different species of bacteria (Chakravorty et al., 2007). These regions are flanked by conserved regions on which universal primers can be designed for their amplification. Because the variation of the hypervariable regions is correlated with the identity of taxa, it is often of no use to analyse the whole 16S rRNA gene when identifying species. This adds to the convenience of using the 16S rRNA gene for identifying bacterial species. Since high-throughput sequencing platforms sequence short segments of DNA, analysis of only these hypervariable regions, which are a few hundred bases long, falls within the scale of massive parallel

sequencing. This has accelerated the generation of 16S rRNA sequences and their entry into public databases.

It is easy for sequence analysis of the 16S rRNA gene to be adopted by many laboratories because it generally requires only PCR and sequencing, which are widely used techniques for many other applications. As a result, there are many studies that have employed sequence analysis of the 16S rRNA gene in taxonomic classification. The computational tools have been employed to identify a wide range of bacteria through the sequence analysis of their 16S rRNA genes. Using this method, the 16S rRNA gene fragments are amplified using PCR method, and bacteria are identified based on 16S rRNA gene sequence similarity based method on the existing microbial databases. According to Barghoutti (2011), when pure PCR products of the 16S gene are obtained, sequenced, and aligned against bacterial DNA data base, then the bacterium can be identified. For bacterial identification, the 16S rRNA gene is regarded as the most widely accepted gene (Song et al., 2003). Signature nucleotides of 16S rRNA genes allow classification and identification of bacterial species even if a particular sequence has no match in the database. The distinctive approach when identifying bacterial species using this method is to perform high-throughout sequencing of 16S rRNA genes, which are then taxonomically classified based on their similarity to known sequences in existing databases (Mizrahi-Man, 2013).

Kumrapich et al. (2011) examined the endophytic bacteria in the internal tissues of sugarcane leaves and stems using molecular methods. They used a nutrient agar medium to cultivate the endophytes, whereupon 107 isolates of bacteria in the internal tissues of sugarcane leaves and stems were selected for analysis and 23 species of bacteria were identified and divided into three groups, based on the 16S rRNA sequences and phylogenetic analysis. The taxa identified were *Sphingobacterium, Bacillus amyloliquefaciens, Bacillus cereus, Bacillus megaterium, Bacillus pumilus, Bacillus subtilis*, *Agrobacterium larrymoorei*, *Burkholderia cepacia, Chromobacterium violaceum*, *Acinetobacter* (one strain), *Enterobacter* (three strains), *Klebsiella* (one strain), *Serratia* (one strain), *Pantoea* (three strains), and *Pseudomonas* (two strains). Based on the amplified 16S rRNA gene sequencing, Bhore et al. (2010) identified bacterial isolates from the leaves of *Gaultheria procumbens* (eastern teaberry, checkerberry, boxberry, or American wintergreen) as *Pseudomonas resinovorans, Paenibacillus polymaxa*, and *Acinetobacter calcoaceticus*. Muzzamal et al. (2012) isolated and identified an array of 76 endophytic bacteria from the roots, stems, and fresh and wilted leaves of various plants in Pakistan. The morphological, biochemical and physiological characterisation and 16S rRNA gene sequence analysis of the selected endophytic isolates led to the identification of different bacterial species belonging to the genera *Bacillus, Pseudomonas, Serratia, Stenotrophomonas* and *Micromonospora*.

Although sequence analysis of the 16S rRNA gene has been by far the most common, reliable and convenient method of bacterial species identification, this technique has some shortfalls. Firstly, with this method it is not possible to differentiate between species that share the same sequence of this gene. Identification of bacterial species based on sequence analysis of the 16S rRNA gene relies on matching the obtained sequence with the existing sequence. Matching with a sequence that was incorrectly identified leads to incorrect identification. Other problems associated with using the 16S rRNA gene are sequencing artefacts and problems with the purity of bacterial isolates which may lead to incorrect identification. Problems associated with sequence analysis of the 16S rRNA gene when identifying bacterial species argue for the use of alternative methods to confirm findings. Among these alternative methods is matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) which relies on exploiting differences in bacterial protein profiles.

MATRIX-ASSISTED LASER DESORPTION/IONISATION TIME-OF-FLIGHT MASS SPECTROMETRY (MALDI-TOF MS)

A rapid, high-throughput identification method, MALDI-TOF MS, has been introduced in bacterial taxonomy. This system has brought reliability, simplicity and convenience. MALDI-TOF is the only polypeptide fingerprinting-based methods even to be used for bacterial identification. The first studies regarding the identification of bacteria by MALDI-TOF were conducted towards the end of the 1990s and technology was made available as a research tool. It was commercialised for use in private and public laboratories in 2008 and the delay was in commercialising MALDI-TOF was because of the lack of robust information tools and efficient databases. The MALDI-TOF MS technique offers easily determinable peptide/protein fingerprints for the identification of bacterial species. This technique has the ability to measure peptides and other compounds in the presence of salts and to analyse complex peptide mixtures, making it an ideal method for measuring non-purified extracts and intact bacterial cells.

Bacterial cultures to be queried are spotted on the MALDI-TOF plate which is placed in the time-of-flight (TOF) chamber. Each sample is spotted at least in duplicate, to verify reproducibility. A control specimen of known identity is included to ensure correct identity. The samples are allowed to air-dry at room temperature, inserted into the mass spectrometer and subjected to MALDI-TOF MS analysis. In addition to the cell-smear and cell-extract methods, additional sample preparation methods, as described previously (Smole et al., 2002), are used on a small number of strains. These include heat treatment (15 min at 95°C) of the cell extracts and cell smears, sonication (30 s, 0.3 MHz) of intact cells and the so-called sandwich method (Williams et al., 2003).

MALDI-TOF MS has been successfully applied to a number of taxa of *Listeria* species (Barbuddhe et al., 2008), *Campylobacter* spp. (Fagerquist et al., 2007; Grosse-Herrenthey et al., 2008), *Streptococcus pyogenes* (Moura et al., 2008), the *Burkholderia cepacia* complex (Vanlaere et al., 2006), *Arthrobacter* (Vargha et al., 2006), *Leuconostoc* spp., *Fructobacillus* spp., and *Lactococcus* spp. (De Bruyne et al., 2011). According to De Bruyne et al. (2011), different experimental factors, including sample preparation, the cell lysis method, matrix solutions and organic solvents may affect the quality and reproducibility of bacterial MALDI-TOF MS fingerprints and this warrants the use of alternative methods to guarantee correct identification. Computational tools for MALDI -TOF are used according to the tasks they perform: Firstly, pre-processing of spectra, then unsupervised data mining methods which can be used for preliminary data examination, then supervised classification applied for example, in biomarker discovery.

A MALDI-TOF dataset represents a set of mass spectra with two spatial coordinates *x* and *y* assigned to each spectrum. Unsupervised data mining, unsupervised methods are used for data mining, can be applied without any prior knowledge, and aim at revealing general data structure. Supervised methods (mainly classification) require specifying at least two groups of spectra which need to be differentiated, for example, by finding *m/z*-values differentiating spectra of tumor regions from spectra of control regions (Alexandrov, 2012). For isolates requiring identification to the species level (*n* _986), correct species identifications is done by the Biotyperand Vitek MS systems and the Saramis database.

BIOLOG

Different methods have traditionally been used to identify bacteria based on biochemical activity. These methods include the oxidase test and the catalase test. The Biolog OmniLog Identification System [or simply “Biolog” (Biolog Inc, Hayward, California)], a system that utilises automated biochemical methodologies, as an instrument (Miller and Rhoden, 1991; Holmes et al., 1994; Morgan et al., 2009) that tests a microorganism's ability to utilise or oxidise a panel of 95 carbon sources. Tetrazolium violet is incorporated in each of the substrates contained in a 96-well microtitre plate. Biolog's patented technology uses each microbe's ability to use particular carbon sources, and uses chemical sensitivity assays to produce a unique pattern or "phenotypic fingerprint" for each bacterial species tested. As a bacterium begins to use the carbon sources in certain wells of the microplate, it respires. With bacteria, this respiration process reduces a tetrazolium redox dye and those wells change colour to purple. The end result is a pattern of coloured wells on the microplate that is characteristic of that bacterial species. A unique biochemical pattern or “fingerprint” is then produced when the results are surveyed. The fingerprint data are analysed, compared to a

database, and identification is generated. The Biolog system was originally created for the identification of Gram-negative bacteria, but since the introduction of this system in 1989, the identification capability of the instrument has broadened to include Gram-positive bacteria (Stager and Davis, 1992).

According to Morgan et al. (2009) isolates are prepared according to the manufacturer's instructions in the OmniLog ID System User Guide (Biolog, Hayward, CA). All isolates, except the *Bacillus* species, are cultured at 35°C on a Biolog Universal Growth (BUG) agar plate with 5% sheep blood. After an incubation period of 18 to 24 h, the bacterial growths are emulsified to a specified density in the inoculating fluid (0.40% sodium chloride, 0.03% Pluronic F-68, and 0.02% gellan gum). *Bacillus* species require a special “dry-tube method” preparation as described by the manufacturer. Colonies are picked with a sterile wooden Biolog Streakerz™ stick and rubbed around the walls of an empty, sterile, glass tube. Inoculating fluid (5 ml) is added to suspend the bacterial film. The suspension is subsequently used to inoculate culture wells of Gram-positive microplates (Biolog, Hayward, CA).

For all isolates, each well of the Gram-positive or Gram-negative microplate is inoculated with 150 µL of the bacterial suspension. Depending on the type of organism, the microplates are incubated at 30 or 35°C for 4 to 24 h. If bacterial identification has not occurred after 22 h, a reading of “no ID” is given. Each metabolic profile is compared with the appropriate GNor GPOmnilog Biolog database (Biolog,Hayward, CA), which contains biochemical fingerprints of hundreds of gram negative and gram positive species (Morgan et al., 2009). Biolog has been applied successfully to a number of taxa such as *Paenibacillus azotofixans* (Pires and Seldin, 1997), *Xanthomonas campestris* pv. *campestris* (Massomo et al., 2003) and *Glycine* spp. (Hung and Annapurna, 2004).

Computational tools such as standard multivariate analysis tools which include cluster analysis, principal component analysis and principal coordinate analysis are available for simple set summurization of numerical taxonomikc traits. Another tool is the co-inertia analysis which is a multivariate statistical method that perform a joint analysis on two data tables and assign egual consideration of both of them. These method is a two table ordination method that facilitate establishment of connections between tables with data domains that contain the same or even different numbers of variable method allow are to connect various standard single-table coordination methods such as principal component analysis and correspondence analysis.Mantel test is a regression procedure in which variables themselves are either distance or dis-similarity matrices, summarising pair similarities among objects. Computations and graphic displays of Mantel test and the co-inertia analysis are obtained using ADE-4 package (Thiolouse et al., 1997). The documentation and downloading of this programme is available in the internet.

The Biolog method indicates potential, but not actual, catabolic activity of a community. Glimm et al. (1997) noticed that an assortment of substrates does not necessarily reflect substrates which are available to bacteria in the soil environment, so one can suspect that some microbial species are incapable of growing on plates because of the lack of proper substrates. According to Morgan et al. (2009) the Biolog system requires pure cultures and the subsequent growth of the bacteria – and pure culture and growth are frequently problematic when it comes to slow-growing, fastidious, unusual, nonviable, or non-culturable bacteria. The turnaround time required for identifying bacterial isolates can be several days to several weeks. The Biolog system is better at identifying fermentative organisms than nonfermenters. However, it should be noted that biochemically active nonfermenters do achieve high identification rates (88%) in the Biolog system, so a different product may be more suitable for inactive nonfermenters. Because of its disadvantages, other bacterial species identification procedures are required.

RIBOTYPING

The identification of bacterial species based on ribotyping exploits sequence differences in rRNA. DNA is extracted from a sample and is digested with restriction enzymes to generate a unique combination of discrete-sized fragments (ribotyping fingerprint) for a particular bacterial species. This pattern is queried in a database containing numerous patterns of different bacterial species. Before a ribotyping fingerprint database had been developed, rRNA fragments produced from restriction digestion would be probed with a known DNA probe for bacterial species identification.

A known ribotyping system, RiboPrinter®, is an automated system used for characterising bacterial samples and is a well regarded method of genotyping pure culture isolates which is often used in epidemiological studies. The basis of ribotyping is the use of rRNA as a probe to detect chromosomal restriction fragment length polymorphisms (RFLPs). The whole DNA of a pure culture is extracted and cleaved into various lengths of fragments using many endonucleases. The resultant fragments are separated by gel chromatography, then probed with labelled rRNA oligonucleotides. Kivanç et al. (2011) used the RiboPrinter® to identify a total of 45 lactic acid bacteria from 10 different boza (a malt drink) samples in Turkey. In a study by Inglis et al. (2002) an automated ribotyping device was used to determine the ribotypes of a collection of *Burkholderia pseudomallei* isolates, and the comparison of automated ribotyping with DNA macrorestriction analysis showed that an *Eco*RI ribotyping protocol can be used to obtain discriminating molecular typing data on all isolates analysed. Optimal discrimination was obtained by analysing gel images of automated *Eco*RI ribotype patterns obtained with BioNumerics software in combination with the results of DNA macrorestriction analysis.

HIGH-THROUGHPUT SEQUENCING TECHNIQUES

There are four sequencing technologies available (capillary sequencing, pyrosequencing, reversible terminator chemistry, sequence-by litigation). The Sanger capillary sequencing is still based on the same general scheme applied in 1977 for the φX174 genome. Roche/454 GS FLX Titanium sequencer was the first of the new high-throughput sequencing platforms on the market and it was released in 2005. It is based on the pyrosequencing approach. Compared to Sanger sequencing, it is based on iteratively complementing single strands and simultaneously reading out the signal emitted from the nucleotide being incorporated. Illumine Genome Analyzer II/IIx is a reversible terminator technology and employs a sequencing-by-synthesis concept that is similar to that used in Sanger sequencing, however the Illumina sequencing requires protocol the sequence to be determined are converted in to special sequencing library, which allows them to be amplified and immobilised for sequencing (Bentley et al., 2008). The SOLiD sequence platform (sequencing-by-litigation) is very different from the rest discussed thus far and the sequence extension reaction is not carried out by polymerases but rather by ligases (Shendure et al., 2005). The Sanger capillary sequencing is a low- throughput method and the sequencing error observed for Sanger sequencing is mainly due to errors in the amplification step (a low rate when done in vivo), natural variance, and contamination in the sample used, as well as polymerase slippage at low complexity sequences like simple repeats (short variable number tandem repeats) and homopolymers (stretches of the same nucleotide). The the high-throughput techniques (pyrosequencing, reversible terminator chemistry, sequence-by litigation) makes bacterial identification easier and even possible for even single research groups to generate large amounts of sequence data very rapidly and at substantially lower costs than traditional Sanger sequencing.

Novel DNA sequencing technologies called high-throughput sequencing (HTS) techniques are capable of generating massive amounts of genetic information with increased speed, accuracy and efficiency. High-throughput genome sequencing provides a more detailed real-time assessment of the genetic traits of bacteria than could be achieved with routine subtyping methods. HTS technologies are used for studying diversity and genetic variations and solving genomic complexities. Approximately 300 complete bacterial genomes had been sequenced by 2010. This has aided and sped identification of bacterial species and these HTS technologies remain useful especially for identification of bacterial species that constitute a population in a sample.

CONCLUSIONS

The traditional identification of bacteria on the basis of phenotypic characteristics is generally not as accurate as identification based on genotypic methods. The more traditional methods whereby bacteria have been identified based on their physical properties, are compound light microscopy in combination with histological staining and electron microscopy. The later is the conventional scanning microscope which generally offers unique advantages such as high resolution and great depth of field. The fatty acid profiles of bacteria, which are determined with the aid of gas chromatography, have also been used extensively for the identification of bacterial species. Bacterial phylogeny and taxonomy have further benefited greatly from the use of the sequence analysis of 16S ribosomal RNA, which makes the identification of rarely isolated, phenotypically anomalous strains possible. Comparison of the bacterial 16S rRNA gene sequence has emerged as a preferred genetic technique. 16S rRNA gene sequence analysis can better identify poorly described, rarely isolated, or phenotypically aberrant strains can be routinely used for identification of mycobacteria and can lead to the recognition of novel pathogens and noncultured bacteria. Cutting-edge technologies such as MALDI-TOF MS, Biolog and the RiboPrinter® has facilitated bacteriological identification even further. The MALDI-TOF MS technique offers easily determinable peptide or protein finger printing for the identification, typing and characterisation of various strains. Biolog has been used to identify various lactic acid bacteria strains. Biolog tests a microorganism's ability to utilise or oxidase a panel of carbon sources and this method is used when characterising bacterial samples within a fixed degree of similarities. The computational tools have been developed for quering the relevant microbial databases that are associated with the bacterial identification methods. From the current review, it is evident that with the increase in the adoption of new technologies and high-throughput sequencing techniques, bacterial identification is becoming easier.

REFERENCES

Adams DJ, Gurr S, Hogge J (2004). Cellular Fatty-Acid Analysis of *Bacillus thuringiensis* Var. *kurstaki* Commercial Preparations. J. Agric. Food Chem., 53(3): 512-517.

Alexandrov T (2012). MALDI imaging mass spectrometry: statistical data analysis and current computational challenges. BMC Bioinformatics 13 (Suppl 16): S11.

Barbuddhe SB, Maier T, Schwarz G, Kostrzewa M, Hof H, Domann E, Chakraborty T, Hain T (2008). Rapid identification and typing of listeria species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol., 74(17): 5402-5407.

Barghoutti SA (2011). A Universal Method for the Identification of Bacteria Based on General PCR Primers. Indian J Microbiol., 51(4): 430- 444.

Bentley DR, Balasubramanian S, Swerdlow HP, et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–9.

Bergmans L, Moisiadis P, Van Meerbeek B, Quirynen M, Lambrecht P (2005). Microscopic observation of bacteria: review highlighting the use of environmental SEM. Int. Endod. J., 38: 775-788.

Bernhardt J, Weibezahn J, Scharf C, Hecker M (2003). *Bacillus subtilis* During Feast and Famine: Visualization of the Overall Regulation of Protein Synthesis DuringGlucose Starvation by Proteome Analysis. *Genome Research,* Vol. 13, pp. 224-237.

Bhore SJ, Nythia R, Loh CY (2010). Screening of endophytic bacteria G isolated from leaves of sambung nyawa [*Gynura procumbens* (Lour.) Merr.] for cytokinin-like compounds. Bioinformation 5(5): 191-197.

Cabeen MT, Jacobs-Wagner C (2005). Bacterial Cell Shape. Nat. Rev. Microbiol., 3(8): 601-610.

Cambray G (2006). Basic and applied microbiology. Eds., Cloete TE, Atlas RM Van Schaik Publishers.

Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods, 69(2): 330-339.

Chanama S (1999). Comparative 16S rRNA sequence analysis. Warasan Wichai Witthayasat Kanphaet, 13: 107-117.

Christopher K, Bruno E (2003). Identification of bacterial species. In Tested studies for laboratory teaching. Ed., O'Donnell MA. Proceedings of the 24th Workshop/Conference of the Association for Biology Laboratory Education, 24: 103-130.

Davis CL, Brlansky RH (1991). Use of Immunogold Labelling with Scanning Electron Microscopy To Identify Phytopathogenic Bacteria on Leaf Surfaces. Appl Environ Microbiol. 1991 October; 57(10): 3052-3055.

De Bruyne K, Slabbincka B, Waegeman W, Vauterin P, De Baets B, Vandamme P (2011). Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. Syst. Appl. Microbiol., 34: 20-29.

De Val C, Rivas E, Tores-Quesada O, Toro N, Jiménez-Zurdo JI (2007). Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. Mol Microbiol., 66(5):1080-1089.

Eren AM, Ferris MJ, Taylor CM (2011). A framework for analysis of metagenomic sequencing data. Pac. Symp. Biocomput., 2011:131-141.

Fagerquist CK, Yee E, Miller WG (2007). Composite sequence proteomic analysis of protein biomarkers of *Campylobacter coli*, *C. lari* and *C. concisus* for bacterial identification. Analyst, 132(10): 1010-1023.

Fogel GB, Collins CR, Li J, Brunk CF (1999). Prokaryotic genome size and SSU rDNA copy number: estimation of microbial relative abundance from a mixed population. Microb. Ecol., 38: 93-113.

Giacomini M, Ruggiero C, Calegari L, Bertone S (2000). Artificial neural network based identification of environmental bacteria by gas-chromatographic and electrophoretic data. J Microbiol Methods. 2000 Dec 1;43(1):45-54.

Glimm E, Heuerh, Engelen B, Smalla K, Backhaus H (1997). Statistical comparisons of community catabolic profiles. J. Microbiol. Methods, 30: 71.

Grosse-Herrenthey A, Maier T, Gessler F, Schaumann R, Böhnel H, Kostrzewa M, Krüger M (2008). Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). Anaerobe, 14(4): 242-249.

Heyrman J, Mergaert J, Denys R, Swings J (1999). The use of fatty acid methyl ester analysis (FAME) for the identification of heterotrophic bacteria present on three mural paintings showing severe damage by microorganisms. FEMS Microbiol. Lett., 181: 55-62.

Hiremath PS, Bannigidad P, Yelgond SS (2013). An Improved Automated Method for Identification of Bacterial Cell Morphological Characteristics. International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 11-16 2 (1): 11-16.

Holmes B, Costas M, Ganner M, On SL, Stevens M (1994). Evaluation of Biolog system for identification of some Gram-negative bacteria of clinical importance. J. Clin. Microbiol., 32: 1970-1975.

Hung PQ, Annapurna K (2004). Isolation and characterization of endophytic bacteria in soybean (*Glycine* sp.). Omonrice, 12: 92-101.

Inglis TJJ, O'Reilly L, Foster N, Adele CA, Sampson J (2002). Comparison of rapid, automated ribotyping and DNA macro restriction analysis of *Burkholderia pseudomallei*. J. Clin. Microbiol., 40(9): 3198-3203.

Janda JM, Abbot SL (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J. Clin. Microbiol., 45(9): 2761-2764.

Kenzaka T, Tani K (2012). Scanning Electron Microscopy Imaging of Bacteria Based on Nucleic Acid Sequences, Scanning Electron Microscopy, Dr. Viacheslav Kazmiruk (Ed.), ISBN: 978-953-51-0092-8, InTech, Available from: http://www.intechopen.com/books/scanning-electron-microscopy/scanningelectron-microscope-imaging-of-bacteria-based-on-nucleic-acid-sequences.

Kivanc M, Vilmaz M, Cakir E (2011). Isolation and identification of lactic acid bacteria from boza, and their microbial activity against several reporter strains. Turk. J. Biol. 35: 313-324.

Kloepper JW, McInroy JA, Bowen KL (1992). Comparative identification by fatty acid analysis of soil, rhizosphere, and geocarposphere bacteria of peanut *(Arachis hypogaea* L.). Plant Soil, 139: 85-90.

Kumrapich B, Klayraung S, Wongkattiya N, Topoonyanont N (2011). Diversity of bacteria isolated from leaves and stems of sugarcane (Saccharum sp. Var. LK92-11). 37th Congress on Science and Technology of Thailand.

Massomo SMS, Nielsen H, Mabagala RB, Mansfeld-Giese K, Hockenhull J, Mortensen CN (2003). Identification and characterisation of *Xanthomonas campestris* pv. *campestris* strains from Tanzania by pathogenicity tests, Biolog, rep-PCR and fatty acid methyl ester analysis. Eur. J. Plant Pathol., 109: 775-789.

Martiny D, Busson L, Wybo I, Ait El Haj R, Dediste A, Vandenberga O. (2012). Comparison of the Microflex LT and Vitek MS Systems for Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry. Journal of Clinical Microbiology p. 1313-1325.

Miller JM, Rhoden DL (1991). Preliminary evaluation of Biolog, a carbon source utilization method for bacterial identification. J. Clin. Microbiol., 29, 1143-1147.

Mizrahi-Man O, Davenport ER, Gilad Y (2013). Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. PLoS ONE 8(1).

Morgan MC, Boyette M, Goforth C, Sperry KV, Greene SR (2009). Comparison of the Biolog OmniLog Identification System and 16S ribosomal RNA gene sequencing for accuracy in identification of atypical bacteria of clinical origin. J. Microbiol. Methods, 79: 336-343.

Moura H, Woolfitt, AR, Carvalho MG, Pavlopoulos A, Teixeira LM, Satten GA, Barr JR (2008). MALDI-TOF mass spectrometry as a tool for differentiation of invasive and noninvasive *Streptococcus pyogenes* isolates. FEMS Immunol. Med. Microbiol., 53(3): 333-342.

Muzzamal H, Sarwar R, Sajid I, Hasnain S (2012). Isolation, identification and screening of endophytic bacteria antagonistic to biofilm formers. Pakistan J. Zool., 44(1): 249-257.

Núñez-Cardona MT (2012). Fatty Acids Analysis of Photosynthetic Sulfur Bacteria by Gas Chromatography, Gas Chromatography - Biochemicals, Narcotics and Essential Oils. Ed. Salih B. InTech.

Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009). Metagenomic pyrosequencing and microbial identification. Clin. Chem., 55(5): 856-866.

Pires MN, Seldin L (1997). Evaluation of Biolog system for identification of strains of *Paenibacillus azotofixans*. Antonie Leeuwenhoek, 71: 195-200.

Purcaro G, Tranchida PQ, Dugo P, La Camera E, Bisignano G, Conte L, Mondello L (2010). Characterization of bacterial lipid profiles by using rapid sample preparation and fast comprehensive two-dimensional gas chromatography in combination with mass spectrometry. J. Sep. Sci., 33(15): 2334-2340.

Sasser M (2011). Identification of Bacteria by Gas Chromatography of Cellular Fatty Acids. MIDI Technical Note #101.

Shah N, Tang H, Doak TG, Ye H (2010). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. Pac. Symp. Biocomput., 2011: 165-176.

Shendure J, Porreca GJ, Reppas (2005). Accurate multiplex polony sequencingof an evolved bacterial genome. Science 309:1728-1732.

Smole SC, King LA, Leopold PE, Arbeit RD (2002). Sample preparation of Gram-positive bacteria for identification by matrix assisted laser desorption/ionization time-of-flight. J. Microbiol. Methods, 48(2-3): 107-115.

Song Y, Liu C, McTeague M, Finegold SM (2003). 16S ribosomal DNA sequence-based analysis of clinically significant gram-positive anaerobic cocci. J Clin Microbiol., 41(4):1363-1369.

Stager CE, Davis JR (1992). Automated systems for identification of microorganisms. Clin. Microbiol. Rev., 5: 302-327.

Vanlaere E, Sergeant K, Dawyndt P, Devreese B, Vandamme P (2006). Identification of *Burkholderia cepacia* complex using MALDI-TOF mass spectrometry. J. Cyst. Fibr., 5(Suppl. 1): S34-S134.

Vargha M, Takats Z, Konopka A, Nakatsu CH (2006). Optimization of MALDI-TOF MS for strain level differentiation of Arthrobacter isolates. J. Microbiol. Methods, 66: 399-409.

Veloo ACM, Knoester M, Degener JE, Kuijper EJ. 2011. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods for the identification of clinically relevant anaerobic bacteria. Clin. Microbiol. Infect. 17:1501–1506.

Welch DF (1991). Application of cellular fatty acid analysis. Clin Microbiology Rev. ; 4(4):422-438.

Williams TL, Andrzejewski D, Lay JO, Musser SM (2003). Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells. J. Am. Soc. Mass Spectrom., 14(4): 342-351.

Woese CR (1987). Bacterial evolution. Microbiol. Rev., 51(2): 221-271.