

APPENDIX A (1)

Manuscripts

OLIVIER I., LOOTS, D.T. 2011. An overview of tuberculosis treatments and diagnostics. What role could metabolomics play? *J. Cell Tissue Res.* 11(1):2655-2671

AN OVERVIEW OF TUBERCULOSIS TREATMENTS AND DIAGNOSTICS. WHAT ROLE COULD METABOLOMICS PLAY?

OLIVIER, I. AND LOOTS, D.T.[?]

School for Physical and Chemical sciences, Centre for Human Metabonomics,
North-West University, Potchefstroom, 2520, South Africa.
E-mail: dutoit.loots@nwu.ac.za

Received: Received: February 1, 2011; Accepted: February 28, 2011

Abstract: *In 2001, the WHO declared TB a global emergency, as one third of the worlds population suffered from latent M. tuberculosis infection. Today, a decade later, millions of people still die worldwide as a result of this disease. The growing TB incidence may be ascribed to a variety of reasons, including, amongst others, the inadequacies associated with the currently available diagnostic methods and TB treatment regimes, especially when considering the growing MDR-TB and HIV epidemics. This review discusses and compares the various TB diagnostic and treatment approaches researched and developed to date, considering these with regards to their advantages and shortcomings. We additionally discuss the potential of a relatively new research approach, termed metabolomics, as a tool for new biomarker discovery, and consequently, better diagnostic approaches. Furthermore, this approach, when used alone or in combination with other “omics” techniques, allows for a better understanding of TB disease mechanisms, which may ultimately lead to improved treatment regimens, bringing us ever closer to eradicating this disease.*

Key words: Tuberculosis, Treatment, Diagnostics, Metabolomics

INTRODUCTION

Annually, approximately 9 million new tuberculosis (TB) cases are recorded, and up to 2 million people die as a result of this disease. TB is most commonly localised in the lungs and is caused by the bacteria, *Mycobacterium tuberculosis*. Although this is a worldwide epidemic, Africa (30%) and Asia (55%) accounts for 85% of all global TB cases. When evaluating recorded adult deaths in low- and middle-income countries, TB is ranked third, after HIV/AIDS and ischemic heart disease [1]. These statistics are disturbing considering the fact that TB can be prevented and is, in most instances, a curable disease. From the 1950s to the 1980s, various first-line anti-TB drugs have been developed, and when using combinations of these, patients with drug-susceptible TB can successfully be cured within 6 months. Approximately 0.4 - 0.5 million annual TB cases are, however, multi-drug resistant (MDR) [2], requiring the use of second-line drugs, which are costly, have

severe side effects and require almost 2 years of continual treatment, resulting in a dismal cure rate of only 50 – 70%. Since the 1980s, the rising HIV pandemic has become a significant obstacle in modern TB control. TB is considered to be the primary cause of death in most HIV infected patients, partly due to the increased TB infection rate in HIV patients, a weakened immune system, and the difficulties associated with diagnosing TB in HIV patients due to a lowered bacterial load in the lungs of these individuals [3,4].

In this overview, we describe the development and mechanism of action of current TB vaccination, diagnostic approaches, and treatment protocols, considering their advantages and drawbacks. We also discuss the potential of a number of newer treatment and diagnostic approaches recently developed. In order to fully understand the mechanisms by which these preventative, treatment and diagnostic approaches function, an understanding

of the pathophysiology of TB is required.

PATHOPHYSIOLOGY OF TUBERCULOSIS

TB is transmitted via aerosolized droplet nuclei, which are *M. tuberculosis* containing particles with a diameter of 1-5 μm . These droplets are produced and expectorated by patients with pulmonary TB during talking, singing, sneezing or coughing and can, due to their small particle size, stay airborne for many hours [5]. Additionally, these droplets can also be created in clinics and hospitals during sputum induction, aerosol treatments, aerosolization during bronchoscopy, or through tissue or sputum processing [6].

After inhalation by a previously uninfected individual, the infectious droplet nuclei lodge in the alveoli of the distal airways of the lung. Alveolar macrophages then engulf these nuclei, which sets off a cascade of events leading to either a successful suppression of the infection or the progression to active TB [7].

M. tuberculosis replicates slowly, but constantly, inside these macrophages and spreads to the hilar lymph nodes via the lymphatic system. In the majority of infected individuals, a cell-mediated immune reaction develops 2-8 weeks after infection. Consequently, necrotic, cheese-like granulomas, containing non-viable *M. tuberculosis*, are formed by activated T-lymphocytes and macrophages, which restrict the further replication and spread of the bacteria [8]. At the cellular level, *M. tuberculosis* infected macrophages interact with T-lymphocytes via a number of vital cytokines. Activated T-lymphocytes then release interferon- γ , another cytokine, which indirectly stimulates phagocytosis of *M. tuberculosis* inside these macrophages [9]. Interferon- γ also stimulates the release of tumor necrosis factor by these macrophages, which, in turn, aid in granuloma formation, and controls the extent of infection [10].

In a host with a strong cell-mediated immunity, a latent infection can be maintained in this form, and active disease may never occur. Even though small amounts of viable *M. tuberculosis* may be present in these granulomas, infected individuals (without active disease) cannot transmit the organism and are, hence, not infectious during this latent phase [6].

Active disease manifests when the host immune

system cannot restrain *M. tuberculosis* replication. Although this primary developing disease most frequently occurs in the parenchyma of the mid and lower lung, it may spread throughout the body and can present in almost any organ system. Several factors may trigger the conversion of latent infection to active disease, with HIV co-infection being the single greatest risk factor for this. Other immunocompromising conditions increasing the risk for active disease development include; diabetes mellitus, chemotherapy, malnutrition, vitamin D or A deficiency, renal failure, and extensive corticosteroid therapy [5].

Due to the insidious onset of TB, symptoms might only occur in later stages of the disease. In the case of an advanced disease, clinical symptoms include cough, fever, night sweats, anorexia, weight loss, chest pain, hemoptysis, and dyspnea [11]. As inflammation and tissue necrosis develop, sputum, which is widely used as diagnostic material, is produced. Hemoptysis, or the coughing up of blood, is typically a result of preceding disease and does not automatically indicate active TB. Hemoptysis may manifest as a result of remaining tuberculous bronchiectasis (irreversible dilation of part of the bronchial tree), breakage of a dilated vessel in a cavity wall, bacterial or fungal infections in a residual cavity, and / or erosion of calcified lesions into the lumen of an airway. TB may cause severe respiratory failure, but dyspnea is atypical, except in the case of extensive disease [6,12]. These symptoms alone are, however, rather non-specific and cannot be used for accurately diagnosing TB, as these correlate with many other lung infections or malignant conditions [11]. This strong, cell-mediated immune response, induced by TB infection, forms the basis on which the currently used TB vaccination procedures function.

TUBERCULOSIS VACCINATION

The first TB vaccination, and successful immunization, was carried out in 1921. This vaccine, the Bacille Calmette-Guerin or BCG, an avirulent *M. bovis* strain, was attenuated for 13 years by serial passages into glycerol imbibed potato slices [13]. Due to the high demand, the original BCG strain was distributed globally, even before the establishment of an appropriate culture protocol. Therefore, numerous BCG strains, with a variety of antigenic and immunological differences, exist today in various parts of the world. Nevertheless, BCG is currently still

the most commonly used TB vaccine worldwide, and BCG immunization is mandatory in high-incidence TB areas [14]. Using 1264 published articles, Colditz *et al.* (1994) did a meta-analysis on the efficiency of the BCG vaccination. This study concluded that, when vaccinated with BCG, the risk to develop TB is reduced by only 50% [15]. Newer approaches in the development of TB vaccines are, like BCG, also based on the induction of a strong cell-mediated immune response, which can control bacterial replication and maintain infection in the latent phase. Some of these attempts include: attenuated *M. tuberculosis* mutants; recombinant BCG strains; recombinant proteins; and DNA vaccines. However, due to the complexity of these studies, it is expected that the results of phase III trials will only be available by 2014-2015 [14].

TUBERCULOSIS TREATMENT

One of the greatest discoveries of the 20th century was that of antimicrobial drugs, which have ever since saved millions of lives and are still the major element in the treatment of infection. The first effective *M. tuberculosis* antimicrobial, streptomycin, was tested as part of the first randomized clinical trial carried out from 1948 until 1988. The publication of the positive outcome of this trial forever changed the national TB treatment policy [16].

Today, many anti-TB drugs are available which can, based on their activity, be classified into 3 groups: those with bactericidal activity, those with sterilizing activity, and those preventing drug resistance. Bactericidal activity is the capacity of a drug to reduce the amount of actively dividing bacilli in the initial therapy stage. Although rifampicin and streptomycin have some bactericidal activity, the most potent bactericidal anti-TB drug is isoniazid [17]. Sterilizing activity, as in the case of rifampicin and pyrazinamide, is the ability of a drug to eliminate the putative subpopulation of dormant bacteria from which a clinical relapse can occur [18]. Most effective treatment regimes consist of two phases: the initial intensive phase, during which a combination of at least two bactericidal anti-TB drugs (isoniazid and rifampicin) are used to kill actively growing *M. tuberculosis* populations, followed by a continuation phase for the elimination of intermittent dividing and dormant bacteria, using sterilizing drugs [5].

In an attempt to reduce the global burden of TB, the WHO formulated the Millennium Development Goals

(MDGs). The target of the MDGs is to halve TB prevalence and death rates by 2015 compared to that in 1990, and furthermore to completely eliminate TB by 2050. To achieve this, the WHO developed the DOTS (directly observed treatment, short-course) strategy in the mid-1990s, which was recommended internationally and consequently expanded worldwide. DOTS is a 6 months therapy regimen consisting of an initial 2 month treatment phase with four first-line drugs (isoniazid, rifampicin, pyrazinamid and ethambutol), followed by a 4 months treatment phase with only isoniazid and rifampicin. The addition of direct observation therapy (DOT), where patients consume each dose of anti-TB drugs under supervision, to the treatment strategy is strongly recommended. This approach maximizes the probability of therapy completion, hence, limiting the emergence of drug-resistance [19]. More or less 90% of all drug-susceptible TB cases are cured when this regimen is fully adhered to. In the case of MDR-TB, with resistance to at least rifampicin and isoniazid, DOTS alone may not succeed. For these cases, the WHO recommends DOTS-plus, which includes adding second-line drugs to the conventional DOTS program. This regimen is, however, costly, takes up to 24 months, and has a much higher level of toxicity as a result of harmful second-line medication [1].

A brief overview of the mechanisms of action and adverse effects of the most commonly used first- and second-line TB drugs will now be discussed.

First-line medications

Streptomycin: In 1944, Selman Waksman discovered and isolated the first effective anti-TB drug, streptomycin, from *Streptomyces griseus*, and for this, Waksman was awarded the Nobel Prize in 1952 for physiology and medicine [20].

Streptomycin is an aminoglycoside antibiotic which interferes with protein synthesis of the TB causing bacteria by inhibiting mRNA translation, resulting in the misreading of the genetic code, and consequently, cell membrane damage [21,22]. The exact binding site of streptomycin was found to be in the small 30S subunit of the ribosome, more specifically at ribosomal protein S12 (*rpsL*) and 16S rRNA (*rrs*) [23].

As early as 1946, only two years after its discovery, Klein and Kimmelman already reported the first streptomycin resistant mutants. These mutants could

be classified into two groups depending on whether or not they demonstrated a high or low level of resistance [24]. Ever since, numerous studies have investigated the underlying molecular mechanism of streptomycin resistance [25,26,27], but despite this, the molecular cause of this resistance is, however, still not yet completely understood.

Mutations within both the *rrs* and *rpsL* genes have been linked to a high-level streptomycin resistance [25,26]. Most of these resistance causing mutations leads to a hyper-accurate phenotype (increased accuracy of translation), compensating for the effect of the drug without having any effect on the interaction between the ribosome and the drug [28,29,30]. These mutations, nevertheless, have only been identified in just more than one half of all clinical streptomycin-resistant *M. tuberculosis* isolates [31,32]. The underlying mechanisms of the other, low-level streptomycin resistant strains, is still, however, obscure.

Recently, Okamoto *et al.* (2007) identified a mutation within the gene, *gidB*, conferring low-level resistance in 33% of *M. tuberculosis* isolates [30]. They furthermore confirmed that *gidB* encodes for a conserved 7-methylguanosine (m7G) methyltransferase (GidB) specific for the 16S rRNA. Further studies on the precise function and role of GidB in the pathogenesis of *M. tuberculosis* is, however, still unknown.

Apart from organisms developing resistance, other complications may also occur regarding the side effects of streptomycin treatment. Hypersensitivity is not unusual in patients treated with streptomycin, and may, depending on the severity, require treatment interruption. Transient dizziness and numbness around the mouth can occur after injections, but does regress and completely disappear when drug administration is stopped. With long-term treatment, chronic toxicity might lead to blackouts, ataxia, ringing in the ears, and even permanent deafness [1].

Isoniazid: Isonicotinic acid hydrazide, or isoniazid, is one of the most efficient anti-TB drugs used to inhibit mycobacterial growth. However, despite an immense number of studies dedicated to elucidating the mode of action of this drug, its exact mechanism is still largely unknown, mainly due to the fact that it may influence a variety of different mycobacterial cellular processes.

Early investigations indicated that, when exposed to isoniazid, *M. tuberculosis* loses its acid-fast nature and viability, which led to the suspicion that the drug functions by altering the cell-wall lipids, specifically by inhibiting mycolic acid synthesis [33]. Several studies have ever since confirmed that isoniazid does indeed disturb cell-wall structure [34,35] and electronmicroscopy scanning of *M. tuberculosis* exposed to isoniazid determined various altered morphological features [36]. Other modes of action of isoniazid include reactivity with bacterial proteins (tyrosine residues specifically) [37] and the formation of reactive oxygen radicals during the activation of the drug [38].

More recently, it has been established that isoniazid is a pro-drug and requires oxidative activation by catalase-peroxidase KatG [39]. An isonicotinoyl radical, the activated form of isoniazid, then reacts with NAD, leading to the covalent adduct INH-NAD, which inhibits the *M. tuberculosis* InhA enzyme [40]. InhA is an enoyl-acyl carrier protein, catalyzing the reduction of the trans double bond, conjugated to the carbonyl group of fatty acyl substrates, with NADH acting as the hydrogen donor for these reactions [41]. Inhibition of InhA will consequently block fatty acid elongation via the FAS II system, which is essential for mycolic acid synthesis, the main building blocks of the mycobacterial envelope [42].

Resistance to isoniazid entails various mutations (insertions, deletions and point mutations) in a number of genes. The major targets are the *katG* and the coding and regulatory area of *inhA*. Other mutations either have minor roles in isoniazid resistance or are compensatory due to the loss of catalase-peroxidase activity [43].

When recommended isoniazid doses are administered, adverse effects are uncommon, unless the patients have a history of previous kidney or liver failure [19]. Hypersensitivity reactions may, however, occur in the first week of treatment. Peripheral neuropathy is the most common adverse effect (20% of cases) of isoniazid treatment, especially in high-risk patients (pregnant women, alcoholic, malnourished, and diabetic patients), but can be prevented with a complementary low dose of vitamin B6 [1]. When isoniazid is used in isolation, hepatitis manifests in 0,6% of cases, and when used in combination with rifampicin, the incidence increases to 2,7%. This incidence increases with age and in

patients with previous liver disease. The fatality rate of isoniazid treated patients due to hepatitis is, however, less than 0.023% [19,44].

Ethambutol: Ethambutol is normally included in initial TB treatment regimes, especially when potential isoniazid resistance is expected [17]. Ethambutol is effective against intracellular and extracellular bacteria, however, the exact mode of action is still unknown. To date, several studies have been done in an attempt to explain this drug's mode of action, however, most of these analyses focused on its actions on altering the mycobacterial cell wall structure [45,46]. Takayama and Kilburn (1989) showed that ethambutol has an inhibitory effect on the transfer of arabinogalactan (arabinosyl transferases) into the *Mycobacterium* cell wall, which in turn leads to the accumulation of trehalose mono- and dimycolates [46]. Silve *et al.* (1993) additionally indicated that ethambutol inhibits the transfer of [D-14C]glucose into the D-arabinose fraction of arabinogalactan [47]. Consequently it was suggested that, due to inefficient arabinogalactan transfer into the cell wall, mycolic acids accumulate in these mycobacteria, leading to the previously observed bacterial declumping and morphological alterations following ethambutol treatment [48]. Furthermore, sequence analyses of ethambutol resistant, clinically isolated mycobacteria, showed that this resistance is primarily linked to a number of missense mutations in the ethambutol resistance determining region of the arabinosyl transferase encoding gene, *embB* [49]. Starks *et al.* (2009) determined that mutations in *embB* codon 306, in particular, are important indicators of ethambutol resistance, and may also be useful for confirming isoniazid resistance in 50-70% of clinical samples [50]. The exact role of the *embB* 306 mutation in the acquisition of isoniazid resistance is, however, still controversial.

A major adverse effect associated with ethambutol treatment, is retrobulbar optic neuritis, with symptoms including: the decline of visual sharpness (blurry vision), dyschromatopsia (colour blindness), and central scotoma (tunnel vision). These effects can be prevented in the case of low dosage treatment for a maximum period of 3 months [1].

Rifampicin: Rifampicin is a powerful anti-TB drug, reducing actively dividing and semi-dormant organisms. This drug primarily inhibits DNA-

dependant RNA-polymerase, an enzyme essential for the transcription of RNA [51]. Resistance to rifampicin is mainly as a result of alterations in the β -subunit of RNA-polymerase due to mutations in the encoding *rpoB* gene. A variety of specific resistance-conferring mutations (accounting for more than 95% of rifampicin-resistant strains) have been described in the 81-bp region of *rpoB* known as the rifampicin-resistance determining region (RRDR) [52]. The majority of these are point mutations, and results in the replacement of aromatic with non-aromatic amino acids. These replacements lead to drug resistance by interrupting the forces that bind rifampicin to RNA polymerase. These mutations also impair the fitness of these bacteria, but the fitness can, however, be restored by secondary mutations [53].

Minor side effects associated with rifampicin treatment include: flu-like symptoms, abdominal pain, fatigue, ataxia, dyspnea, and anorexia and, in most cases, do not warrant the discontinuation of treatment. In combined-treatment procedures, however, exanthema (a rash-like reaction) can occur, which requires immediate discontinuation. Furthermore, cholestatic hepatitis occurs in 2.7% of patients treated with a combination of isoniazid and rifampicin and in up to 1.1% when rifampicin is combined with other anti-TB drugs [1,19,44].

Pyrazinamide: Pyrazinamide, together with isoniazid, rifampicin, and ethambutol, plays an important role in initial TB treatment regimes. As it is active in an acidic pH environment, hence, able to kill semi-dormant bacteria not killed by other TB drugs, the inclusion of pyrazinamide in these regimens shortens TB therapy from 9 to 6 months [19]. Even though pyrazinamide is widely used today, its specific mechanism of action, as is the case with most other TB medication, is still largely unknown. Pyrazinamide, a prodrug, is converted to its active form, pyrazinoic acid (POA) by the bacterial enzyme nicotinamidase/pyrazinamidase (PZase) [54], and it has been shown that various mutations in the PZase encoding gene, *pncA*, leads to pyrazinamide resistance in *M. tuberculosis* [55]. All bacteria are equipped with PZase, nevertheless, *M. tuberculosis* is uniquely susceptible to pyrazinamide. This susceptibility is due to a deficient pyrazinoic acid efflux mechanism in *M. tuberculosis*, in contrast to the natural pyrazinamide resistant *M. smegmatis*, which rapidly extrudes pyrazinoic acid out of the bacterial cell [56]. Various studies have implied that

pyrazinoic acid does not have a specific bacterial target and most likely kills *M. tuberculosis* due to its weak acid nature [57]. More recent studies, however, suggest that the bacterial membrane is de-energized by pyrazinoic acid and pyrazinamide as a result of a collapsing membrane potential, and that these compounds also have an effect on membrane transport function in an acidic environment [58].

Despite its comparatively better mode of action for eliminating TB, pyrazinamide does have rather severe side effects. Pyrazinamide treatment may be associated with pruritus, exanthema, or rhabdomyolysis with kidney failure, and myoglobunuria. Pyrazinamide is, additionally considered to be the most hepatotoxic of all first-line anti-TB drugs and treatment should be replaced or, at least, temporarily discontinued when any of the above side effects are noticed. [1,44].

Second-line medication

D-Cycloserine: As is the case with many anti-TB drugs, D-cycloserine (D-4-amino-isoxazolidone), a cyclic structural analogue of D-alanine, acts by inhibiting cell wall synthesis in mycobacteria [59]. D-Amino acids (D-alanine, D-glutamate, and D-aminopimelatein, in particular) are important components of the backbone of the bacterial cell wall, peptidoglycan. Alanine is typically accessible as the L-stereoisomer and is converted to D-alanine via D-alanine racemase (Alr) in the initial step of alanine branching during peptidoglycan biosynthesis [60]. The dipeptide D-alanyl-D-alanine, is then formed in a D-alanine:D-alanine ligase (Ddl) catalyzed reaction. D-cycloserine blocks mycobacterial cell wall synthesis by inhibiting Alr and Ddl [59].

Caceras *et al.* (1997) determined that the over expression of *alr* in *M. smegmatis* and *M. bovis* led to a D-cycloserine resistant phenotype [61]. This resistance was observed to a far higher degree than was seen during Ddl over expression, leading to the suspicion that the primary target of D-cycloserine might be Alr [62]. Furthermore, an increased sensitivity to D-cycloserine was observed when *alrA* [63] or *ddl* [64] was inactivated. Even though it is presumed that D-cycloserine resistance results due to mutations in *alrA* and *ddl*, the exact mechanism of resistance is still unclear [65].

Despite the fact that D-cycloserine is a successful antimycobacterial drug, it is rarely prescribed and only

used in combined therapies due to its severe adverse effects. Neurological effects including: headaches, vertigo, memory deficiency and mental confusion only to name a few, in addition to psychiatric effects including: depressive and paranoid reactions, and psychotic states, are common in patients receiving D-cycloserine treatment [65]. These adverse effects are the result of D-cycloserine binding to neuronal N-methyl aspartate receptors [66], and the inhibition of the enzymes involved in the metabolism and synthesis of the neurotransmitter γ -aminobutyric acid [67].

Ethionamide: Being structurally similar to isoniazid, ethionamide is also considered a prodrug, requiring activation by the bacterial cell. When activated, ethionamide, like isoniazid, disrupts cell wall biosynthesis by inhibiting a mutual cellular target, the enoyl-acyl carrier protein, *inhA* [68,69]. Gene array studies, indicating similar patterns of gene expression induced by both isoniazid and ethionamide, verifies this shared site of action [70]. The fact that isoniazid resistance does not always result in ethionamide resistance, has led to the suspicion that different enzymes may activate these drugs. Little was known about this enzyme, until only recently, when two different research groups almost simultaneously identified *etaA*, a monooxygenase [71,72]. *EtaA*, a FAD-enclosing enzyme, oxidizes ethionamide to its analogous S-oxide, which, in turn, is oxidized to 2-ethyl-4-amidopyridine by an unstable oxidized sulfinic acid intermediate [69]. Resistance to ethionamide may occur due to various mutations in this activating enzyme, *etaA*, and the target, *inhA* [73].

Ethionamide is classified as a second-line drug due to its severe side effects. Extreme gastrointestinal associated complications, including intense salivation, nausea, loss of appetite, vomiting, and abdominal pain are common after ethionamide treatment. These symptoms can be reduced by administering the drug with food or at bedtime. As is the case with isoniazid, ethionamide can cause hepatotoxicity, especially in patients with liver disease and a history of alcoholism. Neurological symptoms such as optic neuritis, anxiety, depression, and hallucinations have been reported in only 1-2% of all treated patients. Other adverse effects of ethionamide include: postural hypotension, alopecia, impotence, hypothyroidism, acne, and photosensitivity, only to name a few [65].

Kanamycin and amikacin: Since the discovery of streptomycin in 1944 [20], aminoglycosides have played a major role in TB therapy. Kanamycin in

(isolated from *Streptomyces kanamyceticus*) and amikacin (a semi-synthetic drug derived from kanamycin), are commonly used in MDR-TB treatment regimens [65,74]. These drugs, like streptomycin and other aminoglycosides, bind to the small 30S ribosomal subunit, in particular the 16S rRNA (*rrs*), consequently inhibiting translation and, hence, protein synthesis [75]. Information regarding the exact mechanism underlying kanamycin resistance in *M. tuberculosis* is scarce. What is known is that similar to streptomycin, mutations in the *rrs* gene, results in high-level kanamycin resistance, and some mutations may also cause cross-resistance to amakacin and other second-line drugs [74]. However, as much as 80% of clinical kanamycin resistant clinical isolates exhibit low-level resistance without *rrs* mutations or cross-resistance. Recently, Zaunbrecher *et al.* (2009), determined that in *M. tuberculosis*, an over expression of the survival protein, Eis, due to *eis* promotor mutations, is common in these low-level kanamycin resistant clinical isolates. Eis, an aminoglycoside acetyltransferase, inactivates both kanamycin and amakacin. Kanamycin is 3-times more efficient than amakacin as a substrate for Eis, explaining the lack of cross-resistance of an *eis* mutation to kanamycin [76].

Ototoxicity, or damage to the ear, is one of the most severe adverse effects of aminoglycosides. Hearing loss is a result of cranial nerve VIII damage, including cochlear and vestibular impairment. Instant discontinuation of aminoglycoside treatment is advised when ototoxicity occurs. As a result of their build-up in renal tubules, aminoglycosides may additionally cause toxic renal effects such as proteinuria, oliguria (low urine output), and decreased creatinine clearance. In rare occasions, aminoglycoside treatment might bring about hypersensitivity (extreme allergic reaction to the drug) or neuromuscular blockage, resulting in respiratory failure [65].

Fluoroquinolones: Nalidixic acid, the first quinolone medication, was acquired in the early 1960's as an impurity during the production of quinine [77]. Many fluoroquinolone derivatives have ever since been tested for their antibacterial activity. Levofloxacin, sparfloxacin, ofloxacin, and ciprofloxacin have shown to be extremely active against *M. tuberculosis* [78], and have therefore been used in TB treatment regimens since 1980's [79].

Fluoroquinolones circulate throughout the body and

have the extraordinary property of functioning intracellularly, reaching mycobacteria inside the macrophages and, hence, leading to a potent treatment outcome [80]. Fluoroquinolones function by inhibiting DNA gyrase (Gyr, a type II topoisomerase), an enzyme essential for reducing the tension when double-strand DNA is unwound during DNA replication, recombination and expression. This inhibition prevents supercoiling of the DNA, leading to uncontrolled mRNA synthesis, exonuclease production, protein synthesis and chromosome degradation, due to free DNA ends [80,81]. When used as a monotherapy, mycobacteria quickly develop resistance to fluoroquinolones due to mutations in the DNA gyrase enzyme (*gyr*). Other resistance conferring mechanisms include; 1) the presence of an efflux system, actively releasing the drug from the bacterial cell or, 2) an altered cell membrane structure, leading to fluoroquinolone impermeability and thus reducing diffusion of the drug into the cell [80]. Cross-resistance between fluoroquinolones and other TB-drugs however, does not occur, and even though cross-resistance between various fluoroquinolones has been described [80,81], moxifloxacin and levofloxacin have been used to successfully treat patients resistant to ofloxacin. Furthermore, a study done in India showed that oxifloxacin, in combination with various first-line drugs, may be effective as a three month, ultra-short course TB treatment regime [82].

Gastrointestinal effects are the most common side effects associated with fluoroquinolone treatment. These effects including; vomiting, anorexia, diarrhea, and abdominal pain, arise in 3 - 17% of treated patients. A small amount (0.9 - 11%) of patients receiving fluoroquinolone treatment may also develop insomnia, tremors, and headaches. Skin rash, erythema, and pruritus may occur in 0.4 - 2.2% of fluoroquinolone treated patients [65].

Table 1 gives a summary of the modes of action, cellular targets, and resistance conferring genes of the drugs used for TB treatment. Although the WHO recommended treatment regimes are highly effective, various studies have demonstrated unwanted interactions between the different anti-TB drugs and between anti-TB drugs and other medications used by TB patients [19].

Three of the four first-line anti-TB drugs in the DOTS program are possibly hepatotoxic and depending on

Table 1: Mode of action of commonly used anti-TB drugs (Adapted from Zhang [22])

Drug	Discovery year	Mechanism of action	Cellular target	Resistance conferring genes
First-line				
Streptomycin	1944	Inhibition of protein synthesis	Ribosomal S12 protein and 16S rRNA	<i>rpsL</i> <i>rrs</i>
Isoniazid	1952	Inhibition of cell wall mycolic acid biosynthesis	Multiple targets including acyl carrier protein reductase (<i>inhA</i>)	<i>inhA</i> <i>katG</i>
Pyrazinamide	1952	Membrane transport disruption	Membrane energy metabolism	<i>pcnA</i>
Ethambutol	1961	Inhibition of cell wall arabinogalactan synthesis	Arabinosyl transferase	<i>embB</i>
Rifampicin	1966	Inhibition of transcription	RNA polymerase β subunit	<i>rpoB</i>
Second-line				
D-cycloserine	1952	Inhibition of peptidoglycan synthesis	D-alanine racemase	<i>alrA</i> , <i>ddl</i>
Ethionamide	1956	Inhibition of cell wall mycolic acids biosynthesis	Acyl carrier protein reductase (<i>inhA</i>)	<i>InhA</i> <i>EtaA</i>
Kanamycin and amikacin	1957	Inhibition of protein Synthesis	16S rRNA Aminoglycoside acetyltransferase	<i>rrs</i> <i>eis</i>
Fluoroquinolones	1963	Inhibition of DNA synthesis	DNA gyrase	<i>gyrA</i> <i>gyrB</i>

race, geographic location, and socioeconomic status, 1-10% of patients treated with these medications develop drug-induced hepatitis [83].

In addition to the severe side effects resulting from these drug interactions, serum drug concentrations are also altered, thereby reducing their efficiency [44]. Severe effects due to anti-TB drugs do, however, emerge because of several factors, and the extent of these effects are influenced by, amongst others, the age of the patient, nutritional status, dosage, time of administration, and pre-existing diseases or dysfunctions (such as HIV co-infection, liver or kidney impairment, and alcoholism) [1,19,44]. Changes in first-line treatment procedures are mainly as a result of severe adverse effects, leading to the use of more toxic, less active, and more expensive second-line TB drugs, usually accompanied by increased hospitalizations and home visits [83]. These side effects additionally contribute to patients disrupting or ceasing treatment, leading to higher levels of treatment failure and, hence, acquired resistance [44].

The shortcomings of current anti-TB drugs and TB treatment regimes are not entirely responsible for the case detection rate of 61% in mid-2010, which fell far short of the global target of 70% [3]. Despite the need for faster acting, less toxic TB treatment procedures, we also urgently require innovative, sensitive and rapid diagnostic approaches, which can manage the concurrent HIV epidemic and rising

incidence of MDR-TB cases.

TUBERCULOSIS DIAGNOSTICS

The emergence of MDR-TB, and the rising HIV pandemic, has challenged the conventional TB diagnostic methods, and the development of new, accurate, sensitive and quick TB diagnostic approaches are now crucial. In the past few years, innovative, high-tech diagnostic procedures such as molecular techniques and rapid cultures have entered the market. These tests are, however, still only commercially used in high-income countries, where the TB incidence is low. Due to their complexity and high costs, these tests are not yet suited for high-burden, low-income settings [84].

The underlying mechanism of the currently available and more recently developed TB diagnostic procedures, considering each of their advantages and disadvantages will subsequently be discussed.

Tuberculin skin test: The tuberculin skin test (TST) is based on the delayed-type hypersensitivity reaction produced in *M. tuberculosis* infected individuals, when antigenic compounds (purified protein derivative (PPD), acquired from heat-killed *M. tuberculosis*) are injected intra-cutaneously into the forearm of an individual [6,84]. T-cells formed as a result of present or prior infection, then travel to the infected skin area where they release lymphokines. The induction of local vasodilatation, fibrin deposition,

edema and the accumulation of other inflammatory cells, as a result of these lymphokines, cause a thickening of the skin at the injection site, which can be measured [6]. The reported sensitivity and specificity of this technique, however, varies due to number of reasons. For instance, in a study conducted by Al *et al.* (2000), patients with a TST with a size smaller than the 5 mm threshold, were found to be less likely to have active TB, however, cases above the threshold were not necessarily indicative of active disease [85]. Additionally, patients with varying disease states gave almost identical results, considering the size and shape of the TST. Furthermore, false-positives regularly occur in patients who were previously vaccinated with BCG or are infected with other, non-tuberculous mycobacteria. False-negative reactions, on the other hand, may also occur due to a variety of technical factors, for example, in patients with compromised immune systems due to immunosuppressive drugs (i.e. steroids), AIDS, cancer, age (newborns and adults over 65 years), and supplementary bacterial, fungal or viral infections [86,84].

Cytokine detection assay: A newer approach used for the detection of latent TB infection, also making use of the cell-mediated immune response, is the cytokine detection assay. In this case, circulating lymphocytes (extracted from a patients blood), are exposed to mycobacterial antigens for 6 to 24 hours. *M. tuberculosis* infected lymphocytes recognize these antigens and subsequently produce cytokines, mostly interferon-gamma (INF- γ), which can be measured [84]. The first INF- γ release assays (IGRAs) used PPD as the antigen of choice. More recently developed tests, however, rely on antigens which are more specific to *M. tuberculosis*, such as early secreted antigen target (ESAT)-6 and culture filtrate protein (CFP)-10 [87]. Currently, there are two commercially available IGRAs termed QuantiFERON®-TB Gold and T-SPOT-TB. Several studies indicated that these assays have a high specificity (>95%), but a lower, variable sensitivity (75-97%), and also, results between the two IGRAs vary, limiting their efficacy in routine clinical settings [87,88]. Considering that both TST and the cytokine detection assay detect markers for possible infection only, follow-up diagnostics are required to determine if the patient with an infection does in fact have active TB.

Radiographic methods: For the diagnosis of active pulmonary TB, chest X-rays (CXRs) can be classified as either 1) typical of TB, with i) the occurrence of

nodular, alveolar, or interstitial infiltrates predominantly affecting the zones above the clavicles or upper zones of the lung, or ii) the presence of cavitations affecting the upper zones or the apical segment of the lower lobes of the lungs; 2) compatible with TB in the case of atelectasis, enlarged hilar nodes, pleural exudate, mass lesion, miliary or pneumonic lesion, or; 3) atypical (all other patterns, including normal CXR) [89]. As summarized by the WHO [84], X-rays alone are still widely used as an important tool in TB diagnostics, despite the fact that various studies have proven that TB shows no unique radiographic patterns [6]. Also, numerous other lung diseases have a similar radiographic appearance, which can easily mimic TB [90], leading to over diagnosis when used alone. When using culture as the reference test, Van Cleeff *et al.* (2005) reported the sensitivity and specificity of CXRs as 80% and 67% respectively [91], whereas Arslan *et al.* (2010) reported these values to be 73% and 94% respectively [92], which is unsatisfactory. CXR can thus be useful for the identification of abnormalities in the lungs, but to ascertain the tubercular aetiology further tests, such as bacteriology, are an absolute necessity [84].

Microscopic examination: Despite recent advantages in TB diagnostics, smear microscopy, as first demonstrated by Robert Koch in 1882, is still the most commonly used screening method for the detection of mycobacteria in clinical sputum specimens [93]. This quick (less than 2 hours), simple and low cost technique is based on the acid-fastness of mycobacteria, and thus the ability of these organisms to retain dye after treatment with an acid-alcohol solution [6]. The characteristic mycolic acids present in the cell walls of all mycobacteria are responsible for this colour reaction, hence, limiting the ability of the method for species identification and drug susceptibility testing. Also, these fatty acids do persist when the bacteria dies, and therefore, this technique cannot discriminate a current, active disease state, from previous *Mycobacterium* infection [94]. Furthermore, numerous quantitative studies have shown that high amounts of bacilli (5000 – 10 000 bacteria mL⁻¹) are required for the detection of bacteria using smear microscopy tests [95], leading to a sensitivity of no more than 35 – 70%. This method has on occasion, however, been reported to detect only 20 – 30% of all TB cases [96], in less advanced disease states, TB-HIV co-infection, and in children [84]. Nevertheless, although a negative smear does not rule out mycobacterium infection, a positive result almost verifies a diagnosis, resulting

in a technique specificity of more than 95%. This high specificity is limited only by cross-contamination, due to environmental mycobacterium, non-tuberculous mycobacterium, and technical inadequacies [94].

Bacteriological culture: The WHO considers bacteriological culture as the gold standard for TB diagnosis, as it detects over 80% of TB cases accurately, with a reported specificity of close to 100% [97,84]. However, it has been reported that negative sputum cultures may occur in 15-20% of adult pulmonary TB cases. Additionally, despite immense anti-contamination procedures, false-positive culture results have been reported to occur in 1-4% of all cases, due to a transfer of bacilli from TB-positive to TB-negative samples during laboratory handling [98].

Either solid or liquid media can be used to grow cultures. Solid media including; egg based media (e.g. Löwenstein–Jensen), or agar based media (e.g. Middlebrook 7H10), are most commonly used due to their low costs. This approach, however, has a diagnostic time of 4-6 weeks [99]. Growth in liquid media is considerably faster (2-4 weeks) and far more sensitive, but does have the disadvantage of higher contamination rates (8-10%) when compared to solid media (3-5%) [94]. The development of automated culture systems, such as BACTEC 460 and mycobacterial growth indicator tube (MGIT) systems was a major improvement in mycobacterium culture diagnosis [6]. Hanna *et al.* (1999) reported a multicenter evaluation of the BACTEC MGIT 960 and BACTEC 460 systems in comparison to solid culture media for mycobacteria isolation. They found that traditional solid media (Löwenstein–Jensen slopes and Middlebrook 7H11 plates) was capable of a *M. tuberculosis* recovery rate of 79%, in a mean time of 24.1 days, in comparison with BACTEC 460 TB and BACTEC MGIT 960, with recovery rates of 90% and 77%, and mean times of 15.2 and 14 days, respectively. A combination of solid media and BACTEC 460 gave the best *M. tuberculosis* recovery rate (97%) [100]. As an added advantage, the above-mentioned methods can differentiate between various clinically important *Mycobacterium* species, including non-tuberculous *Mycobacterium* [101]. These automated systems are, however, expensive and requires expensive infrastructure and maintenance, and skilled and experienced staff [84].

Bacteriological culture is also the conventionally used method for drug susceptibility testing (DST) in TB. When using solid media, the growth of *M. tuberculosis* in the presence of anti-TB medication can be detected by one of three methods including: proportions, resistance ratios, or absolute concentrations [43]. In the proportions method, which is considered the reference standard, the growth of organisms on a drug-free media is compared to the growth on media containing an anti-TB drug. The resistance ratio method, on the other hand, determines the minimum inhibitory concentration of a drug-susceptible, reference strain, as compared to the patient's strain. Lastly, the absolute concentrations method utilizes media containing various dilutions of the anti-TB drug in order to determine the lowest concentration of the medication necessary to inhibit growth. These results are, however, only available within 2-3 months after sample collection [84].

The gold standard for culture DST, for both first- and second line medication, is the semi-automated radiometric BACTEC 460 TB system [102,103], where bacilli are cultivated in liquid media containing various concentrations of anti-TB drugs. This method has been used effectively for over 25 years, and has reduced the time required for DST to a mere 4-13 days [102]. Nevertheless, the BACTEC 460 TB system makes use of radioactivity, leading to many concerns regarding safety during use and disposal after use. To overcome this problem, the BD BACTEC MGIT 960 SIRE (testing for Streptomycin, Isoniazid, Rifampin, and Ethambutol) assay has been developed and has shown to be an excellent alternative to the radiometric assay [103]. This automated, non-radiometric method uses fluorometric technology to accurately detect the consumption of O₂ in the presence of anti-TB medication, leading to a diagnostic result within 4-12 days [102]. It does, however, suffer the disadvantage of higher contamination rates than the BACTEC 460 TB assay [102,103]. This contamination may be due to either the richness of the medium (which is unlikely to change due to the methods grow detection principle) or the use of screw caps instead of rubber septa, which could be resolved in time [103]. These methods are considered extremely expensive and require highly trained personnel, a limitation to developing countries [84]. A newer liquid-media-based DST method, the microscopic observation drug susceptibility (MODS) assay, is however, more affordable and does not require radioactive isotopes or fluorescent indicators.

With MODS, inverted-light microscopy is used to detect early growth of *M. tuberculosis* as strings and tangles of bacterial cells in Middlebrook 7H9 broth medium in the absence or presence of anti-TB drugs, with a diagnostic result in less than two weeks [104]. Six to eight mL of sputum per specimen is, however, required for this analysis, which is already difficult to obtain from adults, not to mention children and immuno-suppressed (TB/HIV co-infected) patients. Further disadvantages associated with this method are; the fact that it is an indirect method for the detection of *M. tuberculosis*, requires daily microscopic observations, and does not allow for mycobacterial species identification [43].

Colourimetric methods used for DST in TB are based on the reduction of an oxidation–reduction indicator, which is added to liquid culture medium after the exposure of *M. tuberculosis* to various anti-TB medications. A colour change, proportional to the number of viable bacteria, is indicative of drug-resistance [105]. These tests are mainly performed on clinical isolates and hence, do not exclude the critical culturing waiting period. Some of these methods have, nevertheless, been tested directly on sputum samples with a sensitivity and specificity ranging between 88% and 100%. These studies did, however, only focus on the detection of rifampicin- and isoniazid-resistant TB, using highly infectious samples (positive with more than 10 acid-fast bacilli per microscopic field) [106]. A big concern of colourimetric tests is the biohazard of aerosol generation due to the manipulation of microtitre plates. Also, phase III and IV diagnostic trails are still required before the clinical implementation of these methods can be considered [105].

High performance liquid chromatography: For species identification, high performance liquid chromatography (HPLC) assays can be completed in a few hours, and have sensitivities and specificities of almost 100%. This technique can identify and distinguish over 50 *Mycobacterium* species, based on their unique sets of mycolic acids and β -hydroxy- α -fatty acids [107]. HPLC cannot, however, distinguish between *M. tuberculosis* and *M. bovis* and requires at least 10^5 pure culture organisms for a diagnostic result, hence not excluding the delay of bacteriological culture [6].

Nucleic acid amplification: A newly developed TB diagnostic approach, nucleic acid amplification tests (NATs), enzymatically amplifies regions of bacterial

DNA specific to the *M. tuberculosis* complex. The most widely used NATs are polymerase chain reaction (PCR), transcription mediated amplification (TMA) and strand displacement amplification (SDA) [84]. Several commercial NATs tests are available, which can be used directly on sputum or other clinical samples [109]. In a study conducted by Catanzaro *et al.* (2000), the clinical performance of a NAT, approved by the Food and Drug Administration in 1995, the enhanced *Mycobacterium tuberculosis* Direct (E-MTD) test, was investigated. The sensitivity of the E-MTD test for low, intermediate, and high clinical TB suspicion was 83%, 75%, and 87% respectively, with a corresponding specificity of 97%, 100% and 100%. The positive prediction values of the E-MTD test were 59% (low TB suspicion), 100% (intermediate TB suspicion), and 100% (high TB suspicion) vs. 36%, 30%, and 94% respectively for smear microscopy, hence, confirming NATs to be potentially helpful for the diagnosis of early stage TB [109]. Furthermore, Pounder *et al.* (2010), developed a genomic deletion assay based on multiplex PCR with melting temperature analysis, to differentiate between six clinically important *Mycobacterium* species, based on the regions of difference (RDs) in the their complete genome sequences. Using a set of 3 primers for each RD, they correctly identified 96% of *Mycobacterium* isolates in pure culture [110]. However, because sequence variation may occur at primer binding sites during evolution, a reselection of the target sequence may be required in future.

Genotyping methods have also been developed for DST of all first-line and most second-line TB drugs by detecting specific resistance linked mutations in target genes of *M. tuberculosis*. PCR-restriction fragment length polymorphism (PCR-RFLP) analysis is a rapid, low-cost method for the detection of polymorphisms at mutated codons of mostly INH and EMB resistant *M. tuberculosis* strains [111,112]. However, due to the presence of PCR inhibitors in sputum samples, researchers still prefer to use pure cultures as the genomic DNA source for PCR amplification, hence, this diagnostic approach still suffers the culturing waiting period [111]. Direct DNA sequencing is the most realistic method to use if all resistant strains have mutations in a specific region of a single target gene (as is the case with the RRDR in the *rpoB* gene of rifampicin-resistant strains). DNA sequencing is, however, impractical when analysing large amounts of specimens, particularly in developing countries [43]. Espasa *et*

al. (2005) evaluated the potential of real-time PCR for rifampicin and isoniazid susceptibility testing in clinical samples. The sensitivity of the test ranged from 30.4 to 35.3% in smear-negative samples and 95.1 to 99.2 % in smear-positive samples, with a 100% specificity. The detection limit of real-time PCR for detecting target mutations in clinical samples was found to be 1.5×10^3 CFU/mL, compared to 10 CFU/mL in culture [113]. Furthermore, genetic alterations in the target gene sequences were absent in 30% of the isoniazid-resistant isolates, hence, resulting in the low sensitivity of this diagnostic approach.

Commercially available NATs, nevertheless, suffer the further disadvantages of high costs and the need for high tech infrastructure and well-trained personnel. The high incidence of false-positive results due to laboratory cross-contamination, also limits its performance under field conditions [84,109].

Serology: The use of serology (detection of antibodies, antigens and immune complexes) in TB diagnostics has, up to now, failed largely in providing the necessary sensitivity and specificity for effective clinical use [84]. Perkins *et al.* (2003), for example, evaluated the performance of a commercial immunochromatographic test kit (ICT Tuberculosis), employing five recombinant *M. tuberculosis* antigens. This kit correctly identified only 64.2% of the smear-positive and of the 46.3% smear-negative, culture confirmed-TB patients [114]. Using a serological approach for TB diagnostics is challenging, as the various stages of the disease e.g. exposure, latent infection, active disease and severe disease, each have their own antibody patterns. Therefore, when developing such a test, one must consider the use of antigens expressed in all stages of the disease by employing, for instance, cocktails of multiple antigens [108]. Results may also vary due to the lack of reproducibility in antigen purifying methods and, additionally, since environmental mycobacteria can cross-react with antibodies in samples, false positive results frequently occur. Therefore, the use of serological TB diagnostic methods is not yet recommended by the WHO [84].

Phage assay: Currently, the only commercially available phage assay for the detection of TB, FASTPlaqueTB, makes use of mycobacteriophages (mycobacteria infecting viruses) to signify the incidence of viable *M. tuberculosis* in clinical

samples. When using bacteriological culture as the reference test, Muzaffar *et al.* (2002) determined the sensitivity and specificity of this assay to be 87.4% and 88.2% for smear-positive sputum samples as opposed to 67.1% and 98.4% respectively in smear-negative, culture confirmed cases [115]. Another study, conducted in South Africa, confirmed the poor sensitivity (48.7%) of this approach in smear-negative, culture confirmed specimens [116]. Although these assay test results can be achieved in 48 hours and is specific for *M. tuberculosis*, it missed other infectious *Mycobacterium* spp., and is not able to detect drug resistance [84].

When weighing up the advantages and disadvantages of the above-mentioned TB-diagnostic approaches, it is obvious that no test currently available meets all the specifications of sensitivity, specificity, speed, safety, robustness, training simplicity, and cost. Hence, the world is in need for a new, innovative TB diagnostic method, which will overcome current limitations, and can easily be implemented in low-income, high-burden counties.

WHERE TO FROM HERE?

The metabolome is the ultimate downstream result of genome transcription, and may be described as a compilation of all the metabolites (small molecular compounds), present in a specific cell or organism, participating in metabolomic reactions during normal cell function, growth, and maintenance. Metabolomics is 'the non-biased identification and quantification of *all* these metabolites in a biological system', using highly sensitive analytical procedures [117]. During active disease and perhaps even latent infection, the infectious organisms, such as TB causing bacilli, disturb the host's biochemical networks and provoke alterations in the quantity and types of metabolites present [118]. Using specialized techniques, this response to pathological stimuli can be measured over time. This approach is the foundation for the discovery of new metabolomic biomarkers for diagnosing disease, better describing the mechanisms of drug resistance, elucidating drug mechanisms and monitoring treatment approaches.

Since TB was discovered in the 1800's, characterization, diagnosis and treatment relied solely on the identification and traits of the causative organism in clinical samples. Ever since then, technical advances in the detection limits of various

analytical techniques have made it possible to diagnose TB from samples containing very low concentrations of bacilli. These analytical techniques do, however, go hand-in-hand with elevated levels of complexity and costs [84]. Metabolomics not only considers the metabolome of the infected organism, but also measures changes in the host metabolism due to infection. Consequently, new biomarkers may be used, not only diagnostically, but also to better characterize disease mechanisms in the host and to monitor treatment outcomes. In the past few years, metabolomics applications for the characterization of a variety of diseases has been described, and these include, amongst others: coronary heart disease [119]; type 2 diabetes [120]; epithelial ovarian cancer [121]; Huntington disease [122]; hypertension [123]; liver cancer [124]; liver failure due to hepatitis B infection [125]; meningitis and ventriculitis [126] Parkinson's disease [127, 128]; pre-eclampsia [129,130], and schizophrenia [131]. Recently, two groups have investigated the potential use of metabolomics for TB diagnostics.

Pavlou et al. [132] completed a pilot study to test the potential of an electronic nose, which uses 14 conducting polymer sensor arrays, to identify *M. tuberculosis* in culture and patient collected sputum samples. Using this approach, they were able to discriminate between *M. tuberculosis* and sterile cultures, with a 100% prediction value, and also, between closely related cultures (*M. avium*, *P. aeruginosa*, *M. tuberculosis*, and *M. tuberculosis* + *M. scrofulaceum*) with a prediction value of 96%. They furthermore built a discriminant model using 36 patient collected sputum samples (*M. tuberculosis* (6), *M. avium* (8), *P. aeruginosa* (8), mixed infection (6) and control (6)), which correctly identified 90% of the unknown samples used to validate this method [132]. The same research group then further investigated this approach, applying it to 330 culture proven, HIV tested sputum samples. They detected TB with a sensitivity of 89% and a specificity of 91% at a detection limit of 1×10^4 bacteria mL⁻¹, which is comparable to smear microscopy. As the nature of sensor volatiles is not yet fully understood, the characteristics of these volatiles can, however, not be identified or quantified, which may be a limitation to using this sensor array technology [133].

Consequently, Phillips et al. [134] tested the hypothesis that volatile organic compounds (VOCs) in breath

might lead to the identification of new TB biomarkers using a GCMS metabolomics approach. This hypothesis was made based on the fact that mycobacteria produce unique VOC patterns *in vitro*, and the observation that increased oxidative stress in TB patients may lead to distinct VOC patterns. In this study, they compared the VOC patterns in the headspace of *M. tuberculosis* cultures to that of sterile growth media. Headspace VOCs were captured on sorbent traps, which were analyzed by automated thermal desorption, gas chromatography and mass spectroscopy (ATD/GC/MS). A set of 130 VOCs, mainly benzene derivatives, naphthalene, and alkanes, were constantly detected exclusively in the *M. tuberculosis* cultured samples [134]. Following the same approach, breath VOCs from 42 patients, suspected of pulmonary TB, were analyzed. Accordingly, this method could distinguish between hospitalized patients (with suspected TB) and healthy controls with a 100% specificity and 100% sensitivity, and between TB culture positive and negative patients with a specificity of 78,9% and sensitivity of 95,7%. Using the characteristic VOCs, this approach was used to diagnose 226 patient collected sputum samples with a reported sensitivity of 84% and specificity of 64,7%, and a positive prediction value of 76%, when compared to smear microscopy results, 68% for sputum culture and 66% for chest X-rays [135]. This study shows the potential of metabolomics towards less invasive TB-diagnostic procedures and may, furthermore, assist in the diagnosis of children and HIV patients, who usually experience difficulty in supplying adequate sputum for current diagnostic procedures. Additionally, these procedures are relatively quick, taking merely a few hours to attain a diagnostic result, in contrast to the current gold standard, bacterial culture, which can take anything from 2-6 weeks [99].

The fact that the final outcome of alterations in the genome results in an altered metabolite profile, additionally makes metabolomics an excellent functional genomics tool [136]. In 2000, Fiehn *et al.* used a metabolomics approach to identify distinct metabolic profiles for four different *Arabidopsis* genotypes, including two homozygous ecotypes and a mutant of each [137]. Since then, several other studies have documented alterations in the metabolite profiles of a number of bacterial species, due to genetic perturbations [136]. Considering this, metabolomics may additionally be used to identify biomarkers specific to drug-resistant TB, contributing to a

better understanding of the underlying mechanisms of drug action and drug resistance, and additionally, potentially resulting in better MDR-TB diagnosis and treatments.

Furthermore, concentrations of newly identified metabolomics biomarkers may be used to monitor treatment outcomes, or may potentially be used for the early detection of relapses. This would enhance treatment strategies and may consequently prevent / lower the incidence of drug-resistance due to non-adherence. Metabolomics may additionally be used to test the performance of new anti-TB drugs using these biomarkers. An example of this is reported by Loots *et al.* (2005), using a metabolomics approach to investigate the effect of combined anti-TB drug therapy (using Rifater, a combination of rifampicin, isoniazid, and pyrazinamide) and melatonin (an antioxidant) on the organic acid and free radical profiles of rats. They indicated that Rifater treatment results in increased hydroxyl radicals and abnormal organic acids, characteristic of a multiple acyl-CoA dehydrogenase defect (MADD). Furthermore, co-administrating melatonin dramatically reverses these side effects [138]. This metabolomics approach was the first of its kind to indicate an altered metabolism due to potential electron-transport chain inhibition by anti-TB medication, and the prevention thereof using the antioxidant, melatonin. In a similar study, Huo *et al.* (2009) successfully differentiated serum metabolic profiles of type 2 diabetes mellitus patients treated with metformin hydrochloride from profiles of untreated patients. They, furthermore, identified a number of compounds, which were characteristic of metformin hydrochloride therapy, as potential, treatment-induced, biomarkers [139]. These studies prove the capacity of metabolomics as a tool to elucidate drug mechanisms and better treatment.

CONCLUSION

Considering the above information, it is clear that the still growing TB epidemic is fuelled by the inadequate performance of currently used TB treatment and diagnostic procedures [140]. Thus, the world is in urgent need of new, less toxic, faster acting TB treatment procedures alongside more sensitive, rapid TB diagnostic methods, which are able to cope with the growing incidence of MDR-TB and the co-existing HIV epidemic. The relatively new research field of metabolomics may lead to new biomarker identification and a holistic view of intra-host changes

during TB infection, active disease, and treatment. This added information may potentially open the door to a new era in TB research, diagnostics, and drug development. The identification of novel TB-biomarkers might be challenging, but should earnestly be pursued in an attempt to eradicate this worldwide pandemic.

REFERENCES

- [1] WHO. The Global Plan to Stop Tb 2011-2015: Transforming the fight towards elimination of tuberculosis. Geneva 27, Switzerland. (2010).
- [2] WHO (c). Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 Global Report on surveillance and response. Geneva, Switzerland. (2010).
- [3] U.S. Global Health Policy. The Global Tuberculosis Epidemic: Facts Sheet June 2010.
- [4] UNAIDS.: http://Data.Unaids.Org/Pub/Factsheet/2006/Tb_Hiv_Qa.Pdf. (2006).
- [5] Frieden, T.R., Sterling, T.R., Munsiff, S.S., Watt, C.J. and Dye, C.: *Lancet.*, 362:887–899 (2003).
- [6] American Thoracic Society And The Centers For Disease Control And Prevention.: *Am. J. Respir. Crit. Care Med.*, 161 (4 Pt 1):1376-1395 (1999).
- [7] Glickman, M.S. and Jacobs, W.R. Jr.: *Cell.*, 104:477–485 (2001).
- [8] Schluger, N.W. and Rom, W.N.: *Am. J. Respir. Crit. Care Med.*, 157:679–691 (1998).
- [9] Ellner, J.J.: *J. Infect. Dis.*, 176:1351-9 (1997).
- [10] Flynn, J.L., Chan, J., Triebold, K.J., Dalton, D.K., Stewart, T.A. and Bloom, B.R.: *J. Exp. Med.*, 178: 2249-2254 (1993).
- [11] Lobue, P.A. and Catanzaro, A.: *Ann. Emerg. Med.*, 43(4):185-246 (1997).
- [12] Murray, H.W., Tuazon, C.U., Kirmani, N. and Sheagren, J. N.: *Chest.*, 73:37-43 (1979).
- [13] Calmette, A.: *Proc. R. Soc. Med.*, 24(11):1481–1490 (1931).
- [14] Delogu, G. and Fadda, G.: *J. Infect. Developing Countries.* 3(1):5-15 (2009).
- [15] Colditz, G.A., Brewer, T.F., Berkey, C.S., Wilson, M.E., Burdick, E., Fineberg, H.V. and Mosteller, F.: *JAMA.*, 271(9):698-702 (1994).
- [16] Medical Research Council.: *Brit. Med. J.*, 2:769-782 (1948).
- [17] Hershfield, E.: *Can. Med. Assoc. J.*, 161:405-11 (1999).
- [18] Davies, G.R. *Tuberculosis.*, 90:171-176 (2010).
- [19] American Thoracic Society, Centers For Disease Control And Prevention, And The Infectious Diseases Society Of America.: *Am. J. Respir. Crit. Care Med.*, 167:603–662 (2003).
- [20] Waksman, S.A.: *Science.*, 118(3062):259-266 (1953).
- [21] Davies, J., Gorini, L. and Davis, B.D.: *Mol. Pharmacol.*, 1:93–106 (1965).
- [22] Zhang, Y.: *Annu. Rev. Pharmacol. Toxicol.*, 45:529–64 (2005).

- [23] Garvin, R.T., Biswas, D.K. and Gorini, L.: *Proc. Nat. Acad. Sci. USA.*, 71(10):3814-3818 (1974).
- [24] Klein, M. and Kimmelman, L.J.: *J. Bacteriol.*, 52: 471-479 (1946).
- [25] Honoré, N. and Cole, S.T.: *Antimicrob. Agents Chemother.*, 38:238-242 (1994).
- [26] Honoré, N., Marchal, G. and Cole, S.T.: *Antimicrob. Agents Chemother.*, 39:769-770 (1995).
- [27] Ogle, J.M. and Ramakrishnan, V.: *Annu. Rev. Biochem.*, 74: 129-177 (2005).
- [28] Montandon, P.E., Wagner, R. and Stutz, E.: *Embo. J.*, 5:3705-3708 (1986).
- [29] Pinard, R., Payant, C., Melancon, P. and Brakier-Gingras, L.: *Faseb. J.*, 7:173-176 (1993).
- [30] Okamoto, S., Tamaru, A., Nakajima, C., Nishimura, K., Tanaka, Y., Tokuyama, S., Suzuki, Y. and Ochi, K.: *Mol. Microbiol.*, 63(4):1096-1106 (2007).
- [31] Gillespie, S.H.: *Antimicrob. Agents Chemother.*, 46:267-274 (2002).
- [32] Ramaswamy, S.V., Dou, S.J., Rendon, A., Yang, Z., Cave, M.D. and Graviss, E.A.: *J. Med. Microbiol.*, 53:107-113 (2004).
- [33] Takayama, K., Wang, L. and David, H.L.: *Antimicrob. Agents Chemother.*, 2(1):29-35 (1972).
- [34] Morris, S., Bai, G.H., Suffys, P., Portillo-Gomez, L., Fairchok, M. and Rouse, D.: *J. Infect. Dis.*, 171:54-60 (1995).
- [35] Quemard, A., Lacave, C. and Lanéelle, G.: *Antimicrob. Agents Chemother.*, 35(6):1035-9 (1991).
- [36] Takayama, K., Wang, L. and Merikal, R.S.: *Antimicrob. Agents Chemother.*, 4(1):62-5 (1973).
- [37] Herman, R.P. and Webber, M.M.: *Antimicrob. Agents Ch.* 17:170-8 (1980).
- [38] Shoeb, H.A., Bowman, Jr B.U., Ottolenghi, A.C. and Merola, A.J.: *Antimicrob. Agents Chemother.*, 27(3):404-7 (1985).
- [39] Johnsson, K. and Schultz, P.G.: *J. Am. Chem. Soc.* 116:7425-7426 (1994).
- [40] Nguyen, M., Quémard, A., Broussy, S., Bernadou, J. and Meunier, B.: *Antimicrob. Agents Chemother.*, 46:2137-2144 (2002).
- [41] Dessen, A., Quémard, A., Blanchard, J.S., Jacobs Jr. W.R. and Sacchettini, J.C.: *Science.* 267:1638-1641 (1995).
- [42] Scior, T., Morales, I.M., Eisele, S.J.G., Domeyer, D. and Laufer, S.: *Arch. Pharm.* 335:511-525 (2002).
- [43] Ahmad, S. and Mokaddas, E.: *Resp. Med.*, 3:51-61 (2010).
- [44] Arbex, M.A., Varella, M.C.L., De Siqueira, H.R. and De Mello, F.A.F.: *J. Bras. Pneumol.*, 36(5):626-640 (2010).
- [45] Beggs, W. H. and Andrews, F.A.: *Antimicrob. Agents Chemother.*, 4:115-119 (1973).
- [46] Takayama, K. and Kilburn, J.O.: *Antimicrob. Agents Chemother.*, 33(9):1493-1499 (1989).
- [47] Silve, G., Valero-Guillen, P., Quemard, A., Dupont, M., Daffe, M. and Laneelle, G.: *Antimicrob. Agents Chemother.*, 37(7):1536-1538 (1993).
- [48] Kilburn, J. O., Takayama, K., Armstrong, E.L. and Greenberg, J.: *Antimicrob. Agents Chemother.*, 19(2):346-348 (1981).
- [49] Belanger, A. E., Besra, G.S., Ford, M.E., Mikušová, K., Belisle, J.T., Brennan, P.J. and Inamine, J.M.: *Proc. Natl. Acad. Sci. USA.*, 93:11919-11924 (1996).
- [50] Starks, A.M., Gumusboga, A., Plikaytis, B.B., Shinnick, T.M. and Posey, J.E.: *Antimicrob. Agents Ch.* 53(3):1061-1066 (2009).
- [51] Wehrli, W.: *Rev. Infect. Dis.*, 5(3):S407-S411 (1983).
- [52] Anthony, R.M., Schuitema, A.R.J., Bergval, I.L., Brown, T.J., Oskam, L. and Klatser P.R.: *Ann. Clin. Microbiol. Antimicrob.*, 4:9 (2005).
- [53] Bergval, I.L., Klatser, P.R., Schuitema, A.R.J., Oskam L., and Anthony, R.M.: *FEMS Microbiol. Lett.*, 275:338-343 (2007).
- [54] Heifets, L.B., Flory, M.A. and Lindholm-Levy, P.J.: *Antimicrob. Agents Ch.*, 33(8):1252-1254 (1989).
- [55] Hirano, K., Takahashi, M., Kazumi, Y., Fukasawa, Y. and Abe, C.: *Tubercle Lung Dis.*, 78(2):117-122 (1998).
- [56] Zhang, Y., Scorpio, A., Nikaido, H. and Sun, Z.: *J. Bacteriol.*, 181(7):2044-2049 (1999).
- [57] Zhang, Y. and Mitchison, D.: *Int. J. Tuberc. Lung Dis.*, 7(1):6-21 (2003).
- [58] Zhang, Y., Wade, M.M., Scorpio, A., Zhang, H. and Sun, Z.: *J. Antimicrob. Chemother.*, 52:790-795 (2003).
- [59] Strych, U., Penland, R.L., Jimenez, M., Krause, K.L. and Benedik, M.J.: *Fems Microbiol. Lett.* 196:93-98 (2001).
- [60] Walsh, C.T.: *The J. Biol. Chem.*, 264(5):2393-2396 (1989).
- [61] Caceres, N.E., Harris, N.B., Wellehan, J.F., Feng, Z., Kapur, V. and Barletta, R.G.: *J. Bacteriol.*, 179:5046-55 (1997).
- [62] Feng, Z. and Barletta R.G.: *Antimicrob. Agents Chemother.*, 47:283-91 (2003).
- [63] Chacon, O., Feng, Z., Harris, N.B., Caceres, N.E., Adams, L.G. and Barletta, R.G.: *Antimicrob. Agents Chemother.*, 46:47-54 (2002).
- [64] Belanger, A.E., Porter, J.C. and Hatfull, G.F.: *J. Bacteriol.*, 182:6854-56 (2000).
- [65] Arbex, M.A., Varella, M.C.L., De Siqueira, H.R. and De Mello, F.A.F.: *J. Bras. Pneumol.*, 36(5):641-656 (2010).
- [66] Thompson, L.T., Moskal, J.R. and Disterhoft, J.F.: *Nature.*, 359:638-641 (1992).
- [67] Wood, J. D., Peesker, S.J., Gorecki, D.K.J. and Tsui, D.: *Can. J. Physiol. Pharmacol.*, 52:62-68 (1978).
- [68] Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., Collins, D., De Lisle, G. and Jacobs, W. R.: *Science.*, 263:227-230 (1994).
- [69] Vannelli, T.A., Dykman, A. and Ortiz De Montellano P.R.: *J. Biol. Chem.*, 277(15):12824-12829 (2002).
- [70] Wilson, M., Derisi, J., Kristensen, H.-H., Imboden, P., Rane, S., Brown, P.O. and Schoolnik, G.K. 1999. *Proc. Natl. Acad. Sci. USA.*, 96(22):12833-12838.

- [71] Baulard Ar, Betts Jc, Engohang-Ndong J, Quan S, Mcadam Ra, Brennan, P.J., Loch, C. and Besra, G.S.: *J. Biol. Chem.*, 275(36):28326–28331 (2000).
- [72] Debarber, A.E., Mdluli, K., Bosman, M., Bekker, L.G. and Barry, C.E. 3rd. *Proc. Natl. Acad. Sci. USA.*, 97(17):9677–9682 (2000).
- [73] Morlock, G.P., Metchock, B., Sikes, D., Crawford, J.T. and Cooksey, R.C.: *Antimicrob. Agents Chemother.*, 47(12):3799–3805 (2003).
- [74] Krüüner, A., Jureen, P., Levina, K., Ghebremichael, S. and Hoffner, S.: *Antimicrob. Agents Chemother.*, 47(9):2971–2973 (2003).
- [76] Zaunbrecher, M.A., Sikes, R.D. Jr., Metchock, B., Shinnick, T.M. and Posey, J.E.: *PNAS*, 106(47):20004–20009 (2009).
- [77] Andersson, M.I. and Macgowan, A.P.: *J. Antimicrob. Chemother.*, 51(Suppl. 1):1–11 (2003).
- [78] Jacobs M.: *Drugs*, 58(Suppl. 2):19–22 (1999).
- [79] Leysen, D.C., Haemers, A. and Pattyn, S. R.: *Antimicrob. Agents Chemother.*, 33(1):1-5 (1989).
- [80] Ginsburg, A.S., Grosset, J.H. and Bishai, W.R.: *Lancet Infect. Dis.*, 3(7):432-42 (2003).
- [81] Di Perri, G. and Bonora, S.: *J. Antimicrob. Chemother.*, 54(3):593-602 (2004).
- [82] Tuberculosis Research Centre C.: *Indian J. Tuberc.*, 49:27–38 (2002).
- [83] Yee, D., Valiquette, C., Pelletier, M., Parisien, I., Rocher, I. and Menzies, D.: *Am. J. Respir. Crit. Care Med.*, 167(11):1472-1477 (2003).
- [84] WHO. Diagnostics for tuberculosis, global demand and market potential. Geneva, Switzerland. 33-49 (2006).
- [85] Al, Z.K., Al, J.H. and Menzies, D.: *Am J Resp. Crit. Care.*, 162:1419-22 (2009).
- [86] Pouchot, J., Grasland, A., Collet, C., Coste, J., Esdaile, J.M. and Vinceneux, P.: *Ann. Inter. Med.* 126:210-4 (1997).
- [87] Pai, M., Kalantri, S. and Dheda, K.: *Expert. Ref. Mol. Diagn.*, 6(3):413-422 (2006).
- [88] Ferrara, G., Losi, M., D'amico, R., Roversi, P., Piro, R., Meacci, M., Meccugni, B., Dori, I.M., Andreani, A., Bergamini, B.M., Mussini, C., Rumpianesi, F., Fabbri, L.M. and Richeldi, L.: *Lancet.*, 367:1328–1334 (2006).
- [89] Tattevin, P., Casalino, E., Fleury, L., Egmann, G., Ruel, M. and Bouvet, E.: *Chest.*, 115(5):1248-1253 (1999).
- [90] Nakamura K., Ohmi, A., Kurihara, T., Suzuki, S. and Tadera M.: *Kekkaku.* 45(4):121-128 (1970).
- [91] Van Cleeff, M.R.A., Kivihya-Ndugga, L.E., Meme, H., Odhiambo, J.A. and Klatser, P.R.: *BMC Infect. Dis.*, 5:111 (2005).
- [92] Arslan, S., Özdemir, L., Demirel, Y. and Akkurt, I.: *Afr. J. Microbiol. Res.*, 4(8):613-617 (2010).
- [93] Drobniewski, F.A., Caws, M., Gibson, A. and Young, D.: *Lancet Infect. Dis.* 3:141–47 (2003).
- [94] Ruiz-Manzano, J., Blanquer, R., Calpe, J.L., Caminero, J.A., Caylà, J., Domínguez, J.A., García, J.M. and Vidal, R.: *Arch Bronconeumol.*, 44(10):551-66 (2008).
- [95] Kox, LFF: *Resp. Med.*, 89:399-408 (1995).
- [96] Urbanczik, R. 1985.: *Zbl. Bakt. Hyg. A.*, 260:81-87 (1985).
- [97] Lee, J.J., Suo, J., Lin, C.B., Wang, J.D., Lin, T.Y. and Tsai, Y.C.: *Int. J. Tuberc. Lung Dis.*: 7(6):569–574 (2003).
- [98] Getahun H., Harrington M., O'brien R. and Nunn P.: *Lancet.* 369:2042-2049 (2007).
- [99] Hannan, A., Chaudhary, S., Saleem, S., Qayyum, A. and Arshad, M.U.: *J. Ayub. Med. Coll. Abbottabad.*, 20(4):67-69 (2008).
- [100] Hanna, B.A., Ebrahimzadeh, A., Elliott, L.B., Morgan, M.A., Novak, A.M., Rusch-Gerdes, S., Acio, M., Dunbar, D.F., Holmes, T.M., Rexer, C.H. and Savthyakumar, C.: *J. Clin. Microbiol.*, 37(3): 748–752 (1999).
- [101] Cruciani, M., Scarpato, C., Malena, M., Bosco, O., Serpelloni, G. and Mengoli, C.: *J. Clin. Microbiol.*, 42(5):2321–2325 (2004).
- [102] Scarpato, C., Ricordi, P., Ruggiero, G. and Piccoli, P.: *J. Clin. Microbiol.*, 42(3):1109–1114 (2004).
- [103] Tortoli, E., Benedetti, M., Fontanelli, A. and Simonetti M.T.: *J. Clin. Microbiol.*, 40(2): 607–610 (2002).
- [104] Caviedes, L., Lee, T., Gilman, R.H., Sheen, P., Spellman, E., Lee, E.H., Berg, D.E., Montenegro-James, S. and The Tuberculosis Working Group In Peru.: *J. Clin. Microbiol.*, 38(3):1203-1208 (2000).
- [105] Palomino, J.C., Martin, A. and Portaels, F.: *Clin. Microbiol. Infect.*, 13:754–762 (2007).
- [106] Martin, A., Panaioto, S., Portaels, F., Hoffner, S., Palomino, J.C. and Angeby, K.: *J. Antimicrob. Chemoth.*, 65:56–64 (2008).
- [107] Butler, W.J. and Kilburn, J.O.: *J. Clin. Microbiol.*, 26(1):50-53 (1988).
- [108] Pai, M., Kalantri, S. and Dheda, K.: *Expert. Ref. Mol. Diagn.*, 6(3):423-432 (2006).
- [109] Catanzaro, A. Perry, S., Clarridge, J.E., Dunbar, S., Goodnight-White, S., Lobue, P.A., Peter, C., Pfyffer, G.E., Sierra, M.F., Weber, R., Woods, G., Mathews, G., Jonas, V., Smith, K. and Della-Latta, P.: *JAMA.*, 283(5) (2000).
- [110] Pounder, J.I., Anderson, C.M., Voelkerding, K.V. and Salfinger, M.: *Diagn. Micr. Infec. Dis.*, 67:101–105 (2010).
- [111] Ahmad, S., Mokaddas, E. and Jaber, A.-A.: *Mol. Cell. Probe.*, 18:299–306 (2004).
- [112] Cockerill, F.R. III and Uhl, J.R., Temesgen, Z., Zhang, Y., Stockman, L., Roberts, G.D., Williams D.L. and Kline, B.C.: *J. Infect. Dis.*, 171:240-5 (1995).
- [113] Espasa, M., González-Martín, J., Alcaide, F., Aragón, L.M., Lonca, J., Manterola, J.M., Salvadó, M., Tudó, G., Orús, P. and Coll, P.: *J. Antimicrob. Chemoth.*, 55:860–865 (2005).
- [114] Perkins, M.D., Conde, M. B., Martins, M. and Kritski, A.L.: *Chest.* 123:107-112 (2003).
- [115] Muzaffar, R., Batool, S., Aziz, F., Naqvi, A. and Rizvi, A.: *Int. J. Tuberc. Lung. Dis.*, 6(7):635–640 (2002).

- [116] Albert, H., Heydenrych, A., Brookes, R., Mole, R.J., Harley, B., Subotsky, E., Henry, R. and Azevedo, V.: *Int. J. Tuberc. Lung. Dis.*, 6(6):529–537 (2002).
- [117] Dunn, W.B., Bailey, N.J.C. and Johnson, H.E.: *Analyst.*, 130:606–625 (2005).
- [118] Parida, S.K. and Kaufmann, H.E.: *Drug Discov. Today.*, 15(3/4):148–157 (2010).
- [119] Brindle, J., Antti, H., Holmes, E., Tranter¹, G., Nicholson, J.K., Bethell, H.W.L., Clarke, S., Schofield, P.M., Mckilligin, E., Mosedale, D.E. and Grainger, D.J.: *Nat. Med.*, 8(12):1439–1444 (2002).
- [120] Yang, J., Xu, G., Honga, Q., Liebich, H.M., Lutz, K., Schmülling, R.-M. and Wahl, H.G.: *J. Chrom. B.*, 813:53–58 (2004).
- [121] Odunsi, K., Wollman, R.M., Ambrosone, C.B., Hutson, A., Mccann, S.E., Tammela, J., Geisler, J.P., Miller, G., Sellers, T., Cliby, W., Qian, F., Keitz, B., Intengan, M., Lele, S. and Alderfer, J.L.: *Int. J. Cancer.*, 113:782–788 (2005).
- [122] Underwood, U., Broadhurst, D., Dunn, W.B., Ellis, D.I., Michell, A.W., Vacher, C., Mosedale, D.E., Kell, D.B., Barker, R.A., Grainger, D.J. and Rubinsztein, D.C.: *Brain.*, 129:877–886 (2006).
- [123] Brindle, J.T., Nicholson, J.K., Schofield, P.M., Grainger, D.J. and Holmes, E.: *Analyst.*, 128:32–36 (2003).
- [124] Yang, J., Xu, G., Zheng, Y., Kong, H., Pang, T., Lva, S. and Yang, Q.: *J. Chrom. B.*, 813:59–65 (2004).
- [125] Yu, K., Sheng, G., Sheng, J., Chen, Y., Xu, W., Liu, X., Cao, H., Qu, H., Cheng, Y. and Li, L.: *J. Proteome Res.*, 6:2413–2419 (2007).
- [126] Coen, M., O'sullivan, M., Bubb, W.A., Kuchel, P.W. and Sorrell, T.: *Clin. Infect. Dis.*, 41:1582–90 (2005).
- [127] Ahmed, Sssj., Santosh, W., Kumar, S. and Christlet, H.T.T.: *J Biomed. Sci.*, 16:63–75 (2009).
- [128] Bogdanov, M., Matson, W.R., Wang, L., Matson, T., Saunders-Pullman, R., Bressman, S.S. and Beal, M.F.: *Brain.*, 131:389–396 (2008).
- [129] Dunn, W.B., Brownb, M., Worton, S.A., Crocker, I.P., Broadhurst, D., Horgan, R. Kenny, L.J., Baker, P/ N., Kell, D.B. and Heazell, A.E.P.: *Placenta.*, 30:974–980 (2009).
- [130] Kenny, L.C., Broadhurst, D., Brown, M., Dunn, W.B., Redman, C.W.G., Kell, D.B. and Baker, R.N.: *Reprod. Sci.* 15(6):591–597. (2008).
- [131] Kaddurah-Daouk, R.: *Plos Med.*, 3(8):1222–1223 (2006).
- [132] Pavlou, A.K., Magan N., Jones, J.M., Brown, J., Klatser, P. and A.P.F. Turner.: *Biosens. Bioelectron.* 20:538–544 (2004).
- [133] Fend, R., Kolk, A.H., Bessant, C., Buijtel, P., Klatser, P.R. and Woodman, A.C.: *J. Clin. Microbiol.*, 44:2039–2045 (2006).
- [134] Phillips, M., Basa-Dalay, V., Bothamley, G., Cataneo, R.N., Lam, P.K., Natividad, M.P.R., Schmitt, P. and Wai, J.: *Tuberculosis.*, 90:145–151 (2010).
- [135] Phillips, M., Cataneo, R.N., Condos, R., Erickson, G.A.R., Greenberg, J., La Bombardi, V., Munawar, M.I. and Tietje, O.: *Tuberculosis.*, 87:44–52 (2007).
- [136] Baran, R., W. Reindl and T.R. Northen. *Curr. Opin. Microbiol.*, 12:547–552 (2009).
- [137] Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. 2000. Metabolite Profiling For Plant Functional Genomics. *Nature Biotechnol.*, 18:1157–1161.
- [138] Loots, D.T., Wiid, I.J., Page, B.J., Mienie, L.J. and Van Helden P.D.: *J. Pineal Res.*, 38(2):100–106 (2005).
- [139] Huo, T., Cai, S., Lu, X., Sha, Y., Yu, M. and Li, F.: *J. Pharmaceut. Biomed.*, 49:976–982 (2009).
- [140] Perkins, M.D. and Small, P.M.: *Int. J. Tuberc. Lung Dis.*, 10(1):1 (2006).

APPENDIX A (2)

Manuscripts

OLIVIER, I., LOOTS, D.T. 2012. A metabolomics approach to characterise and identify various *Mycobacterium* species. *J. Microbiol. Meth.* 88:419-426.



A metabolomics approach to characterise and identify various *Mycobacterium* species

Ilse Olivier, Du Toit Loots *

School for Physical and Chemical Sciences, Centre for Human Metabonomics, North-West University, Potchefstroom, 2520, South Africa

ARTICLE INFO

Article history:

Received 13 September 2011

Received in revised form 13 January 2012

Accepted 18 January 2012

Available online 24 January 2012

Keywords:

Chemometrics

GC–MS

Metabolomics

Tuberculosis

ABSTRACT

We investigated the potential use of gas chromatography mass spectrometry (GC–MS), in combination with multivariate statistical data processing, to build a model for the classification of various tuberculosis (TB) causing, and non-TB *Mycobacterium* species, on the basis of their characteristic metabolite profiles. A modified Bligh–Dyer extraction procedure was used to extract lipid components from *Mycobacterium tuberculosis*, *Mycobacterium avium*, *Mycobacterium bovis*, and *Mycobacterium kansasii* cultures. Principle component analyses (PCA) of the GC–MS generated data showed a clear differentiation between all the *Mycobacterium* species tested. Subsequently, the 12 compounds best describing the variation between the sample groups were identified as potential metabolite markers, using PCA and partial least-squares discriminant analysis (PLS-DA). These metabolite markers were then used to build a discriminant classification model based on Bayes' theorem, in conjunction with multivariate kernel density estimation. This model subsequently correctly classified 2 “unknown” samples for each of the *Mycobacterium* species analysed, with probabilities ranging from 72 to 100%. Furthermore, *Mycobacterium* species classification could be achieved in less than 16 h, and the detection limit for this approach was 1×10^3 bacteria mL^{-1} . This study proves the capacity of a GC–MS, metabolomics pattern recognition approach for its possible use in TB diagnostics and disease characterisation.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In 2008, there were an estimated 9.4 million new tuberculosis (TB) cases and 1.8 million TB related deaths reported globally, with the highest incidence rate recorded in sub-Saharan Africa (WHO, 2010). Additionally, a large proportion of TB causing mycobacteria, other than *Mycobacterium tuberculosis* (e.g. *Mycobacterium bovis*), and non-TB causing *Mycobacterium* (e.g. *Mycobacterium avium*, *Mycobacterium kansasii* etc.), are still encountered frequently in clinical specimens (Míguez-Burbano et al., 2006; Olivier, 1998). The ability of a diagnostic technique, not only to detect TB, but also to differentiate between the various *Mycobacterium* species, is of great importance, as different treatment regimes are followed for each of the above mentioned mycobacterial infections (Grange, 2001). Additionally, the detection limits of these techniques are also of particular significance, especially when one considers the low bacterial loads usually associated with samples collected from patients with early TB infection, TB/HIV

co-infection, and from TB infected children (Getahun et al., 2007; Marais and Pai, 2007; WHO, 2006).

The current diagnostic gold standard for TB diagnostics, bacteriological culture, is considered extremely sensitive, with a reported detection limit of only $10\text{--}100$ bacteria mL^{-1} when using solid or liquid culture medium (Getahun et al., 2007; WHO, 2006). Two of the major disadvantages of this TB diagnostic method is: 1) that 15–20% of all adult TB cases are falsely reported as negative (Frieden et al., 2003; Getahun et al., 2007) and 2) the fact that 2–6 weeks are required before bacterial colonies are visually detected (WHO, 2006), leading to an unnecessary delay prior to the patient commencing treatment. Additionally, because immense anti-contamination procedures are required for culturing, 1–4% of all diagnosed cases are false positives (Getahun et al., 2007). A major improvement in mycobacterial culture diagnostics was the development of automated culture systems, such as BACTEC 460, which is able to identify and differentiate between clinically important *Mycobacterium* species within 15 days (Cruciani et al., 2004). These systems are, however, rather costly and require expensive maintenance and infrastructure, in addition to highly trained and experienced personnel (WHO, 2006), limiting its implementation in developing countries such as South Africa.

As summarised by the World Health Organizations (WHO) report on TB diagnostics (WHO, 2006), smear microscopy is still the TB diagnostic method most commonly used worldwide, due to its simplicity and affordability. Nevertheless, this technique has an extremely poor detection limit of $5000\text{--}10000$ bacteria mL^{-1} (Kox, 1995; WHO, 2006), and is consequently only able to detect 60–70%

Abbreviations: AMDIS, Automated Mass Spectral Deconvolution and Identification System; ATD–GC–MS, automated thermal desorption–gas chromatography mass spectrometry; DNA, deoxyribonucleic acid; GC–MS, gas chromatography mass spectrometry; NAA, nucleic acid amplification; PC, principle component; PCA, principle component analyses; PCR, polymerase chain reaction; PLS-DA, partial least-squares discriminant analysis; TB, tuberculosis; VIP, variable influence on the projection.

* Corresponding author. Tel.: +27 18 299 1818; fax: +27 18 299 2316.

E-mail address: dutoit.loots@nwu.ac.za (D.T. Loots).

of all adult, culture-confirmed TB cases, even when done with extreme care (Colebunders and Bastian, 2000). The number of identified TB cases can, however, be as low as 20–30%, when this method is performed under sub-optimal conditions (Urbanczik, 1985; WHO, 2006). Additionally, smear microscopy is also incapable of differentiating between the various *Mycobacterium* species responsible for the infection (Ruiz-Manzano et al., 2008).

A more recent and still rapidly improving TB diagnostic strategy, nucleic acid amplification (NAA), claims to detect TB infection using less than 10 bacteria mL⁻¹ sputum, however, with a reported sensitivity of only 60–70% in smear-negative, culture-positive samples (Moore and Curry, 1995; WHO, 2006). Pounder et al. (2010) developed a genomic deletion assay based on a multiplex polymerase chain reaction (PCR) with melting temperature analysis, which correctly identified 96% of the six clinically important *Mycobacterium* species from pure culture samples. However, a reselection of the target sequence may, in future, be required, since sequence variation may occur at primer binding sites due to the evolution of these organisms. Furthermore, as a result of the occurrence of PCR inhibitors in sputum samples, pure cultures are still the preferred source of genomic DNA for NAA assays, hence, long culturing times are still a limitation to these methods (Ahmad et al., 2004). An additional restriction associated with these assays is the high incidence of false positive results (Noordhoek et al., 1994) and, furthermore, due to the presence of residual genetic material, positive results may still be obtained after effective treatment and these assays can thus not be used in patients with previous *Mycobacterium* infections, or for the monitoring treatment response (Frieden et al., 2003).

Using HPLC, the MIDI Research and Development Laboratory developed pattern recognition software (The Sherlock Mycobacteria Identification System, MYCO-LCS) for the identification of a number of various mycobacterial species, on the basis their varying cellular mycolic acid profiles (MIDI, inc., 2009). This approach however, cannot distinguish between *M. tuberculosis* and *M. bovis*, and requires at least 1 × 10⁵ cultured cells in order to achieve a successful diagnostic outcome (ATC and CDC, 1999).

The metabolome is the final downstream result of genome transcription, and may be defined as a collection of all the metabolites (small molecular compounds) present in a specific cell or organism, participating in metabolic reactions during normal cell function, growth and maintenance. Using specialised software and multivariate statistical analyses, these metabolite profiles can be compared, in order to differentiate between the various biological systems investigated, such as the *Mycobacterium* species used in this study, and identify potential metabolite markers characteristic of these systems (Dunn et al., 2005; Olivier and Loots, 2011). This relatively new research field termed 'metabolomics' has successfully been used over the past few years to characterise a variety of disease states (Schoeman and Loots, 2011). Using an electronic nose (gas sensor array), and following a metabolomics approach, Fend et al. (2006) were able to differentiate between various *Mycobacterium* species and *Pseudomonas aeruginosa*, by analysing the headspace of both pure cultures and culture spiked control sputum, at a detection limit of 1 × 10⁴ bacteria mL⁻¹. This study proved the viability of metabolomics as a tool for TB diagnosis. However, as the nature by which these sensors function is not yet fully understood, the identity and quantities of the differentiating volatiles cannot be determined, which may be seen as a limitation to using this sensor array technology.

Building on these results, we investigated the capacity of GC–MS metabolomics, using chemometrics data analyses for identifying pattern differences, to differentiate between various cultured *Mycobacterium* species, based on their cellular fatty acid profiles. Various previous studies have used gas chromatography as a means to identify a variety of *Mycobacterium* species based on their characteristic lipid constituents (Jantzen et al., 1989; Lambert et al., 1986; Mosca et al., 2007). The current study is, however, the first of

its kind to use a true GC–MS, metabolomics research approach to 1) identify lipid markers characterising *Mycobacterium* species, 2) build a fully automated, multivariate statistical diagnostic model using these markers and 3) prove that this approach has the capacity to repeatedly detect these markers over time.

2. Materials and methods

2.1. Reagents and chemicals

Potassium hydroxide and nonadecanoic acid (C19:0) were obtained from Merck (Darmstadt, Germany). All organic solvents used were ultra pure Burdick & Jackson brands (Honeywell International Inc., Muskegon, MI, USA) and were used without further purification.

2.2. Sample collection and preparation

All organisms used in this study were obtained from the Royal Tropical Institute, Amsterdam, The Netherlands. All bacteria (*M. tuberculosis*, *M. avium*, *M. bovis*, and *M. kansasii*) were cultured (originally from patient sputum samples) in Middlebrook 7H9 medium with oleic acid–albumin–dextrose–catalase enrichment. The bacteria were incubated at 37 °C while shaking at 120–150 rpm until an optical density (420 nm) of 0.30 (≈ 2 × 10⁸ bacteria mL⁻¹) was reached. The cells were washed once with PBS, collected via centrifugation, killed in a water bath at 110 °C and stored at –80 °C until extraction and GC–MS metabolomics analysis.

For the purpose of investigating the differentiation capacity and potential diagnostic ability of this approach, 10 cultured samples of each of the 4 infectious bacterial species groups (*M. tuberculosis*, *M. avium*, *M. bovis*, and *M. kansasii*), were prepared at a concentration of 1 × 10⁸ bacteria mL⁻¹ in ddH₂O, prior to lipid extraction, GC–MS analyses and chemometrics data processing. Two samples of each of the bacterial species groups, grown from different patient sputum samples, were analysed in the same manner, 6 months later, and used to validate the metabolite pattern recognition model built using the prior 10 samples from each of the 4 groups.

In order to determine the potential detection limit of this metabolomics method, 6 repeats of a *M. tuberculosis* culture sample was re-suspended in ddH₂O, at various dilutions, in order to create a concentration gradient ranging from 1 × 10² to 1 × 10⁸ bacteria mL⁻¹, including a blank (ddH₂O only), and subsequently analysed. In order to verify the differentiation capacity of this method at the determined detection limit, the 10 cultured samples of all 4 of the selected *Mycobacterium* species groups, were once again extracted and analysed, only this time the samples were diluted to the concentration identified as the detection limit. The concentrations of the various bacterial pre-extraction sample solutions were determined spectrophotometrically by using the optical density measurements at 420 nm together with McFarland standards.

2.3. Extraction procedure

A modified Bligh–Dyer extraction procedure (Rezwan et al., 2007) was used to extract the lipid components of these cultured samples. As internal standard, 100 µL of C19:0 was added to 1 mL of the bacterial suspension, followed by the addition of 2 mL chloroform, 4 mL methanol, and 1.6 mL ddH₂O. The sample mixture was then sonicated for 2 min, vortexed for 30 s, and allowed to stand for 2 h at room temperature, in order to facilitate phase separation. Chloroform and ddH₂O (2 mL of each) were again added and the sample was centrifuged for 15 min at 550 × g. The organic phase was collected and the water phase was re-extracted with 2 mL of chloroform. The combined organic phase was washed with 5 mL ddH₂O, and once again left to stand for 2 h at room temperature, in order to facilitate phase separation, followed by centrifugation for 20 min at 550 × g. The

newly formed organic phase was collected and dried under nitrogen. Subsequently, in order to methylate these extracts, 0.5 mL chloroform, 0.5 mL methanol, and 1 mL methanolic KOH (0.2 M) were added. The mixture was incubated for 30 min at 60 °C followed by the addition of 2 mL hexane, 200 µL glacial acetic acid (1 M), and 2 mL ddH₂O. The organic phase was again collected and transferred to a new tube. The water phase was re-extracted 3 times with hexane. The combined organic phases from each extraction were dried under nitrogen and re-suspended in 100 µL hexane prior to injection.

2.4. GC–MS parameters

Analyses were done in the splitless mode by injecting 1 µL of the extract on an Agilent 7890A gas chromatograph coupled to an Agilent 5975 mass selective detector equipped with a 7683B injector, 7683 auto sampler, and VF1-MS capillary column (30 m × 250 µm i.d., 0.25 µm film thickness) purchased from Agilent (Atlanta, GA). The injector temperature was held constant at 250 °C for the entire run time. The initial GC oven temperature was held at 50 °C for 1 min after injection, followed by an increase of 10 °C min^{−1} to 240 °C. The oven temperature was then increased at a rate of 20 °C min^{−1} to a final temperature of 300 °C and maintained for 7 min. Helium was used as the carrier gas and pressure programmed as such that the helium flow was kept constant at 1.2 mL min^{−1}. Detection was achieved by using MS detection in full scan mode (*m/z* 50–550).

2.5. Data-acquisition

AMDIS software (Automated Mass Spectral Deconvolution and Identification System, version 2.65) was used to deconvolute and analyse the raw GC–MS data at a threshold of 0.01% of the total signal. A new library was created, compiled from the mass spectra of all the compounds detected in each of the 4 bacterial groups. Each compound was given a code corresponding to the retention time at which it was first detected. This library was manually inspected and duplicate compound mass spectra identified by AMDIS due to GC–MS spectral shifting, were deleted. Subsequently, each sample was analysed using this library containing the mass spectra of all possible compounds present in the study cohort for the purpose of creating an output file containing the detected peak areas of the identified compounds, as well as the compounds in the library not detected for each sample. This data were used to compile a data matrix comprising of each compound, normalised using the internal standard, for each of the analysed samples of the *Mycobacterium* species groups. This data matrix was then subjected to multivariate statistical data processing in order to develop a model capable of identifying these species, using their characteristic metabolite profiles.

2.6. Statistical data analyses

The statistical packages, “R” (version 2.13.0) and Statistica (version 10) were used for all the statistical data analyses. To lessen the weight of the compounds present in elevated concentrations, data were scaled using a non-parametric transformation function (Koekemoer and Swanepoel, 2008), after which mean centering was applied. Principle component analyses (PCA) were performed to determine whether or not a natural grouping exists between the various sample groups. PCA involves a mathematical procedure that transforms a number of possibly related variables (in this case metabolites) into a smaller number of unrelated variables known as principal components (PC's). PC 1 accounts for the most variance in the data and each subsequent PC (PC 2, PC 3 etc.) accounts for the next highest variance of the remaining data. Using one PC per axis, the PCA can then be visualized as a scoresplot. Using the data generated from the species differentiation analyses at a concentration

of 1×10^8 bacteria mL^{−1}, the metabolites were then ranked according to their respective modelling powers.

Additionally, a partial least square discriminant analysis (PLS-DA) model was built in order to identify those compounds that contribute most to the separation between the sample groups, by ranking the compounds according to the variable influence on the projection (VIP) parameter. VIP is a weighted sum of squares of the PLS-DA weights, indicating the importance of the metabolite to the total model.

To reduce the dataset into a set of potentially relevant metabolite markers, we created a combined list of the markers with the highest modelling powers (PCA) and VIP values (PLS-DA). Using the mass fragmentation patterns generated by the MS, together with their respective GC retention times, the identities of these metabolite markers were determined using libraries generated from previously injected standards. In order to test this approach's potential diagnostic capacity, using only the identified metabolite markers, a discriminant model based on Bayes' theorem, in conjunction with multivariate kernel density estimation, using the first three PCs as input, was built. This model was developed for the purpose of estimating the class membership probabilities of an unknown bacterial sample based on the presence of the aforementioned metabolite markers. To validate the model, two new, “unknown” sputum cultures for each of the above mentioned bacterial groups, derived from patients previously diagnosed with these infectious organisms, were used in order to determine if the model would correctly identify or allocate these to their respective groups.

In order to determine the potential detection limit of this method, a PCA was performed on the data obtained after GC–MS analyses of the *M. tuberculosis* concentration gradient, as described in Section 2.2. The sample group with the lowest bacterial concentration, not overlapping with the blank sample, was regarded as the detection limit. To test the capacity of this method to still discriminate between the *Mycobacterium* species at the detection limit of this approach, a PCA was once again performed, using data collected from the analyses of the 4 bacterial groups, prepared at a concentration determined by the proposed detection limit for *M. tuberculosis*.

3. Results and discussion

Although the Bligh and Dyer extraction method (water-methanol-chloroform) (Rezwan et al., 2007) and the synthesis of fatty acid methyl esters (FAME) (Mosca et al., 2007; Lambert et al., 1986) are well-known procedures used for extracting and derivatising lipids from various biological samples, before the analyses and metabolomics data processing of the collected mycobacterial diagnostic samples, the authors evaluated the above mentioned extraction approach for metabolomics purposes by initially processing 4 × 1 mL of the isolated *M. tuberculosis* sample repeats, re-suspended in ddH₂O at a concentration of 1×10^8 bacteria mL^{−1}. The data collected from the GC–MS analyses of these repeats indicated a coefficient of variation (CV) value for the determined internal standard area to be 4.9%, and the CV values for the major fatty acids detected below the 20% CV cut-off recommended by the FDA for target compound analyses (t'Kindt et al., 2009).

3.1. Bacterial lipid content

Considering the compounds detected after extraction and GC–MS analysis of the 10 cultured samples of each of the 4 *Mycobacterium* species, 347 mass spectra were added to the library. The majority of the mass spectra in this library were identified as fatty acid methyl esters, however, the unmethylated form of these compounds will subsequently be discussed, as this represents the compounds present in these bacterial samples in their native state. Fig. 1 represents examples of the total ion chromatograms obtained after GC analyses

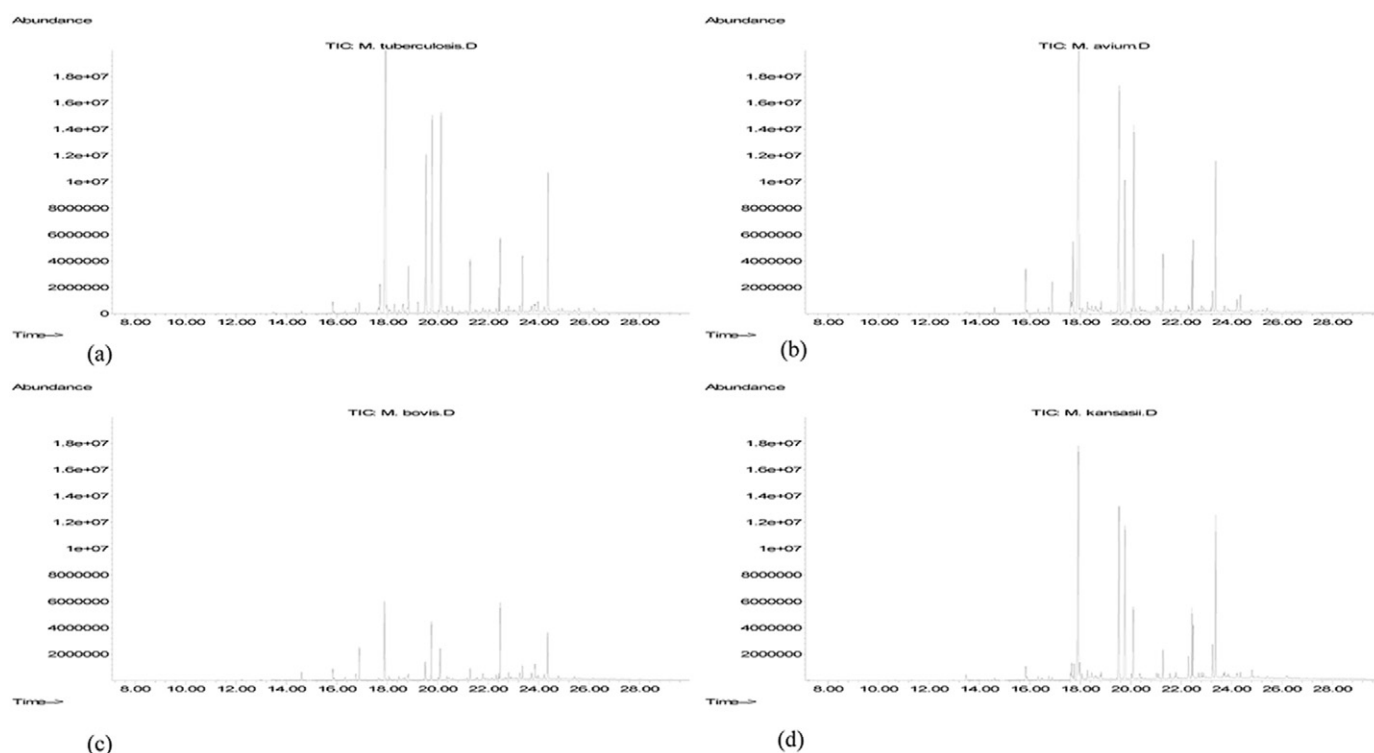


Fig. 1. Example of the total ion chromatograms obtained after chromatographic analyses of the four *Mycobacterium* species investigated. (a) represents *M. tuberculosis*, (b) *M. avium*, (c) *M. bovis*, and (d) *M. kansasii*.

of each of the four *Mycobacterium* species investigated, showing a clear visual difference between the species.

The most abundant compounds detected in each of the species groups are given in Table 1.

Corresponding to that previously determined (Lambert et al., 1986; Mosca et al., 2007), the most abundant fatty acids detected in the *Mycobacterium* species samples were hexadecanoic acid (C16:0) and oleic acid (C18:1 ω9c). Additionally, as can be expected, we detected the well-known *Mycobacterium* biomarker, tuberculostearic acid (TBSA) (Lambert et al., 1986; Larsson et al., 1987; Mayakova et al., 1995; Mosca et al., 2007; Odham et al., 1979; Stopforth et al., 2004) as one of the major compounds in all *Mycobacterium* species

groups. Other shorter-chain fatty acids detected in relatively high concentrations, included: tetradecanoic acid (C14:0), palmitoleic acid (C16:1 ω7c), and heptadecanoic acid (C17:0), which are consistent with previous findings (Lambert et al., 1986; Luquin et al., 1991). Long-chain fatty acids including: eicosanoic acid (C20:0), docosanoic acid (C22:0), tetracosanoic acid (C24:0), and hexacosanoic acid (C26:0) were also identified as part of the most abundant lipid components using this approach. According to Guerrant et al. (1981), these fatty acids are formed as a result of heat cleavage of the mycolic acids in the injection port of the gas chromatograph at injector temperatures exceeding 235 °C, as was the case in our study. Mycolic acids are unique, high-molecular-weight, α-alkyl, β-hydroxy, very-long fatty acids (C70–

Table 1

Mean concentrations of the most abundant lipid components found in the various *Mycobacterium* species. Standard deviations are given in brackets.

Species	RT (min)	m/z	<i>M. tuberculosis</i>	<i>M. avium</i>	<i>M. bovis</i>	<i>M. kansasii</i>
Compound ^a			Concentration (μg mg ⁻¹ sample)			
C16:0	17.9213	270	234.8 (6.3)	457.0 (13.9)	178.7 (57.0)	262.5 (9.3)
C18:1 ω9c	19.5315	297	121.9 (3.7)	275.3 (91.6)	49.7 (9.2)	152.1 (48.3)
TBSA	20.0940	312	134.6 (0.7)	203.5 (9.4)	69.6 (14.5)	52.0 (2.1)
C16:1 ω7c	17.7062	268	14.9 (0.6)	60.4 (2.4)	2.3 (0.3)	12.1 (0.7)
BPA	17.9579	292	7.4 (0.1)	8.1 (3.5)	6.2 (4.9)	13.5 (3.3)
C17:0	18.8380	284	23.0 (0.8)	8.6 (0.8)	0.2 (0.6)	T
Hexadecane	14.6032	226	2.0 (0.6)	5.6 (0.6)	21.2 (7.4)	3.4 (1.0)
C26:0	24.3723	410	85.6 (2.5)	12.3 (4.1)	103.2 (20.4)	3.9 (0.2)
C24:0	23.3584	382	23.3 (0.8)	116.9 (4.1)	23.8 (3.5)	102.8 (6.0)
C20:0	21.2870	326	23.0 (0.6)	7.3 (6.2)	20.8 (3.1)	17.7 (0.7)
C14:0	15.8386	242	6.4 (2.2)	37.5 (2.4)	28.7 (5.8)	14.2 (2.0)
C22:0	22.4286	254	9.3 (0.2)	34.9 (0.9)	9.9 (1.1)	36.1 (1.9)
Isopropyl-C14:0	16.8897	270	0	23.0 (6.2)	85.1 (10.4)	0
C24:1 ω9c	23.2519	380	0	13.3 (1.6)	0	21.5 (1.1)
Squalene	23.8618	410	0.3 (0.8)	1.7 (0.8)	32.2 (25.1)	1.4 (0.6)
C22:1 ω9c	22.2917	352	0	8.1 (0.3)	T	14.7 (0.7)

^a TBSA, tuberculostearic acid. BPA, benzenepropanoic acid. T, trace amount.

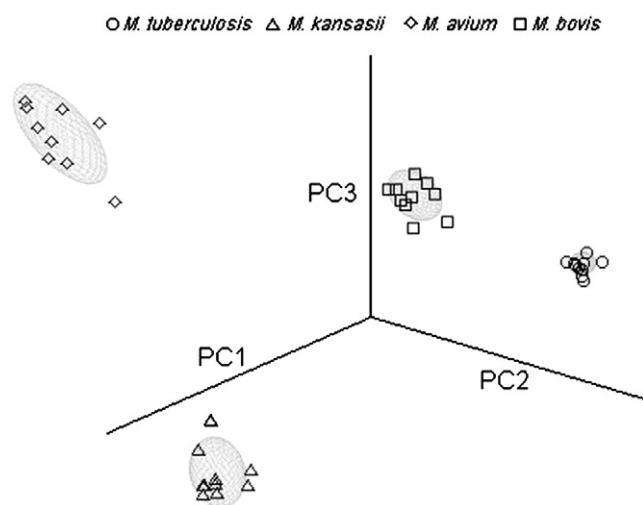


Fig. 2. Three-dimensional PCA scoresplot showing a clear differentiation between *Mycobacterium* sample groups at a concentration of 1×10^8 bacteria mL^{-1} .

C90), representing 40–60% of the dry cell weight of *Mycobacterium* and *Corynebacterium* species (Brennan, 2003). The concentrations of these mycolic acid derived products are known to vary consistently among different *Mycobacterium* species (Knisley et al., 1985; Lambert et al., 1986; Luquin et al., 1991). Tetracosanoic acid (C24:0), for instance, was detected in higher amounts in *M. avium* and *M. kansasii*, and hexacosanoic acid (C26:0) was detected in comparatively higher amounts in *M. tuberculosis* and *M. bovis*. Furthermore, we also detected minor concentrations of C26:0 in *M. kansasii*, which to date has not yet been reported, and isopropyl tetradecanoic acid (isopropyl-C14:0) was seen in higher amounts in *M. bovis* and in minor concentrations in *M. avium*. Lastly, the alkanes detected in major concentrations (hexadecane and heneicosane) reflect the ability of this method to additionally extract these compounds.

3.2. *Mycobacterium* species differentiation

Prior to multivariate statistical data analyses, a 50% filter was applied to the data, excluding those compounds which do not appear in at least 50% of the samples, in one or more of the sample groups. This filter, together with the removal of compounds with no detected variation between the groups, led a reduction of the total number of compounds from 347 to 151. Using these 151 variables, a clear differentiation of all of the sample groups, at a concentration of 1×10^8 bacteria mL^{-1} , was achieved when using the first three PCs of the PCA (Fig. 2). The total amount of variance explained by the first three PCs (R^2X cum) was

71%, of which PC 1 explained 35%, PC 2 explained 21.3%, and PC 3 explained 14.7%.

The PLS-DA model also used three components, with the modelling parameter R^2Y (cum) being 98.4%, indicative of the total explained variation of the response Y . Q^2 (cum), the cross-validated variation explained by the response Y , was 97.8%. These high values indicate that the model had an excellent ability to explain the data.

By selecting those metabolites with the highest modelling powers for PCA and the highest VIP values for the PLS-DA, 12 compounds were identified as potential metabolite markers for differentiating these bacterial species. These metabolite markers, together with their concentrations (normalised using the internal standard) are listed in Table 2.

As seen in Table 2, the majority of these metabolite markers are not necessarily novel to a particular bacterial group, but were identified due to the constant differences in their concentrations between the groups. Of the 12 PCA identified metabolite markers described here, 6 were also detected as part of the most abundant compounds in the individual species groups, and their occurrence have therefore already been discussed in Section 3.1. The unsaturated fatty acids, 10-heptadecenoic acid (C17:1 ω 7c) and gondoic acid (C20:1 ω 9c), were detected as potential metabolite markers, despite the fact that they were not detected in the most abundant concentrations. Consistent with our results, Mosca et al. (2007), reported 10-heptadecenoic acid (C17:1 ω 7c) to be in comparatively higher concentrations in *M. avium* than in the other *Mycobacterium* species tested, however, gondoic acid has not yet been described as a characteristic marker for these species. Furthermore, the mycobacterial lipid, C32 mycocerosic acid, which has previously been detected to be present in high concentrations in *M. tuberculosis* (relative to other *Mycobacterium* species) with a characteristic m/z of 494 (Redman et al., 2009), has also been identified as a potential metabolite marker.

Other potential metabolite markers detected included two as yet unknown compounds, with molecular masses of 408 and 422 respectively, which were classified as unknown when using the conventional NIST mass spectral reference library.

An interesting observation is the absence of the almost universally detected octadecanoate, and 2,4-dimethyl tetradecanoate in *M. kansasii*. Although these compounds were detected, their concentrations were not of such to be listed in the top 10 highest abundances, nor unique enough for either to be selected as differentiating markers using the statistical methods described.

3.3. Data modelling for diagnostic purposes

Using the 12 identified metabolite makers, a discriminant model based on Bayes' theorem, in conjunction with multivariate kernel density estimation, was built in order to predict the class membership

Table 2

Mean concentrations of identified metabolite markers for the various *Mycobacterium* species. Standard deviations are given in brackets.

Species	RT (min)	m/z	<i>M. tuberculosis</i>	<i>M. avium</i>	<i>M. bovis</i>	<i>M. kansasii</i>
Compound ^a			Concentration ($\mu\text{g mg}^{-1}$ sample)			
Isopropyl-C14:0	16.8897	270	0	23.0 (6.2)	85.1 (10.4)	0
C16:1 ω 7c	17.7062	268	14.9 (0.6)	60.4 (2.4)	2.3 (0.3)	12.1 (0.7)
C17:1 ω 7c	18.6051	282	3.5 (0.2)	4.5 (0.4)	0	1.7 (0.5)
C17:0	18.8380	284	23.0 (0.8)	8.6 (0.8)	0.2 (0.6)	T
TBSA	20.0940	312	134.6 (0.7)	203.5 (9.4)	69.6 (14.5)	52.0 (2.1)
C20:1 ω 9c	21.1010	324	0	4.7 (0.7)	0	3.5 (0.3)
C22:1 ω 9c	22.2917	352	0	8.1 (0.3)	T	14.7 (0.7)
C24:1 ω 9c	23.2519	380	0	13.3 (1.6)	0	21.5 (1.1)
Unknown 422	23.9792	422	4.2 (0.2)	0	5.6 (1.1)	0
Unknown 408	24.2235	408	0	11.6 (1.5)	0	2.0 (0.5)
Tricosane	25.4079	324	0	3.3 (0.8)	0	T
C32 mycocerosic acid	26.1988	494	2.7 (0.5)	0	0	1.8 (0.4)

^a TBSA, tuberculostearic acid. T, trace amounts.

of the newly extracted “unknown” validation samples (2 cultures of each species). The G built model can be defined as: $P_i^{(x,y,z)} = \sim f_{XYZ}^{(i)}(x,y,z) / \sum_{j=1}^G \sim f_{XYZ}^{(j)}$, where G is the number of groups (in this case 4) and $\sim f_{XYZ}^{(i)}$ is a multivariate kernel density estimate of the first 3 PCs (labelled X, Y, and Z) for group i . This model gives a probability like estimate ($P_i^{(x,y,z)}$) of the group membership of an “unknown” validation sample. The group membership of the validation sample is consequently assigned to the group with the highest probability.

Fig. 3 is a scatter plot of the predicted bacterial group probabilities as determined by the aforementioned discriminant model. All the previously “unknown” samples could be correctly assigned to their respective groups with probabilities of 72–100%.

From these results, it is evident that such a model, using only 12 marker compounds, has the potential to identify various *Mycobacterium* species from cultured sputum samples, in an estimated time of 16 h from sample collection, extracting 24 samples simultaneously. In 1989, Jantzen et al. also made use of 12 characteristic lipid constituents (cellular fatty acids and alcohols identified via capillary gas chromatography) to develop a diagnostic scheme for the identification of various *Mycobacterium* species. In this study, however, mycobacterial species are identified using a powerful automated, pattern recognition data processing method, which can be completed within a few seconds, eliminating the need for manual inspection by the analyst and, hence, enhancing the possibility of an accurate diagnosis.

The routine diagnostic procedure for patient collected sputum in most South African TB diagnostic laboratories is to firstly subject it to smear microscopy analyses. In the case of a positive smear outcome (requiring 5000 – 10 000 bacteria mL^{-1} sputum (WHO, 2006)), the sample is sent for PCR analyses in order to identify the specific causative *Mycobacterium* species. Although, in theory, PCR may be performed directly on sputum samples, clinicians do however, prefer pure cultures as the source for genomic DNA due to the presence of PCR inhibitors in sputum samples (Ahmad et al., 2004). Hence, these smear positive samples are usually cultured for 1–2 days before PCR analyses, in an attempt to ensure sufficient DNA specimens. In the case of a smear-negative outcome, however, sputum samples are usually cultured for up to 5 weeks. If a culture sample becomes smear positive within this time, it is subsequently sent for PCR analyses for species identification and, if not, the sample is considered negative and no further diagnostic procedures are required. As the metabolite pattern recognition approach described in this manuscript also requires pure cultures, it comparable to PCR for *Mycobacterium* species identification in smear-positive samples.

In an attempt to lower the prevalence of TB, treatment should commence as soon as possible after clinical manifestation of the disease,

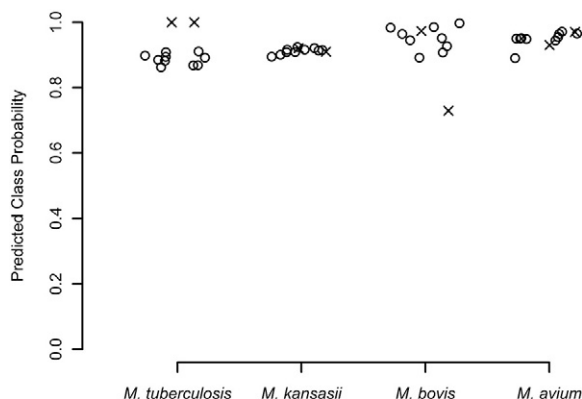


Fig. 3. Probabilities for the “unknown” validation samples (2 of each group) correctly identified for: *M. tuberculosis* are: 100% in both cases; for *M. kansasii*: 92% and 91%; for *M. bovis*: 97% and 72%; and for *M. avium*: 97% and 93%. (o = original samples used to build the model, x = “unknown” validation samples).

0 = blank; 2 = 1×10^2 ; 3 = 1×10^3 ; 4 = 1×10^4 ; 5 = 1×10^5 ; 6 = 1×10^6 ; 7 = 1×10^7 ; 8 = 1×10^8 bacteria/mL

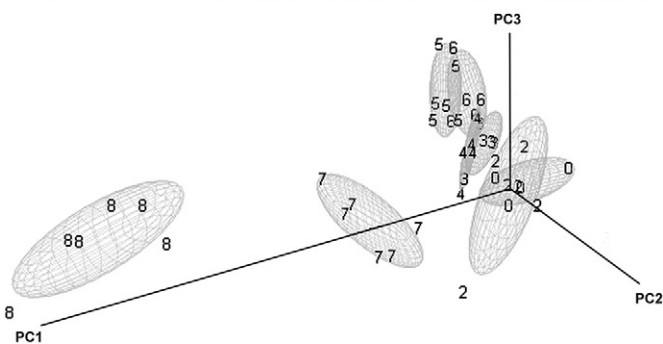


Fig. 4. Three-dimensional PCA scoresplot showing the grouping of the different *M. tuberculosis* dilutions. Individual numbers represent the different analysed sample repeats of the various bacterial concentrations, and a clustering of the replicates further from the blank sample group indicates a higher degree of differentiation and, hence, a more significant difference between the metabolite profiles of the specific concentration group and the blank samples. All the groups, except that containing 1×10^2 bacteria mL^{-1} could be distinguished from the blank group, determining a potential detection limit of 1×10^3 bacteria mL^{-1} .

leaving less time to spread the infection. Therefore, diagnostic techniques should be fast, and the time required for culturing before species identification analyses should ideally be shortened, especially in smear-negative cases. To achieve this, a method should comprise of the lowest possible detection limit, and for that reason, we investigated the potential detection limit of this metabolite pattern recognition approach for *Mycobacterium* species classification.

3.4. Detection limit

Fig. 4 illustrates the PCA scoresplot of the data generated from the *M. tuberculosis* sample repeats prepared in the concentration gradient (1×10^2 to 1×10^8 bacteria mL^{-1}). As seen in Fig. 4, a differentiation between the various *M. tuberculosis* concentration groups was attained and the sample replicates with higher bacterial concentrations clustered farther from the blank sample group, indicating a higher degree of differentiation and, hence, a more significant difference between the metabolite profiles of the higher concentration groups and the blank samples. From these results, it is clear that the concentration group of 1×10^2 bacteria mL^{-1} overlaps with

○ *M. tuberculosis* △ *M. kansasii* ◇ *M. avium* □ *M. bovis*

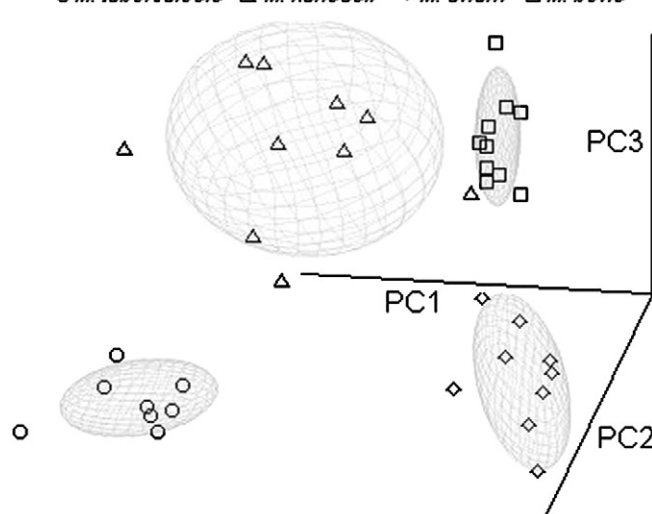


Fig. 5. Three-dimensional PCA scores plot showing a clear differentiation between *Mycobacterium* sample groups at a concentration of 1×10^3 bacteria mL^{-1} .

the blank ($0 \text{ bacteria mL}^{-1}$), indicative that the possible detection limit for this approach would be in the range of $1 \times 10^3 \text{ bacteria mL}^{-1}$, which could still be clearly differentiated from the blank sample group.

In order to determine if this approach is able to differentiate between the various TB causing *Mycobacterium* species at the above mentioned detection limit for *M. tuberculosis*, we repeated the analysis of the 4 *Mycobacterium* species, only this time at a concentration of $1 \times 10^3 \text{ bacteria mL}^{-1}$. Once again, three PCs were used. The PCA scoresplot of the processed data (Fig. 5) indicates that all the infectious species groups once again differentiated from one another, leading to a potential requirement of merely 1000 cells per sample for species identification. The total amount of variance explained by the first three PCs ($R^2 \text{X cum}$) was 68.7%, of which PC 1 explained 27.7%, PC 2 explained 24.6%, and PC 3 explained 16.4%.

This potential detection limit is far less than that of the only commercially available mycobacteria pattern recognition system, MYCO-LCS, which requires at least 1×10^5 cultured cells in order to achieve a successful diagnostic outcome (ATC and CDC, 1999). Although, a pure culture is required, this method still compares significantly well to the currently used PCR methods, however, it requires no more than 16 h to obtain a diagnostic result, extracting 24 samples simultaneously. In future, this time may dramatically be shortened using faster extraction procedures and by analysing only the 12 relevant metabolite markers. Furthermore, as a smear-negative, culture positive sputum sample contains a minimum of 10 cells per sample (Getahun et al., 2007; WHO, 2006), these samples would only require a culturing step of approximately 7 days, considering that the population of these organisms doubles every 15–20 h. This performance is a significant improvement on the 5 weeks of culturing currently required for smear-negative samples before species identification can be done. However, this may not be entirely relevant to sampling a slow growing culture, where an undetermined cell mass (likely not measurable by eye or OD), may or may not be ready for analysis on an accelerated timeline, and this should be confirmed using an alternative approach.

Furthermore, even though the preliminary performance of this metabolomics approach is similar to that of PCR, the cost per analysis is far less, as it does not require the use of expensive DNA primers. On average, the cost of a PCR machine typically used for TB diagnostics is in the range of US\$10 000–20 000 (Scherer et al., 2009), the BACTEC culturing system, US\$ 19 500–38 950 (FIND, 2011) and a GC-FID (which would be used for routine analysis of the TB markers identified, once implemented) US\$ 11 000–26 000 (Buck Scientific, 2011), depending on the make and model. Furthermore, the actual cost of analysis is about US\$ 13 for PCR (Scherer et al., 2009), US\$ 12 for the BACTEC system (Drancourt and Raoult, 2007) and US\$ 2.50 for a lipid analysis using GC-FID (Sasser, 2009). The use of non-viable bacterial cells for the current species identification approach described in this manuscript may be considered a further advantage from a safety perspective. However, before this conceptual approach can be implemented in the clinical environment, its robustness needs to be assessed using far larger sample groups, with a larger variety of *Mycobacterium* species and strains.

4. Conclusions

Using a metabolomics research approach, we were able to prove that GC-MS lipid analysis, followed by multivariate statistical data processing, has the capacity to identify metabolite markers characterising various TB causing and non-TB *Mycobacterium* species, following a culturing step from sputum. The powerful, automated pattern recognition model developed, using only 12 marker compounds, sets this method apart from seemingly similar mycobacterial FAME profiling efforts. Furthermore, taking the potential detection limit into account, added refinement of this approach may lead to faster species

identification, as shorter culturing times could potentially be used to obtain sufficient amounts of diagnostic material.

Acknowledgements

Gerhard Koekemoer (Statistical Consultation Services, North-West University, Potchefstroom, South Africa) for assistance with the statistical data analyses.

The financial assistance of the National Research Foundation (NRF) of South Africa towards this research is hereby acknowledged.

References

- Ahmad, S., Mokaddas, E., Jaber, A.-A., 2004. Rapid detection of ethambutol-resistant *Mycobacterium tuberculosis* strains by PCR-RFLP targeting embB codons 306 and 497 and iniA codon 501 mutations. *Mol. Cell. Probes* 18, 299–306.
- American Thoracic Society and the Centers for Disease Control and Prevention, 1999. Diagnostic standards and classification of tuberculosis in adults and children. *Am. J. Respir. Crit. Care Med.* 161 (4 Pt 1), 1376–1395.
- Brennan, P.J., 2003. Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis* 83, 91–97.
- Buck Scientific, 2011. Industrial GC with FID & TCD [Web:] <http://bucksci.com/atomic-absorption-spectrophotometers/details/3073/4/gas-chromatographs/industrial-gc-with-fid-&-tcd.html> 2011 [Date of use: 08 Dec. 2011].
- Colebunders, R., Bastian, I., 2000. A review of the diagnosis and treatment of smear-negative pulmonary tuberculosis. *Int. J. Tuberc. Lung Dis.* 4, 97–107.
- Cruciani, M., Scarparo, C., Malena, M., Bosco, O., Serpelloni, G., Mengoli, C., 2004. Meta-Analysis of BACTEC MGIT 960 and BACTEC 460 TB, with or without solid media, for detection of *Mycobacteria*. *J. Clin. Microbiol.* 42, 2321–2325.
- Drancourt, M., Raoult, D., 2007. Cost-effectiveness of blood agar for isolation of mycobacteria. *PLoS Negl. Trop. Dis.* 1 (2), e83. doi:10.1371/journal.pntd.0000083.
- Dunn, W.B., Bailey, N.J.C., Johnson, H.E., 2005. Measuring the metabolome: current analytical technologies. *Analyst* 130, 606–625.
- Fend, R., Kolk, A.H., Bessant, C., Buijtel, P., Klatzer, P.R., Woodman, A.C., 2006. Prospects for clinical application of electronic-nose technology to early detection of *Mycobacterium tuberculosis* in culture and sputum. *J. Clin. Microbiol.* 44, 2039–2045.
- FIND, 2011. FIND prices for BACTEC and MGIT and country list [Web:] http://www.finddiagnostics.org/about/what_we_do/successes/find-negotiated-prices/bactec-mgit.html 2011 [Date of use: 09 Dec. 2011].
- Frieden, T.R., Sterling, T.R., Munsiff, S.S., Watt, C.J., Dye, C., 2003. Tuberculosis. *Lancet* 362, 887–899.
- Getahun, H., Harrington, M., O'Brien, R., Nunn, P., 2007. Diagnosis of smear-negative pulmonary tuberculosis in people with HIV infection or AIDS in resource-constrained settings: informing urgent policy changes. *Lancet* 369, 2042–2049.
- Grange, J.M., 2001. *Mycobacterium bovis* infection in human beings. *Tuberculosis* 81, 71–77.
- Guerrant, G.O., Lambert, M.A., Moss, C.W., 1981. Gas-chromatographic analysis of mycolic acid cleavage products in mycobacteria. *J. Clin. Microbiol.* 13, 899–907.
- Jantzen, E., Tangen, T., Eng, J., 1989. Gas chromatography of mycobacterial fatty acids and alcohols: diagnostic applications. *APMIS* 97, 1037–1045.
- Knisley, C.V., Damato, J.J., McClatchy, K., Brennan, P.J., 1985. Rapid and sensitive identification of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 22, 761–767.
- Koekemoer, G., Swanepoel, J.W.H., 2008. A semi-parametric method for transforming data to normality. *Stat. Comput.* 8, 241–257.
- Kox, L.F.F., 1995. Tests for detection and identification of mycobacteria. How should they be used? *Respir. Med.* 89, 399–408.
- Lambert, M.A., Moss, C.W., Silcox, V.A., Good, R.C., 1986. Analysis of mycolic acid cleavage products and cellular fatty acids of *Mycobacterium* species by capillary gas chromatography. *J. Clin. Microbiol.* 23, 731–736.
- Larsson, L., Odham, G., Westerdahl, G., Olsson, B., 1987. Diagnosis of pulmonary tuberculosis by selected-ion monitoring: improved analysis of tuberculoesterate in sputum using negative-ion mass spectrometry. *J. Clin. Microbiol.* 25, 893–896.
- Luquin, M., Ausina, V., López Calahorra, F., Belda, F., García Barceló, M., Celma, C., Prats, G., 1991. Evaluation of practical chromatographic procedures for identification of clinical isolates of *Mycobacteria*. *J. Clin. Microbiol.* 29, 120–130.
- Marais, B.J., Pai, M., 2007. New approaches and emerging technologies in the diagnosis of childhood tuberculosis. *Paediatr. Respir. Rev.* 8, 124–133.
- Mayakova, T.I., Kuznetsova, E.E., Kovaleva, M.G., Plyusnin, S.A., 1995. Gas chromatographic-mass spectrometric study of lipids and rapid diagnosis of *Mycobacterium tuberculosis*. *J. Chromatogr. B* 672, 133–137.
- MIDI, inc., 2009. Sherlock@Mycobacteria Identification System [Web:] http://www.midi-inc.com/pdf/Sherlock_MYCO_LCS_Brochure.pdf 2009 [Date of use: 08 Dec. 2011].
- Miguez-Burbano, M.J., Flores, M., Ashkin, D., Rodriguez, A., Granada, A.M., Quintero, N., Pitcheik, A., 2006. Non-tuberculous mycobacteria disease as a cause of hospitalization in HIV-infected subjects. *Int. J. Infect. Dis.* 10, 47–55.
- Moore, D.F., Curry, J.I., 1995. Detection and identification of *Mycobacterium tuberculosis* directly from sputum sediments by Amplicor PCR. *J. Clin. Microbiol.* 33, 2686–2691.
- Mosca, A., Russo, F., Miragliotta, L., Iodice, M.A., Miragliotta, G., 2007. Utility of gas chromatography for rapid identification of mycobacterial species frequently encountered in clinical laboratory. *J. Microbiol. Methods* 68, 392–395.
- Noordhoek, G.T., Kolk, A.H.J., Bjune, G., Catiy, D., Dale, J.W., Fine, P.E.M., Godfrey-Faussett, P., Cho, S., Shinnick, T.S.B., Svenson, S.B., Wilson, S., van Embden, D.A.,

1994. Sensitivity and Specificity of PCR for detection of *Mycobacterium tuberculosis*: a blind comparison study among seven laboratories. *J. Clin. Microbiol.* 32, 277–284.
- Odham, G., Larsson, L., Mårdh, P., 1979. Demonstration of tuberculostearic acid in sputum from patients with pulmonary tuberculosis by selected ion monitoring. *J. Clin. Invest.* 63, 813–819.
- Olivier, K., 1998. Nontuberculous mycobacterial pulmonary disease. *Curr. Opin. Pulm. Med.* 4, 48–53.
- Olivier, I., Loots, D.T., 2011. An overview of tuberculosis treatments and diagnostics. What role could metabolomics play? *J. Cell Tissue Res.* 11, 2655–2671.
- Pounder, J.L., Anderson, C.M., Voelkerding, K.V., Salfinger, M., 2010. *Mycobacterium tuberculosis* complex differentiation by genomic deletion patterns with multiplex polymerase chain reaction and melting analysis. *Diagn. Microbiol. Infect. Dis.* 67, 101–105.
- Redman, J.E., Shaw, M.J., Mallet, A.I., Luisa Santos, A., Roberts, C.A., Gernaey, A.M., Minnikin, D.E., 2009. Mycocerosic acid biomarkers for the diagnosis of tuberculosis in the Coimbra Skeletal Collection. *Tuberculosis* 89, 267–277.
- Rezwan, M., Lanéelle, M., Sander, P., Daffé, M., 2007. Breaking down the wall: fractionation of mycobacteria. *J. Microbiol. Methods* 68, 32–39.
- Ruiz-Manzano, J., Blanquer, R., Calpe, J.L., Caminero, J.A., Caylà, J., Domínguez, J.A., García, J.M., Vidal, R., 2008. Diagnosis and treatment of tuberculosis. *Arch. Bronconeumol.* 44, 551–566.
- Sasser, M., 2009. Microbial identification by gas chromatography analysis of fatty acid methyl esters (GC-FAME)[Web:] www.midi-inc.com 2009 [Date of use: 08 Dec. 2011].
- Scherer, L.C., Sperhake, R.D., Ruffino-Netto, A., Rossetti, M.L.R., Vater, C., Klatser, P., Kritski, A.L., 2009. Cost-effectiveness analysis of PCR for the rapid diagnosis of pulmonary tuberculosis. *BMC Infect. Dis.* 9, 216. doi:10.1186/1471-2334-9-216.
- Schoeman, J.C., Loots, D.T., 2011. Improved disease characterisation and diagnostics using metabolomics: a review. *J. Cell Tissue Res.* 11, 2673–2683.
- Stopforth, A., Tredoux, A., Crouch, A., van Helden, P., Sandra, P., 2004. A rapid method of diagnosing pulmonary tuberculosis using stir bar sorptive extraction-thermal desorption-gas chromatography-mass spectrometry. *J. Chromatogr. A* 1071, 135–139.
- t'Kindt, R., Morreel, K., Deforce, D., Boerjan, W., Van Bocxlaer, J., 2009. Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: repeatability and sample pre-treatment. *J. Chromatogr. B* 877, 3572–3580.
- Urbanczik, R., 1985. Present position of microscopy and of culture in diagnostic Mycobacteriology. *Zentralbl. Bakteriol. Mikrobiol. Hyg. A* 260, 81–87.
- WHO, 2006. Diagnostics for Tuberculosis, Global Demand and Market Potential. WHO Press, Geneva, Switzerland. pp. 33–49.
- WHO, 2010. The Global Tuberculosis Epidemic: fact sheet June 2010 <http://www.kff.org/globalhealth/7883.cfm> 2010.

APPENDIX A (3)

Manuscripts

OLIVIER, I., LOOTS, D.T. Altered fatty acid metabolism due to rifampicin-resistance conferring mutations in the *rpoB* gene of *M. tuberculosis*. *OMICS*. Submitted. Manuscript nr. OMI-2012-0028.

1 **Title:** Altered fatty acid metabolism due to rifampicin-resistance conferring mutations in
2 the *rpoB* gene of *M. tuberculosis*

3

4 **Authors:** Ilse Olivier¹, Du Toit Loots¹

5

6 ¹Centre for Human Metabonomics,
7 School for Physical and Chemical Sciences
8 North-West University
9 Potchefstroom
10 2520
11 South Africa.

12

13 Ilse Olivier

14 Tel: +27 84 299 2052

15 Fax: +27 18 299 2316

16 E-mail: ilseolivier111@gmail.com

17

18 Du Toit Loots (Corresponding author)

19 Centre for Human Metabonomics,

20 Tel: +27 18 299 1818

21 Fax: +27 18 299 2316

22 E-mail: dutoit.loots@nwu.ac.za

23

24 **Keywords:** gas chromatography mass spectrometry; rifampicin-resistance; metabolomics;
25 *M. tuberculosis*

26

Abstract

We investigated the use of gas chromatography mass spectrometry (GC-MS) metabolomics to potentially better characterize rifampicin-resistance by comparing the fatty acid metabolomes of two *rpoB* mutant *M. tuberculosis* strains (S522L and S531L) to that of a fully susceptible wild-type parent strain. Using the generated GC-MS metabolite data, principal component analysis (PCA) showed a clear differentiation between all three sample groups analyzed. We subsequently identified those metabolites contributing most to the variation in the data using PCA and partial least squares discriminant analysis (PLS-DA). The altered metabolite markers detected in the *rpoB* mutants propose a decreased synthesis of various 10-methyl branched-chain fatty acids and cell wall lipids, and an increased use of the shorter-chain fatty acids as carbon sources. Furthermore, the *rpoB* S531L mutant, previously reported to occur in well over 50% of all clinical rifampicin-resistant *M. tuberculosis* strains, potentially showed a better capacity for using these alternative energy sources, in comparison to the less frequently detected *rpoB* S522L mutant. This study additionally proves the capacity of metabolomics to identify new metabolite markers, describing previously unknown metabolic alterations induced by mutations in the *rpoB* gene of *M. tuberculosis*.

1 **Introduction**

2

3 More than 9 million new tuberculosis (TB) cases are recorded annually, of which 0.4 - 0.5
4 million are multi-drug resistant (MDR) (WHO, 2010). MDR-TB is defined as resistance to
5 at least rifampicin and isoniazid, the two most important drugs currently used to treat
6 *Mycobacterium tuberculosis* infection. As almost 90% of rifampicin-resistant strains are
7 also resistant to isoniazid, resistance to rifampicin is used as a surrogate marker for
8 detecting MDR-TB (Aziz et al., 2006).

9 Rifampicin functions by inhibiting RNA-polymerase, an enzyme essential for the
10 transcription of bacterial DNA to RNA (Wehrli, 1983). Resistance to rifampicin is primarily
11 caused by an alteration in the β -subunit of RNA-polymerase, owing to mutations in the
12 encoding *rpoB* gene. Various specific resistance-conferring mutations (accounting for over
13 95% of the rifampicin-resistant strains), occur in an 81-bp region of *rpoB*, known as the
14 rifampicin-resistance determining region (RRDR) (Ahmad et al., 2000; Anthony et al.,
15 2005). The majority of these are point mutations, resulting in a replacement of aromatic
16 with non-aromatic amino acids. These replacements lead to drug resistance by
17 interrupting the forces that bind rifampicin to RNA polymerase, but do, however, also
18 initially impair the fitness of these bacteria, which may be restored by secondary mutations
19 (Gagneux et al., 2006; Mariam et al., 2004). Since the *rpoB* gene codes for the β -subunit
20 of RNA-polymerase, mutations in this gene are thought to lead to disruptions to many
21 parts of the bacterial metabolism (Gagneux et al., 2006), which to date have not yet been
22 elucidated.

23 In recent years, the integrated systems biology research approach (genomics,
24 transcriptomics, proteomics and metabolomics) has gradually become an essential tool for
25 better understanding the mechanisms behind various biological processes (Debnath et al.,
26 2011; Chen, 2011; Schnackenberg et al., 2006). As metabolomics (the study of small-

1 molecular weight compounds) reflects changes in gene expression (due to, for example,
2 mutations), it is considered an important functional genomics tool, which, when interpreted
3 as part of a systems biology approach, gives a more holistic description of the expression
4 profile (Olivier and Loots, 2011). Fiehn et al. (2000) successfully used metabolomics as a
5 tool for functional genomics, to compare four *Arabidopsis* genotypes, including two
6 homozygous ecotypes and a mutant of each ecotype. A clustering of the individual
7 samples into each of the respective genotype groups was observed after principal
8 component analysis (PCA). This grouping was ascribed to the distinct variations in the
9 metabolic profiles of the individual samples in the respective groups. Since then, other
10 research groups have also identified variances in the metabolite profiles of a variety of
11 bacterial species, due to genetic perturbations (Baran et al., 2009). Proving the capacity of
12 metabolomics as a tool to investigate previously unknown mechanisms of the
13 mycobacterial metabolome, in a lipidomics study, Jain et al. (2007), indicated that *M.*
14 *tuberculosis* undergoes a metabolism shift to preferentially use host lipids as a carbon
15 source during infection, consequently increasing their virulence associated lipid anabolism.
16 Similarly, using an untargeted metabolic profiling approach, de Carvalho et al. (2010)
17 successfully proved that *M. tuberculosis* can simultaneously catabolize multiple carbon
18 sources.

19 Considering this, using a similar metabolomics approach, we aimed to differentiate and
20 potentially better characterize two rifampicin-resistant mutants by comparing their fatty
21 acid metabolite profiles to that of a fully drug-sensitive *M. tuberculosis* parent strain. This
22 was done for the purpose of identifying new metabolite markers, in order to potentially
23 better explain the altered metabolism induced by these rifampicin-resistance conferring
24 mutations. The two rifampicin-resistant strains selected are the previously described *rpoB*
25 S531L mutant, which is present in well over 50% of all clinical isolates, and the *rpoB*
26 S522L mutant, which is mostly found only in laboratory isolates (Gagneux et al., 2006).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

Materials and Methods

Reagents

Potassium hydroxide, nonadecanoic acid and glacial acetic acid were obtained from Merck. Chloroform, methanol and hexane were ultra-purity Burdick & Jackson brands (Honeywell International Inc.).

Bacterial cultures

All the organisms used in this study were obtained from the Royal Tropical Institute (KIT), Amsterdam, The Netherlands. The rifampicin-resistant *M. tuberculosis* strains used are spontaneous mutants derived from the wild-type parent strain (MTB72, ATCC35801), and were picked after one round of selection with rifampicin. For this selection, bacteria were cultured in Middlebrook 7H9 (Difco) broth for 14 days in a shaking incubator at 37 °C. Then, 0.5 ml of broth was plated out onto Middlebrook 7H11 (Difco) media containing rifampicin (8 µg ml⁻¹), and selected colonies were once again streaked on rifampicin containing media (8 µg ml⁻¹) to confirm the resistant phenotype (Anthony et al., 2005). Mutations in the *rpoB* gene were identified by sequencing a 271-bp region containing the 81-bp RRDR hotspot and if no mutation was detected in this region, an additional 365-bp region at the N-terminus (Anthony et al., 2005). The mutations in the two rifampicin-resistant strains used in this study were identified as S522L and S531L. The rifampicin resistant and wild-type parent strains were then cultured in Middlebrook 7H9 supplemented with oleic acid-albumin-dextrose-catalase, in a shaking incubator at 37 °C. For each strain, liquid starting cultures were made by inoculating pure colonies from

slopes into 10 ml culture medium until these cultures reached the logarithmic growth phase. Prior to extraction, bacteria were isolated from each culture (n = 12) for each of the 3 strains, and resuspended in ddH₂O at a concentration of 1 X 10⁸ bacteria ml⁻¹.

Extraction procedure

As an internal standard, 100 µl of nonadecanoic acid (14 µg ml⁻¹) was added to 1 ml of the above mentioned sample suspensions, followed by a modified Bligh-Dyer fatty acid extraction procedure (Olivier and Loots, 2012). Briefly, 4 ml of methanol, 2 ml of chloroform, and 1.6 ml of ddH₂O was added. The mixture was subsequently sonicated for 2 min, vortexed for 30 sec, and allowed to stand for 2 hours (at room temperature) for the phases to separate. Chloroform and ddH₂O (2 ml of each) were then added to the mixture, followed by centrifugation at 550 x g for 15 min. Hereafter, the organic phase was collected and the aqueous phase was re-extracted with 2 ml of chloroform. To the collective organic phases, 5 ml of ddH₂O was added and the mixture was once again left to stand for 2 hours (at room temperature) for the phases to separate, followed by centrifugation at 550 x g for 20 min. The newly formed organic phase was then collected and dried under a light stream of nitrogen. Subsequently, 0.5 ml of methanol, 0.5 ml of chloroform, and 1 ml of methanolic KOH (0,2 M) was added to the dried sample. The mixture was then incubated at 60 °C for 30 min followed by the addition of 2 ml of hexane, 2 ml of ddH₂O, and 200 µl of glacial acetic acid (1 M). The organic phase was again collected and transferred to a new tube. The remaining aqueous phase was re-extracted 3 times with hexane. The combined organic phases from each extraction step were dried under nitrogen and re-suspended in 100 µl hexane prior to injection.

Gas chromatography mass spectrometry (GC-MS)

1
2 The extracts were analyzed by a 1 μ l splitless injection on an Agilent 7890A gas
3 chromatograph coupled to an Agilent 5975 mass selective detector equipped with a 7683B
4 injector, 7683 auto sampler and VF1-MS capillary column (30 m x 250 μ m i.d., 0.25 μ m
5 film thickness). The injector temperature was held constant at 250 $^{\circ}$ C for the entire run.
6 The initial GC oven temperature was set at 50 $^{\circ}$ C for 1 min, followed by an increase of 10
7 $^{\circ}$ C min⁻¹ to 240 $^{\circ}$ C. This was then followed by an increase in oven temperature of 20 $^{\circ}$ C
8 min⁻¹ to a final temperature of 300 $^{\circ}$ C, at which the temperature was maintained for 7 min.
9 Helium was used as carrier gas and pressure programmed such that helium flow was kept
10 at a constant 1.2 ml min⁻¹. Detection was achieved by using MS detection in full scan
11 mode (m/z 50-550).
12 The raw GC-MS data were deconvoluted and analyzed using AMDIS software (Automated
13 Mass Spectral Deconvolution and Identification System, V2.65) at a threshold of 0.01% of
14 the total signal. A library was then compiled containing the mass spectra of all the
15 compounds detected in all the individual samples of the three respective sample groups.
16 After the deletion of duplicate peaks, the generated library was used to reanalyze each of
17 the individual samples. These analyses resulted in a data matrix containing the relative
18 concentrations (normalized with the internal standard) for all the detected compounds, in
19 each of the 12 samples, for each of the 3 strains. Using the mass fragmentation patterns
20 generated by the MS, together with their respective GC retention times, compounds were
21 identified using libraries generated from previously injected standards.

22

23 *Statistical data analyses*

24

25 The statistical packages, "R" (version 2.13.0) and Statistica (version 10) were used for all
26 the statistical data analyses. Prior to multivariate data analyses, a 50% filter was applied

1 to the data, excluding those compounds which do not appear in at least 50% of the
2 samples, in one or more of the sample groups. Principal component analysis (PCA) and
3 partial least squares discriminant analysis (PLS-DA) were used to analyze the generated
4 data (Madsen et al., 2010). PCA is an unsupervised, multivariate statistical method used
5 for pattern recognition where the sample data are not assigned to specific groups prior to
6 the analysis, but are analyzed independently, in order to determine if a natural grouping or
7 differentiation exists. In order to identify those compounds which best describe the
8 variance in the data, the metabolites were ranked according to their respective modelling
9 powers. PLS-DA, on the other hand, is a supervised classification procedure, where class
10 membership information is used to build a discrimination and classification model. PLS-
11 DA may therefore be used to identify those compounds which best characterize the
12 differentiated the sample groups, by ranking the metabolites according to the variable
13 influence on the projection (VIP) parameter. VIP is a weighted sum of squares of the
14 PLS-DA weights, indicating the importance of the metabolite to the total model. Scaling
15 the data prior to PCA and PLS-DA is important, as variables (metabolites) with higher
16 concentrations will dominate the analyses. For this purpose, a non-parametric
17 transformation function was used (Koekemoer & Swanepoel, 2008), after which mean
18 centring was applied.

19

20 **Results and Discussion**

21

22 *Clustering of samples into three genotype groups*

23

24 The process of manually correlating changes in metabolite levels to genetic perturbations
25 is challenging, mainly due to the complexity and interconnectivity of biochemical pathways,
26 and the fact that each analyzed sample actually represents a separate biological system.

1 Therefore, multivariate statistical methods, such as PCA and PLS-DA, are highly
2 recommended for this purpose Fiehn et al. (2000). PCA involves a mathematical
3 procedure that transforms a number of possibly related variables (in this case,
4 metabolites) into a smaller number of unrelated variables known as principal components
5 (PC's). Using one PC per axis, the PCA can then be visualized as a scores plot. Fig. 1
6 represents the PCA scoresplot using the metabolite data generated after GC-MS analyses
7 of the wild-type and the two rifampicin-resistance conferring *M. tuberculosis rpoB* mutant
8 samples groups (S533L and S531L), and shows a clear clustering of individual samples
9 into their respective genotypes. An apparent differentiation between not only the wild-type
10 and mutant sample groups, but also between the two different rifampicin-resistant strains,
11 was also evident. The fact that these sample groups are closely related and only differ by
12 a single *in vitro* selection step, highlights the capacity of this metabolomics approach to
13 detect minor metabolic changes induced by genetic alterations.

14
15 **Figure 1 here**
16

17 Similar to PCA, PLS-DA also reduces the dataset into a number of smaller components.
18 PLS-DA is, however, aimed at finding a well defined borderline between samples groups,
19 in a supervised manner (i.e., individual samples are allocated to their respective sample
20 groups before the analyses) by building a prediction model. In this case, the built PLS-DA
21 model used two components, with the modeling parameter R^2Y (cum) being 47.6%,
22 indicative of the total explained variation of the response Y. Q^2 (cum), the cross-validated
23 variation explained by the response Y, was 92.1%. The top ranked metabolites identified
24 by each of the above mentioned multivariate statistical methods, are listed in Table 1.

25
26 **Table 1 here**

1

2 When considering the results in Table 1, it is clear that the differentiation of the wild-type
3 and rifampicin-resistant strains are dependant not only on the occurrence of compounds
4 novel to a particular sample group, but also due to constant concentration differences
5 between compounds common to two or more of the sample groups. However, two novel
6 branched-chain fatty acids (10-methyl-pentadecanoic acid (10Me-C15:0) and 10-methyl-
7 hexadecanoic acid (10Me-C16:0), were identified exclusively in the wild-type *M.*
8 *tuberculosis* parent strain group. Furthermore, an unknown fatty acid of molecular mass
9 394 (Unknown 394), was identified exclusively in the wild-type and S531L mutant sample
10 group. These unique markers and the other branched-chain and straight-chain fatty acids
11 identified as potential metabolite markers (due to the constant differences in their
12 concentrations between the respective sample groups) will subsequently be discussed in
13 an attempt to better characterize the two rifampicin-resistance strains investigated.

14

15 *Altered branched-chain fatty acid levels due to mutations in the rpoB gene*

16

17 Fig. 2 is a schematic representation of our results, incorporating the identified metabolite
18 markers into known biochemical pathways of *M. tuberculosis*.

19

20 ***Figure 2 here***

21

22 Two branched-chain fatty acids, 10Me-C15:0 and 10Me-C16:0, were identified exclusively
23 in the wild-type *M. tuberculosis* parent strain samples, and in undetectable concentrations
24 in both *rpoB* rifampicin resistant mutants. Interestingly, the well known characteristic
25 *Mycobacterium* branched-chain fatty, tuberculostearic acid (TBSA, 10-methyl-
26 octadecanoic acid), was detected in comparatively elevated concentrations in both the

1 resistant strains. In *Mycobacterium* species, these saturated mid-chain methyl-branched
2 fatty acids are formed via methylation of a delta-9-unsaturated fatty acid, with S-
3 adenosylmethionine (SAM) acting as the methyl donor (Fig. 2). The resulting residue is
4 then reduced to the 10-methyl compound, with NADPH acting as co-substrate for this
5 reaction (Ramakrishnan et al., 1972). The synthesis of the above mentioned delta-9-
6 unsaturated fatty acids, requires the dehydrogenation of the corresponding saturated fatty
7 acyl-CoA precursor, and it has been shown that, unlike other aerobic bacteria,
8 mycobacteria require Fe^{2+} and a flavin (FAD or FMN), in addition to NADPH and O_2 (Fulco
9 & Bloch, 1974) for this reaction. Furthermore, in prokaryotes, RNA polymerases use
10 ribonucleoside 5'-triphosphates (NTPs) including ATP, GTP, CTP, and UTP, as substrates
11 to synthesize mRNA. Additionally, the degradation of mRNA leads to a release of free
12 nucleotide bases (adenine, guanine etc.), which may subsequently be used to re-
13 synthesise these NTPs via phosphoribosyltransferases (Garrett & Grisham, 2005),
14 ensuring an mRNA - NTP equilibrium. Hence, it is not unlikely that a change in the
15 configuration of RNA polymerase, as is the case in the *rpoB* mutants used in this study,
16 may lead to a disturbance in this equilibrium. Furthermore, as GTP is a precursor of
17 riboflavine (FAD and FMN) synthesis (KEGG, 2011), this disturbance may ultimately
18 influence the synthesis of the mid-chain methyl-branched fatty acids (Fig. 2), explaining
19 the reduced amounts of 10-Me-C15:0 acid and 10-Me-C16:0 detected for the *rpoB*
20 mutants investigated. Additionally, as SAM is synthesised by methionine
21 adenosyltransferase, using both L-methionine and ATP as substrates (González et al.,
22 2003), the proposed decrease in NTPs, including ATP, as discussed, would be expected
23 to contribute to the reduction of the 10-methyl branched-chain fatty acids in these mutant
24 *M. tuberculosis* strains.

25 High amounts of ATP (Shi et al., 2010), together with a number of SAM-dependent
26 enzymes (Barry et al., 1998), are also required for mycolic acid synthesis, and

1 consequently, lowered ATP levels would be expected to have an effect on mycolic acid
2 synthesis in these *rpoB* mutations. The framework of the mycobacterial cell-wall,
3 mycolylarabinogalactan peptidoglycan, consists of a mycolic acid complex embedded with
4 lipoarabinomannan (LAM) (Chatterjee, 1997). Hence, the proposed reduction of mycolic
5 acid synthesis, would in turn lead to a complementary decreased use of LAM in this
6 framework, but not necessarily in its synthesis, and consequently, elevating circulating
7 LAM. The elevated levels of C16:0 and TBSA, the major fatty acid components of LAM
8 (Nigou et al., 2003) in the rifampicin-resistant *M. tuberculosis* strains, supports this
9 hypothesis, and explains why TBSA is elevated, while the other 10-methyl branched-chain
10 fatty acids previously described are reduced.

11

12 *Altered straight-chain fatty acids levels due to mutations in the rpoB gene*

13

14 Two unsaturated fatty acids, oleic acid (C18:1 ω 9c) and palmitoleic acid (C16:1 ω 7c),
15 were detected in elevated concentrations in the *rpoB* mutants (Table 1). As described
16 above, and indicated in Fig. 2, these fatty acids serve as substrates for the synthesis of the
17 saturated mid-chain methyl-branched fatty acids via SAM (Ramakrishnan et al., 1972).
18 Hence, the previously proposed deficiencies in SAM in the *rpoB* mutants would be
19 expected to result in an accumulation in the immediate upstream delta-9-unsaturated fatty
20 acids, as seen in our investigation, and supporting the previous findings.

21 Furthermore, according to the metabolite rankings from both the PCA and PLS-DA, the
22 compound considered most important for differentiating the sample groups was
23 heptadecanoic acid (C17:0). This saturated straight-chain fatty acid was detected in
24 reduced amounts in the S522L mutant samples and only in trace concentrations the
25 S531L mutant samples, as compared to the wild-type strain. During conditions of stress,
26 *M. tuberculosis* are known to up-regulate the synthesis of the glyoxylate shunt enzyme,

1 isocitrate lyase (Icl), allowing these bacteria to survive on acetate or fatty acids as their
2 primary carbon sources (Glickman & Jacobs, 2001, Shi et al., 2010). As these *rpoB*
3 mutant rifampicin-resistant *M. tuberculosis* strains are associated with a competitive fitness
4 cost and a reduced growth rate (Gagneux et al., 2006), they would be expected to react in
5 the same manner, explaining the large reduction in the C17:0 observed, and further
6 confirming the reduction in the 10-methyl branched-chain fatty acids detected.
7 Additionally, as the oxidation of these odd-chain and methyl-branched fatty acids, when
8 used as alternative energy, results in elevated propionyl-CoA and acetyl-CoA (Muñoz-
9 Elías et al., 2006), the elevated concentrations of benzenepropionic acid, a propionic acid
10 derivative in the S531L mutants, further confirms these observations.
11 The straight-chain fatty acid, hexadecanoic acid (C16:0) on the other hand, was detected
12 in elevated concentrations in both resistant strains. As previously mentioned, C16:0 is an
13 important component of LAM (Nigou et al., 2003) and the proposed increased
14 concentrations of circulating LAM due to the *rpoB* mutation explains the increased
15 concentrations of C16:0 (and the previously described increased TBSA detected), in these
16 rifampicin-resistant mutants.

17

18 *Comparison of the two rpoB mutants*

19

20 From Table 1 it is also evident that a number of differences in the concentrations of the
21 detected metabolite markers exist when comparing the two *rpoB* mutant strains.

22 Gagneux et al. (2006) compared the competitive fitness of various clinical and laboratory
23 generated rifampicin resistant *M. tuberculosis* mutant strains, by allowing drug-resistant
24 and the drug-susceptible organisms to compete for limited resources in identical
25 environments. They determined that all the laboratory mutant strains tested had a
26 significant fitness cost, and that this cost was noticeably less in the S531L mutant

1 compared to the other rifampicin resistant strains. When tested on clinically derived
2 mutants, no fitness cost was observed for the S531L strain, as opposed to considerable
3 fitness defects detected for all other strains. According to our results, we hypothesize that
4 the differences seen in the fitness costs or growth of these strains are not necessarily
5 because the specific S531L *rpoB* mutation has a less severe effect on the overall fitness of
6 these bacteria, but due to the fact that this strain has other characteristics, allowing for a
7 better comparative capacity to use alternative energy sources, such as the fatty acids
8 described above. This hypothesis is supported by the fact that, although the
9 concentrations of the detected compounds serving as substrates for these alternative
10 energy pathways (specifically C17:0), were reduced in both the resistant strains
11 investigated, the reduction was more severe in the S531L mutant samples as compared to
12 the S522L mutant strain. This would also explain the higher concentrations of the
13 propionic acid derivative, benzenepropionic acid, detected in the S531L mutant strain, as
14 compared to that in the S522L mutant strain, as this is a product of the β -oxidation of odd-
15 chain and methyl-branched fatty acids.

16

17 **Conclusions**

18

19 As stated by Bergval et al. (2007): “Mutations in a gene as critical for bacterial survival as
20 *rpoB*, most likely have effects on many areas of cellular function” and “perhaps all strains
21 carrying mutations in *rpoB* have an altered expression of a wide range of genes”.
22 Considering this, this study is the first of its kind to use metabolomics to indicate the
23 effects of two *rpoB* mutations and the role of the β -subunit of RNA-polymerase, on the fatty
24 acid metabolism of rifampicin resistant *M. tuberculosis*. Furthermore, this study proves the
25 capacity of a metabolomics research approach to identify previously unknown metabolite
26 markers, never before associated with rifampicin-resistance.

1 Further prospects of these methods and findings suggest that by using the current
2 approach, which specifically targeted fatty acid alterations due to *rpoB* mutations, one
3 could similarly investigate other groups of metabolites (e.g. purines, pyrimidines, amino
4 acids, carbohydrates, fatty acids etc.) using other extraction methods, prior to
5 metabolomics data processing. This could potentially allow for the identification of
6 additional biomarkers and metabolic pathways, which may be influenced by the *rpoB* gene
7 region, which will ultimately broaden our understanding of rifampicin resistance and the
8 role of *rpoB* on all other aspects of the bacterial metabolome. Furthermore, many of the
9 hypothesized mechanistic alterations proposed by the newly identified metabolite markers
10 detected using this untargeted fatty acid metabolomics approach, could potentially also be
11 investigated in a more targeted manner e.g. the proposed reduced mycolic acids and
12 NTPs. This may in time lead to a better characterization of these *Mycobacterium* mutants
13 in a quest for improved diagnostics and treatments.

14

15 **Acknowledgements**

16

17 We thank the Royal Tropical Institute (KIT), Amsterdam, The Netherlands for providing
18 access to the bacterial samples, Hans de Ronde for performing bacterial culture, and
19 Richard Anthony for useful discussions.

20 The financial assistance of the National Research Foundation (NRF) of South Africa
21 towards this research is hereby acknowledged. Opinions expressed and conclusions
22 arrived at are those of the authors and are not necessarily to be attributed to the NRF.

23

24 **Author Disclosure Statement**

25

26 All authors declare that there are no conflicts of interest.

27

References

- AHMAD, S., ARAJ, G.F., AKBAR, P.K., FARES, E., CHUGH, T.D., and MUSTAFA, A.S. (2000). Characterization of *rpoB* mutations in rifampin-resistant *Mycobacterium tuberculosis* isolates from the Middle East. *Diagn Microbiol Infect Dis* **38**, 227–232.
- ANTHONY, R.M., SCHUIITEMA, A.R.J., BERGVAL, I.L., BROWN, T.J., OSKAM, L., and KLATSER, P.R. (2005). Acquisition of rifabutin resistance by a rifampicin resistant mutant of *Mycobacterium tuberculosis* involves an unusual spectrum of mutations and elevated frequency. *Ann. Clin Microbiol Antimicrob* **4**, 9.
- AZIZ, M. A., WRIGHT, A., LASZLO, A. DE MUYNCK, A., PORTAELS, F., VAN DEUN, A., et al. (2006). Epidemiology of antituberculosis drug resistance (the Global Project on Anti-tuberculosis Drug Resistance Surveillance): an updated analysis. *Lancet* **368**, 2142–54.
- BARAN, R., REINDL, W., and NORTHEM, T.R. (2009). Mass spectrometry based metabolomics and enzymatic assays for functional genomics. *Curr Opin Microbiol* **12**, 547–552.
- BARRY, C.E. III, LEE, R.E., MDLULI, M., SAMPSON, A.E., SCHROEDER, B.G., SLAYDEN, R.A., and YUAN, Y. (1998). Mycolic acids: structure, biosynthesis and physiological functions. *Prog. Lipid Res.* **37**, 143-179.
- BERGVAL, I.L., KLATSER P.R., SCHUIITEMA, A.R.J., OSKAM, L., and ANTHONY, R.M. (2007). Specific mutations in the *Mycobacterium tuberculosis rpoB* gene are associated with increased *dnaE2* expression. *FEMS Microbiol Lett* **275**, 338–343.
- CHATTERJEE, D. (1997). The mycobacterial cell wall: structure, biosynthesis and sites of drug action. *Curr Opin Chem Biol* **1**, 579-568.
- CHEN, S. (2011). An “Omics” Approach to Determine the Mechanisms of Acquired Aromatase Inhibitor Resistance. *OMICS* **15**, 347-352.

1 DEBNATH, M., PANDEY, M., and BISEN, P.S. (2011). An omics approach to understand the plant abiotic
2 stress. OMICS **15**, 739-762.

3

4 DE CARVALHO, L.P.S., FISCHER, S.M., MARRERO J., NATHAN, C., EHRT, S., and RHEE, L.Y. (2010).
5 Metabolomics of *Mycobacterium tuberculosis* reveals compartmentalized co-catabolism of carbon
6 substrates. Chem Biol **17**, 1122-1131.

7

8 FIEHN, O., KOPKA, J., DÖRMANN, P., ALTMANN, T., TRETHEWEY, R.N., and WILLMITZER, L. (2000).
9 Metabolite profiling for plant functional genomics. Nature Biotechnol **18**, 1157-1161.

10

11 FULCO, A.J., and BLOCH, K. (1964). Cofactor Requirements for the Formation of Δ^9 -Unsaturated Fatty
12 Acids in *Mycobacterium phlei**. J Biol Chem **239**, 993-997.

13

14 GAGNEUX, S., LONG, C.D., SMALL, P.M., VAN, T., SCHOOLNIK, G.K., and BOHANNAN, B.J.M. (2006).
15 The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science **312**, 1944-1946.

16

17 GARRETT R.H., and GRISHAM C.M. (2005). *Biochemistry*, 3rd edn. Thomas Brooks/Cole. Belmont, CA. p.
18 860-862.

19

20 GLICKMAN, M.S., and JACOBS, W.R, Jr. (2001). Microbial Pathogenesis of *Mycobacterium tuberculosis*:
21 Dawn of a discipline. Cell **104**, 477–485.

22

23 GONZÁLEZ, B., PAJARES, M.A., HERMOSO, J.A., GUILLERM, D., GUILLERM, G., and SANZ-APARICIO,
24 J. (2003). Crystal structures of methionine adenosyltransferase complexed with substrates and products
25 reveal the methionine-ATP recognition and give insights into the catalytic mechanism. J Mol Biol **331**, 407–
26 416.

27

28 JAIN, M., PETZOLD, C.J., SCHELLE, M.W., LEAVELL, M.D., MOUGOUS, J.D., BERTOZZI, C.R., et al.
29 (2007). Lipidomics reveals control of *Mycobacterium tuberculosis* virulence lipids via metabolic coupling.
30 PNAS **104**, 5133–5138.

31

32 KEGG. Kyoto Encyclopedia of Genes and Genomes. (2011). <http://www.genome.jp/kegg/pathway>.

1

2 KOEKEMOER G., and SWANEPOEL J.W.H. (2008). A semi-parametric method for transforming data to
 3 normality. *Stat Comput.* **18**, 241-257.

4

5 MADSEN, R., LUNDSTEDT, T., and TRYGG, J. (2010). Chemometrics in metabolomics - A review in human
 6 disease diagnosis. *Anal Chim Acta* **659**, 23–33.

7

8 MARIAM, D.H., MENGISTU, Y., HOFFNER, S.E., and ANDERSSON, D.I. (2004). Effect of *rpoB* mutations
 9 conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **48**,
 10 1289–1294.

11

12 MUÑOZ-ELÍAS, E.J., UPTON, A.M., CHERIAN, J., and MCKINNEY, J.D. (2006). Role of the methylcitrate
 13 cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence. *Mol Microbiol* **60**, 1109–
 14 1122.

15

16 NIGOU, J., GILLERON, M., and PUZO, G. (2003). Lipoarabinomannans: from structure to biosynthesis.
 17 *Biochimie* **85**, 153–166.

18

19 OLIVIER, I., and LOOTS, D.T. (2011). An overview of tuberculosis treatments and diagnostics. What role
 20 could metabolomics play? *J Cell Tissue Res* **11**, 2655–2671.

21

22 OLIVIER, I., and LOOTS, D.T. (2012). A metabolomics approach to characterise and identify various
 23 *Mycobacterium* species. *J Microbiol Meth* **88**, 419-426.

24

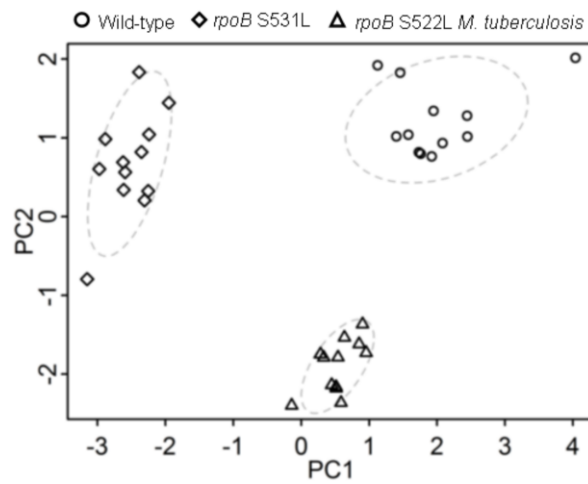
25 RAMAKRISHNAN, T. MURTHY, P.S., and GOPINATHAN, K.P. (1972). Intermediary metabolism of
 26 mycobacteria. *Bacteriol Rev* **36**, 65-108.

27

28 REZWAN, M., LANÉELLE, M., SANDER, P., and DAFFÉ M. (2007). Breaking down the wall: Fractionation of
 29 mycobacteria. *J Microbiol Methods* **68**, 32-39.

30

1 SCHNACKENBERG, L.K., JONES, R.C., THYPARAMBIL, S., TAYLOR, J.T., HAN, T., TONG, W., et al.
2 (2006). An integrated study of acute effects of valproic acid in the liver using metabonomics, proteomics, and
3 transcriptomics platforms. OMICS **10**, 1-14.
4
5 SHI, L., SOHASKEY, C.D., PFEIFFER, C., DATTA, P., PARKS, M., MCFADDEN, et al. (2010). Carbon flux
6 rerouting during *Mycobacterium tuberculosis* growth arrest. Mol Microbiol **78**, 1199–1215.
7
8 WEHRLI, W. (1983). Rifampin: Mechanisms of Action and Resistance. Rev Infect Dis **5**, S407-S411.
9
10 WHO. (2010). Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 global report on surveillance
11 and response. Geneva, Switzerland.
12



1
2 **FIG. 1.** PCA scores plot showing principal component 1 vs. principal component 2 of the
3 *M. tuberculosis* wild-type parent strain and the two rifampicin-resistant conferring *rpoB*
4 mutants (12 samples of each strain), subsequent to fatty acid extraction and GC-MS
5 analyses, showing differentiation of the 3 *M. tuberculosis* strains due to the variation in the
6 underlying metabolite profiles of these strains. The total amount of variance explained by
7 the first three PCs (R^2X cum) was 65.37 %, of which PC 1 explained 29.66 %, PC 2
8 explained 16.38 %, and PC 3 explained 10.33.

9

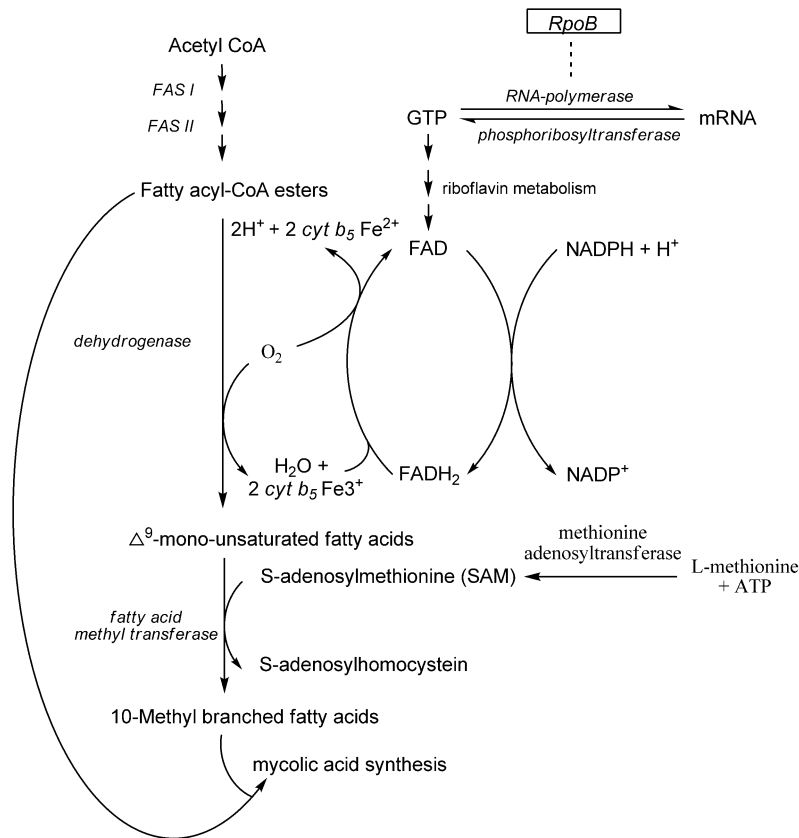


FIG. 2. The role of the *rpoB* gene in the biosynthesis of 10-methyl-branched fatty acids in *Mycobacterium*. FAS I and FAS II generate fatty acyl-CoA esters from acetyl-CoA, which are dehydrogenated to their corresponding Δ^9 -mono-unsaturated straight-chain fatty acids, and ultimately methylated to form 10-methyl branched-chain fatty acids, with S-adenosylmethionine acting as the methyl donor. The dehydrogenation step requires Fe^{2+} , GTP derived FMN, NADPH and O_2 . The availability of GTP for this process is dependent on its equilibrium with mRNA, regulated by phosphoribosyltransferase and RNA polymerase (*rpoB* expression).

1 **TABLE 1. MEAN RELATIVE CONCENTRATIONS ($\mu\text{g ml}^{-1}$ SAMPLE) AND RANKINGS OF METABOLITE MARKERS,**
2 **AS IDENTIFIED BY PCA AND PLS-DA. STANDARD DEVIATIONS (SD) ARE GIVEN IN PARENTHESES.**

PCA ranking	PLS-DA ranking	Compound	Wild-type <i>M. tuberculosis</i>		Rifampicin-resistant mutants			
					<i>rpoB</i> S522L		<i>rpoB</i> S531L	
					Mean concentration (SD)			
1	1	C17:0	2.01	(0.18)	1.4	(0.09)	T	
2	7	10Me-C16:0	0.20	(0.02)	0		0	
3	2	Unknown 422	1.63	(0.10)	0.17	(0.02)	0.76	(0.10)
4	6	Unknown 394	0.22	(0.02)	0		0.19	(0.03)
5		10Me-C15:0	0.04	(0.003)	0		0	
6		C22:0	0.12	(0.009)	0.10	(0.02)	0.24	(0.04)
7	8	C20:0	0.94	(0.04)	0.69	(0.02)	1.01	(0.08)
8	9	C24:0	0.33	(0.05)	0.28	(0.04)	0.54	(0.06)
9		C18:2n6c	0.11	(0.03)	0.19	(0.02)	0.21	(0.11)
10		C14:0	0.82	(0.09)	0.79	(0.15)	0.98	(0.12)
11		Unknown 459	0.09	(0.01)	0.10	(0.01)	0.16	(0.03)
12		C14:0, 1-ethyl,ethylester	0.98	(0.06)	0.98	(0.06)	0.78	(0.05)
	3	TBSA	6.71	(0.95)	8.75	(0.87)	9.79	(1.52)
	4	C16:0	19.48	(1.40)	25.68	(1.14)	25.41	(2.19)
	5	C26:0	1.21	(0.32)	0.9	(0.08)	1.46	(0.17)
	10	BPA	1.78	(0.21)	1.54	(0.29)	2.00	(0.30)
	11	C16:1 ω7c	1.43	(0.26)	1.74	(0.10)	1.45	(0.20)
	12	C18:1 ω8c	13.86	(1.94)	15.79	(0.84)	14.83	(1.50)

BPA: Benzenepropanoicacid, Me: Methyl, T: trace amounts

APPENDIX A (4)

Manuscripts

OLIVIER, I., LOOTS, D.T. 2012. A comparison of two extraction methods for differentiating and characterising various *Mycobacterium* species and *Pseudomonas aeruginosa* using GC-MS metabolomics. *Afr. J. Microbiol. Res.* 6(13): 3159-3172.

Full Length Research Paper

A comparison of two extraction methods for differentiating and characterising various *Mycobacterium* species and *Pseudomonas aeruginosa* using GC-MS metabolomics

Ilse Olivier and Du Toit Loots*

School for Physical and Chemical sciences, Centre for Human Metabonomics, North-West University,
Potchefstroom, 2520, South Africa.

Accepted 23 February, 2012

We investigated the capacity of a metabolomics research approach to characterise and differentiate between various infectious *Mycobacterium* species and *Pseudomonas aeruginosa*, and compared two extraction procedures; 1) extracting the fatty acid metabolome, and 2) extracting the total metabolome, prior to gas chromatography mass spectrometry (GC-MS) and statistical data analyses. Both extraction procedures, as part of a metabolomics study, were able to successfully differentiate between all bacterial groups investigated. The total metabolome extraction method proved the better of the two methods due to its comparative: simplicity; speed (taking less than 4 h), repeatability; extraction capacity (considering the range of compounds extracted and their relative concentrations), and; ability to extract those compounds which allow a better differentiation and characterisation of the investigated sample groups.

Key words: Tuberculosis, *Mycobacterium*, metabolomics, gas chromatography mass spectrometry (GC-MS), biosignature.

INTRODUCTION

Although tuberculosis (TB) was declared a global public health emergency by the World Health Organization (WHO) almost 20 years ago, it is still considered a major health threat today, with 9.4 million newly diagnosed cases and 1.8 million TB related deaths reported worldwide in 2008 alone (WHO, 2010). Although this is a universal epidemic, Africa (30%) and Asia (55%) account for 85% of all global TB cases, and when evaluating recorded adult deaths in low- and middle-income countries, TB is ranked third, after HIV/AIDS and ischemic heart disease (WHO, 2010).

Although *Mycobacterium tuberculosis* is the primary cause of TB, other mycobacterial strains may also result in opportunistic infections in humans and are frequently

encountered in clinical specimens (Miguez-Burbano, 2006). These species include: *Mycobacterium avium*, which has the potential to cause pulmonary infection in patients with chronic lung disease; *Mycobacterium kansasii*, which may cause skin and soft tissue infections, skeletal infections, lung infections, surgical site infections and disseminated disease (Davis, 2007) and; *Mycobacterium bovis*, the organism responsible for causing TB in cattle (Todar, 2005). Additionally, other lung pathogens, such as *Pseudomonas aeruginosa*, a genus from the phylum Proteobacteria and a common cause of iatrogenic infections, may also cause symptoms similar to that of pulmonary TB, when intruding the respiratory passages of patients with immuno-compromising illnesses (Todar, 2005).

The relatively new research field of metabolomics encompasses 'the non-biased identification and quantification of all the metabolites in a biological system', using highly sensitive analytical procedures (Dunn et al., 2005).

*Corresponding author. E-mail: dutoit.loots@nwu.ac.za. Tel: +27 (0)18 299 1818. Fax: + 27 (0)18 299 2316.

Lately, metabolomics has successfully been used as a tool to characterise a variety of disease states, in order to better understand the underlying mechanisms of these diseases (Schoeman and Loots, 2011). Only one metabolomics study to date, using electronic nose metabolomics analyses, was capable of differentiating between various infectious *Mycobacterium* species (*M. tuberculosis*, *M. avium*, and *Mycobacterium scrofulaceum*) and *P. aeruginosa* (Fend et al., 2006). However, this approach, using sensory array detection, isn't capable of identifying those compounds best explaining the variation detected, and hence, no marker metabolites characterising these species were determined.

Consequently, we developed and compared two rapid extraction procedures and investigated their respective capacities to extract those compounds which best differentiate the above mentioned infectious *Mycobacterium* species and *P. aeruginosa*, using a GC-MS metabolomics approach. Given that the fatty acid composition of these bacterial species is a well-known characteristic feature, which may potentially differentiate these groups (Lambert et al., 1986; Mosca et al., 2006), the first extraction procedure developed was aimed at selectively isolating these fatty acids. The second extraction procedure, comparatively investigated, was however aimed at extracting a far greater variety of metabolites, considering that the definition of metabolomics is the "unbiased" identification of "all" metabolites present in a biological system (Dunn et al., 2005). Furthermore, using multivariate statistical methods, those compounds contributing most to the variation between the infectious bacterial groups, as detected via each of the extraction methods investigated, prior to GC-MS analyses, were identified. The biological importance of these compounds was then discussed, in order to further validate their use for metabolomics applications. These potential metabolite makers may be of great scientific value, considering their ability to better characterise these organisms and their potential to be used diagnostically.

MATERIALS AND METHODS

Chemicals and reagents

N-methyl-N-(tert.-butyldimethylsilyl)-trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS), pyridine, potassium hydroxide (KOH), methyl-nonadecanoic acid (Me-C19), glacial acetic acid, and trimethylpentane were purchased from Merck (Darmstadt, Germany). Methoxyamine hydrochloride and 3-phenyl butyric acid were purchased from Sigma-Aldrich (St. Louis, Mo., USA). All organic solvents used were ultra purity Burdick & Jackson brands (Honeywell International Inc., Muskegon, MI, USA) and were used without any further purification.

Cultures

All the cultured organism samples used in this study were supplied by the Royal Tropical Institute, Amsterdam, Netherlands. All the

bacteria (*M. tuberculosis*, *M. avium*, *M. bovis*, *M. kansasii*, and *P. aeruginosa*, 12 repeats of each) were cultured in Middlebrook 7H9 medium with oleic acid-albumin-dextrose-catalase enrichment. The bacteria were incubated at 37°C while shaking at 120-150 rpm until an optical density (420 nm) of 0.30 ($\approx 2 \times 10^8$ bacteria/ml) was reached. The cells were washed once with PBS, collected via centrifugation, snap frozen and stored at -80°C until extraction and GC-MS metabolomics analysis. For determining the comparative species differentiation capacities of the two respective extraction methods (fatty acid metabolome and total metabolome extractions), samples were re-suspended in ddH₂O at a concentration of 1×10^8 bacteria/ml, prior to extraction, GC-MS analysis and statistical data processing.

Extraction procedure 1 – Fatty acid metabolome extraction

An extraction solvent mixture (1.25 ml), consisting of chloroform, methanol, and ddH₂O, in a ratio of 1:3:1, respectively, was added to 250 µl of the above mentioned sample suspensions in a microcentrifuge tube. The extraction was performed using an MM 400 vibration mill (Retsch GmbH and co. KG, Haan, Germany) at a frequency of 30 Hz for 5 min with a 3 mm tungsten carbide bead (Retsch GmbH and co. KG) for increased extraction efficiency. Chloroform and ddH₂O (250 µl of each) were subsequently added to the extract, followed by centrifugation for 5 min at 550 x g. The organic phase was collected, transported to a kimax tube, and dried under a light stream of nitrogen. Thereafter, 0.5 ml of chloroform, 0.5 ml of methanol, and 1 ml of methanolic KOH (0.2 M) was added to the dried extract. The mixture was then incubated for 30 min at 60°C followed by the addition of 2 ml of hexane, 200 µl of glacial acetic acid (1 M), and 2 ml of ddH₂O. After centrifugation at 550 x g for 5 min, the organic phase was once again collected and transferred to a new kimax tube. The remaining water phase was re-extracted 3 times with hexane and the combined organic phases from each of the previous extractions were dried under nitrogen and re-suspended in 50 µL of external standard (Me-C19) dissolved in trimethylpentane (15.6 ng/ml).

Extraction procedure 2 - Total metabolome extraction

As an internal standard, 50 µl of 3-phenyl butyric acid (26.25 mg/ml) was added to 250 µl of the above mentioned sample suspensions in a microcentrifuge tube, followed by the addition of 1.25 ml of an extraction solvent mixture containing chloroform, methanol, and ddH₂O, in a 1:3:1 ratio. The extraction was again performed using an MM 400 vibration mill at a frequency of 30 Hz, for 5 min, after adding a 3 mm tungsten carbide bead to each tube for increased extraction efficiency. After centrifugation, the supernatant was collected, transferred to a GC-MS sample vial and dried under a light stream of nitrogen. The dried supernatant was subsequently derivatized with 50 µl of methoxyamine hydrochloride in pyridine (15 mg/ml) at 50°C for 90 min. Hereafter, the extract was trimethylsilylated with 50 µl of MSTFA with 1% TMCS at 50°C for 60 min.

GC-MS parameters

The above mentioned prepared extracts were analysed by injecting 1 µl on an Agilent 7890A gas chromatograph (Agilent, Atlanta, GA) coupled to an Agilent 5975 mass selective detector equipped with a 7683B injector, 7683 auto sampler and VF1-MS capillary column (30 m x 250 µm i.d., 0.25 µm film thickness) in the splitless mode. Helium was used as the carrier gas and the pressure was programmed as such that the helium flow was kept constant at 1.2 ml/min. Compounds were detected using the MS in full scan mode

(m/z 50-550 for the fatty acid extracts and m/z 60-800 for the total metabolite extracts). For the fatty acid metabolome extracts, the most optimal analytical conditions were an injector temperature of 250°C, for the total run time, and an initial GC oven temperature of 50°C for 1 min followed by a primary increase in oven temperature of 20°C/min to 160°C. This was then followed by an increase in oven temperature of 10°C/min to 190°C and then a further increase of 5°C/min to a final temperature of 300°C, at which the temperature was maintained for 5 min. For the total metabolome extracts, the most optimal analytical conditions were an injector temperature of 270°C, for the total run time, and an initial GC oven temperature of 70°C for 2 min, followed by an increase in oven temperature of 4°C/min to a final temperature of 300°C, at which the temperature was held for 2 min.

Data-acquisition

The raw GC-MS data were deconvoluted and analysed using AMDIS software (Automated Mass Spectral Deconvolution and Identification System, V2.65). Alignment of the detected compounds across the samples analysed was achieved by creating a new reference library in AMDIS, which contained the mass spectra of all the compounds detected above a threshold of 0.01% of the total signal, for all the samples analysed. Each analysed sample was subsequently processed using the aforementioned reference library, and the resulting output of each sample was combined into a data matrix containing the relative concentrations (normalised with the internal standard) for all compounds present or absent in each sample. This was done for each of the GC-MS data sets collected from each of the 2 extraction methods investigated.

Statistical data analysis

The statistical packages, "R" (version 2.13.0) and Statistica (version 10) were used for all the statistical data analyses. Data were normalised relative to the respective internal standards in order to compensate for sample loss during the extraction or chromatographic injection. This relative normalisation was used purely as a measure to identify variations in the abundances of the detected compounds between individual samples and sample groups, and not for absolute quantification purposes.

To prevent variables with high concentrations from dominating the multivariate statistical analyses, data pre-treatment using a non-parametric transformation function (Koekemoer and Swanepoel, 2008) was used to scale the data prior to statistical data analyses, after which mean centring was applied.

The repeatability of the 2 extraction methods investigated were compared using the distribution of the calculated coefficient of variation (CV) values of the relative concentrations of all the compounds detected subsequent to GC-MS analysis of the extracted sample repeats.

In order to determine whether or not a natural grouping (differentiation) exists between the investigated sample groups, an unsupervised classification procedure known as principal component analysis (PCA) was applied. PCA reduces the dimension of the input data matrix by using a weighted sum of the compound concentrations, hereby summarising the variation in the data matrix into a model plane, generating scores. A scatter plot of these scores provides an overview of the samples and how they relate to each other.

Partial least squares discriminant analysis (PLS-DA) is a supervised classification procedure, where class membership information is used to build a discrimination model, and may therefore be used to identify those compounds which best characterise the differentiated sample groups, by ranking the metabolites according to the variable influence on the projection

(VIP) parameter. VIP is a weighted sum of squares of the PLS-DA weights, indicating the importance of the metabolite to the total model, and consequently, the highest ranked compounds are then identified as metabolite markers for the respective sample groups (Madsen et al., 2010). Using the mass fragmentation patterns generated by the MS, together with their respective GC retention times, the identities or compound names for these metabolite markers were determined by using libraries generated from previously injected standards. Using these libraries, a value, representative of the degree of similarity between the mass spectra of the detected compound to that of a previously injected standard in the library, at a corresponding retention time, was then assigned to each compound. A similarity value higher than 80% was considered a positive identification and the corresponding compound name was assigned. All spectra with a similarity lower than 80% were identified as unknown compounds.

RESULTS AND DISCUSSION

Although the purpose of this investigation wasn't aimed at directly comparing any of the two described extraction methods to those conventionally used for similar purposes, it is of value to mention that the use of a vibration mill as described in this study, significantly improves the extraction efficiency of any extraction procedure, especially when extracting compounds from more robust matrixes (such as cell or tissues samples). This increased extraction efficiency allows for faster extraction times and the analyses of 24 samples simultaneously, using far less solvent volumes, in a 2 ml centrifuge tube.

The fatty acid extraction method, for instance, can be completed within 5 h, including the drying and derivitization steps, using a total solvent volume of only 1.75 ml, which is a vast improvement on the 16 h required for the completion of the original Bligh- Dyer method, which also requires far more solvent.

The absence of an organic solvent washing step for the specific extraction of fatty acids makes the total metabolome extraction method even simpler and faster, requiring even smaller amounts of solvent material (1.25 ml) for completion in less than 4 h.

Extraction capacity

The AMDIS generated reference library (a compilation all the compounds extracted from all 5 sample groups) for the fatty acid metabolome extraction procedure contained 270 compounds, predominantly fatty acids and various hydrocarbons while the AMDIS generated reference library for the total metabolome extraction procedure, contained a total 1194 compounds, of various polarities, belonging to a variety of compound classes, including amongst others: fatty acids, amino acids, alcohols, organic acids, monosaccharides, alkenes, alkanes, purines, pyrimidines, etc. Table 1 indicates the number of compounds extracted via each of the compared extraction methods, for each of the infectious disease sample groups. Considering this, the total metabolome extraction procedure detected, on average, 358 more

Table 1. Number of compounds in the AMDIS generated reference libraries originating from each of the bacterial sample groups (12 sample repeats of each), for the two investigated extraction methods. The standard deviation (SD) values, representing the variation in the number of compounds detected between individual samples, are given in parenthesis.

Bacterial sample group	Number of compounds (SD)	
	Fatty acid metabolome extraction method	Total metabolome extraction method
<i>M. tuberculosis</i>	180 (10.09)	461 (10.93)
<i>M. avium</i>	178 (6.12)	501 (13.24)
<i>M. bovis</i>	174 (13.12)	388 (14.40)
<i>M. kansasii</i>	191 (5.87)	641 (14.94)
<i>P. aeruginosa</i>	161 (11.31)	679 (19.66)
Total number of compounds in library	270	1194

compounds per sample group, which could be expected, as the fatty acid metabolome extraction method is a more targeted approach, and is not aimed at extracting all compounds present in the cell matrix.

It should be noted that this evaluation of the number of detected compounds is merely for descriptive purposes, and may not necessarily be an indication of which of the two described methods performs better for metabolomics applications and the identification of markers characterising the above mentioned infectious diseases. Although, as per definition of metabolomics, the total metabolome extraction procedure should better fulfil the required criteria for identifying "all compounds in an unbiased manner", the increased dimensionality of the resulting extract may potentially overwhelm the dimensionality limits of the GC-MS, leading to co-elution and poor repeatability and consequently larger variation or unwanted "noise" in the collected data, hampering its applications for metabolomics research purposes. Knowing this beforehand, however, GC-MS flow rates and temperature programming was optimised, in order to minimise the chance of this happening.

Repeatability

As previously mentioned, the repeatability of the 2 extraction methods investigated were compared using the distribution of the calculated CV values of the relative concentrations of all the compounds detected subsequent to GC-MS analysis of the extracted sample repeats (Figure 1).

Comparatively, for all the organism groups, an averaged $50 \pm 6.7\%$ of all the compounds detected, via the fatty acid extraction procedure, as compared to $62 \pm 4.6\%$ for that of the total metabolome extraction, fell below a 50% CV value. This indicates that the total metabolome extraction procedure, despite extracting a larger number and variety of compounds, with potentially larger variations in their comparative concentrations,

which could in theory overwhelm the dimensionality of the GC-MS with regards to dynamic range and chromatographic separation, still showed better comparative repeatability.

As the most optimal GC-MS conditions were selected for analysing the extracts derived from both extraction methods, the better repeatability of the total metabolome extraction procedure is most easily explained by its comparative simplicity, resulting in a far lower likelihood for analytical variation than the fatty acid metabolome extraction method, which requires an additional step. An interesting observation was that in both cases, the *M. bovis* sample group showed the lowest repeatability of all the bacterial sample groups analysed. When evaluating these results, one needs to keep in mind that the CV value of a compound is directly correlated to its detected concentration (the lower the concentration, the worse the CV) (Rocke and Lorenza, 1995). Using both extraction methods, all identified compounds were detected in significantly lower concentrations in *M. bovis*, as compared to the other bacterial species investigated, possibly explaining the resulting high CV values. *M. tuberculosis* and *M. avium* were extracted with the best comparative repeatability using the fatty acid and total metabolome extraction procedures, respectively.

Differentiation capacities

Considering the application of these methods to comparative metabolomics, not only is the repeatability and extraction capacity as described above of importance, but also the methods' capacity to extract those compounds most important for differentiating of the groups being compared, in order to answer the biological question at hand. Figure 2 represents examples of the total ion chromatograms obtained after GC analyses of each of the bacterial species investigated, showing a clear visual difference between the species using both of the tested extraction methods.

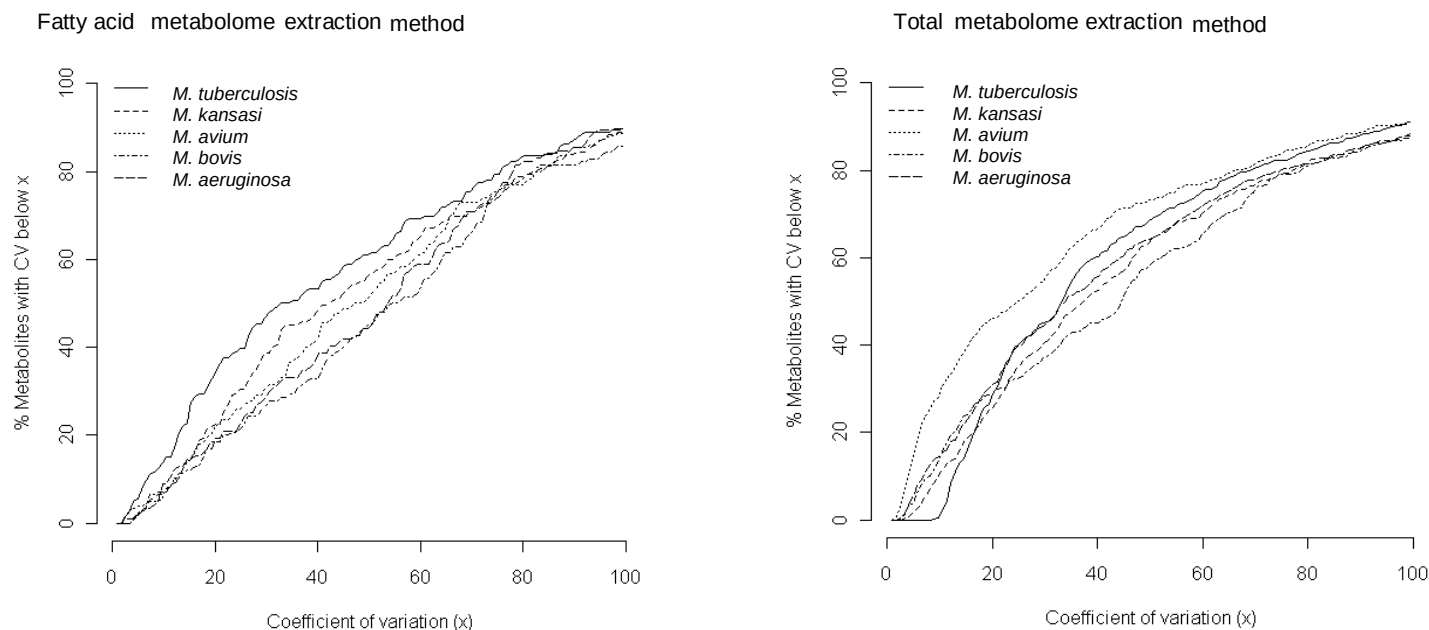


Figure 1. Distribution of the calculated coefficient of variation (CV) values for all the GC-MS detected compounds from each infectious organism group (12 sample repeats of each), extracted using the two extraction methods comparatively.

To determine the ability of the compared extraction procedures to extract those compounds which best explain the variance between the investigated infectious organism groups, two multivariate statistical data analyses (PCA and PLS-DA) were used to summarise and interpret the collected data. Prior to these analyses, a 50% filter was applied to the data collected, in order to exclude those compounds which do not appear in at least 50% of the samples, in one or more of the sample groups, as a data pre-processing step for eliminating unnecessary "noise" in the data.

The PCA scores plot using the fatty acid metabolome extraction procedure's GC-MS data is shown in Figure 3, and indicates a clear grouping of the individual samples of the respective infectious bacterial groups and a differentiation between these groups. Furthermore, all the related *Mycobacterium* species groups were grouped together, relative to the *P. aeruginosa* group, indicating that organisms with similar genetic characteristics group together due to the similarities in their genomes and resultant metabolite profiles.

Figure 4 illustrates the PCA scores plot for the GC-MS data acquired using the total metabolome extraction procedure. Similar to that of the fatty acid metabolome extraction method's PCA, the related *Mycobacterium* species groups were, once again, grouped together, relative to the *P. aeruginosa* group. This time however, a visibly smaller inter-sample, and hence, within group variation was detected, comparative to that of the fatty acid metabolome extraction procedure, resulting in a better differentiation of the infectious bacterial groups investigated. This smaller within group variation can be

explained by the comparatively better extraction repeatability of the total metabolome extraction procedure for those compounds which best described the variation between the infectious organism groups investigated.

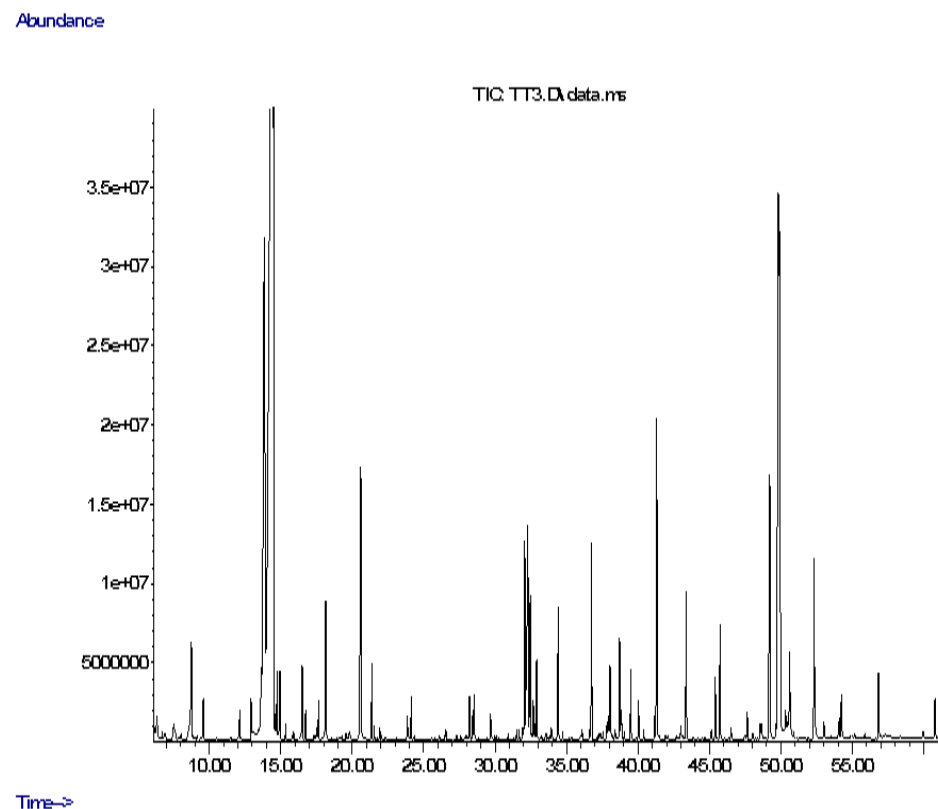
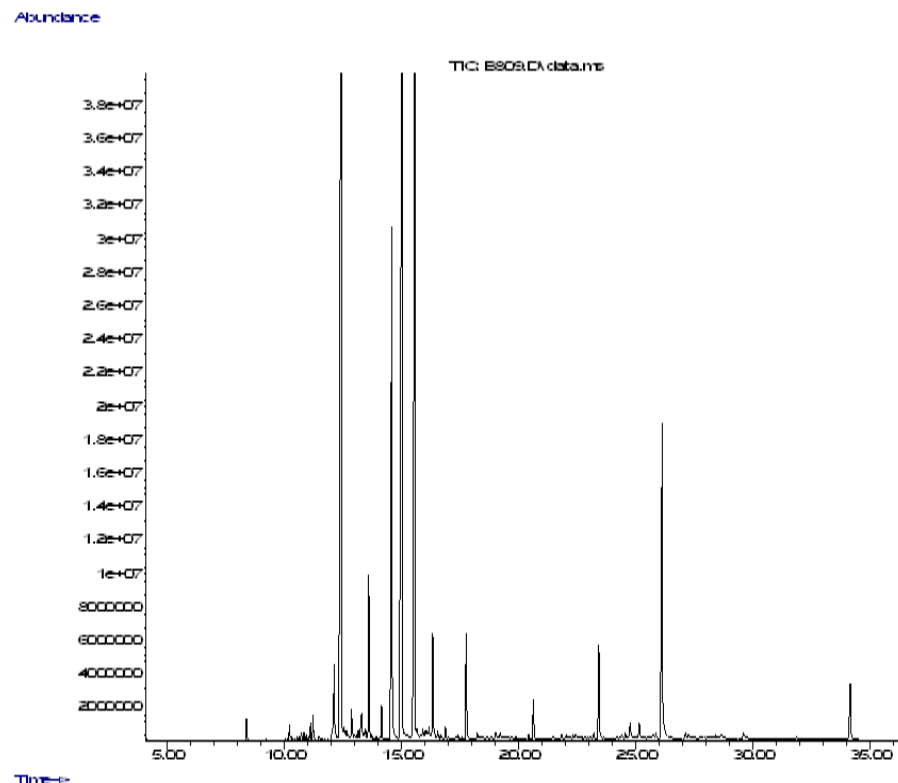
Metabolite marker identification

One of the primary functions of a metabolomics research approach is not only determining if a differentiation of various disease states exists on the basis of varying metabolite profiles, but also to identify those metabolite markers which may be used to better characterise the compared sample groups. Considering this, those metabolites with the highest VIP values (identified via PLS-DA), extracted by each of the investigated methods, were identified and evaluated from a functional biology perspective. This was done in order to ensure that the variation detected is in fact due to the extraction of those compounds of biological relevance. In Table 2, the mean relative concentrations of those metabolite markers identified using the fatty acid metabolome extraction approach are presented. The PLS-DA model used for identifying these had a modelling parameter R^2Y (cum) of 93.7%, indicative of the total explained variation of the response Y. Q^2 (cum), the cross-validated variation explained by the response Y, was 89.1%.

From the data in Table 2, it is clear that it is not only the compounds totally unique to a particular group which may be considered as important for differentiating these infectious organisms, but also those compounds which are common to two or more sample groups, showing

Fatty acid metabolome extraction method

Total metabolome extraction method

M. tuberculosis

constant differences in their concentrations.

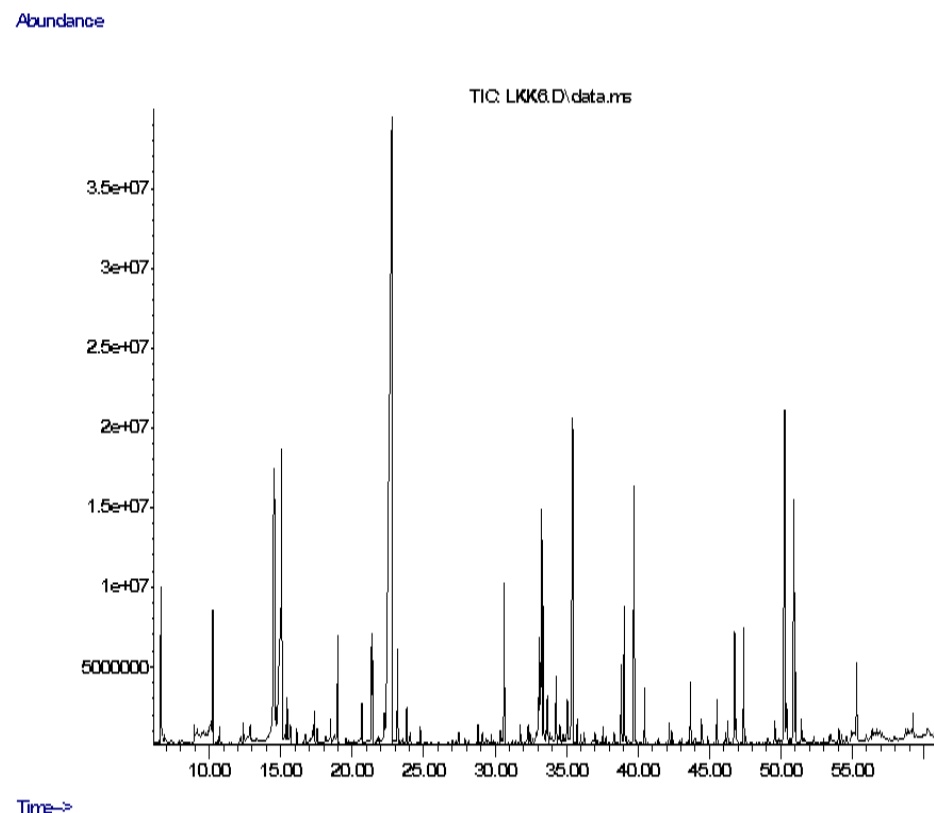
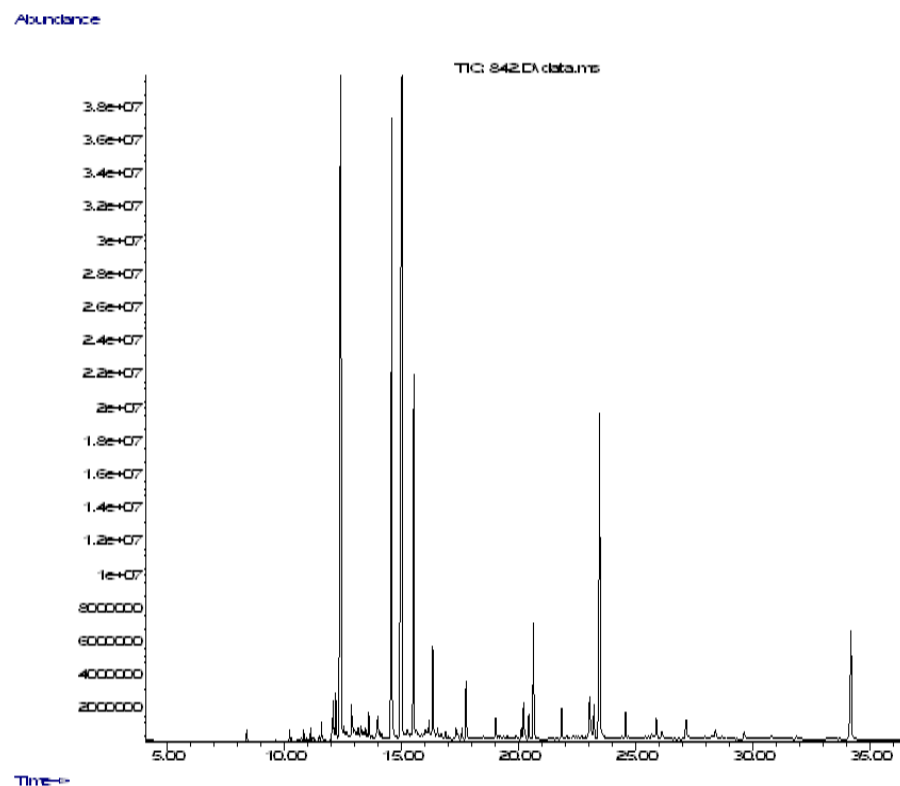
Of the two methods investigated, the fatty acid extraction procedure was selected based on the knowledge that *Mycobacterium* species contain rather unique fatty acids, which are involved in a number of structural and functional

processes, including their pathogenicity (Knisley et al., 1985). Consistent with previous reports, we identified the well known characteristic fatty acid, tuberculostearic acid (TBSA), as well as oleic acid (C18:1 ω 9c), which is known to be one of the most abundant fatty acids in these bacteria

(Lambert et al., 1986; Mosca et al., 2006), as novel metabolite markers for all *Mycobacterium* species groups (Larsson et al., 1987; Stopforth et al., 2004), confirming that this extraction procedure is well suited for characteristic marker metabolite identification, using a metabolomics

Fatty acid metabolome extraction method

Total metabolome extraction method

M. kansasii

approach. Additionally, the mycolic acid cleavage products (MACPs), eicosanoic acid (C20:0), docosanoic acid (C22:0), which are formed as a result of the chromatographic heat cleavage of the characteristic *Mycobacterium* cell wall components, mycolic acids (Guerrant et al., 1981),

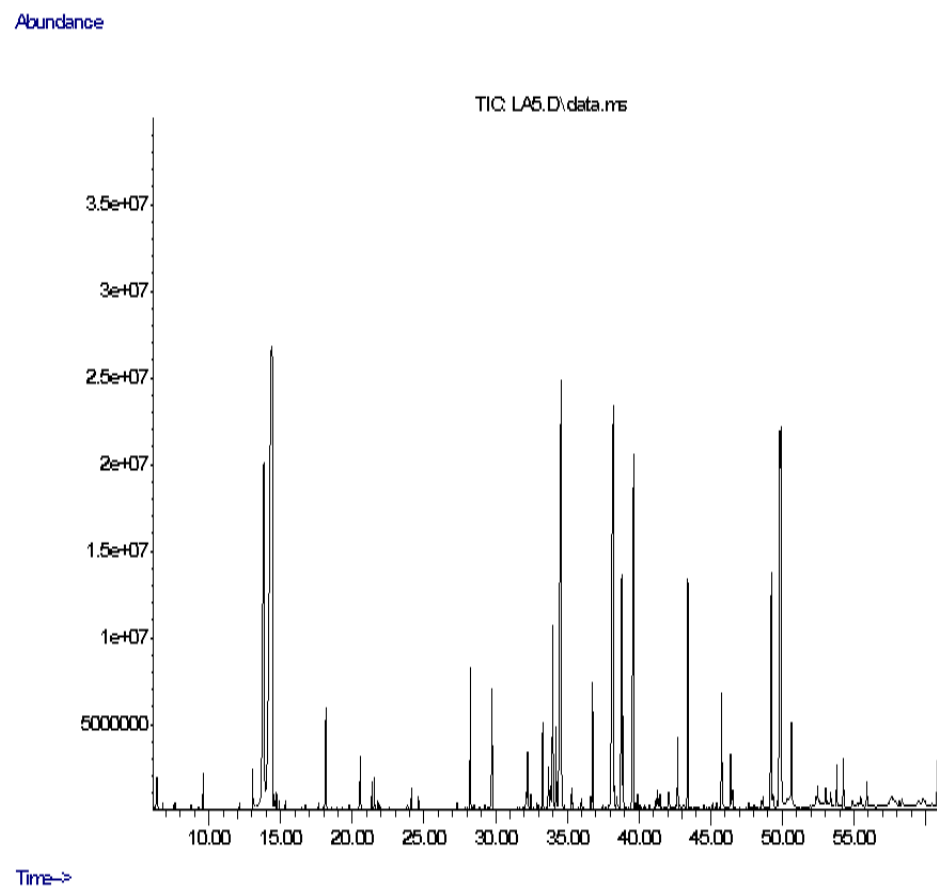
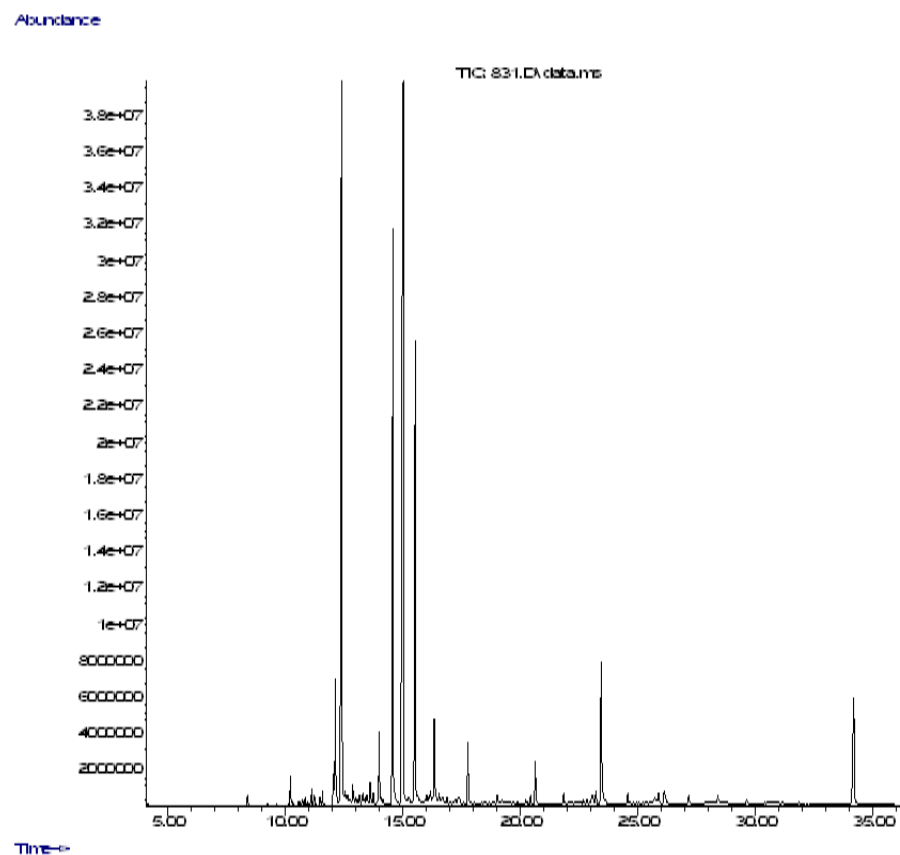
were detected in relatively abundant concentrations in all the *Mycobacterium* species, as opposed to only trace amounts in the *P. aeruginosa* sample group.

Another MACP, octacosanoic acid (C28:0), together with gondoic acid (C20:1 ω 9c) and two

unknown compounds with molecular masses of 466 and 494 respectively, may potentially characterise *Mycobacterium* species, as these were only detected in the *Mycobacterium* species tested, with the exception of *M. bovis*. All identified compounds were detected in significantly

Fatty acid metabolome extraction method

Total metabolome extraction method

M. avium

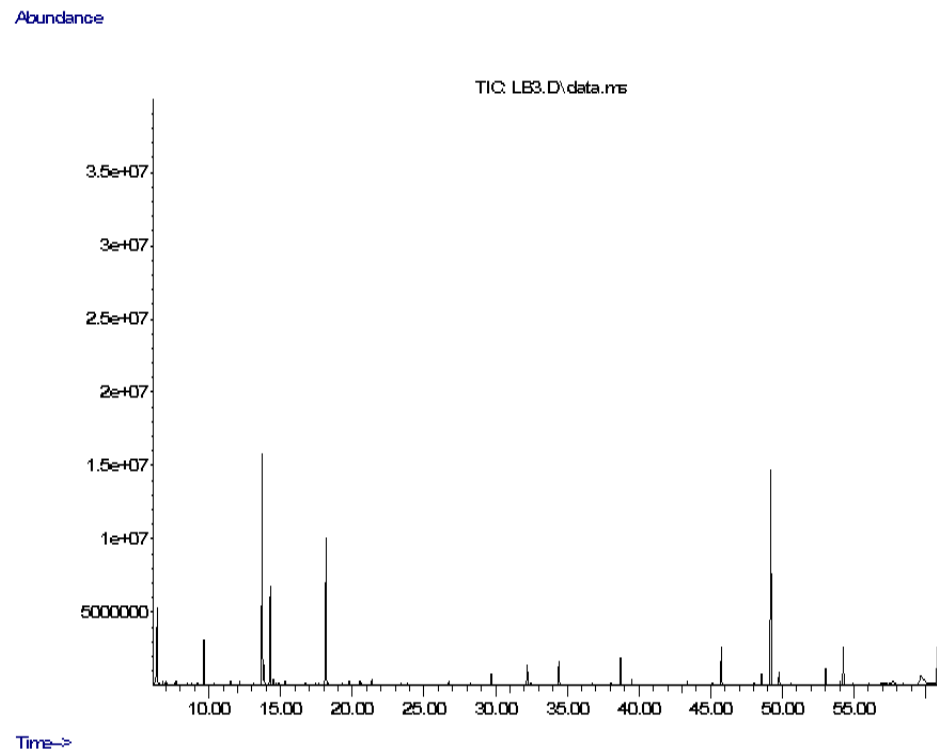
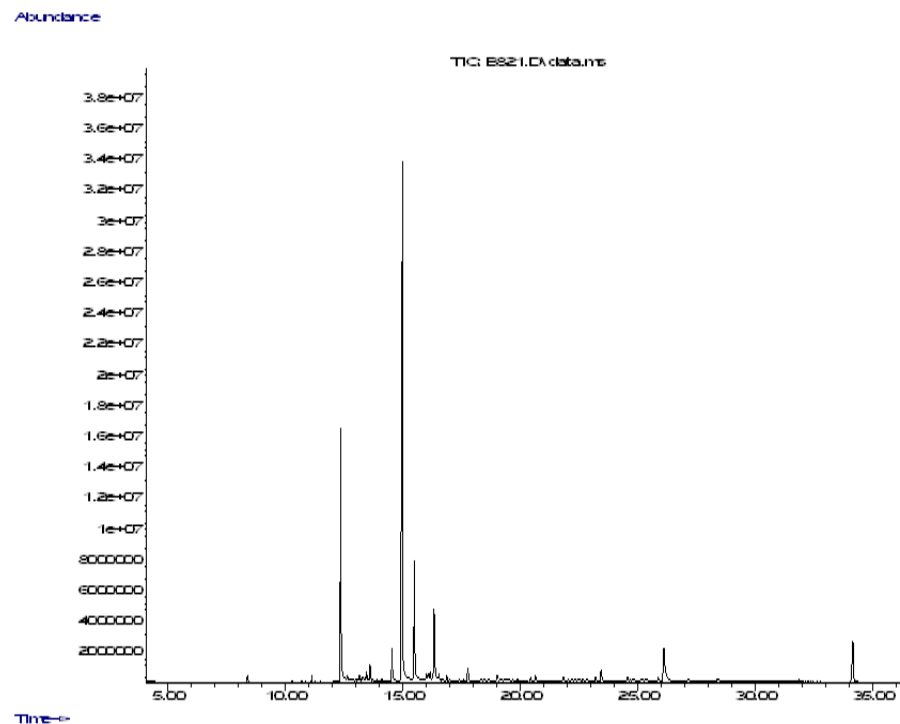
lower concentrations in the *M. bovis* sample group, in comparison to the other bacterial groups tested and, hence, these makers might potentially

be present in these samples, however at concentrations below the detection limit. Two unsaturated long-chain fatty acids, erucic acid

(C22:1 ω 9c) and nervonic acid (C22:1 ω 9c), were detected only in the *M. kansasii* and *M. avium* sample groups, whereas pentacosanoic acid

Fatty acid metabolome extraction method

Total metabolome extraction method

M. bovis

(C25:0) were detected exclusively in the *M. tuberculosis* and *M. bovis* sample groups. Furthermore, as previously identified, 2,4 dimethyl-tetradecanoic (2,4-DM C14:0) was detected as a novel metabolite for *M. kansasii* (Lambert et al., 1986; Mosca et al., 2006)

We additionally detected another unknown compound, with a molecular mass of 340,

specifically characterising *M. tuberculosis* and two compounds, 2-octyl-cyclopropaneoctanoic acid (2-octyl-CPOA) and 2-hexyl-cyclopropaneoctanoic acid (2-hexyl-CPOA) as unique characteristic metabolites of *P. aeruginosa*.

The PLS-DA model used for identifying the characteristic marker metabolites extracted using the total metabolome extraction procedure had a

modelling parameter R^2Y (cum) of 97.8%, with a Q^2 (cum) of 98.8%. This extraction method was developed for the purpose of extracting the entire metabolome and consequently the metabolites identified included various fatty acids, amino acids, alcohols, organic acids, monosaccharides, alkenes, alkanes, purines, pyrimidines, etc. These metabolite markers identified are listed in Table 3.

Fatty acid metabolome extraction method

Total metabolome extraction method

P. aeruginosa

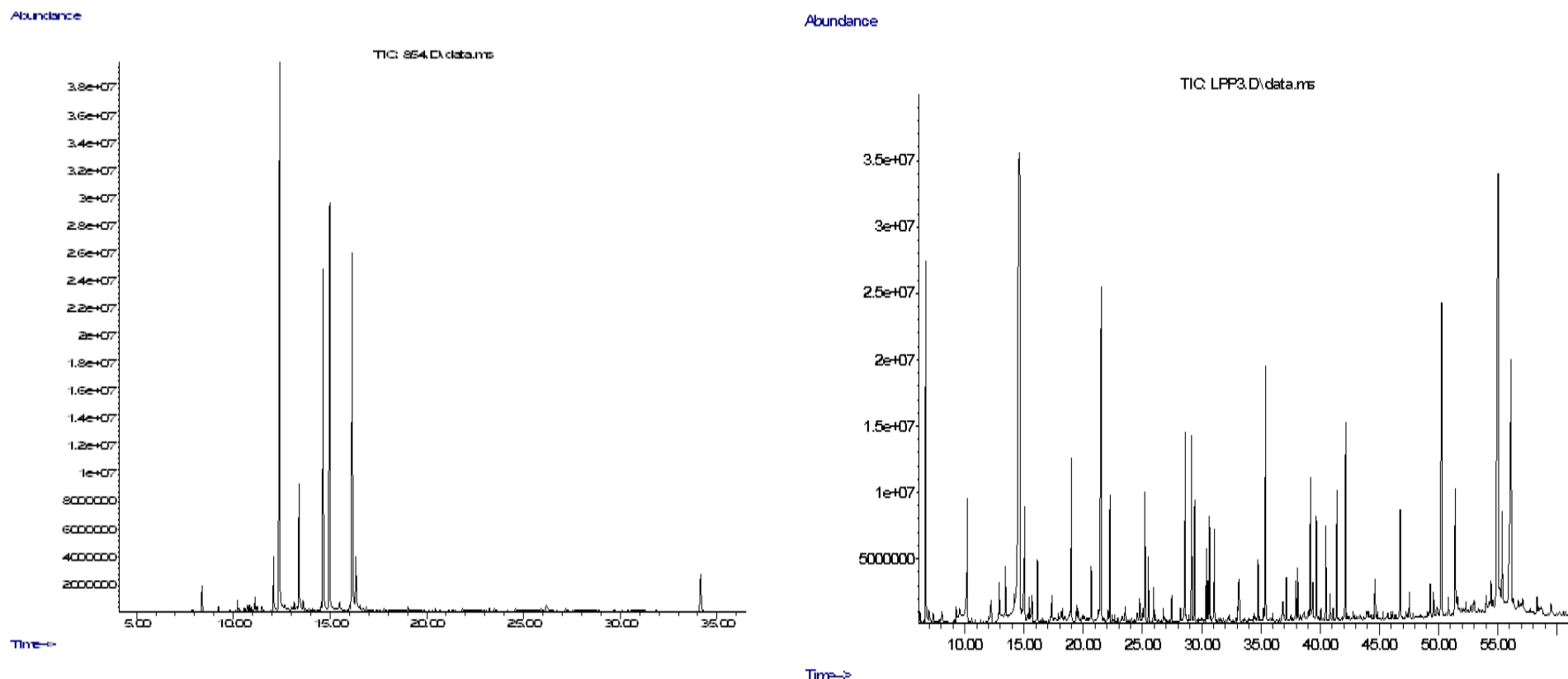


Figure 2. Examples of the total ion chromatograms of the bacterial species investigated, obtained after chromatographic analyses following sample extraction using the two tested extraction methods.

Although both of the investigated methods have the capacity to extract fatty acids, a more

complete fatty acid profile was attained using the fatty acid metabolome extraction approach, due to

the use of more selective solvents. The total metabolome extraction procedure, on the other

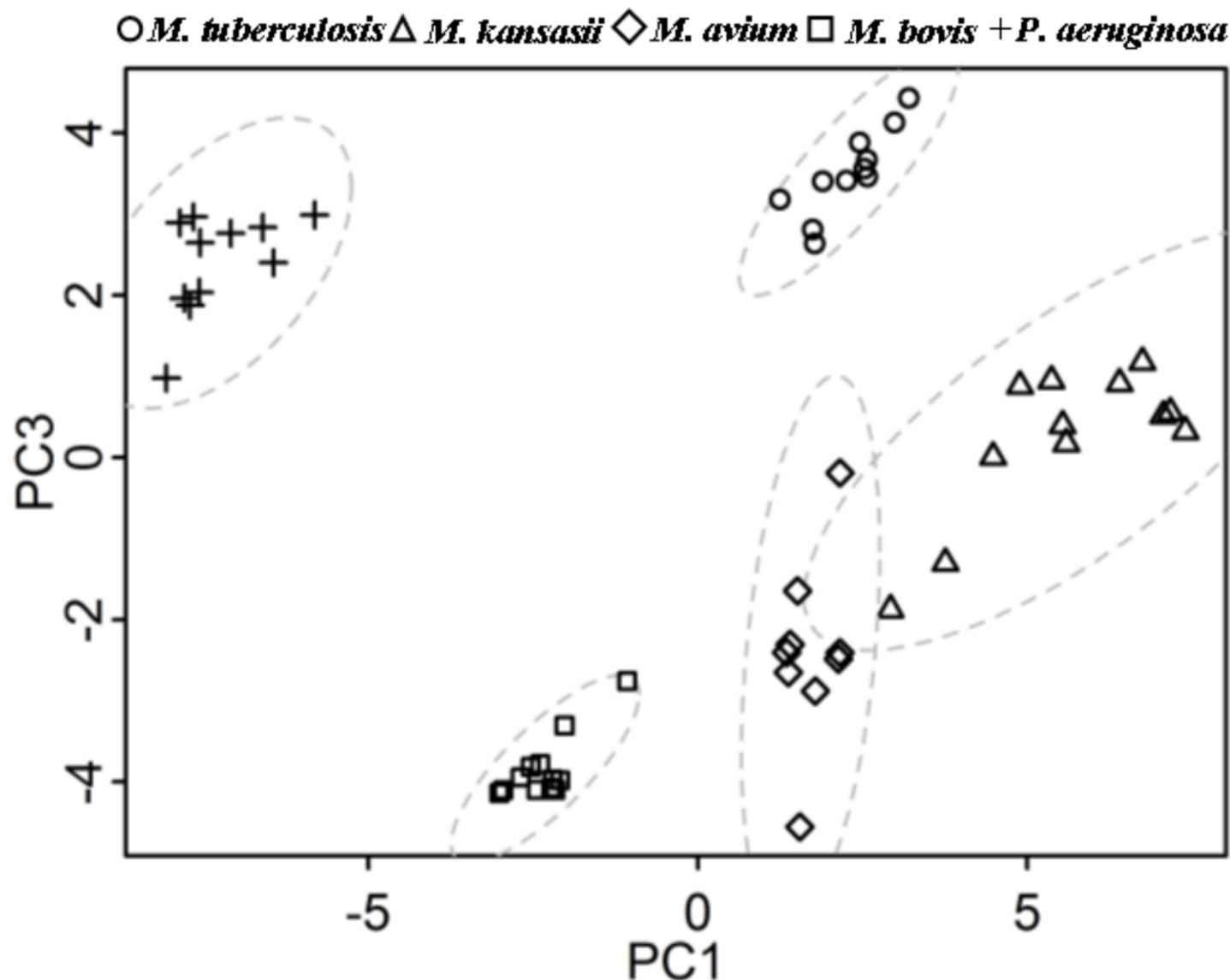


Figure 3. PCA scores plot (PC1 vs. PC3), of the GC-MS data acquired using the fatty acid metabolome extraction method, indicating a grouping of the individual samples into their respective organisms groups on the basis of the variation in the detected fatty acid metabolome. Three PCs were extracted, explaining a total of 64.4% of the variation in the data (R^2X cum), of which PC 1 explained 34.8%, PC 2 explained 15.7%, and PC 3 explained 13.9%. The ellipses represent the 95% confidence interval of the modelled variation.

hand, has the capacity to extract a larger variety of compounds from all compounds classes, including fatty acids, and hence has the advantage of detecting a greater spectrum of characteristic markers. When investigating mycobacteria, however, this may be seen as a minor limitation, due to the fact that some characteristic fatty acids might not be detected in these complex chromatograms. Interestingly, only palmitoleic acid (C16:1 ω 7c) was identified as a common metabolite marker using both extraction methods, and universally occurred in the highest comparative concentration in the *M. avium* bacterial sample group, and in the lowest concentration in the *M. bovis* sample group. The highest ranked

metabolite marker extracted using the total metabolome extraction procedure, and novel to *P. aeruginosa*, was indole-acetic acid. This metabolite has previously been identified in this organism and in other related plant growth-promoting rhizobacteria (Karnwal, 2009). Other novel metabolite markers identified *P. aeruginosa* using this extraction method included: cadaverine, an intermediate in the alternative decarboxylation and catabolism of L-lysine (Fothergill and Guest, 1977; Stewart, 1970); putrescine, an intermediate used in the L-arginine decarboxylase catabolic pathway (Mercenier et al., 1980); purine; valeric acid; and two unknown compounds with molecular masses of 343 and 373. Two

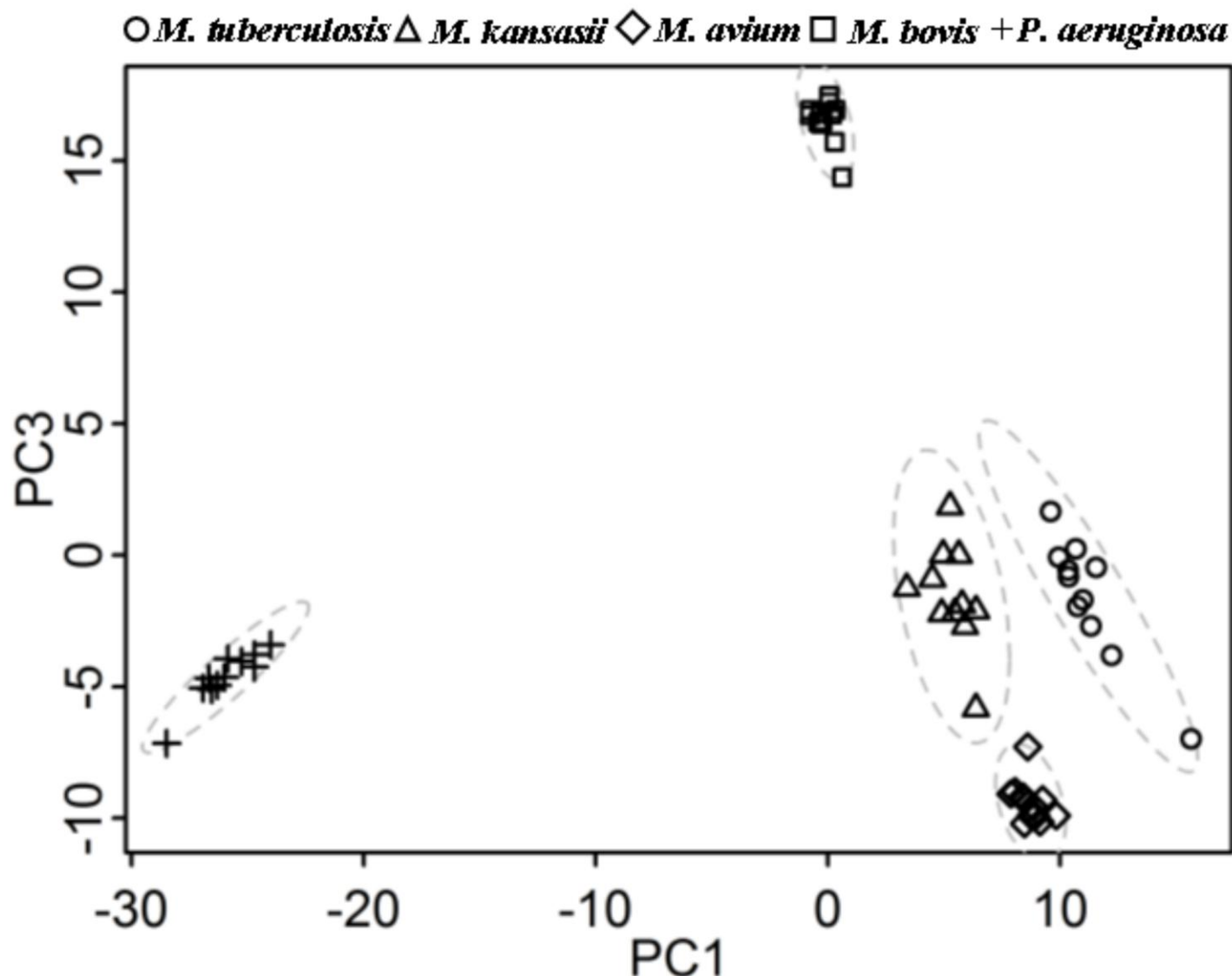


Figure 4. PCA scores plot (PC1 vs. PC3), of the GC-MS data acquired using the total metabolome extraction method, indicating a grouping of the individual samples into their respective organisms groups, with a comparatively smaller within group variation, and hence, bigger inter-group variation, as opposed to the fatty acid extraction procedure. Three PCs were extracted, explaining a total of 48.8% of the variance in the data (R^2X cum), of which PC 1 explained 27.2%, PC 2 explained 14.5%, and PC 3 explained 7.08%. The ellipses represent the 95% confidence interval of the modelled variation.

markers novel to the related *Mycobacterium* species, namely inositol and its stereoisomer, myo-inositol, were also identified using this approach. As summarised by Brown et al. (2007), the myo-inositol supply in mycobacteria is believed to be sustained by *de novo* synthesis initiated by the conversion of glucose-6-phosphate to inositol-1-phosphate followed by dephosphorylation, and is a novel characteristic of these organisms. Succinic acid and an unknown compound with a molecular mass of 268 were identified as novel *M. kansasii* markers, in addition to elevated concentrations of erythritol and citric acid, in comparison to *M. tuberculosis*, *M. avium* and *M. bovis*, which has never

before been documented. It is also important to note that an unknown compound of mass 541 was uniquely detected in *M. tuberculosis*, which could potentially be investigated further for possible diagnostic applications.

Conclusions

Considering these results, the total metabolome extraction procedure has advantages over that of the total fatty acid extraction procedure investigated, considering that: 1) it is simpler; 2) faster; 3) has a comparatively better repeatability, consequently resulting

Table 2. Mean relative concentrations (ug/mg sample) of metabolite markers identified for the various infectious organism groups (12 sample repeats of each) following the fatty acid metabolome extraction method. Standard deviation (SD) values are given in parenthesis.

VIP ranking	Species Compound	<i>M. tuberculosis</i>	<i>M. kansasii</i>	<i>M. avium</i>	<i>M. bovis</i>	<i>P. aeruginosa</i>
		Concentration (SD)				
1	TBSA	273.9 (28.2)	171.3 (50.4)	139.8 (27.3)	32.7 (8.7)	0
2	C22:0	14.1 (2.7)	63 (26.1)	17.7 (5.4)	2.4 (0.6)	T
3	2-octyl-CPOA	0	0	0	0	195.9 (25.8)
4	C17:0	43.2 (8.4)	8.7 (3)	6.3 (0.9)	4.5 (1.2)	3.6 (0.6)
5	C22:1 ω9c	0	21.3 (9)	1.5 (0.3)	0	0
6	C24:1 ω9c	0	25.2 (11.7)	2.4 (0.06)	0	0
7	2,4-DM C14:0	0	3.3 (1.8)	0	T	0
8	Unknown 466	0.6 (0.3)	4.8 (2.7)	0.3 (0.6)	0	0
9	C16:1 ω7c	18 (4.5)	20.7 (8.7)	25.5 (8.4)	0.3 (0.3)	19.5 (6.9)
10	C20:0	36.9 (6.9)	29.4 (11.1)	20.7 (3)	5.4 (1.2)	1.2 (0.3)
11	C14:0	4.8 (2.7)	17.4 (10.8)	8.7 (4.2)	0.3 (0.3)	2.1 (1.5)
12	C28:0	3.3 (0.9)	2.7 (1.2)	T	0	0
13	Unknown 494	1.5 (0.3)	1.5 (0.6)	0.9 (0.3)	0	0
14	Unknown 422	5.7 (1.2)	0	0	0.6 (0.3)	0
15	C25:0	4.2 (0.9)	0	0	0.3 (0.03)	0
16	C18:1 ω9c	192.6 (32.1)	309 (86.1)	193.5 (24.9)	10.5 (2.7)	0
17	2-hexyl-CPOA	0	0	0	0	50.1 (8.4)
18	C20:1 ω9c	0.9 (0.3)	5.4 (2.4)	0.9 (0.6)	0	0
19	Unknown 340	1.8 (0.3)	0	0	0	0

TBSA, tuberculostearic acid; CPOA, cyclopropaneoctanoic acid; DM, dimethyl; T, trace amounts.

Table 3. Mean relative concentrations (ug/mg sample) of metabolite markers identified for the various infectious organism groups (12 sample repeats of each) following the total metabolome extraction procedure. Standard deviation (SD) values are given in parenthesis.

VIP ranking	Species Compound	<i>M. tuberculosis</i>	<i>M. kansasii</i>	<i>M. avium</i>	<i>M. bovis</i>	<i>P. aeruginosa</i>
		Concentration (SD)				
1	Indole-acetic acid	0	0	0	0	123.6 (0.6)
2	Unknown 343	0	0	0	0	216.6 (36.6)
3	Citric acid	0	213.3 (55.2)	0	4.5 (0.3)	0
4	Cadaverine	0	0	0	0	84.3 (15.6)
5	Purine	0	0	0	0	399 (67.5)
6	Inositol	262.8 (138.6)	6.9 (6.0)	168.6 (20.1)	0.3 (0.6)	0
7	5'-Adenylic acid	0	0	55.2 (7.8)	0	735 (111.3)
8	C16:1 ω7t	0.3 (0.6)	0.9 (1.2)	62.1 (5.7)	0	0.3 (1.2)
9	Myo-Inositol	227.1 (149.1)	27.9 (18.0)	348 (50.4)	1.5 (1.8)	0
10	Succinic acid	0	119.4 (19.8)	0	0	0
11	Putrescine	0	0	0	0	191.1 (42.9)
12	C10:0	0	T	19.8 (1.5)	0	53.1 (13.5)
13	Unknown 268	0	22.5 (0)	0	0	0
14	C16:1 ω7c	6.6 (2.2)	7.2 (4.5)	303 (32.1)	0	53.1 (9.9)
15	Erythritol	2.4 (1.2)	19.2 (5.4)	12 (0.9)	0	0
16	Unknown 367	0	0.6 (0.9)	15.9 (1.8)	0	39 (7.2)
17	Unknown 373	0	0	0	0	60.6 (14.7)
18	Unknown 541	84 (42)	0	0	0	0
19	Valeric acid	0	0	0	0	114.9 (15)

in less inter-sample variation, and; 4) isolates the total metabolome, increasing the chance for identifying unique metabolite markers from a variety of compound classes. Despite this however, both these methods proved to extract characterising compounds which best differentiate the various TB causing mycobacteria from one another, and from other unrelated lung pathogens, which are known to induced a similar disease state and symptoms. Hence, both methods could potentially play a valuable role in identifying the metabolite markers which could be used to successfully detect, diagnose, and better characterise various infectious diseases.

REFERENCES

- Brown AK, Meng G, Ghadbane H, Scott DJ, Dover, LG, Nigou, J, Besra GS, Fütterer K (2007). Dimerization of inositol monophosphatase *Mycobacterium tuberculosis* SuhB is not constitutive, but induced by binding of the activator Mg²⁺. *BMC Struct. Biol.*, 7: 55.
- Davis R (2007). Tuberculosis and the intensivist. *Curr. Anaesthesia Crit. Care*, 18: 76-85.
- Dunn WB, Bailey NJC, Johnson HE (2005). Measuring the metabolome: current analytical technologies. *Analyst.*, 30: 606-625.
- Fend R, Kolk AH, Bessant C, Buijtsels P, Klatser PR, Woodman, AC (2006). Prospects for clinical application of electronic-nose technology to early detection of *Mycobacterium tuberculosis* in culture and sputum. *J. Clin. Microbiol.*, 44: 2039-2045.
- Fothergill JC, Guest JR (1977). Catabolism of L-Lysine by *Pseudomonas aeruginosa*. *J. Genl. Microbiol.*, 99: 139-155.
- Guerrant GO, Lambert MA, Moss CW (1981). Gas-chromatographic analysis of mycolic acid cleavage products in mycobacteria. *J. Clin. Microbiol.*, 13: 899-907.
- Karnwal A (2009). Production of indole acetic acid by fluorescent *Pseudomonas* in the presence of L-Tryptophan and rice root exudates. *J. Plant Pathol.*, 91(1): 61-63.
- Knisley CV, Damato JJ, McClatchy K, Brennan PJ (1985). Rapid and sensitive identification of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.*, 22(5): 761-767.
- Koekemoer G, Swanepoel JWH (2008). A semi-parametric method for transforming data to normality. *Stat. Comput.*, 18: 241-257.
- Lambert MA, Moss CW, Silcox VA, Good RC (1986). Analysis of mycolic acid cleavage products and cellular fatty acids of *Mycobacterium* species by capillary gas chromatography. *J. Clin. Microbiol.*, 23: 731-736.
- Larsson L, Odham G, Westerdahl G, Olsson B (1987). Diagnosis of pulmonary tuberculosis by selected-ion monitoring: improved analysis of tuberculostearate in sputum using negative-ion mass spectrometry. *J. Clin. Microbiol.*, 893-896.
- Li X, Xu Z, Lu X, Yang X, Yin P, Kong H, Yu Y, Xu G (2009). Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabonomics: Biomarker discovery for diabetes mellitus. *Anal. Chim. Acta.*, 633: 257-262.
- Madsen R, Lundstedt T, Trygg J (2010). Chemometrics in metabolomics—A review in human disease diagnosis. *Anal. Chim. Acta.*, 659: 23-33.
- Mercenier A, Simon J, Haas D, Stalon V (1980). Catabolism of L-Arginine by *Pseudomonas aeruginosa*. *J. Genl. Microbiol.*, 116: 381-389.
- Mosca A, Russo F, Miragliotta L, Iodice MA, Miragliotta G (2006). Utility of gas chromatography for rapid identification of mycobacterial species frequently encountered in clinical laboratory. *J. Microbiol. Methods.*, 86: 392-395.
- Miguez-Burbano MJ, Flores M, Ashkin D, Rodriguez A, Granada AM, Quintero N, Pitchenik A (2006). Non-tuberculous mycobacteria disease as a cause of hospitalization in HIV-infected subjects. *Int. J. Infect. Dis.* 10: 47-55.
- Rocke DM, Lorenza S (1995). Two-component model for measurement error in analytical chemistry. *Technometrics.*, 37(2): 176-184.
- Schoeman JC, Loots DT (2011). Improved disease characterisation and diagnostics using metabolomics: a review. *J. Cell Tissue Res.*, 11: 2673-2683.
- Stewart DJ (1970). The decarboxylation of L-lysine by *Pseudomonas aeruginosa*. *J. Appl. Bact.*, 33: 501-504.
- Stopforth A, Tredoux A, Crouch A, van Helden P, Sandra P (2004). A rapid method of diagnosing pulmonary tuberculosis using stir bar sorptive extraction-thermal desorption-gas chromatography-mass spectrometry. *J. Chromatogr. A.*, 1071: 135-139.
- Todar K (2005). *Todar's Online Textbook of Bacteriology*. University of Wisconsin-Madison Department of Bacteriology. <http://www.textbookofbacteriology.net/tuberculosis.html>
- WHO (2010). The global plan to stop TB 2011-2015: transforming the fight towards elimination of tuberculosis. WHO Press. Geneva, Switzerland.

APPENDIX B

Registered preliminary patent

LOOTS, D.T., OLIVIER, I. 2011. Method of distinguishing between different pathogens. Preliminary patent: ZA, 2011/03029. 54 p.

REPUBLIC OF SOUTH AFRICA

PATENTS ACT, 1978

PROVISIONAL SPECIFICATION

(Section 30(1) – Regulation 27)

OFFICIAL APPLICATION NO		AGENT'S REFERENCE	
21	01	P42069ZAPO	
FULL NAME(S) OF APPLICANT(S)			
71	NORTH-WEST UNIVERSITY		
FULL NAMES OF INVENTOR(S)			
	LOOTS, Du Toit OLIVIER, Ilse		
TITLE OF INVENTION			
54	METHOD OF DISTINGUISHING BETWEEN DIFFERENT PATHOGENS		

METHOD OF DISTINGUISHING BETWEEN DIFFERENT PATHOGENS

INTRODUCTION AND BACKGROUND TO THE INVENTION

This invention relates to a method of distinguishing between different
5 pathogens in a biological sample.

Tuberculosis (TB) is a disease caused by *Mycobacterium tuberculosis* and
infects approximately one third of the world's population. In 2008 there were
an estimated 9.4 million newly identified TB incidences and 1.8 million deaths
10 globally attributable to TB.

A relatively high prevalence of TB causing mycobacteria other than *M.*
tuberculosis such as *M. bovis*, and non-tuberculous mycobacteria (NTM)
such as *M. avium* and *M. kansasii*, is frequently encountered in clinical
15 specimens and sputum samples. Different treatment regimes are followed
for different bacterial infections and it is becoming more important for
diagnostic techniques to differentiate between various *Mycobacterium*
species.

20 Detection limits further play an important role in the early diagnostic phase of
TB, as only a small amount of bacteria are present in sputum samples, which
often avoids detection and consequently cause the misdiagnosis of TB in
patients. Additionally, HIV co-infection and childhood infections are
frequently left undiagnosed, also due to lower sputum bacterial loads and the
25 lacking analytical detection limits of many of the current TB diagnostic
methods.

In accordance with a report by the World Health Organisation (WHO) on TB diagnostics, smear microscopy is still regarded as the TB diagnostic method most commonly used worldwide, due to its simplicity and affordability.

5 However, a disadvantage associated with the smear microscopy method is that it has a detection limit of 5 000 to 10 000 bacteria ml^{-1} sputum, consequently detecting only 60 to 70% of all TB incidences in adults, as confirmed by the bacteriological culture method discussed below.

10 A further disadvantage associated with this method is that the number of identified cases can be as low as 20 to 30%, when performed under sub-optimal conditions, attributable to inexperienced personnel, for example.

A further disadvantage associated with this method is that it is unable to
15 differentiate between various *Mycobacterium* species in the patient collected sputum, hence, it can only detect the presence of mycobacteria in general, but cannot establish which *Mycobacterium* species is responsible for infection.

20 Another known TB diagnostic method, which has hitherto been regarded as the golden standard for TB diagnostics, is the bacteriological culture method. This method is far more sensitive than the smear microscopy method, and has a detection limit of 10 to 100 bacteria ml^{-1} using solid or liquid media.

25 A major disadvantage associated with this method is that it requires two to six weeks before bacterial colonies can visually be detected, hence leading to a long waiting period prior to initial treatment of the patient.

A further disadvantage associated with the bacteriological culture method is that 15 to 20% of all adult TB incidences are reported to have negative sputum cultures, consequently leading to a false negative result.

5

Yet a further disadvantage associated with this method is that 1 to 4% of all incidences present a false positive, due to poorly executed anti-contamination procedures.

- 10 A more recently developed method for TB diagnostics is nucleic acid amplification, which can obtain a positive result with less than 10 bacteria ml^{-1} , accompanied by a sensitivity of virtually 100% in positive sputum smears. Although this method also has the ability to differentiate various *Mycobacterium* species, it has the disadvantage of having a sensitivity of
15 only 60 to 70% in smear-negative, culture-positive samples.

A further disadvantage associated with the nucleic acid amplification method is that there is a high incidence of false positive results due to cross contamination.

20

A further disadvantage associated with this method is that due to the residual genetic material in the patients, positive results may still be obtained after effective treatment.

25

Yet a further disadvantage associated with this method is that it cannot be used in patients with previous *Mycobacterium* infections and for monitoring of

a patient's response to therapy.

Yet a further disadvantage of this method is that in the clinical practice environment, this method requires two days before a diagnostic result can be
5 obtained, and in most cases, additionally requires prior culturing of the organism, in order to overcome the previously mentioned disadvantages. As a result, an additional two to six weeks diagnostic time is required.

According to the WHO, no serology TB diagnostic assays have yet been
10 approved for clinical use due to a lack of adequate sensitivity and specificity of these methods.

WO 2004/059280 discloses a method for measuring a plurality of different organisms in a sample (mucus, nasal, pharyngeal or genital discharge
15 sample). The measurement includes detecting the presence or absence of a marker indicative of an organism, quantifying the amount of marker, identifying a known marker and determining the identity of a previously unknown marker.

20 The patent further discloses an extraction method for extracting one or more markers from the matrixes of one or more samples and/or rendering the samples more suitable for analysis.

A disadvantage of this method is that it identifies a single marker, therefore
25 being very specific to the presence of one compound in particular. This method is thus not specific enough for differentiating between various

Mycobacterium species, or their respective drug resistant strains.

The method described in **WO 2004/059280** will thus not distinguish between drug resistant and non-resistant strains of TB and will furthermore not be
5 directed at differentiating *Mycobacterium tuberculosis* from other similar infectious *Mycobacterium* species (*M. avium*, *M. kansasii*, *M. bovis* and the like).

WO 2006/079846 discloses a method of detecting and identifying bacteria
10 comprising the steps of collecting volatile bacterial products, subjecting the volatile products to a gas chromatography system employing a surface acoustic wave detector, establishing chromatographic profiles for different bacteria, and comparing the chromatographic profile of the biological sample (including sputum, breath and blood) with profiles in the library. The system
15 is said to be suitable for use in the detection of TB and indicates that mycolic acids could be potential biomarkers. However, no specific markers for TB are described.

A disadvantage of this method is that it was found in practise not to be highly
20 reliable and the results are non-specific, hence, it does not differentiate between the various *Mycobacterium* species or their respective drug resistant strains.

WO 2009/045116 discloses a method for detecting a *Mycobacterium*
25 microorganism by analysing a gas mixture for one or more identified

biomarkers selected from a group consisting of methyl phenyl acetate, methyl p-anisate, methyl nicotinate, and o-phenylanisole.

A disadvantage of this method is also that it is not highly reliable and the results are non-specific, hence it doesn't differentiate between the various *Mycobacterium* species or their respective drug resistant strains.

WO 2009/091375 relates to a method for identifying the presence of *Mycobacterium tuberculosis* in a sample by detecting at least one volatile organic compound indicative of the presence of *M. tuberculosis* in the sample.

The disadvantage associated with **WO 2009/091375** is that the marker compound used for diagnosis is not specific enough to differentiate between the various *Mycobacterium* species.

OBJECT OF THE INVENTION

It is accordingly an object of the present invention to provide a method of distinguishing between different pathogens in a biological sample with which the aforesaid disadvantages could be overcome or at least minimised.

SUMMARY OF THE INVENTION

According to a first aspect of the invention, there is provided a reference system for distinguishing between different pathogens in biological samples (both specific and non-specific) including:

- a reference library of the relative concentrations of carbon based

compounds present in extracts of the said different pathogens in combination;

- a data matrix comprising a set of values of replicate samples of all the said pathogens, each set representing the relative concentrations of all the said carbon based compounds in a sample of each of the said pathogens;
- processing means for converting each set of values into a single representative benchmark value; and
- an output means for expressing the respective single representative benchmark values graphically.

Further according to the invention, the processing means includes a microprocessor for applying principle component analysis (PCA) calculations on each set of values for each replicate sample of each organism.

Further according to the invention, the output means includes a second processing means for preparing a scatter plot of the respective single representative benchmark values of the replicate samples of each organism from the PCA data.

Further according to the invention, the second processing means is adapted to identify, from the PCA calculations, a list of biomarkers or so-called variables of importance (VIPs); being the compounds contributing most the natural separation in the PCA analysis.

The second processing means may be adapted further to identify and select

additional biomarkers, using effect size calculations for all possible group comparisons, based on uniqueness to a particular pathogen, above a predetermined concentration threshold.

- 5 The second processing means may be adapted further to build a discriminant model, using the VIPs or biomarkers identified, for predicting the pathogen group membership of a new sample, the discriminant model being selected from the group consisting of ordinary discriminant models, canonical discriminant models, partial least squares discriminant models, support
10 vector machines, neural networks, logistic regression models and discriminant models based on kernel density estimations.

Further according to the invention, the output means includes a processing means for preparing a scatter plot of the predicted pathogen group
15 probabilities for each of the benchmark pathogen samples using only the biomarkers or VIPs processed via the discriminant model selected.

The output means may include a third processing means for indicating the distance measure limit of each predicted benchmark pathogen group sample,
20 to the particular pathogen group centre, as an indication of true group membership.

According to a second aspect of the invention, there is provided a method of distinguishing between different pathogens in a biological sample (both
25 specific and non-specific) in a system, according to the first aspect of the invention, including the steps of:

- obtaining a biological sample containing at least one pathogen;
 - extracting at least one group of carbon based compounds from at least one pathogen in the sample using at least one extraction solvent; and
 - analysing the entire extracted group of carbon based compounds to
- 5 prepare a data matrix comprising a set of values representing the relative concentrations of all the said VIP or biomarker compounds in the sample.

The method may include the further step of applying the said discriminant model to predict the class membership and calculate the distance measure of

10 the patient collected biological sample using only the selected VIPs or biomarkers compounds.

The step of extracting carbon based compounds from the pathogen may include the step of neutralising microorganisms other than the said

15 pathogens, isolating the surviving pathogens from the neutralised microorganisms and from the remainder of the biological sample, culturing the pathogens and pyrolysing harvested cells.

Alternatively, the step of extracting carbon based compounds from the

20 pathogens may include the step of extracting the compounds directly from the biological sample containing a plurality of microorganisms and the said pathogens.

Further alternatively, the step of extracting carbon based compounds from

25 the pathogens may include the step of neutralising microorganisms other

than the said pathogen, and isolating the neutralised microorganisms and surviving pathogen from the biological sample.

Further according to the invention, the step of extracting at least one group of

5 carbon based compounds includes the steps of:

- adding one or more extraction solvents to the sample to form an extraction mixture;
- breaking up the cell walls of the pathogens to release and extract the carbon based compounds from the pathogens; and
- 10 - subjecting the extraction mixture to centrifugation to separate the liquid fraction from the solid fraction.

The extraction solvent may be selected from the group consisting of acetic acid, acetone, benzene, 1-butanol, 2-butanol, chloroform, chlorobenzene, 15 cyclohexane, 1,2-dichloroethane, diethyl ether, diethyl glycol, diethylene glycol dimethyl ether, dimethyl-formamide (DMF), dimethyl sulfoxide (DMSO), ethanol, ethyl acetate, ethylene glycol, glycerine, hexane, methanol, pentane, 1-propanol, toluene, tetrahydrofuran (THF), pyridine and water.

20 The step of extracting at least one group of carbon based compounds may include the further step of drying the liquid fraction followed by derivitisation by adding a derivatisation solvent.

The derivatisation solvent may be selected from the group consisting of
 25 methoxyamine hydrochloride, pyridine, *N*-Methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA), (N,O-bis(trimethylsilyl)

trifluoroacetamide (BSTFA), trimethylchlorosilane (TMCS), methanolic HCl and other derivatization reagents or combinations thereof.

The carbon based compounds may be selected from the group consisting of
5 fatty acids, sugars, esters, ketones, amines, sulphur compounds, alcohols, carboxylic acids, and hydrocarbons such as alkenes, alkanes and aromatic compounds, and other carbon based compounds and combinations thereof.

Further according to the invention, the pathogens are selected from the
10 group consisting of bacterial species, fungal species and yeast species, preferably bacterial species.

More specifically, but not exclusively, the pathogens are selected from the group consisting of TB causing mycobacteria including *M. tuberculosis*, *M.*
15 *bovis*, non-tuberculous mycobacteria (NTM) such as *M. avium* and *M. kansasii*, and other lung pathogens including, *Pseudomonas aeruginosa*. The pathogens may also include drug-resistant strains of the above mentioned pathogen groups.

20 The method may include the further step of applying principle component analysis (PCA) calculations on each set of values for each replicate sample of each organism.

The method may include the further step of preparing a scatter plot of the
25 respective single representative benchmark values of the replicate samples of each organism from the PCA data.

The method may include the further step of identifying, from the PCA calculations, a list of biomarkers or so-called variables of importance (VIPs); being the compounds contributing most the natural separation in the PCA analysis.

The method may include the further step of identifying and selecting additional biomarkers, using effect size calculations for all possible group comparisons, based on uniqueness to a particular pathogen, above a predetermined concentration threshold.

The method may include the further step of building a discriminant model, using the VIPs or biomarkers identified, for predicting the pathogen group membership of a new sample, the discriminant model being selected from the group consisting of ordinary discriminant models, canonical discriminant models, partial least squares discriminant models, support vector machines, neural networks, logistic regression models and discriminant models based on kernel density estimations.

The method may include the further step of preparing a scatter plot of the predicted pathogen group probabilities for each of the benchmark pathogen samples using only the biomarkers or VIPs processed via the discriminant model selected.

The step of applying the discriminant model may include the further step of indicating the distance measure limit of each predicted benchmark pathogen

group sample, to the particular pathogen group centre, as an indication of true group membership.

BRIEF DESCRIPTION OF THE FIGURES

- 5 The invention will now be described further, by way of non-limiting example only, with reference to the accompanying figures wherein:

figure 1: is a three dimensional Principal Component Analysis (PCA) scores plot of Principal Components (PCs) 1 to 3 illustrating the natural differentiation of the various pathogen groups using the entire extracted group of carbon based compounds, wherein 1 represents *M. tuberculosis*, 2 represents *M. kansasii*, 3 represents *M. avium*, 4 represents *M. bovis*, and 5 represents *P. aeruginosa*;

15 figure 2: is a three dimensional PCA scores plot of PCs 1 to 3 illustrating the natural differentiation of the respective pathogens using a combination of biomarkers selected from the PCA in figure 1 with the additional biomarkers selected according to effect sizes and uniqueness to the particular pathogens wherein 1 represents *M. tuberculosis*, 2 represents *M. kansasii*, 3 represents *M. avium*, 4 represents *M. bovis*, and 5 represents *P. aeruginosa*;

figure 3: illustrates the individual predicted pathogen group assignment probabilities of each benchmark pathogen sample in each of

the individual pathogen groups, as determined by the discriminant model based on kernel density estimates (the majority of samples analysed falling within a 90% probability for belonging to the correct pathogen group). Group 1 represents *M. tuberculosis*, 2 represents *M. kansasii*, 3 represents *M. avium*, 4 represents *M. bovis*, and 5 represents *P. aeruginosa*;

5

figure 4: is a scatter plot of the individual Mahalanobis distances for each of the benchmark pathogen samples, relative to the centre of the predicted pathogen group. The line at the Mahalanobis distance of 4, is the square root of the 99.9% chi-square quantile with 3 degrees of freedom, and thus indicates the distance control limit for correct identification within each particular group. Group 1 represents *M. tuberculosis*, 2 represents *M. kansasii*, 3 represents *M. avium*, 4 represents *M. bovis*, and 5 represents *P. aeruginosa*;

10

15

20

figure 5: is a scatter plot of the predicted pathogen group probabilities as determined by the aforementioned discriminant model (figure 3) based on kernel density estimates, where the previously unknown pathogen samples are correctly assigned to their respective pathogen groups (indicated by X). Group 1 represents *M. tuberculosis*, 2 represents *M. kansasii*, 3 represents *M. avium*, 4 represents *M. bovis*, and 5 represents

P. aeruginosa;

figure 6: is a scatter plot of the individual Mahalanobis distances for each of the benchmark pathogen samples, relative to the centre of the predicted pathogen group as described in figure 4. The line at the Mahalanobis distance of 4, is the square root of the 99.9% chi-square quantile with 3 degrees of freedom, and thus indicates the distance limit for correct identification within each particular group. Group 1 represents *M. tuberculosis*, 2 represents *M. kansasii*, 3 represents *M. avium*, 4 represents *M. bovis*, and 5 represents *P. aeruginosa*.

figure 7: is a three dimensional PCA scores plot of PC 1 to 3 illustrating the natural differentiation of the various *M. tuberculosis* concentration groups using the entire extracted group of carbon based compounds wherein: 0 = blank or 0 cells, 1 = 2.5, 2 = 2.5 X 10¹, 3 = 2.5 X 10²; 4 = 2.5 X 10³; 5 = 2.5 X 10⁴ and 6 = 2.5 X 10⁵ cells, indicating that the detection limit of this approach is 250 bacteria cells; and

20

figure 8 is a three dimensional PCA scores plot of PCs 1 to 3 illustrating the natural differentiation of the various pathogen groups using the entire extracted group of carbon based compounds, at the detection limit concentration of 250 cells, wherein 1 represents

M. tuberculosis, 2 represents *M. kansasii*, 3 represents *M. avium*, 4 represents *M. bovis*, and 5 represents *P. aeruginosa*.

DESCRIPTION OF PREFERRED EMBODIMENTS OF THE INVENTION

- 5 According to a preferred embodiment of the invention there are provided a system and methods for distinguishing between different pathogens in a biological sample.

According to a first preferred embodiment of the invention there is provided a
10 reference system for distinguishing between different pathogens in biological samples (both specific and non-specific) including:

- a reference library of the relative concentrations of carbon based compounds present in extracts of the said different pathogens in combination;
- 15 - a data matrix comprising a set of values of replicate samples of all the said pathogens, each set representing the relative concentrations of all the said carbon based compounds in a sample of each of the said pathogens;
- processing means for converting each set of values into a single representative benchmark value; and
- 20 - an output means for expressing the respective single representative benchmark values graphically.

The processing means includes a microprocessor for applying principle component analysis (PCA) calculations on each set of values for each

replicate sample of each pathogen sample.

The output means includes a processing means for preparing a scatter plot of the respective single representative benchmark values of the replicate
5 samples of each pathogen from the PCA data.

The second processing means is adapted to identify, from the PCA calculations, a list of biomarkers or so-called variables of importance (VIPs); being the compounds contributing most the natural separation in the PCA
10 analysis.

The second processing means is adapted further to identify and select additional biomarkers, using effect size calculations for all possible group comparisons, based on uniqueness to a particular pathogen, above a
15 predetermined concentration threshold.

The second processing means is adapted further to build a discriminant model, using the VIPs or biomarkers identified, for predicting the pathogen group membership of a new sample, the discriminant model being selected
20 from the group consisting of ordinary discriminant models, canonical discriminant models, partial least squares discriminant models, support vector machines, neural networks, logistic regression models and discriminant models based on kernel density estimations.

25 The output means includes a processing means for preparing a scatter plot of the predicted pathogen group probabilities for each of the benchmark pathogen samples using only the biomarkers or VIPs processed via the

discriminant model selected.

The output means includes a third processing means for indicating the distance measure limit of each predicted benchmark pathogen group sample, to the particular pathogen group centre, as an indication of true group membership.

According to a second preferred embodiment of the invention there is provided a method of distinguishing between different pathogens in a biological sample (both specific and non-specific) in a system according to the first embodiment of the invention, including the steps of:

- obtaining a biological sample containing at least one pathogen;
- extracting at least one group of carbon based compounds from at least one pathogen in the sample using at least one extraction solvent;
- analysing the entire extracted group of carbon based compounds to prepare a data matrix comprising a set of values representing the relative concentrations of all the said VIP or biomarker compounds in the sample;
- applying the said discriminant model to predict the class membership and calculate the distance measure of the patient collected biological sample.

The step of extracting carbon based compounds from the pathogen includes the step of neutralising microorganisms other than the said pathogens, isolating the surviving pathogens and the neutralised microorganisms and from the remainder of the biological sample; culturing the pathogens; and pyrolysing harvested cells.

Alternatively, the step of extracting carbon based compounds from the pathogens includes the step of extracting the compounds directly from the biological sample containing a plurality of microorganisms and the said
 5 pathogens.

Further alternatively, the step of extracting carbon based compounds from the pathogens includes the step of neutralising microorganisms other than the said pathogen, and isolating the neutralised microorganisms and
 10 surviving pathogen from the biological sample.

The step of extracting at least one group of carbon based compounds includes the steps of:

- adding one or more extraction solvents to the sample to form an
 15 extraction mixture;
- breaking up the cell walls of the pathogens to release and extract the carbon based compounds from the pathogens; and
- subjecting the extraction mixture to centrifugation to separate the liquid fraction from the solid fraction.

20

The extraction solvent is selected from the group consisting of acetic acid, acetone, benzene, 1-butanol, 2-butanol, chloroform, chlorobenzene, cyclohexane, 1,2-dichloroethane, diethyl ether, diethyl glycol, diethylene glycol dimethyl ether, dimethyl-formamide (DMF), dimethyl sulfoxide

(DMSO), ethanol, ethyl acetate, ethylene glycol, glycerine, hexane, methanol, pentane, 1-propanol, toluene, tetrahydrofuran (THF), pyridine and water.

The step of extracting at least one group of carbon based compounds includes the further step of drying the liquid fraction followed by derivitisation by adding a derivatisation solvent.

The derivatisation solvent may be selected from the group consisting of methoxyamine hydrochloride, pyridine, *N*-Methyl-*N*-trimethylsilyltrifluoroacetamide (MSTFA), (N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA), trimethylchlorosilane (TMCS), methanolic HCl and other derivatization reagents or combinations thereof.

The carbon based compounds are selected from the group consisting of fatty acids, esters, ketones, amines, sulphur compounds, alcohols, carboxylic acids, and hydrocarbons such as alkenes, alkanes and aromatic compounds, and other carbon based compounds and combinations thereof.

The pathogens are selected from the group consisting of bacterial species, fungal species and yeast species, preferably bacterial species. More specifically, but not exclusively, the pathogens are selected from the group consisting of TB causing mycobacteria including *M. tuberculosis*, *M. bovis*, non-tuberculous mycobacteria (NTM) such as *M. avium* and *M. kansasii*, and other lung pathogens, such as *P. aeruginosa*.

25

The method includes the further steps of:

- applying principle component analysis (PCA) calculations on each set of values for each replicate sample of each organism;
- preparing a scatter plot of the respective single representative benchmark values of the replicate samples of each organism from the PCA data;
- identifying, from the PCA calculations, a list of biomarkers or so-called variables of importance (VIPs); being the compounds contributing most the natural separation in the PCA analysis;
- of identifying and selecting additional biomarkers, using effect size calculations for all possible group comparisons, based on uniqueness to a particular pathogen, above a predetermined concentration threshold;
- building a discriminant model, using the VIPs or biomarkers identified, for predicting the pathogen group membership of a new sample, the discriminant model being selected from the group consisting of ordinary discriminant models, canonical discriminant models, partial least squares discriminant models, support vector machines, neural networks, logistic regression models and discriminant models based on kernel density estimations; and
- preparing a scatter plot of the predicted pathogen group probabilities for each of the benchmark pathogen samples using only the biomarkers or VIPs processed via the discriminant model selected.

The step of applying the discriminant model may include the further step of indicating the distance measure limit of each predicted benchmark pathogen

group sample, to the particular pathogen group centre, as an indication of true group membership.

It will be appreciated, however, that the system and method could be adapted to distinguish between most known organisms and differentiate between various drug resistant strains of these also.

The above system, method and individual steps of the method are described in more detail below, by way of examples.

10

EXAMPLE 1

In accordance with a first non-limiting example of a preferred embodiment of the invention according to the invention, the following reagents and chemicals are provided:

15

- MSTFA, TMCS, pyridine, potassium hydroxide, 3-phenyl butyric acid, ethanol and trimethylpentane purchased from Merck (Darmstadt, Germany);
- Methoxyamine hydrochloride purchased from Sigma-Aldrich (St. Louis, Mo., USA); and
- the organic solvents are ultra pure Burdick & Jackson brands (Honeywell International Inc., Muskegon, MI, USA).

20

Biological sample preparation

In accordance with this example, the step of obtaining a biological sample containing at least one pathogen includes:

25

- liquefying the sputum by adding an equal volume of liquefying solution (0.5 N NaOH and 0,0003M N-acetyl-cysteine), followed by vortexing for 60 seconds and incubation for 15 minutes at room temperature. Cells
5 are isolated by centrifugation at 3000 rpm for 15 minutes and washed twice with distilled water;
- obtaining the following pure bacterial cultures, namely *M. tuberculosis*, *M. avium*, *M. bovis*, *M. kansasii* and *P. aeruginosa* originally obtained from patient collected sputum, as described above, and then cultured in
10 Middlebrook 7H9 medium with oleic acid-albumin-dextrose-catalase enrichment;
- culturing cells at 37 degrees Celsius while shaking at 120 to 150 rpm until an optical density (420 nm) of 0.30 ($\approx 2 \times 10^8$ bacteria ml^{-1}) is reached; and
- 15 • washing the cells once with PBS (phosphate buffered saline), and collection via centrifugation, killing or inactivating the microorganisms in a water bath at 110 degrees Celsius for 40 minutes.

Additionally, the carbon based compound extraction can be done directly
20 after bacterial cell isolation following the liquefying step, provided that an internal standard is added.

Furthermore, the carbon based compound extraction can be done on patient collected sputum without a prior liquefying step, provided sample is
25 homogenised using 30% ethanol, and dried in a speedvac, prior to the

carbon based compound extraction. This is accomplished by adding 450 µL pure ethanol and 550 µL H₂O the collected patient sputum sample, followed by a shaking 50 Hz for 2 min in an MM 400 vibration mill (Retsch GmbH & co. KG, Haan, Germany). Hereafter, the homogenised sputum suspension is
5 completely dried in a speedvac and the dried sputum sample is extracted.

Carbon based compound extraction

Further in this example, the step of extracting at least one group of carbon based compounds from the pathogens in the sample using at least one
10 extraction solvent includes the following:

- preparing twelve repeats of an amount of $2,5 \times 10^5$ cells of each sputum cultured species in order to prepare the said system;
- determining the detection limit of the system procedure by preparing six
15 repeats of the cultured *M. tuberculosis*, ranging from 0 to 2.5×10^5 mycobacteria cells, to be extracted and analysed in the same manner;
- adding 50 µl of 3-phenylbutyric acid to the previously mentioned biological samples in a microcentrifuge tube as a reference standard to determine the relative concentrations of all the said carbon based
20 compounds in a sample of each of the said pathogens;
- a protein determination can also be used in addition to or in combination with the internal standard for quantification purposes.
- adding 1,25 mL of an extraction solvent mixture made up of chloroform/Methanol/ddH₂O (2:6:2);
- 25 • performing the extraction using an MM 400 vibration mill (Retsch GmbH

& co. KG, Haan, Germany) at a frequency of 30 Hz s⁻¹ for 5 min;

- adding a 3-mm tungsten carbide bead (Retsch GmbH & co. KG) to each tube for increased extraction efficiency;
- collecting the total liquid fraction (both organic and water fractions) following centrifugation, transferring it to a GC-MS sample vial and drying it under a light stream of nitrogen;
- derivatising the liquid fraction by adding 50 µl of methoxyamine hydrochloride (15 mg mL⁻¹) in pyridine to the sample followed by an incubation period of 90 min at 50 degrees Celsius; and
- trimethylsilylating the extract for 60 min at 50 degrees Celsius after the addition of 50 µl of MSTFA with 1% TMCS.

It should be noted that the volumes of the aforementioned derivatisation reagents can be modified in accordance to the protein determination for each sample.

Gas chromatography-mass spectrometry (GC-MS) analysis

Also in this example, the step of analysing the entire extracted group of carbon based compounds to prepare a data matrix comprising a set of values representing the relative concentrations of all the said carbon based compounds in the sample includes the following:

- analysing the prepared extracts by injecting 1 µL on an Agilent 7890A gas chromatograph (Agilent, Atlanta, GA) coupled to an Agilent 5975 mass selective detector equipped with a 7683B injector, 7683 auto sampler and VF1-MS capillary column (30 m x 250 µm i.d., 0.25 µm film

thickness) in the splitless mode; and

- maintaining the injector temperature at 270 degrees Celsius for the total run time, with the initial GC oven temperature being maintained at 70 degrees Celsius followed by an increase in oven temperature to a final temperature of 300 degrees Celsius, using helium as a carrier gas and keeping the flow of the helium gas at a constant 1.2 mL/min, and using MS detection in full scan mode.

It will be appreciated that instead of GC-MS analysis, GCxGC-TOFMS analysis or other hyphenated GC or liquid chromatography (LC) techniques could be applied. This could also be applied to these and other hyphenated chromatography techniques without derivatisation.

Further analysis and findings

In executing the method according to the preferred embodiment of the invention, use was made of a so-called Automated Mass Spectral Deconvolution and Identification System V2.65 (AMDIS) in the deconvolution and analysis of the raw GC-MS data. AMDIS is a computer program that deconvolutes the individual components as analysed by GC/MS in order to develop a reference library containing the characteristic mass spectra of a total of a 1198 compounds, including: fatty acids, amino acids, sionols, silanes, alcohols, organic acids and monosaccharides, amongst others, which are present in the aforementioned pathogens.

It will be appreciated that instead of AMDIS being used as a deconvolution

tool, other software options could also be considered for deconvolution and alignment of data generated by the GC-MS or other hyphenated chromatographic techniques.

- 5 Also in this example, the step of comparing the compound profile with benchmark profiles to identify the respective organisms in the sample includes:
- characterising the results obtained from the extracted group of carbon based compounds in data matrixes consisting of all detected compounds
10 in order to develop a reference library according to the invention for identifying any of the aforementioned pathogens from a patient sample;
 - the data matrix comprise a set of values of replicate samples of all the said pathogens, each set representing the relative concentrations of all the said carbon based compounds in a sample of each of the said
15 pathogens;
 - the aforementioned pathogen samples are aligned in this database according to the said carbon based compounds present and retention indexes of all the said carbon based compounds in each sample;
 - analysing the obtained data matrixes via processing means known as
20 principal component analysis (PCA). The processing means converts each set of concentration values for all of the said carbon based compounds into a single representative benchmark value termed a principal component (PC) for each benchmark pathogen sample analysed;
 - 25 • (part of PCA) pre-treating data using a non-parametric transformation

function to limit variables (metabolites) with high concentrations from dominating the PCA, after which mean centring is applied;

- extracting four PCs elucidating 58% (this is variable – unique to this example) of the variation in the scaled data, from the data matrixes obtained from the GC-MS analysed data of the extraction procedure;
- building a PCA model using the AMDIS generated data matrixes;
- preparing a scatter plot of the single representative benchmark values for each of the benchmark pathogen samples, thus providing an overview of the biological samples and how they relate to each other (figure 1);

Referring to figure 1, it is evident from the PCA output data that all the bacterial groups are clearly differentiated from one another using all detected compounds.

- extracting a VIP (variables importance in the PC projection) list from the PCA loadings (a quantitative number representing the contribution of each variable to the separation between samples in the model plane).

These included the following fatty acids: 10 Me-C16:0, 10-Me C17:0, 10Me-C15:0, 2,4-DM C14:0, 25-Me C27:0, C10:0, C14:0, 1-methyl ethylester of C14:0, , C16:0, C16:1 w7c , C16:1 w7t, C17:0, C18:1 w8c, C18:1 w9c, C18:1 w9t, C18:2n6c, C20:0, C20:1 w9c, C22:0, C22:1 w9c, C24:0, C24:0, C24:1 w9c, C25:0, C26:0, C28:0, 2-ethyl hexadecyl ester of C6, benzenepropanoicacid, 2-hexyl-cyclopropaneoctanoic acid, 2-octyl- cyclopropaneoctanoic acid, and tuberculostearic acid; and other compounds including: 05'-Adenylic acid, Butyl phthalate, cadaverine, citric acid, erythritol, heptacosane, hexacosane, indole-acetic acid,

inositol, myo-Inositol, purine, putrescine, succinic acid, tetradecane, tricosane, Unknown 268, Unknown 340, Unknown 343, Unknown 367, Unknown 373, Unknown 394, Unknown 394, Unknown 422, Unknown 459, Unknown 466, Unknown 494, Unknown 503, Unknown 541, valeric acid. (The mass spectra of unknown biomarker compounds are included in Appendix 1).

- An additional VIP or Biomarker list was compiled on the basis of effect size calculations and uniqueness to the aforementioned pathogens.
- These included: for *M. tuberculosis* (UnknownTB1); *M. avium* (Unknown406); *M. kansasii* (Glycerine, Unknown466); *P. aeruginosa* (2(1H)-Pyrimidinone, DL-Ornithine, Capric acid, Unknown pseudo1, Unknown341, Unknown237, Unknown315, Unknown516, and Unknown319) (The mass spectra of unknown biomarkers compounds are included in Appendix 1).
- The combined biomarkers were then used to build a biomarker PCA model in order to determine if a natural separation between the said pathogen groups is conserved (figure 2). Referring to figure 2, illustrating the PCA output from data generated using the single representative benchmark values from the selected biomarkers, it is evident that all the bacterial groups are clearly differentiated from one another.
- One or more combinations of the above mentioned PCA biomarkers are used to build a discriminant model based on Bayes' theorem in conjunction with multivariate kernel density estimation. This model is developed for the purpose of estimating the class membership

probabilities of an unknown pathogen sample based on the presence of the aforementioned VIP or biomarkers (figures 3 and 6).

- 5 In figure 5 the probabilities for the two unknown samples correctly identified in group 1 is: 94% and 95%; for group 2: 100% in both cases; group 3: 93% and 96%; group 4: 98% in both cases; group 5: 100% in both cases. Referring to figure 6 the previously unknown infectious samples previously shown to be correctly assigned to their respective pathogen groups in figure 5, also fall within the Mahalanobis distance control limit of 4, hence, confirming absolute correct identification for

10 each patient sample.
- 15 Additionally, it should be noted that the said discriminant model will always assign an unknown sample to one of the benchmark pathogen groups. Due to this occurrence, the Mahalanobis distance measure of the unknown sample to the centre of the predicted benchmark pathogen group is calculated. In a case where the unknown sample lies within the square root of an appropriate chi-square quantile for the predicted benchmark group, is indicative of a correct classification. In a case where the unknown sample lies beyond the square root of an appropriate chi-square quantile from the predicted group it would most likely be

20 indicative for an unknown pathogen to the model (figures 4 and 6).

DETECTION LIMITS

For determining the detection limit of this diagnostic model, a series of samples were prepared including: 0 = blank or 0 *Mycobacterium* cells, 1 =

25 2.5 *Mycobacterium* cells, 2 = 2.5×10^1 *Mycobacterium* cells, 3 = 2.5×10^2

Mycobacterium cells; 4 = 2.5×10^3 *Mycobacterium* cells; 5 = 2.5×10^4 *Mycobacterium* cells; 6 = 2.5×10^5 *Mycobacterium* cells, and analysed in the same manner as described above.

5 The sample with the lowest number of cells which doesn't overlap with the blank is considered the lowest number of cells required for a diagnosis, and hence, the detection limit of the analysis is predicted to be 250 cells as seen in figure 7.

10 Referring to figure 8 the detection limit was confirmed by analysing the various pathogen groups (*M. tuberculosis*, *M. avium*, *M. bovis*, *M. kansasii*, and *P. aeruginosa*), each at the detection limit concentration of 250 cells per analysis, using the entire extracted group of carbon based compounds.

15 **EXAMPLE 2**

In implementing the method of Example 1 in distinguishing between different microorganisms in a biological sample, the method is repeated, using the unknown biological sample as the source of the microorganisms instead.

20 The aforementioned discriminant model developed from the biomarker PCA model of the said benchmark pathogen samples in Example 1, was used to predict class membership of two such unknown samples - derived from separately grown cultures - of each of the bacterial species, further validating the model (figures 4 and 5).

25 Using the methodology explained above, two separate cultured samples were extracted in the same manner, and processed as described above in

order to validate or test the model. All of these validation samples were correctly identified with very high class membership probabilities, and the additional Mahalanobis distances for each of the unknown samples were within the control limit, i.e. the model was able to correctly diagnose different
5 strains of mycobacterium species, grown from separate patient sputum samples.

It is therefore evident that the above mentioned extraction procedure, and the metabolomics data analysis that followed, has the capacity to be developed
10 into a feasible, quick diagnostic method, for the identification of various *Mycobacterium* species from culture and potentially directly from sputum also.

The applicant has found that the method according to the invention has the
15 following advantages over the prior art method:

- This diagnostic method uses the unique metabolome profiles of the most characteristic biomarker compounds and their comparative ratios, for differentiating or identifying various infectious *Mycobacterium* species (e.g. *M. bovis*, *M. avium*, *M. kansasii*, *M. tuberculosis*), and potentially
20 also drug resistant strains of these organisms.
- It is able to differentiate between various *Mycobacterium* species in a single analysis, which is unique in the sense that most of the current methods used currently for diagnosing these organisms (smears).
- Duration for diagnostic identification, excluding the required culturing
25 step, is 7 hours. PCR also requires culturing before hand, however,

effectively takes two days to get a diagnostic result.

- The detection limits for this analysis is extremely low only requiring 250 cells for a positive identification or differentiation of an infectious organism, hence, the required culture time prior to analysis, comparative to other methods, would potentially also be less than that required by other methodologies.
 - This diagnostic approach also has an advantage for identifying unique disease specific biomarkers, which may in turn lead to further refinement of the diagnostic method, leading to even faster, simpler, more sensitive, inexpensive diagnostic kits, which can be used without the need of an analytical or diagnostic laboratory.
 - The biomarkers may also be used for understanding the mechanisms underlying disease progression, drug resistance and co-infection.
- It will be appreciated further that variations in detail are possible with a system and method for distinguishing between different pathogens in a biological sample according to the invention without departing from the scope of this disclosure.

DATED THIS 21 DAY OF APRIL 2011

MARIUS LE ROUX

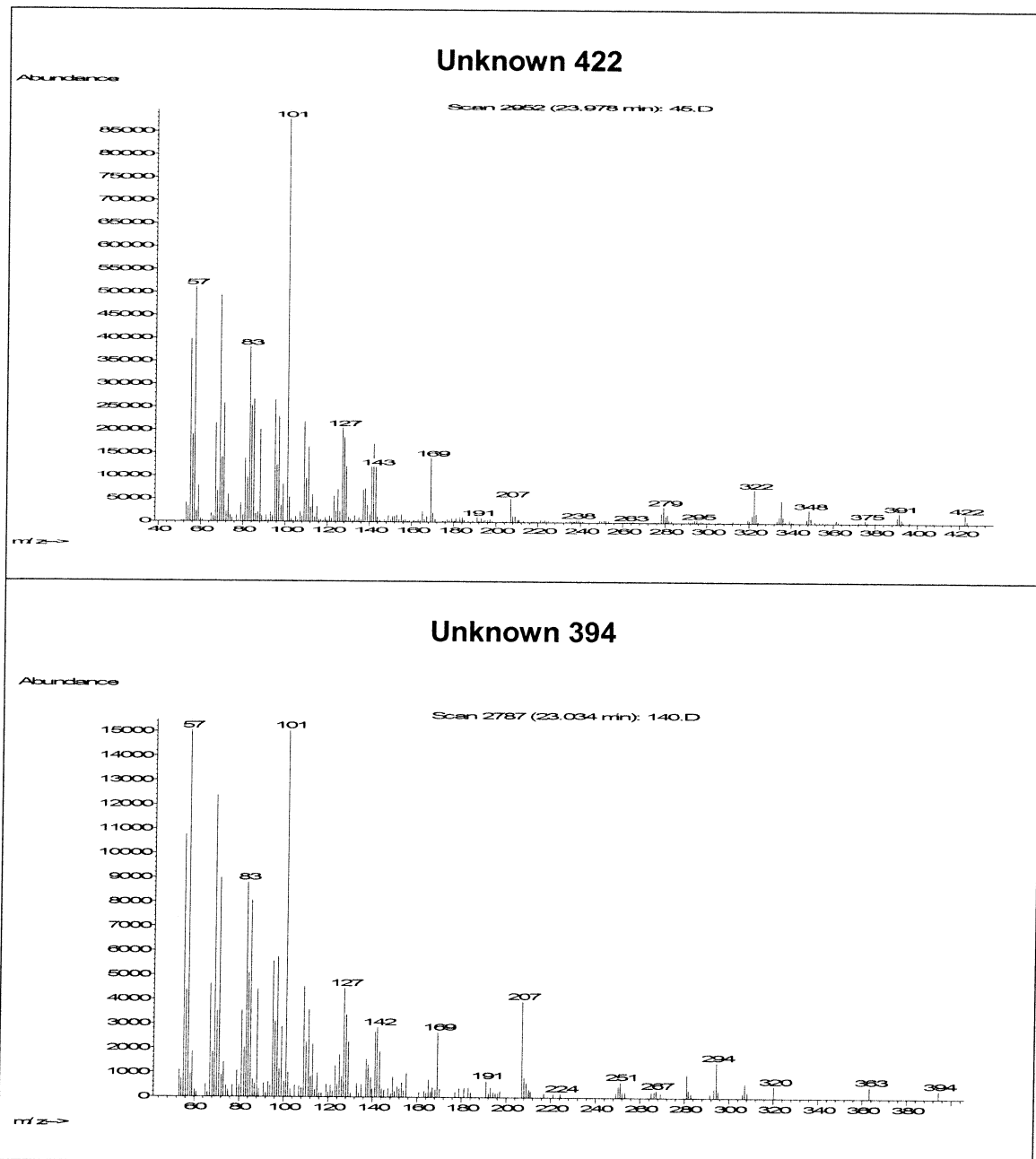
DM KISCH INC

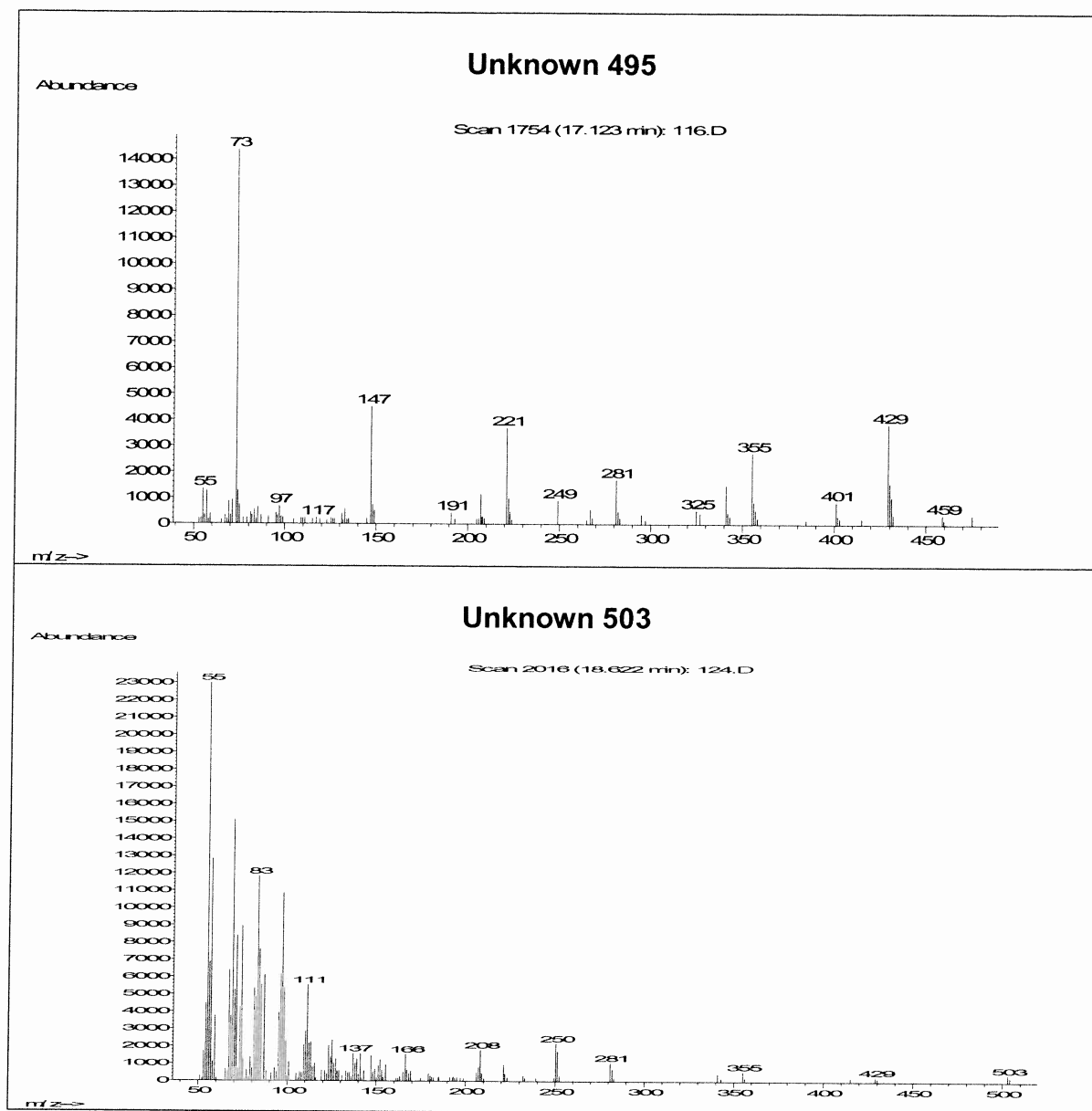
PATENT ATTORNEY FOR APPLICANT

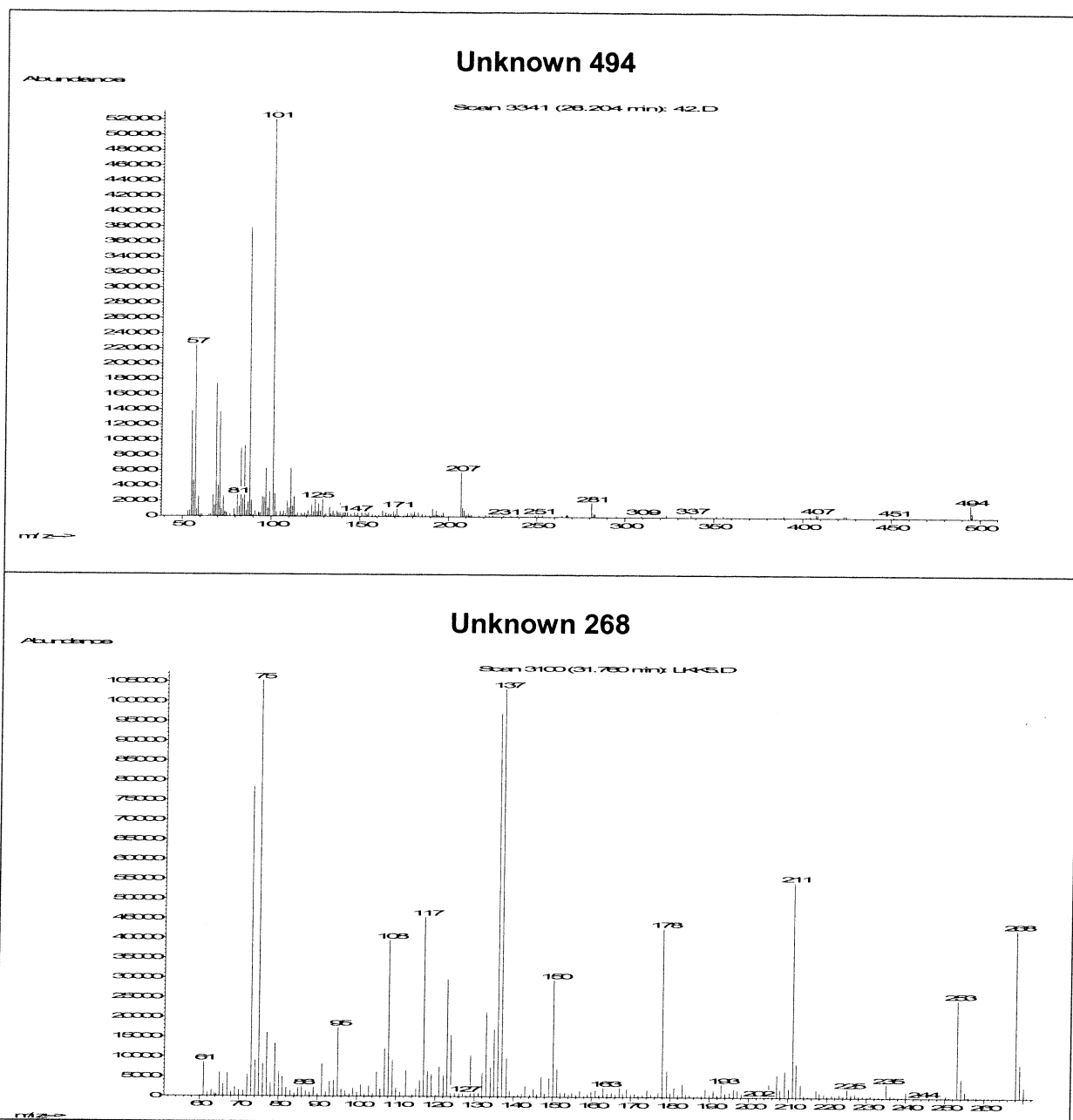
APPENDIX 1

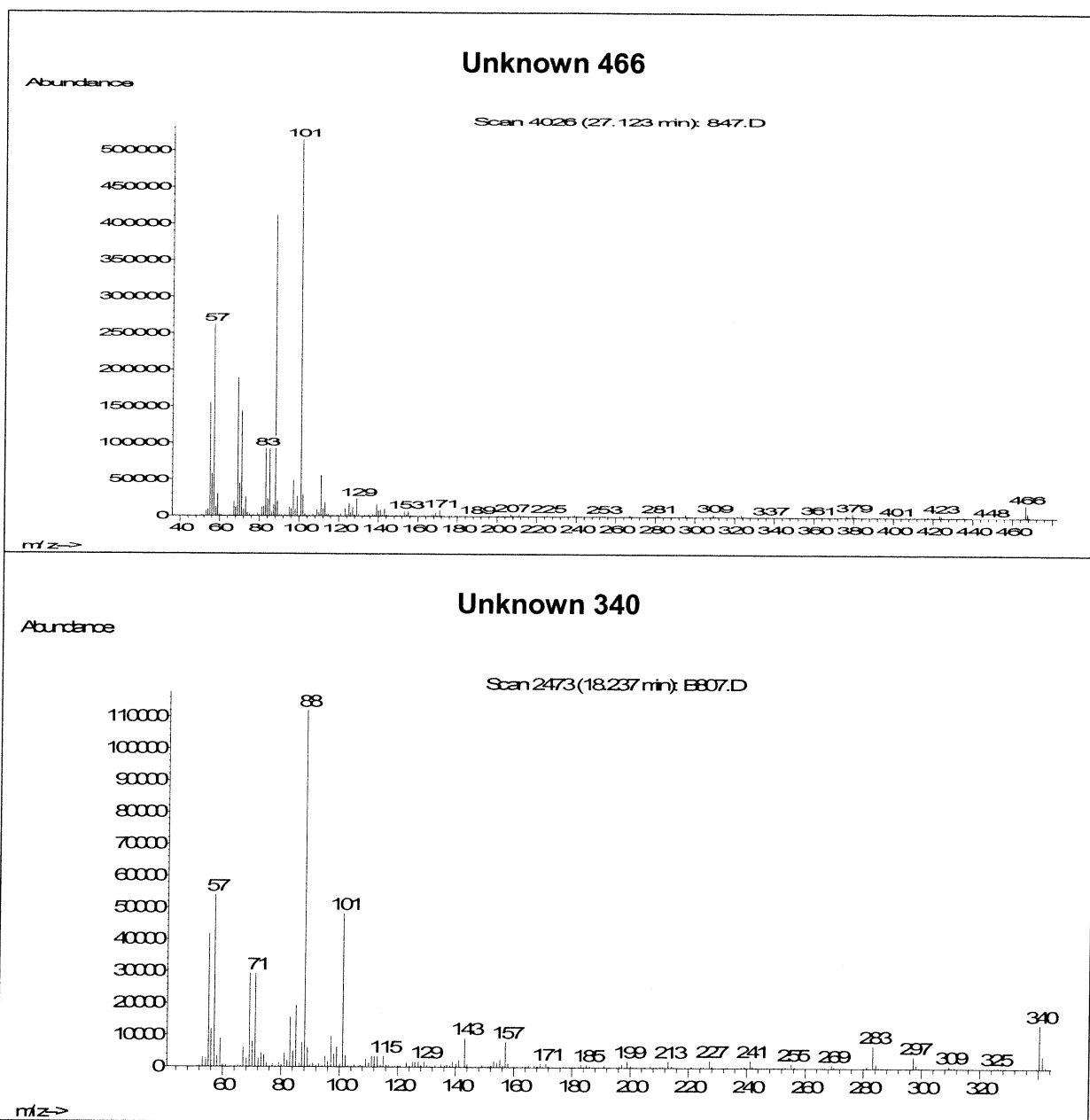
Unknown compounds identified as VIPs using PCA loadings

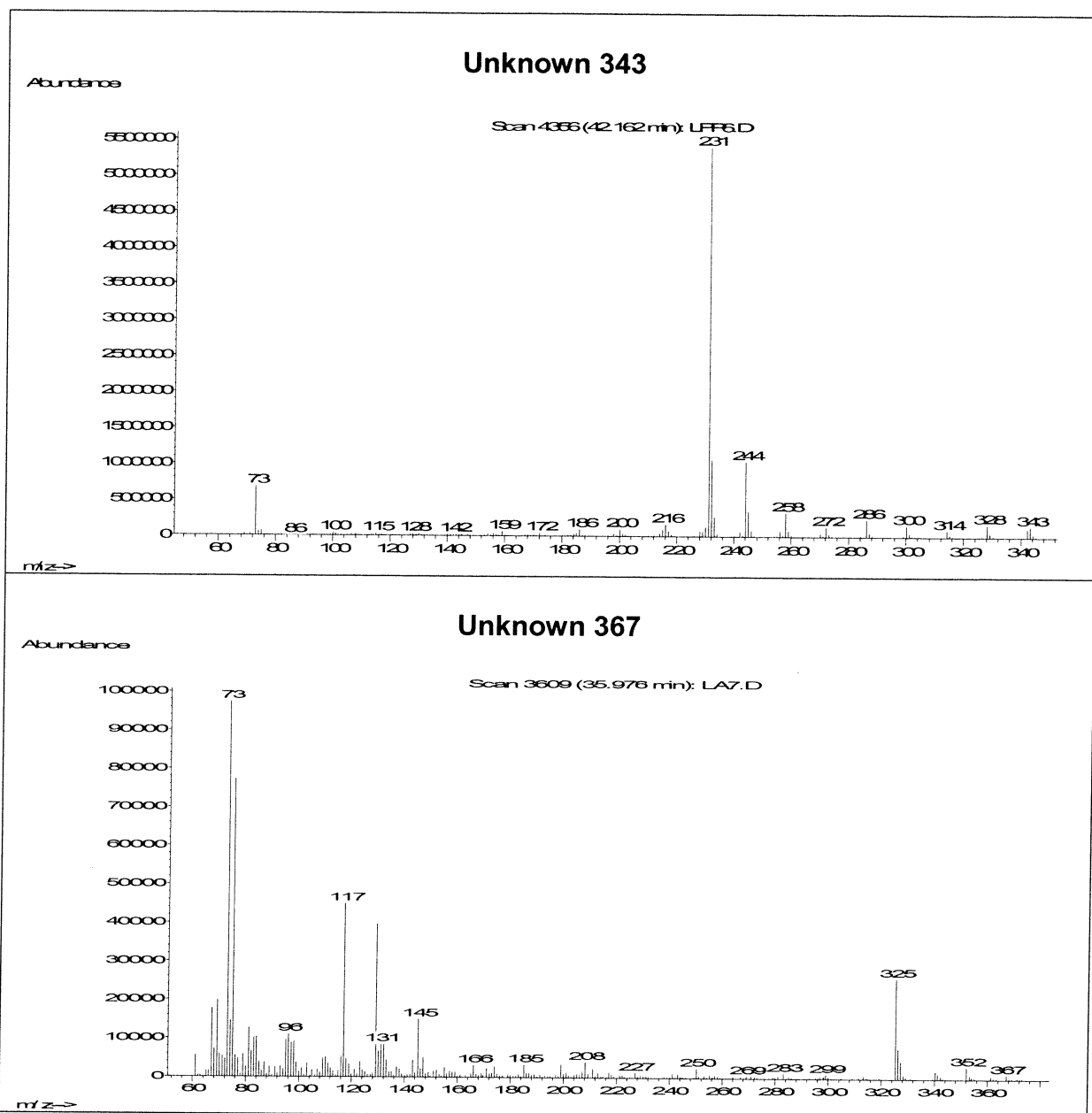
5

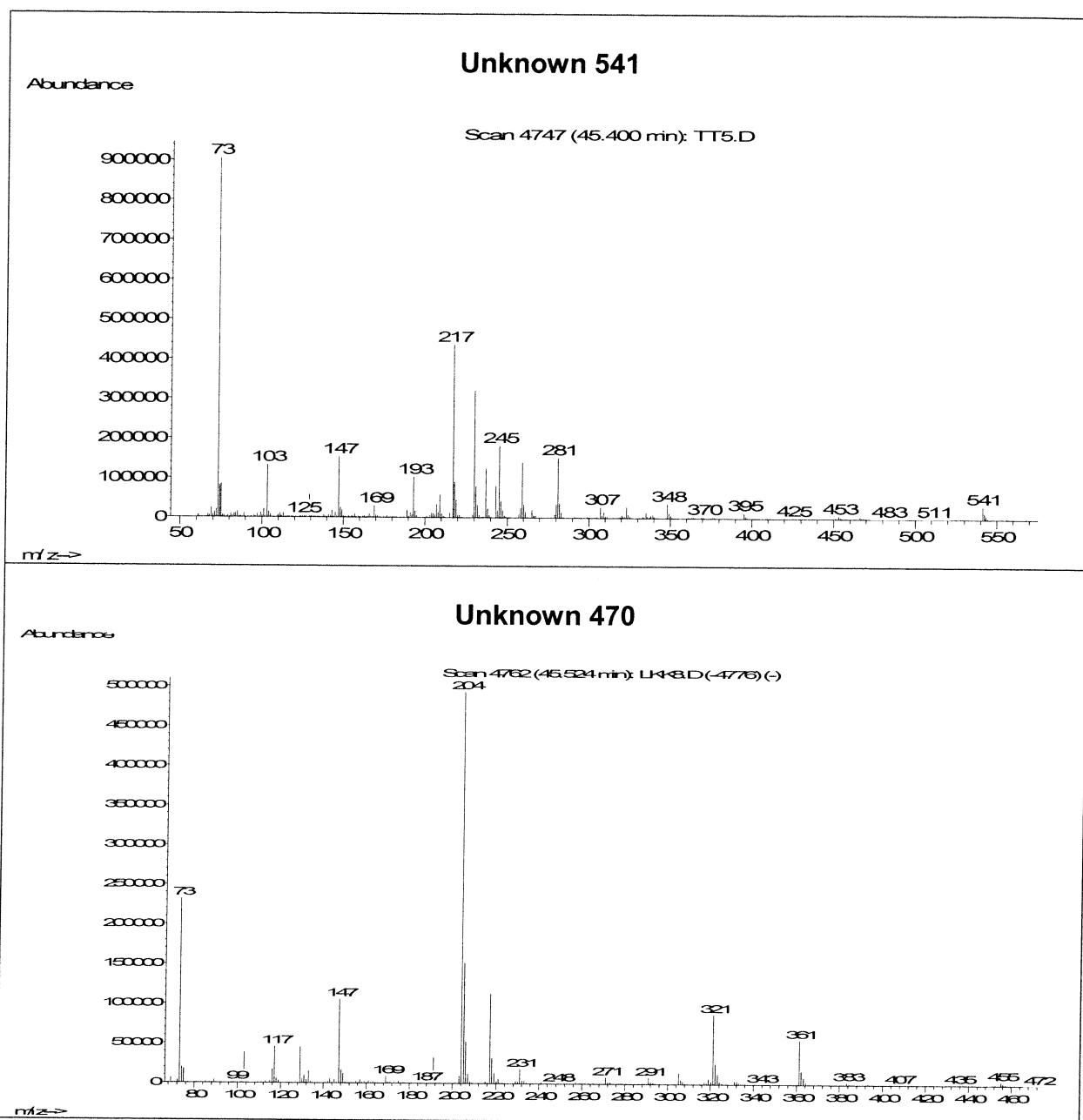


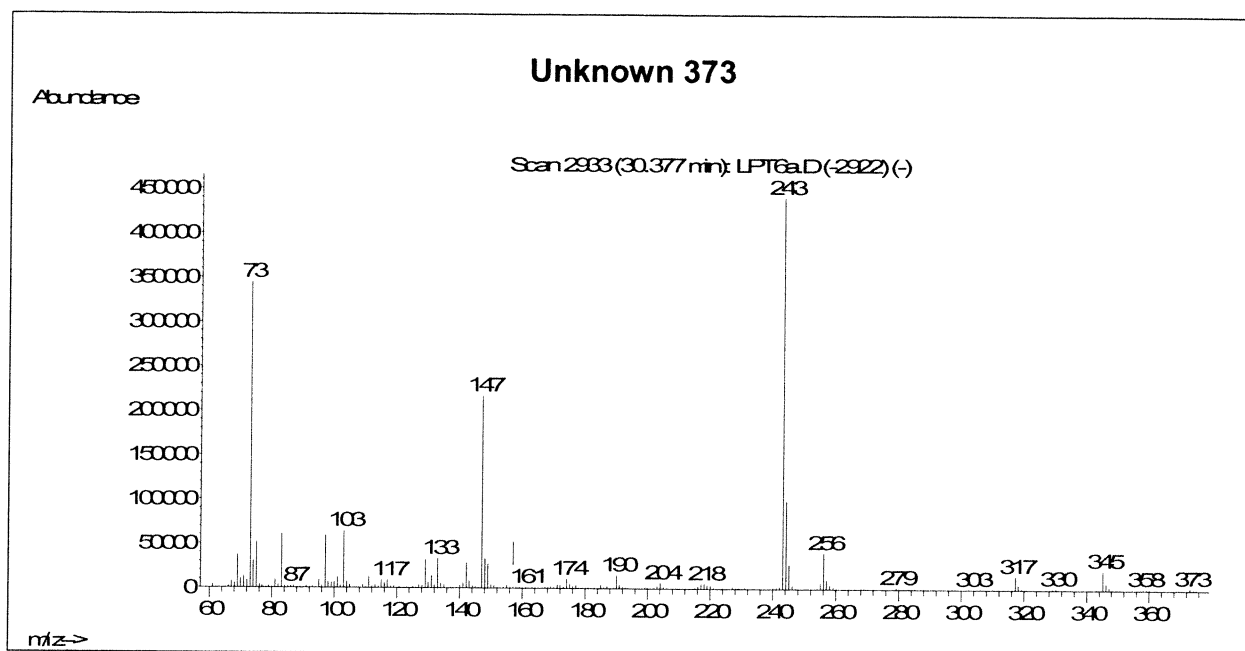




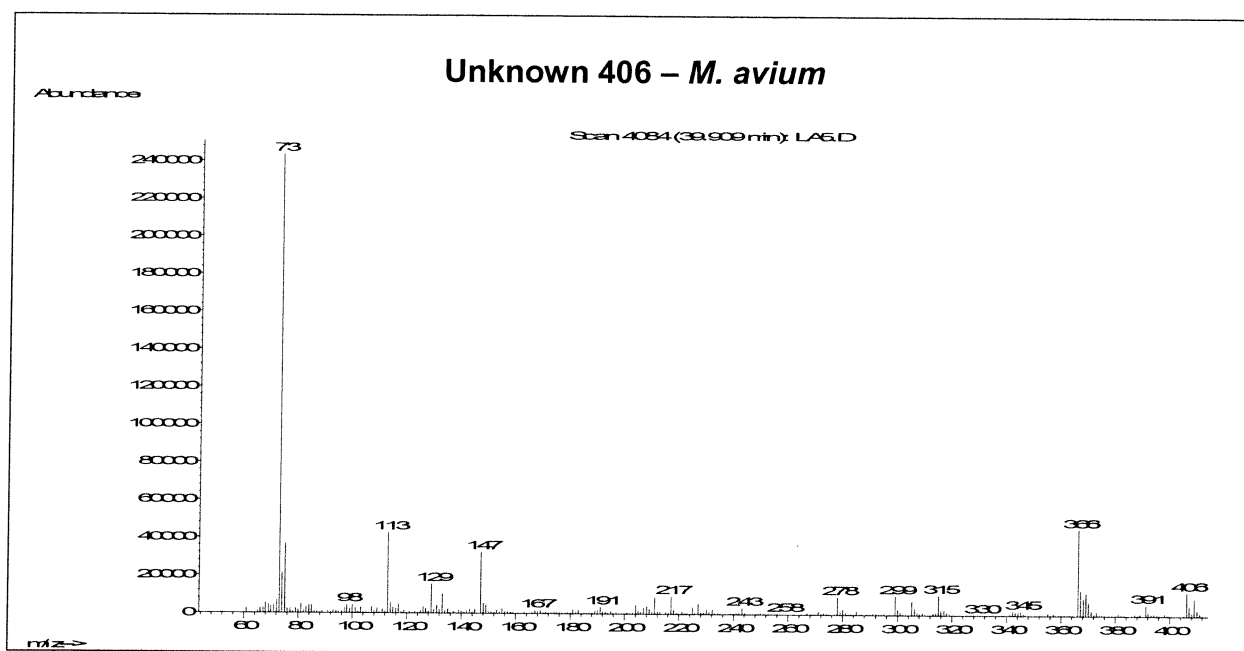


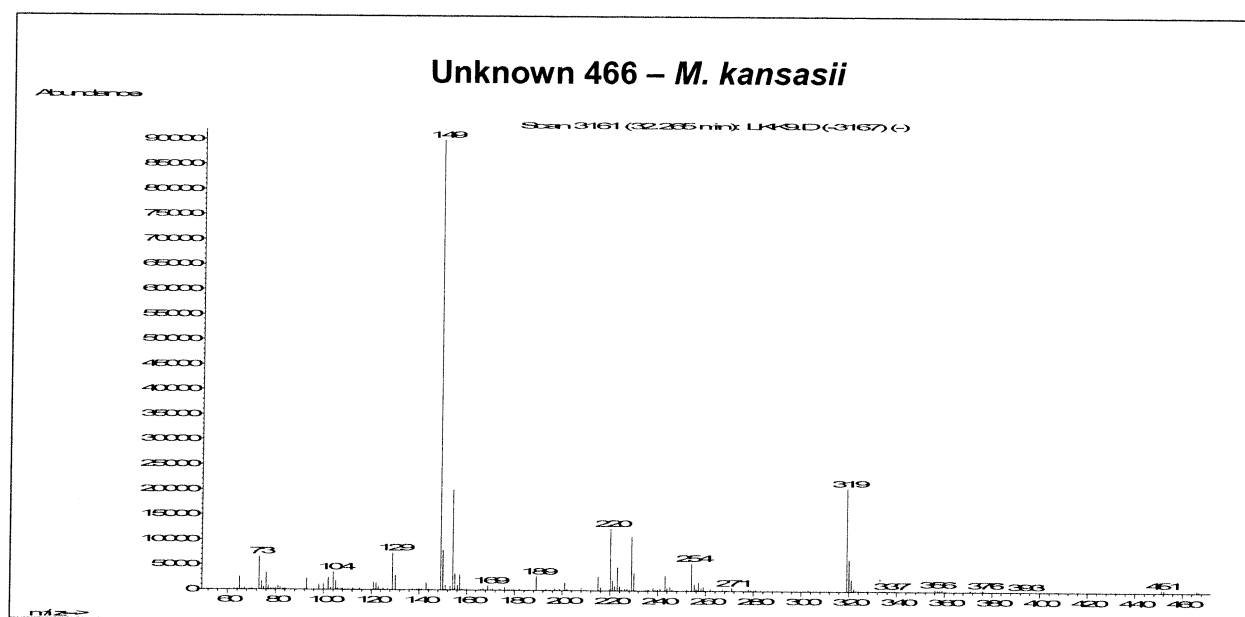
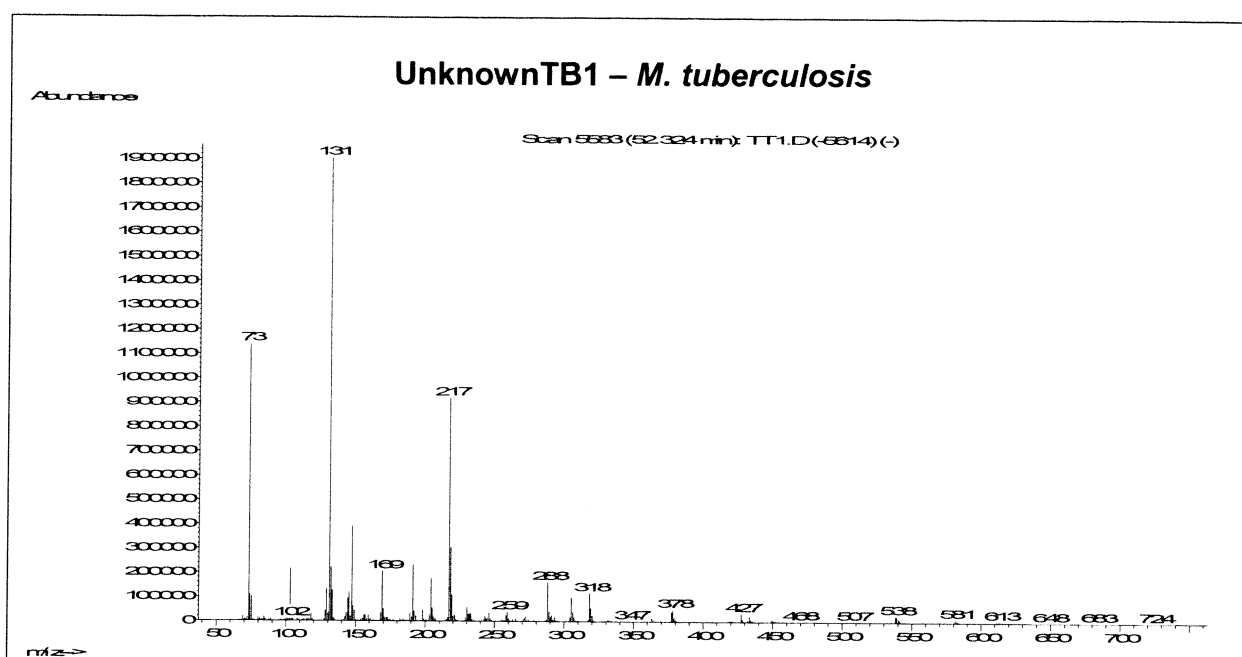


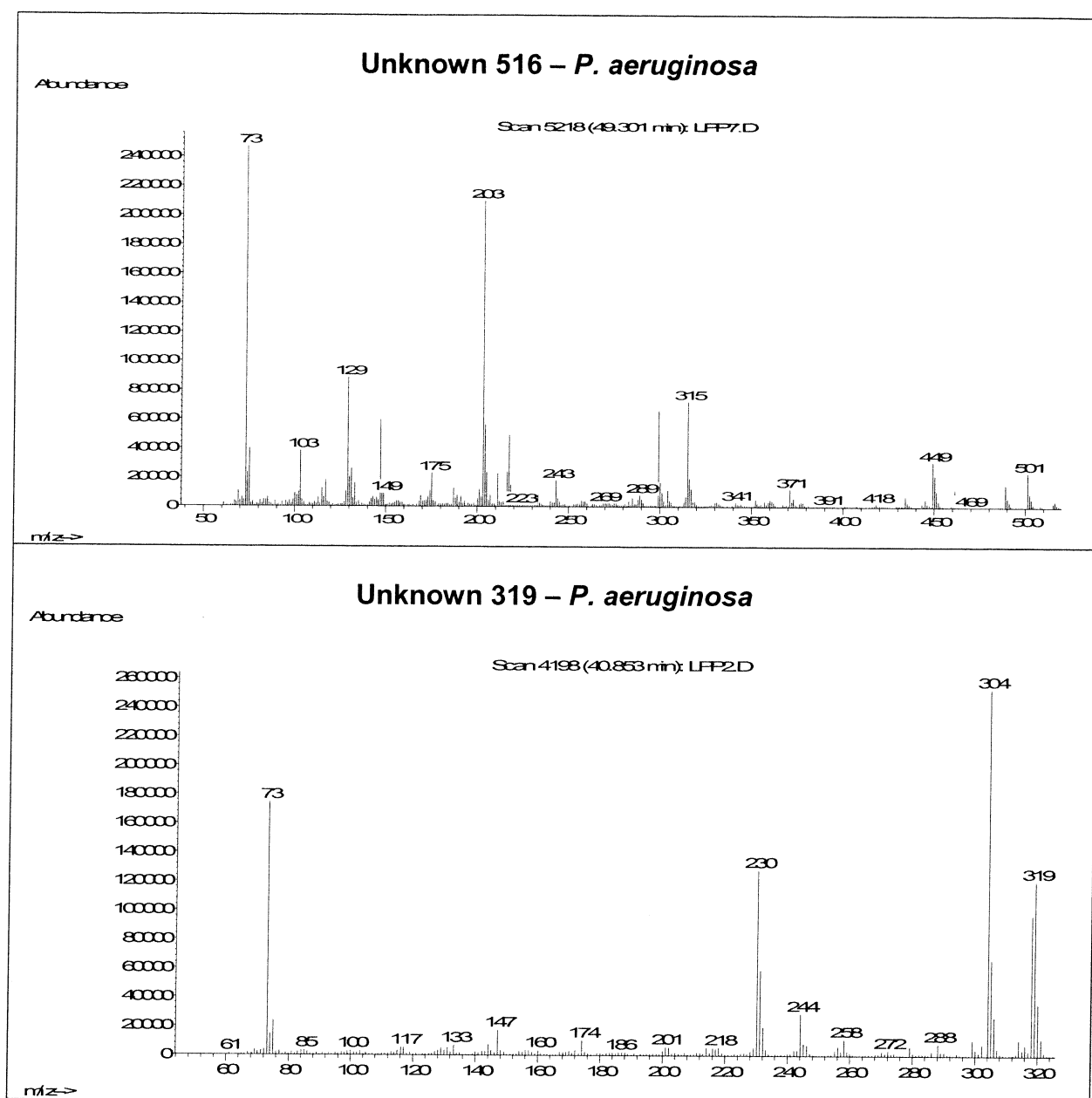


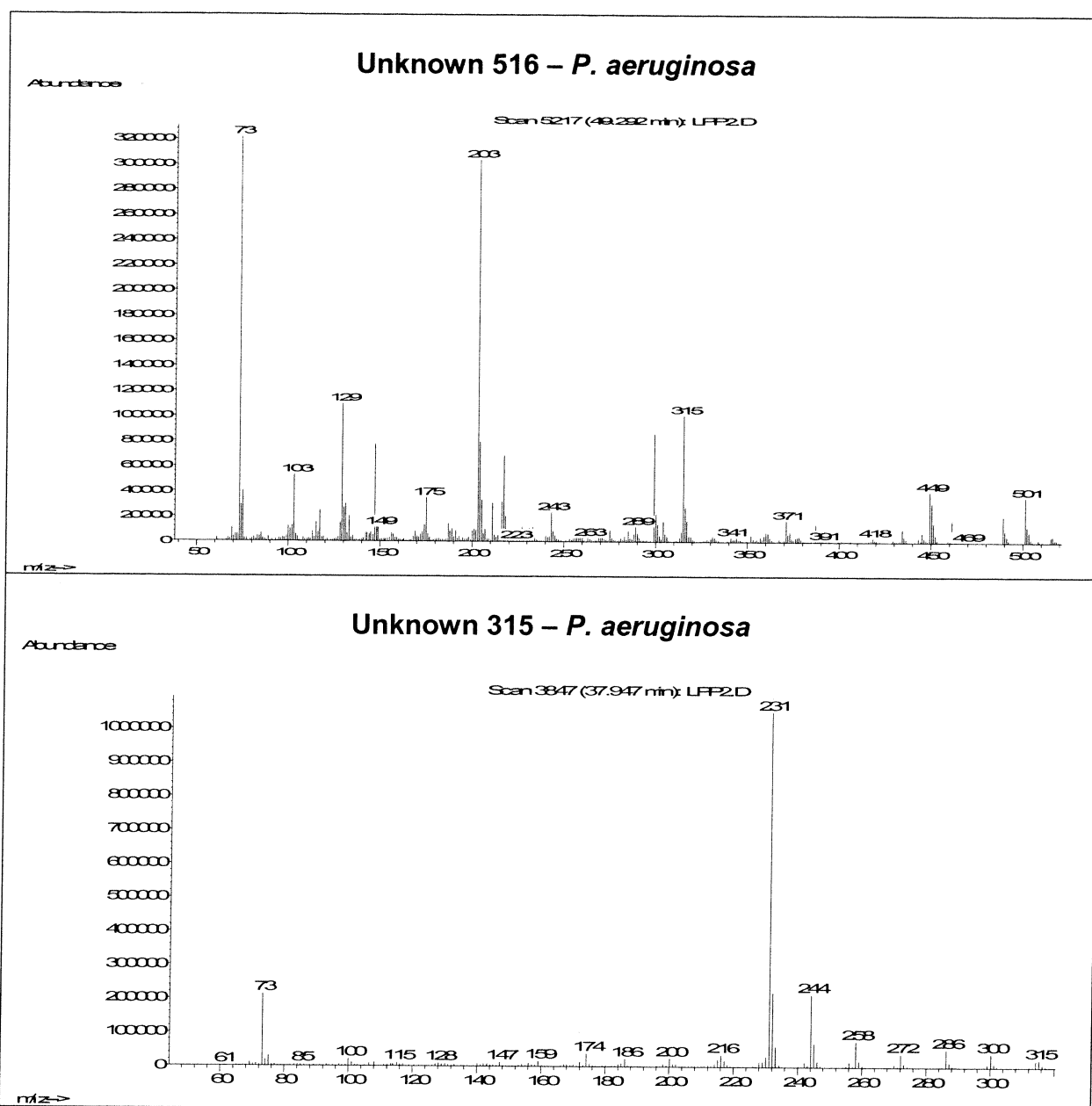


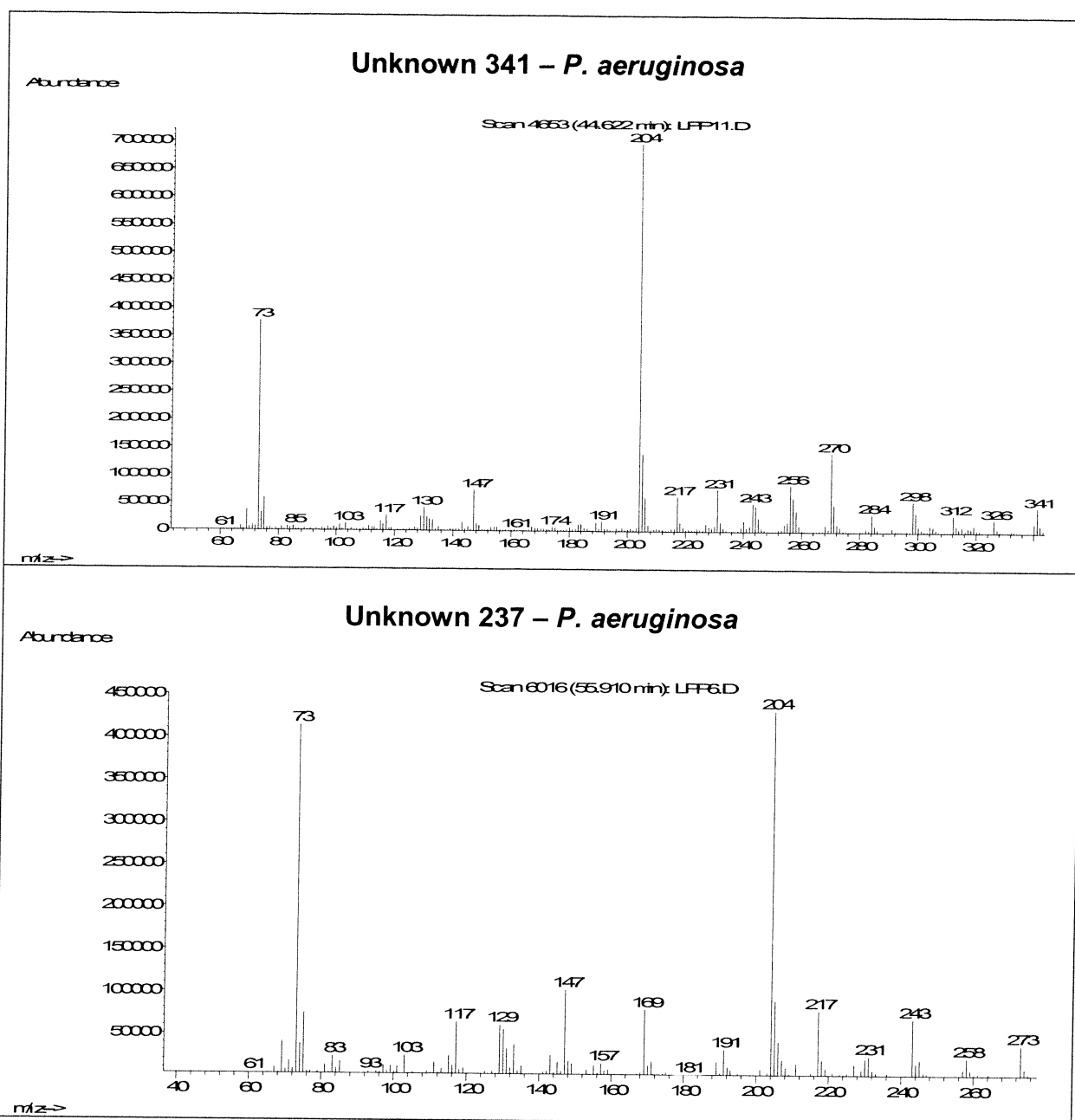
- 5 Unknown compounds identified based on effect size calculations and its uniqueness in species

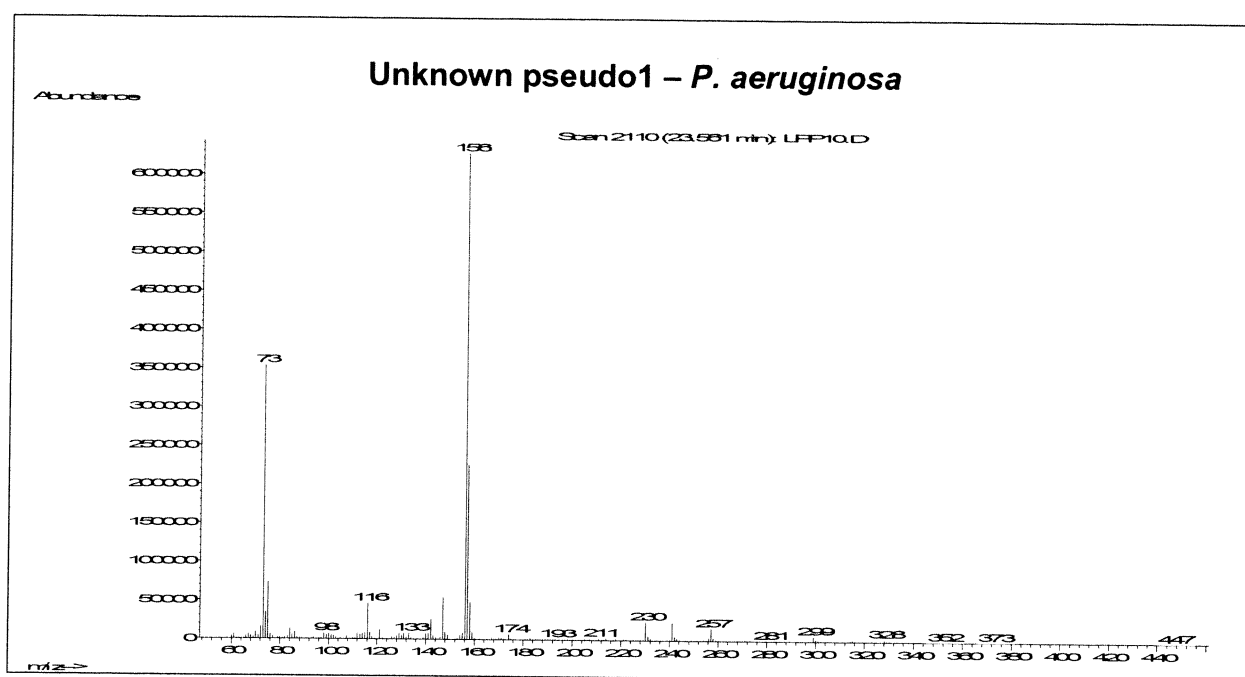












Dated this 21 day of APRIL 2011

MARIUS LE ROUX

Patent Attorney / Agent for the Applicant



FIGURE 1

MARIUS LE ROUX

DM KISCH INC

PATENT ATTORNEYS FOR THE APPLICANT
0494-10-JW

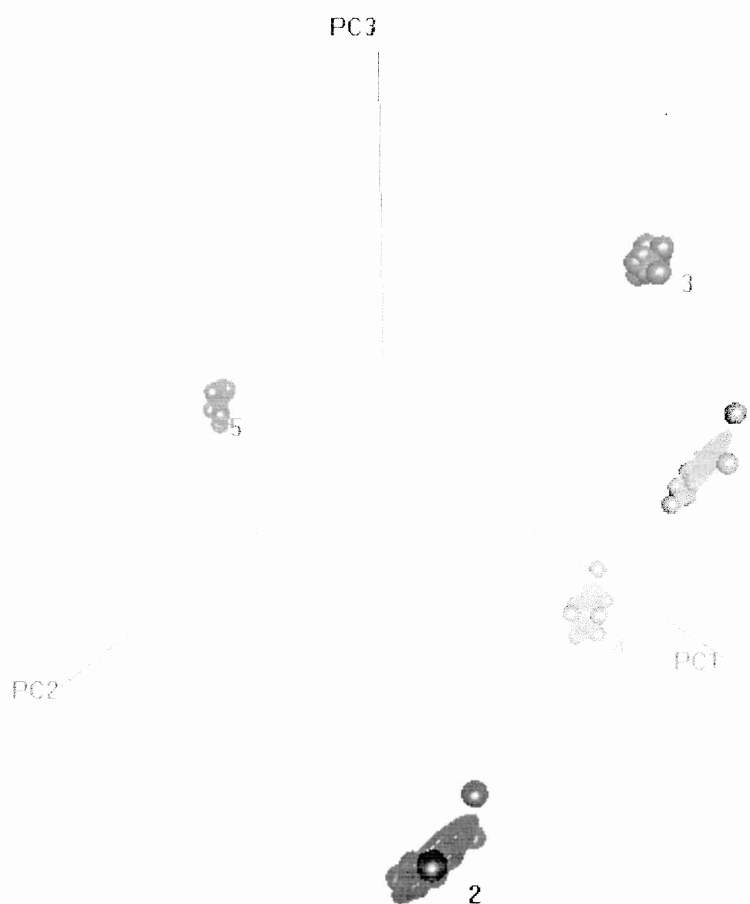


FIGURE 2

MARIUS LE ROUX

DM KISCH INC

PATENT ATTORNEYS FOR THE APPLICANT
0494-10-JW

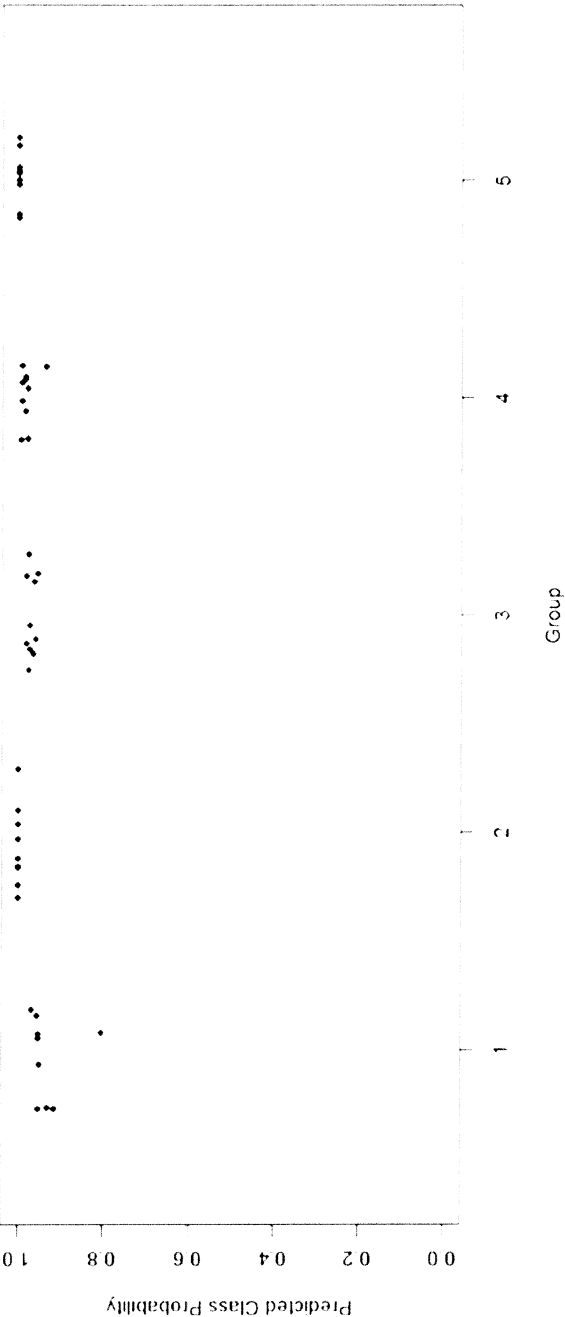


FIGURE 3

MARIUS LE ROUX

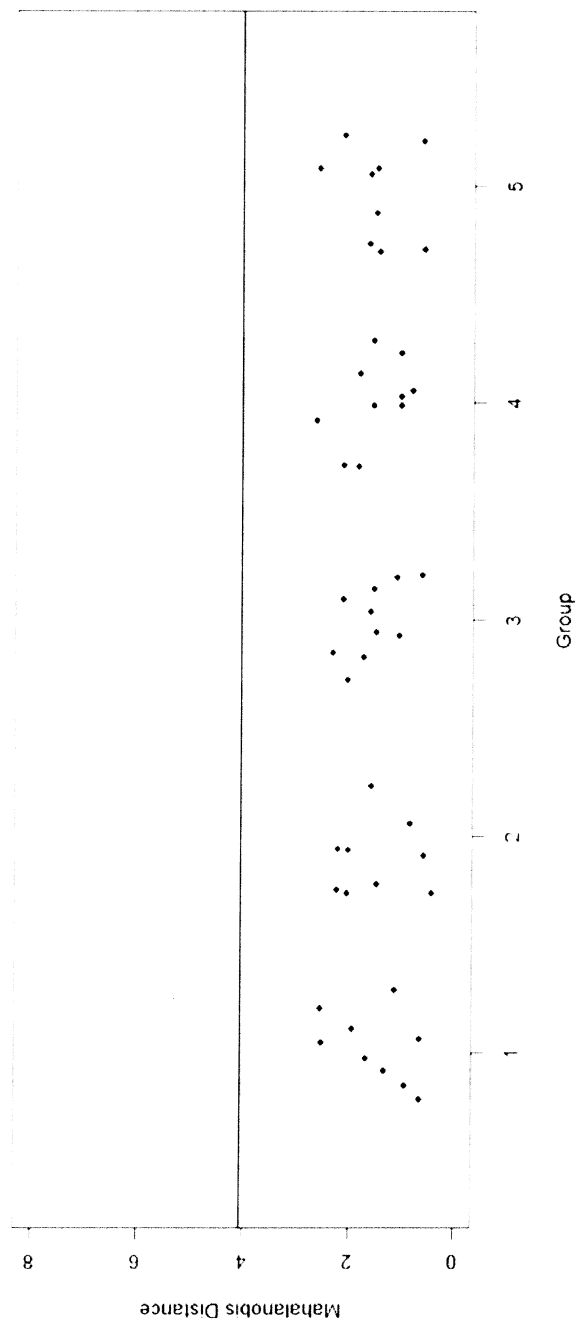


FIGURE 4

MARIUS LE ROUX

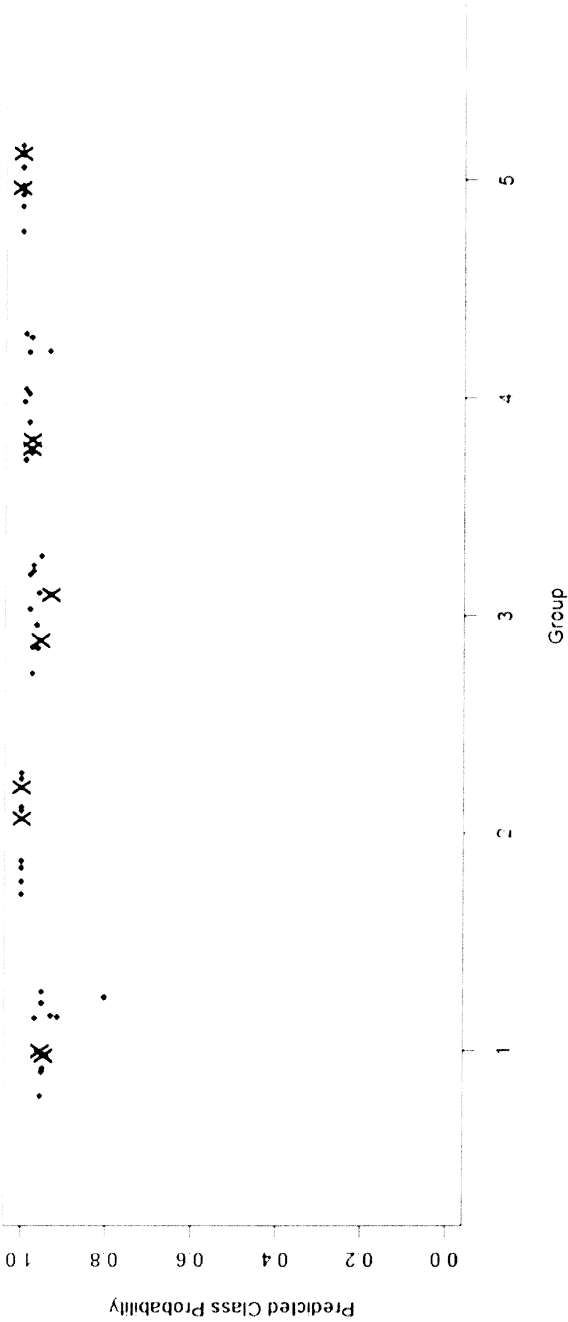


FIGURE 5

MARIUS LE ROUX

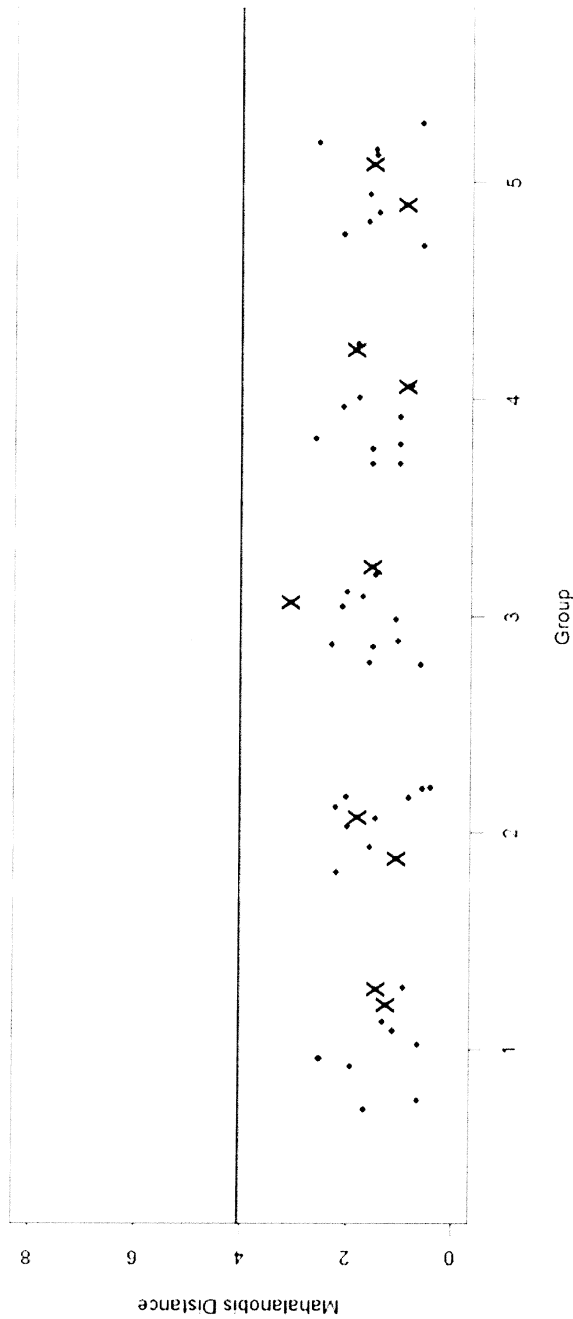


FIGURE 6

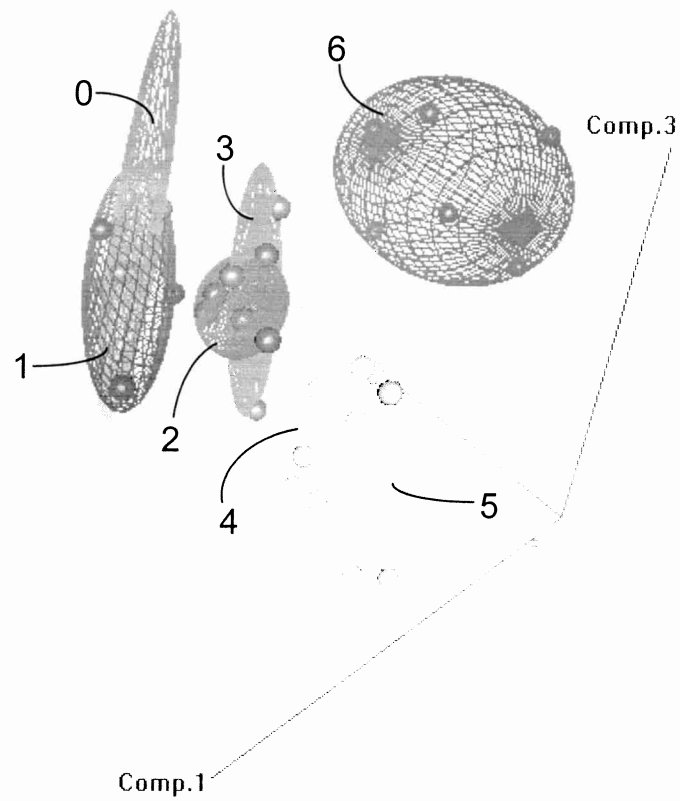


FIGURE 7

MARIUS LE ROUX

DM KISCH INC

PATENT ATTORNEYS FOR THE APPLICANT
0494-10-JW

FIG 8

3

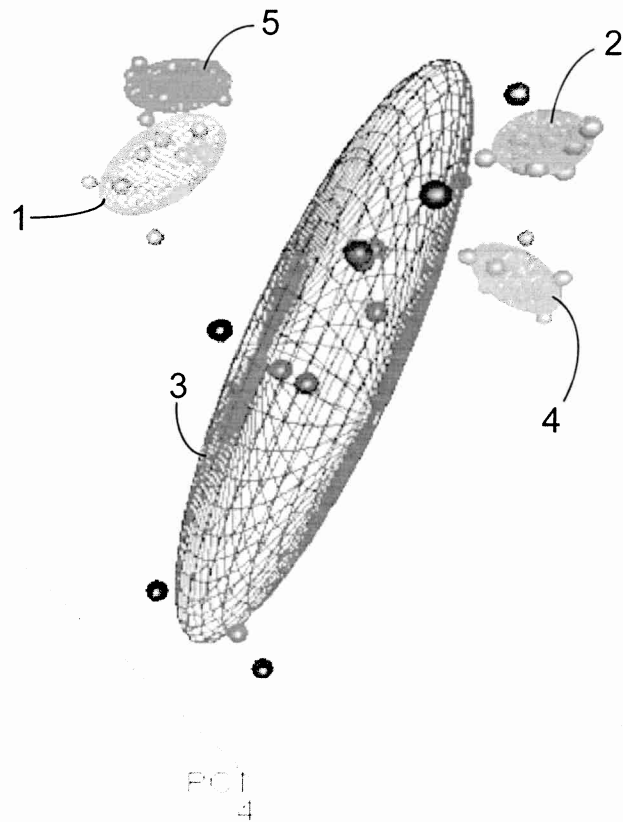


FIGURE 8

MARIUS LE ROUX

DM KISCH INC

PATENT ATTORNEYS FOR THE APPLICANT
0494-10-JW