Metabolomics of Bilharziasis

Gontse P. Moutloatse

BSc Hons Biochemistry

20212100

Dissertation submitted in fulfilment of the requirement for the degree of Master in Science in Biochemistry at the Potchefstroom campus of the North-West University.

Supervisors:

1. **Biochemistry: Prof. C. J. Reinecke**
2. **Bioinformatics: Dr. G. Koekemoer**

May 2012

ABSTRACT

Bilharziasis, a disease that is a major health problem in tropical and subtropical countries, is caused by worms of the genus *Schistosoma.* The main disease-causing species are *S. haematobium, S. mansoni* and *S. japonicum*. Bilharziasis is endemic in South Africa, mostly located in the north-east and covers one quarter of the country, with *S. haematobium* being the most common species. In this investigation we investigated the schistosome-induced changes in human hosts. We systematically investigated the dynamic metabolic profile of humans infected with *S. haematobium* using an untargeted gas chromatography–mass spectrometry (GC-MS) metabolomics approach, including univariate and multivariate data analysis.

The analysis of host urinary composition is a well suited approach to understand the holistic metabolic responses to infections, since metabolomics is a branch of science concerned with the metabolite composition of biological systems and its dynamic response to both endogenous (i.e. physiology and development) and exogenous (i.e. environmental factors and xenobiotics) stimuli. As a holistic approach, metabonomics detects, quantifies and catalogues metabolic processes of an integrated biological system. In this investigation we selected the organic acid component of the metabolome for the metabolic profiling. Organic acids were determined from urine samples obtained from humans infected with *S. haematobium* and a control group of non-infected humans. These metabolites were quantified and identified using an automated mass spectral deconvolution and identification system (AMDIS) from which complex two-dimensional data-matrix sets were created, including assessment of the repeatability in generating a metabolomics matrix. Data matrices were analyzed by principal component and partial least square discriminant analyses (PCA and PLS-DA) to investigate which perturbations existed between the two experimental groups. The biochemical interpretation of the information from these analyses indicated that the main biochemical effects of a *S. haematobium* infection in humans consisted of reduced energy metabolism, liver-function disturbances and perturbations in the gut microbial population common to infections caused by other schistosoma species. Alterations of metabolites of the phenylalanine-tyrosine pathway, including aspects of catecholamine metabolism seems to be novel to a *S. haematobium*

infection and hasn't been reported in current literature. Finally, proposals were formulated for future investigations on *S. haematobium* infection.

Key words: Bilharziasis, *S. haematobium*, Metabolomics, GC-MS analyses, Reproducibility, Multivariate analysis, Biomarker identification

OPSOMMING

Bilharzia, 'n infeksie siekte wat 'n belangrike gesondheidsprobleem in tropiese en subtropiese lande skep, word veroorsaak deur wurms van die genus *Schistosoma*. Die belangrikste siekte-veroorsakende spesies is die *S. haematobium*, *S. mansoni* en *S. japonicum*. Bilharzia is endemies in Suid-Afrika en kom hoofsaaklik voor in die Noord-Oostelike gedeelte van die land, met *S. haematobium* synde die mees algemene spesie. In hierdie navorsing het ons ondersoek ingestel na die geïnduseerde veranderinge in mens weens *S. haematobium*. Ons het die dinamiese metaboliese profiel van mense wat met hierdie spasie besmet is, stelselmatig ondersoek met gebruikmaking van 'n gaschromatografie-massaspektrometrie (GC-MS) gebaseersde metabolomiese benadering, wat sowel eenveranderlike en meerveranderlike data-analise insluit.

Die ontleding van die urine van geïfekteerde individue is 'n geskikte medium om die metaboliese reaksies weens infeksies holisties te verstaan, aangesien metabolomika 'n tak van die wetenskap is met die metaboliet-samestelling van biologiese sisteme en hul dinamiese reaksie op beide endogene (dws fisiologie en ontwikkeling) en eksogene (dws omgewingsfaktore en xenobiotika) stimuli as studieveld. As 'n holistiese benadering ontdek, kwantifiseer en karateriseer metabonomika metaboliese prosesse van 'n geïntegreerde biologiese stelsel. In hierdie ondersoek het ons verkies om die organiese suurkomponent van die metaboloom vir die metaboliese profilering te gebruik. Hiervoor is die organiese sure in urinemonsters kwantitatief bepaal, verkry van mense wat besmet is met *S. haematobium*, en vanaf 'n kontrolegroep van ongeïnfekteerde mense. Die metaboliete is gekwantifiseer en geïdentifiseer deur gebruik te maak van AMDIS, wat komplekse twee-dimensionele data-matrikstelle geskep het, insluitend 'n beoordeling van die herhaalbaarheid van die generering van 'n metabolomiese matriks. Datamatrikse is ontleed deur hoofkomponent-analise en parsiële kleinste kwadrate diskriminant-analise (PCA en PLS-DA) om ondersoek in te stel na steurings wat bestaan het tussen die twee eksperimentele groepe. Die biochemiese interpretasie van die inligting van hierdie ontleding het aangedui dat die hoofbiochemiese uitwerking van 'n *S. haematobium* infeksie in die mens bestaan uit verlaagde energiemetabolisme, lewer-funksie versteurings en versteurings in die mikrobiotoka van die ingewande, wat ooreenkom met infeksies veroorsaak deur ander schistosoma-spesies. Veranderings van metaboliete van die fenielalanien-tirosien-weg, insluitend aspekte van

katesjolamien-metabolisme weens 'n *S. haematobium*-infeksie is nog nie in die huidige literatuur gerapporteer nie. Ten slotte is voorstelle geformuleer vir toekomstige ondersoeke op die gebied van *S. haematobium*-infeksie.

Sleutelwoorde: Bilharzia-parasiet, *S. haematobium*, Metabolomika, GC-MS-ontleding, Herhaalbaarheid, Meerveranderlike analise, Biomarker-identifikasie

ACKNOWLEDGEMENTS

"That I may make the voice of thanksgiving heard and may tell of all your wondrous works."

Psalm 26:7

I would like to thank the following people for their inspiring contribution to my study.

To Prof Carools Reinecke my supervisor thank you for your guidance, supervision and untiring support.

To Dr Gerhard Koekemoer thank you for your support and assistance as my co-supervisor and the statistical contributions to this study.

To Prof Japie Mienie thank you for your valuable assistance and advice.

To Mr Peet Jansen van Rensburg thanks for your support on the experimental aspect of this study and your tireless laboratory assistance. Your efforts have been greatly appreciated.

To Mr Abby Ndlovu of the National Health Laboratory Services (NHLS) thank you for the help on the material-sample collection for this investigation.

Thanks to Prof. Schalk Vorster (Accredited member of the South African Translations' Institute: SATI) for the language editing of my dissertation.

Sincerest thanks to the National Research Foundation (NRF) and the Technology Innovation Agency (TIA) for their financial assistance towards this research.

I would like to give special thanks to my dearest family and the one true to my heart, whom without their unending love and support my journey to achieving my Masters degree, would've been hard-hitting. To my mother thank you for your timeless guidance love and sacrifices you've made on my behalf. For teaching me that "a woman is like a tea bag- you never know how strong she is until she gets in hot water" and to never give up when the road seems tough. To my sister for being my rock and motivator, and for just being you.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

APPENDIX I

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

APPENDIX I

LIST OF ABBREVIATIONS

A

AIDS	Acquired immune deficiency syndrome
AMDIS	Automated mass spectral deconvolution and identification system

B

BN	Bilharziasis negative
BP	Bilharziasis positive
BSTFA	O-bis(trimethylsisyl)-trifluoracetamide

C

CCA	Circulating cathodic antigen
CE/MS	Capillary electrophoresis mass spectrometry
CV	Coefficient of variation

D

DALY	Disability adjusted life year
DNA	Deoxyribonucleotide acid
Da.	Dalton

F

FDA	Food and drug administration, United States of America

G

GC/MS	Gas chromatography mass spectrometry
GC-TOF	Gas chromatography time-of-flight

H

HCl	Hydrochloric acid
HIV	Human immunodeficiency virus
HPLC/MS	High performance liquid chromatography mass spectrometry
^{1}H NMR	Proton nuclear magnetic resonance

I

IEM	Inborn errors of metabolism

L

LC/MS	Liquid chromatography mass spectrometry

M

MB	Mega-byte
MC	Metabolic laboratory controls
MRC	Medical Research Council
MW	Molecular weight

N

Na_2SO_4	Sodium sulphate anhydrous
NHLS	National Health Laboratory Services
NIST	National institute of standards and technology
NMR	Nuclear magnetic resonance
NTD	Neglected tropical diseases
NWU	North-West university

O

O-PLS-DA	Orthogonal-projection to latent structure-discriminant analysis

P

PAG	Phenylacetylglutamine
PCA	Principal component analysis
PLIEM	Potchefstroom laboratory for inborn errors of metabolism
PLS-DA	Partial least square-discriminant analysis

Q

QC	Quality control

R

R&D	Research and development
RNA	Ribonucleic acid

S

S.A.I.M.R	The South African Institute of Medical Research
S. haematobium	*Schistosoma haematobium*
S. japonicum	*Schistosoma japonicum*
S. mansoni	*Schistosoma mansoni*

T

TB	Tuberculosis: *tubercle bacillus*
TCA	Tricarboxylic acid cycle
TMCS	Trimethylchlorosilane
TMS	Trimethylsilyl

U

UP/LC	Ultra performance liquid chromatography

V

VIP	Variables important in projection

W

WHA	World Health Assembly
WHO	World Health Organization

“Data does not equal information

Information does not equal knowledge

And most importantly of all knowledge does not equal wisdom

We have oceans of data

Rivers of information

Small puddles of knowledge, and the odd drop of wisdom.”

……. Henry Nix 1990

Metabolomics is an Art.

CHAPTER 1

INTRODUCTION

By 1980, as the international tide turned decisively against the apartheid regime, South Africa became progressively isolated from the global community in certain sectors such as scientific research and economic development, and was forced to look inward to develop its own research and development (R&D) capacity (Watson, 2007). Thus, following the transition in 1994 to a full democratic dispensation, South Africa became one of today's leading sub-Saharan African countries in the development of competence and forthcoming R&D. Under the heading, "*South Africa – blazing a trail for African biotechnology*", the leading scientific role of South Africa was highlighted in a special issue of Nature-Biotechnology on biotechnological innovations in developing countries (Motari, *et al.,* 2004). Realistically, the editor of that special issue cautions that one of the most important lessons learnt from these biotechnological enterprises, was that it has been "*a long hard slog"* (Marshall, 2004). Nevertheless, one of the main focuses in the country's health biotechnology division is to develop a way toward addressing urgent public health problems of which a good starting point would be to identify a problematic disease prevalent in the community, and a technology that can be applied to address the problem and further develop infrastructure to support research and development in that area (Motari *et al*., 2004). This dissertation is an example of such an initiative – using metabolomics, one of the later developments in the field of omics technologies (Goodacre, 2010), to investigate an aspect of the metabolite profile that emerges in humans suffering from bilharziasis[1].

Bilharziasis is a complex of acute and chronic infections caused by the trematode flatworm of the genus *Schistosoma* which is widespread in tropical and sub-tropical environments and is also endemic in the north-eastern part of South Africa (Gear *et al.,* 1980; Saathoff *et al*., 2004). The transmission cycle requires contamination of surface water by excreta, specific freshwater snails as intermediate hosts, and humans exposed to contact with the

[1] I used the term "bilharziasis" for the infectious disease, also termed as schistosomiasis, caused by various *Schistosoma* species. The term "bilharzia" is also in use for the parasites and /or hosts that cause the bilharziasis disease.

contaminated water. Bilharziasis is one of the most prevalent parasitic infections in the world, and continues to be a global public health concern in the developing world. The main infectious disease-causing species are *S. haematobium, S. mansoni* and *S. japonicum*. According to the recent weekly epidemiological record of the World Health Organisation (WHO), bilharziasis is endemic in 76 countries and territories (WHO, 2011), and it is estimated that more people are infected by the *S. haematobium* species than with the other two combined (Rinaldi *et al.,* 2011). An estimated 650 million people live at risk of infection, and 200 million people are affected, particularly the rural poor living in sub-Saharan Africa, where >85% of the global burden is concentrated (WHO, 2002). A survey published in 1995 estimated that more than four million South Africans were infected with *S. haematobium* (WHO, 1995). Bilharziasis is characterized by focal epidemiology and over-dispersed population distribution, with higher infection rates in children than in adults. It is particularly linked to agricultural and water development schemes and is typically a disease of the poor who live in low socio-economic conditions that favour transmission and who have no access to proper care or effective prevention measures. Although the distribution of bilharziasis has changed over the past 50 years and there has been successful control programmes, the number of people estimated to be infected or at risk of infection remains unchanged (King, 2009). Despite major advances in control and a substantial decrease in morbidity and mortality, bilharziasis continues to spread to new geographic areas. Environmental changes that result from the development of water resources and the growth and migration of population can facilitate the spread of this infectious disease (Hagan *et al*., 1991).

Acute bilharziasis, a feverish syndrome, is often seen in travellers after primary infection. Chronic bilharziasis affects mainly individuals with long-standing infections in poor rural areas. Immuno-pathological reactions against schistosome eggs trapped in the tissues lead to an inflammatory and obstructive disease in the urinary system (*S. haematobium*) or to an intestinal disease (*S. mansoni*, *S. japonicum*), causing hepatosplenic inflammation and liver fibrosis. The need for accurate diagnostic tests is essential for diagnosis at the individual level and for efficient disease surveillance and control at the population level, particularly to monitor large scale therapeutic programs. Microscopic-based analysis of schistosome eggs in urine samples (*S. haematobium*) or faeces (S*. mansoni* and *S. japonicum*) remains the

most commonly used diagnostic method in endemic areas. Although this method has a high specificity, is affordable and requires relatively standard laboratory equipment and training, it's capabilities for sensitivity, predominantly in settings with low infection intensities are weak (Doenhoff *et al*., 2004). During the past decade, progress has been made with immunological methods, namely circulating antigen detection and antibody detection. Although immunoassays display some advantages over parasitological microscopy, wider applications in both clinical and epidemiological settings remain unclear (Doenhoff *et al*., 2004).

Since 2004, a few key investigations have been published on the metabolomics of bilharziasis, using nuclear magnetic resonance (NMR) technology, as will be shown below. **Metabolomics**, is defined as *"*the quantitative measurement of all low-molecular-weight (MW) metabolites (according to general convention, MWs <1 000D) in an organism at a particular time under specific environmental conditions*"* (Nicholson, 2006). The total complement of metabolites is designated as the **metabolome** of an organism. The metabolites in an organism, tissue or cell types at a specified time under specific environmental conditions, have been shown to be an effective tool for disease diagnosis and characterization of biological pathways, thereby capturing the status of the diverse biochemical pathways at a particular moment in time, defining all/any metabolic perturbations (Nicholson *et al*., 2002; Claudino *et al*., 2007; Witkoff *et al*., 2007). Enhancement of our current understanding of a host's metabolic response to a parasitic infection, like that of a human host to the bilharziasis infection, is a promising approach for biomarker identification. In order to investigate whether metabolic perturbations due to the infection by bilharziasis can be identified for diagnostic purposes, the ideal biomarker should be one that can be identified with first-class sensitivity and specificity in patient biological samples, obtained in the least invasive manner i.e. biofluids such as blood, saliva or urine. Through **metabolomic profiling** approaches, which have been increasingly utilised, important quantitative differences in the human metabolome and the investigation of candidate biomarkers from differences within a vast number of endogenous metabolites can now be easily explained (Wu *et al*., 2009). The application of metabolomics for class discrimination and biomarker identification in parasitic infections clarifies the potential of

this methodology for accurate disease diagnosis and as a tool for improved disease surveillance.

Metabolic fingerprinting, which is another subsection of the metabolomics technology, does not attempt to identify or precisely quantify all the metabolites in the sample. Rather, it regards the entire profile, or fingerprint, as a distinctive outline providing a biochemical snapshot of the small molecules produced during cellular metabolism in a particular biofluid, cell line or tissue (Allen *et al*., 2003). Samples such as blood and urine are the most frequently used biofluids for exploring the systematic alteration in the human metabolome, referring to the complete set of small-molecule metabolites as mentioned above (Wu *et al*., 2009). It may also include substances such as metabolic intermediates, hormones and other signalling molecules, and secondary metabolites. Compared with blood samples, utilization of urine samples is preferred as it enables one of the most non-invasive monitoring procedures for metabolomic changes and for investigations and it would probably be the most suitable bio-fluid for investigation of *S. Haematobium* infection, being the urinary variant of bilharziasis.

Metabolomic studies are known to generally employ techniques such as NMR, high-performance liquid chromatography/mass spectrometry (HPLC/MS), liquid chromatography/mass spectrometry (LC/MS), gas chromatography/mass spectrometry (GC/MS) and capillary electrophoreses/mass spectrometry (CE/MS) (Dunn *et al*., 2005). Each technique is characteristic for its own strengths and weaknesses. Among them GC/MS is still the preferred technique for its sensitivity and reproducibility, and has been proposed as an ideal tool for metabolomic profiling of urine samples (Zhang *et al*., 2007). GC/MS is a collective system where volatile and thermally stable compounds are primarily separated by GC, where-after the eluted compounds are detected by the electron impact MS. Coupled with data-reduction techniques, GC/MS offers a powerful approach to generate and analyze high information density metabolic data on biofluids. This approach is capable of simultaneously detecting a wide range of small molecule metabolites, thus providing a molecular characteristic of biofluids under investigation.

Bioinformatics is the term most often used to describe the indispensable contribution of statistics in metabolomics studies. Multivariate statistical analysis along with univariate

analysis, including linear projection methods, principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) are algorithmic methods that can be applied to complex spectral data to assist visualization and characterization of changes relating to a biological perturbation. Metabolomics has consequently become a well-established analytical structure that has been utilized in diverse fields with successful applications, such as the study of disease progression and drug toxicity (Robosky *et al*., 2002) and the detection of metabolites of inborn errors (Wang *et al*., 2004). Current technological advances have enabled researches the ability to perform analysis of more subtle metabolic responses to challenges such as nutritional intervention (Solanky *et al*., 2003), dissimilarities in hormonal cycles, gender differences and diurnal variation (Nicholson *et al*., 2002).

Metabolic profiling by, using NMR-technology has been successfully applied to the investigation of altered urinary metabolites induced by an infection with either *S. mansoni* or *S. japonicum* in mouse and hamster models respectively, as well as on *S. mansoni* infections in humans. This will be discussed in detail in Chapter 2. The urinary metabolite profile in all these studies indicated a stimulated glycolysis process, reduced levels of the tricarboxylic acid cycle intermediates, altered amino acid metabolism and disturbance of the population and/ or activities of gut microbiota, first noted by Wang *et al*. 2004. Here I report the first investigation on the urinary variant of bilharzia by comparing the metabolite profiles of individuals infected by *S. haematobium* with those from non-infected individuals (humans) who served as our controls.

As this is the first study of its kind, a major emphasis was on the methodology of metabolite profiling as a consequence of the infectious disorder which was studied, rather than primarily attempting to identify a specific biomarker for bilharziasis. Thus the dissertation is presented firstly with an overview on bilharziasis, including the metabolomics studies referred to above (Chapter 2). The basic research question and three aims of this investigation are defined at the end of Chapter 2. This is followed by two methodological presentations (Chapter 3 on methodology and Chapter 4 on repeatability), followed by a case–control study of possible changes in the metabolite profile due to bilharziasis (Chapter 5). The significance of this investigation is discussed in Chapter 6, including some suggestions for further studies.

CHAPTER 2

LITERATURE REVIEW

2.1 Bilharziasis

Bilharziasis, also known as schistosomiasis, caused by various species of *Schistosoma,* a genus of the parasitic digenetic trematodes belonging to the phylum *Platyhelminthes*, is among the most severe parasitic diseases in terms of morbidity and mortality and has been highlighted for control by the World Health Organization (WHO) (WHO, 2002; WHO, 2012). Bilharziasis chronically infects more than 200 million people in developing countries, with probably more than 95% of the human infection and the estimated mortality owing to *S. mansoni* and *S. haematobium* in sub-Saharan Africa being 280 000 per year (van der Werf *et al*., 2007; King *et al.*, 2009).

2.1.1 The history of bilharziasis and its geographical distribution

Historical descriptions (Sandbach, 1976) referring to the 'bloody urine' typical of vesicle bilharziasis are already found in the manuscripts of mediaeval Arab physicians. The characteristic haematuria occurring at the end of micturition is likewise mentioned in accounts written by Portuguese doctors working on African trading stations in the 16th and 17th centuries and in reports penned by surgeons attached to the French expeditionary corps in Egypt in 1808. However, it was not until 1851 that the German parasitologists Theodor Bilharz, in the course of a post-mortem examination, discovered the parasite responsible for the disease which now bears his name (Bilharz *et al*., 1853; Eltawil *et al*., 2011). Bilharziasisis is thus a disease which dates back to antiquity. In 1910 Ruffer found calcified bilharzia eggs in a mummy originating from the 20th dynasty, i.e. from 2000-1000 B.C. For a disease "whose destiny is to be born, to live, and to die" – as Charles Nicolle (Cox, 2002) once put it –bilharziasis, despite its hoary age, is still displaying remarkable vitality today; in fact, in almost every environment where it can prevail, it is rapidly spreading hand-in-hand with the development and extension of irrigation

systems. Moreover, attempts to control bilharziasis have so far been handicapped by a lack of really effective therapeutic agents or, at least, of drugs capable of being used on a wide scale.

The persistence and spread of the disease have been enhanced by the following factors: often the impossibility of ensuring strict personal hygiene among the inhabitants of endemic areas; the inadequacy of the drugs available for curing or suppressing the disease; and the impracticality of the measures devised for controlling the vectors, as well as the wide geographical distribution of bilharziasis species as shown in Figure 2.1. Not only is the disease tending to invade new areas, but its prevalence is also increasing in those regions where it is already established (Utzinger *et al*., 2010).

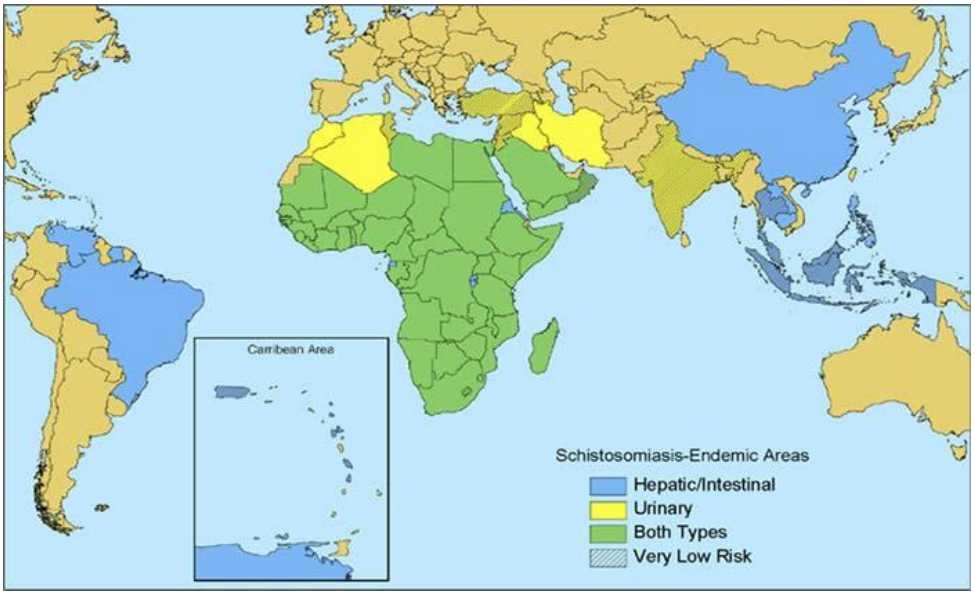


Figure 2.1 Global distribution of bilharziasis*: S. mansoni* and *S. japonicum causing a hepatic/intestinal disorder and S. haematobium*, causing the urinary disorder, are the species of schistosoma that cause the majority of this human disease; they predominate in different areas of the world (Map courtesy of US Centers for Disease Control and Prevention 2008).

Most infections in humans, having significant public health and economic consequences, are caused by the three major bilharziasis species: *S. haematobium, S. mansoni and S. japonicum* (Keiser *et al*., 2002). The consequences of two minor species causing bilharziasis, *S. mekgoni* and *S. intercalatum,* are rather insignificant. *S. haematobium* is endemic in Africa and the Eastern Mediterranean, *S. mansoni* is endemic in Africa and South America, while *S. japonicum* is endemic mainly in China and Philippines. Thus excluding purely intestinal parasitoses, bilharziasis constitutes, after malaria, the second most important epidemiological problem facing the world today.

A pie chart, constructed by and reported in the WHO Preventive Chemotherapy and Transmission Control databank (WHO, 2011), provides a revised report on countries considered to be endemic. It is shown in Figure 2.2 that more than 90% of the population infected and requiring preventative chemotherapy for bilharziasis live in the WHO African region.

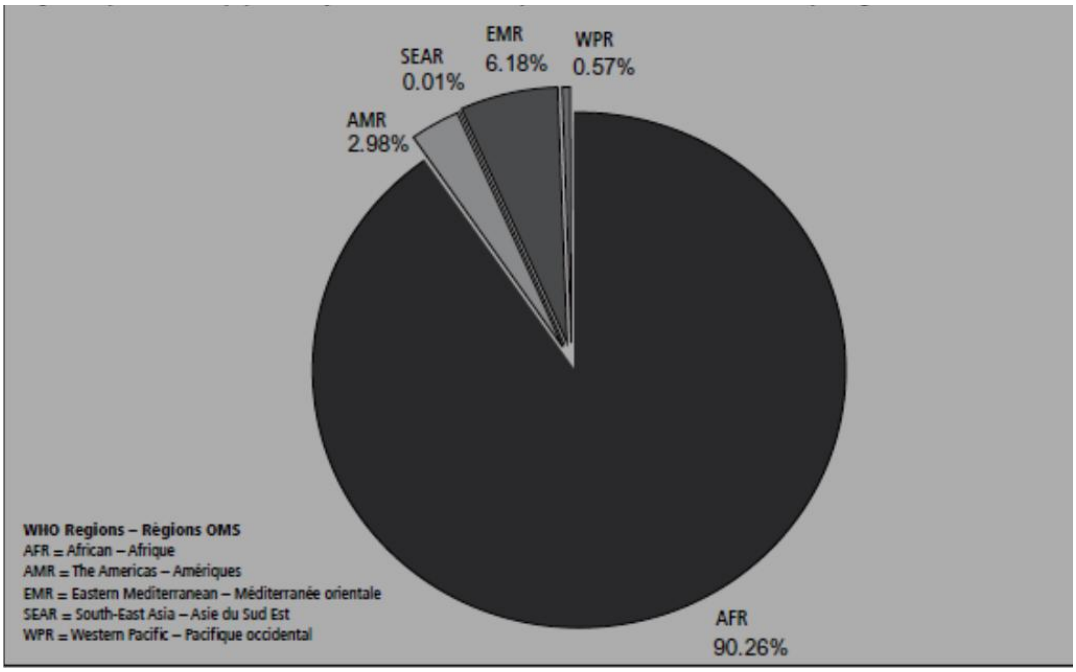


Figure 2.2 Distribution of population requiring preventive chemotherapy for bilharziasis per WHO region, 2009. AFR= African, AMR= The Americas, EMR= Eastern Mediterranean, SEAR= South East Asia, WPR= Western Pacific (WHO 2011).

Unlike the "Big Three": TB, malaria and HIV/AIDS, bilharziasis is one of the 13 World Health Organisations' and other international health agencies' identified neglected tropical diseases (NTDs) (Hotez *et al*., 2006; Hotez *et al*., 2007; Utzinger *et al*., 2010). This is owing to the reality that there are no global funding organizations presently available for combating bilharziasis and other NTDs and unfortunately their priority in U.S. and European pharmaceutical manufacturers market is extremely low. Bilharziasis, perceived as a "poor man's disease" does not occur in the industrialised world or even the substantially wealthy and middle-class in developing countries. According to the WHO (WHO, 2011), bilharziasis is amongst the under-reported NTDs (Table 2.1). Taken from this report, the numbers indicated that on a global scale of the endemic countries only 26.7% reported infection and of those who did report infection, only 8.2% of the estimated infected people were treated in 2009.

Table 2.1 Neglected tropical diseases: Bilharziasis by WHO region, 2009

	WHO Region						
Number of countries and people[a]	**African**	**The Americans**	**South East Asia**	**European**	**Eastern Mediterranean**	**West Pacific**	**Global**
Number of endemic countries	42	10	3	1	14	6	76
Number of countries reporting	14	2	0	0	4	1	21
Number of people treated	14 498 101	30 418	ND	ND	2 550 763	2 491 689	19 570 971

Table obtained from WHO weekly epidemiological record, No. 9, 25 February 2011

ND = no data.

[a] This is the number of countries where bilharziasis is considered endemic according to control of schistosomiasis: second report of the WHO Expert Committee. Geneva, World Health Organization, 1993 (WHO Technical Report Series, No. 830).
(*Also available at* http://whqlibdoc.who.int/trs/WHO_TRS_830.pdf).

It was estimated by the WHO in 2002 that 27 000 people die annually from bilharziasis (case fatality rate of about 0.0014 percent). Many other investigators, however, disagree and believe that this figure is an underestimation, given how poorly reported bilharziasis is (see Table 2.1). Crompton (1999) approximated that 155 000 deaths occurred annually whereas Van der Werf (2003), estimated that bilharziasis mortality alone was at 280 000 per year (case fatality rate of 0.014 percent) having used limited data resources obtained from Africa. Consequently it seems the difference between the approximations for bilharziasis-associated mortality is more than 10-fold (Jamison *et al*., 2006). However, bilharziasis causing so much more chronic disability and morbidity shouldn't be measured by mortality alone. Table 2.2 gives implication of the death toll from annual deaths from neglected tropical diseases.

Table 2.2 Annual deaths from neglected tropical diseases

Annual Deaths from the Neglected Tropical Diseases[1]	
Kala-azar	51 000
African Trypansomiasis	48 000
Schistosomiasis	15 000
Chagas disease	14 000
Soil-transmitted Helminthiases	12 000
Leprosy	6 000
Lymphatic Filariaisis	0
Onchocerciasis	0
Guinea Worm	0
Total Neglected Diseases	146 000

Table courtesy of WHO, world health report 2004; Hotez *et al.,* 2006.

[1]World Health Report 2004.

Ascariasis, Hookworm, and Trichuriasis.

Some estimates indicate that African trypanosomiasis causes 100 000 deaths, leishmaniasis 100 000 deaths, hookworm 65 000 deaths and **schistosomiasis 150 000-280 000** deaths annually. Therefore more than 500 000 deaths annually may result from NDTs.

While it is easier to understand the implication of death tolls and mortality rates, it is more complex to quantify the chronic disability and morbidity effects bilharziasis and other NTDs have in a value that is understood by public health officials and public advocates (Hotez *et al*., 2006). However, bilharziasis and other NTDs are explained and still frequently reported in terms of their disease burden, using the disability–adjusted life year (DALY) as a metric (Murray *et al*., 1996). From the measurement in DALYs it appears (Table 2.3) that bilharziasis and other NTDs only account for approximately one-quarter of the global disease burden from HIV/AIDS. However, even these high disease-adjusted life year values are questioned by emerging studies that are producing even greater figures indicating that these current values are underestimations (King *et al*., 2005; Hotez *et al.,* 2006).

Table 2.3 Neglected tropical diseases ranked by disease burden (DALYs 000)[1]

Lymphatic Filariasis	5 654
Soil-transmitted Helminathiases[2]	4 706
Kala-azar	2 357
Trachoma	2 329
Schistosomiasis	1 760
African Sleeping Sickness	1 598
Onchocerciasis	987
Chagas Disease	649
Leprosy	177
Buruli Ulcer	<100
Guinea Worm	<100
Total Neglected Diseases	20 217
Total HIV/AIDS[3]	84 458

Table courtesy of WHO, World Health Report 2002; Hotez *et al*., 2006)

[1]World Health Report 2002.

[2]Ascariasis, Hookworm, and Trichuriasis.

[3]World Health Report 2004.

2.1.2 Bilharziasis in South Africa

Prior to 1900 urinary bilharziasis was recognized to be only prevalent in the Eastern Cape, in the Natal provinces and the Rustenburg district in the Transvaal (land north or beyond the Vaal river) (Doumenge *et al*., 1987). A few years and surveys later it became evident that the infection had spread to south of the Limpopo river in the Transvaal (Wolmeranstad, Klerksdorp and Potchefstroom and Harts river valley), the north and coastal belt of Natal and further south on the Eastern coast of Cape Province, from Port Edward to Humansdorp (Doumenge *et al*., 1987).

Presently the geographical distribution of bilharziasis lies largely within the eastern half of South Africa, and involves parts or all of six provinces namely Limpopo, North West, Gauteng, Mpumalanga, KwaZulu-Natal and the Eastern Cape (Gear *et al*., 1980; Mqoqi *et al.,* 1996; Moodley *et al*., 2003; Saathoff *et al*., 2004). *S haematobium*, which is responsible for urinary bilharziasis, is much more widespread than *S. mansoni* which causes intestinal (colon-rectal) bilharziasis. Prevalence of *S. haematobium* in children on the eastern escarpment is highest (>70%) in the lowlands and near the coast, but decreases with increasing altitudes to around 1 000m. Transmission is also common across the eastern Highveld with outliers west of Johannesburg in part of the North West Province (Wolmarans *et al*., 2006). Transmission in the Eastern Cape is patchy. A 2001 outbreak in the Jeffreys Bay area, west of Port Elizabeth, was attributed to infected migrant fishermen from further north. Intestinal or rectal bilharziasis (*S. mansoni*) is limited to the low-lying areas of the three eastern provinces, namely Limpopo, Mpumalanga and KwaZulu-Natal. Its distribution overlaps partially with that of *S. haematobium* so that in these three provinces children may be infected with both these parasites (Schutte *et al.,* 1995). The South African Institute of Medical Research (S.A.I.M.R.) and the Medical Research Council (MRC) Unit for Snail Research at the previous Potchefstroom University for Christian Higher Education, compiled an extensive map of the distribution of bilharzia snails in South Africa, including both *S. haematobium* and *S. mansoni* (Figure 2.3). This was published as a detailed atlas in 1980 and documents a high prevalence of *S. haematobium* which is endemic in 67 districts in South Africa (Gear *et al*., 1980).

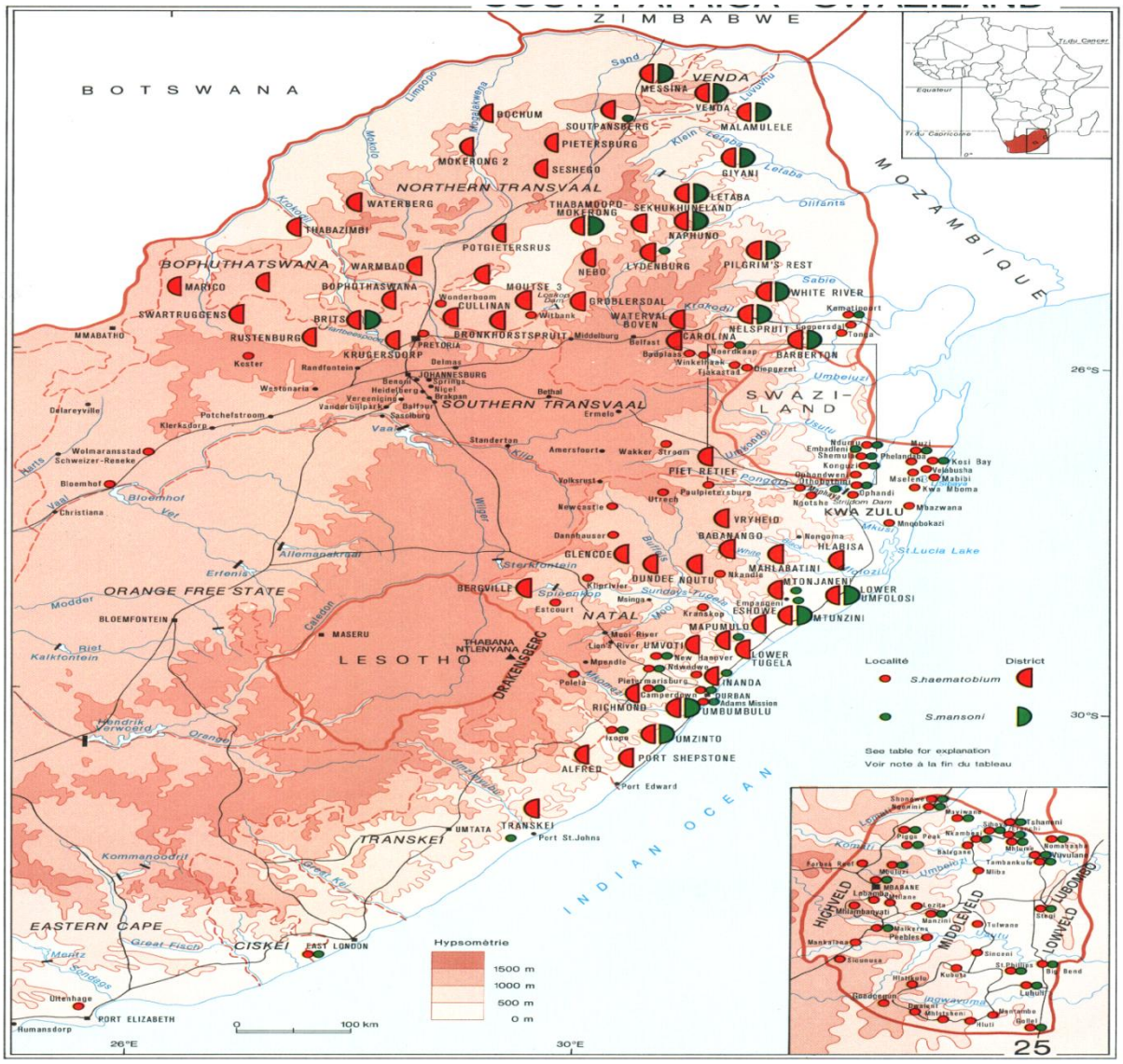


Figure 2.3 Atlas of Bilharziasis. Johannesburg, South African Institute for Medical Research, 48 maps ● ***S. haematobium*** ● ***S. mansoni*** (Courtesy of Gear J.H.S., Pitchford R. J., Van Eeden J.A. and other co-workers compiling the atlas in 1980).

2.2 Life-cycle and physiology of the parasite

It has been increasingly recognized that diagnosis of bilharziasis, particularly in the early stages of the infection, is a problem. However, before consideration of diagnostic tests, it is necessary to recall the life cycle of the bilharzia parasite and the clinical manifestations of the disease. *Schistosoma*, a genus of all the schistosomes that mature in man does not multiply in the human body.

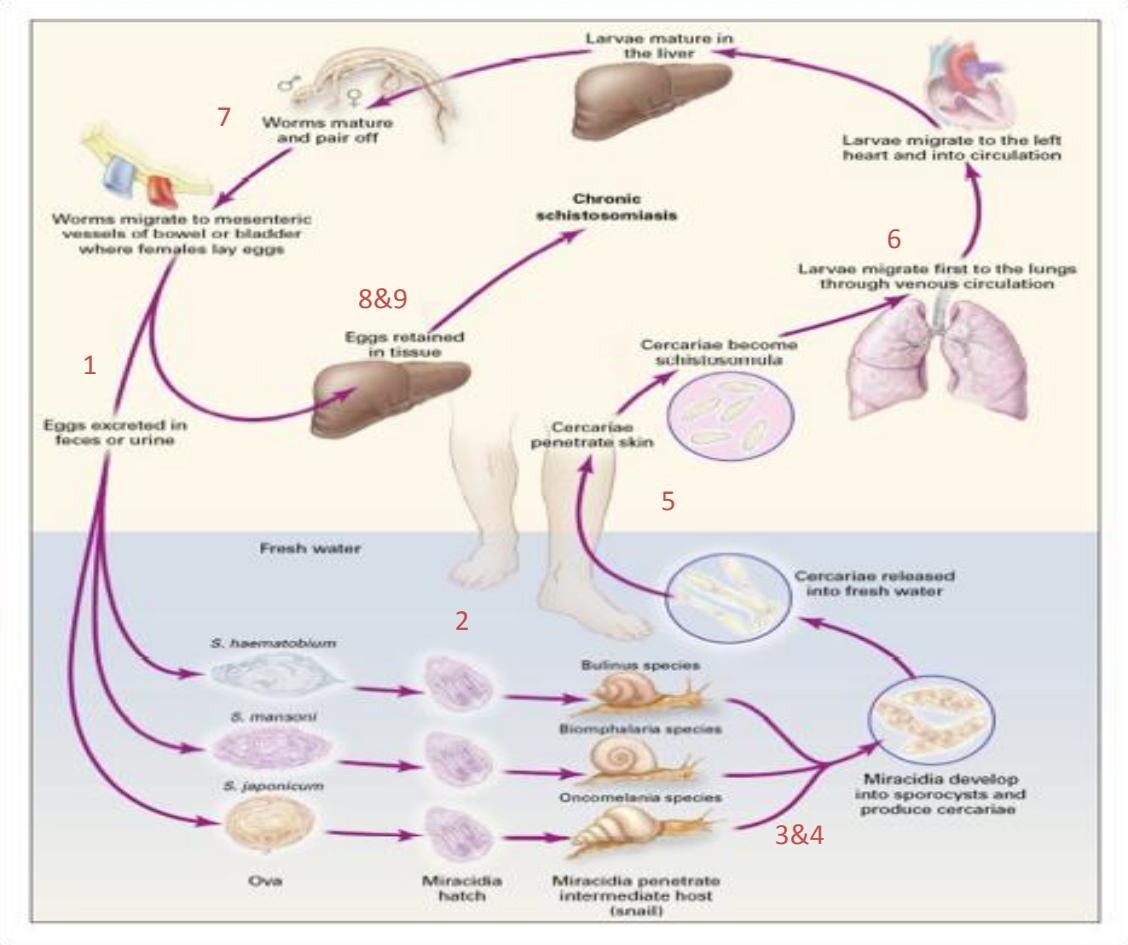


Figure 2.4 *Schistosoma* life-cycles (Diagram taken from: Ross *et al*., 2002)

Eggs from infected humans are released via urine/faeces into the water (1). On contact with water the eggs hatch, releasing miracidia (2) which infect fresh water snails (3). Sporocysts migrate to the snail's hepatopancreas (4) and asexually produce free swimming cercariae (5), which in turn penetrates the skin of the animal/man (6). The cercariae now termed

schistosomula migrate via the lungs to the liver (7), where they mature in the host (8) and lay eggs (9, 10) which are either trapped in the adjacent tissues or are released into the environment via faeces or urine.

The life cycle of schistosome is illustrated in Figure 2.4. It includes a free swimming form called the cercariae (no. 5 Figure 2.4). The stage which is infective to man, the cercariae, characteristically has a forked tail and is shed from the intermediate host snails. Probably more are shed during day-time than at night and they are also more active in daylight, with their forked tails acting like propellers (Gryseels *et al*., 2006). The release of *S. haematobium* cercariae is particularly high when the daily average maximum temperatures reach 30°C or higher for six to seven months of the year, which is the case in the most northerly low-altitude tropical and plain regions of Southern Africa (Appleton, 1977). Cercariae can survive for 8 hours after emerging from their mollusc hosts, during which time they must find their definitive host, man, or they die. Cercariae secrete elastase originating from special glands in their heads, which enable them to penetrate the skin of the host or of animals that come into contact with them in water. If successful in their pursuit, they shed their tails and actively pierce the skin of their host gaining entrance (Wilson, 1987).

The cercariae then transform their trilaminate tegument, covered by a glycocalyx, replacing the cercarial lipid bilayer and glycocalyx with a double lipid bilayer of the schistosomulum along with various physiological changes, such as a change from aerobic to anaerobic respiration and the acquisition of host molecules, predominantly lipids, some of which are incorporated into the tegument, which form part of an adaptation to the definitive host environment (Wilson, 1987). Now referred to as schistosomula, they travel via the lymphatics and veins through the human body. Through the blood stream they reach the right side of the heart and then reach the lungs where it appears that they pass through the capillaries to the pulmonary veins, and then to the left side of the heart (Wilson, 1987). From there via the blood stream they are widely distributed within the host where ultimately, some reach the intra-hepatic branches of

the portal vein by a route which has not yet been clearly defined. It is in the liver where they develop to maturity (Gryseels *et al*., 2006). There are separate sexes. The shorter and broader male encloses the smaller but longer female in a groove made by the ventral folding of the sides of the body, called the gynaecophoric canal (see Figure 2.5). The adult worms are about 1cm long, and the male has a deep ventral groove or schist (hence the term '*schistosome*'), in which the female worm resides permanently *in copula* (Wilson, 1987).

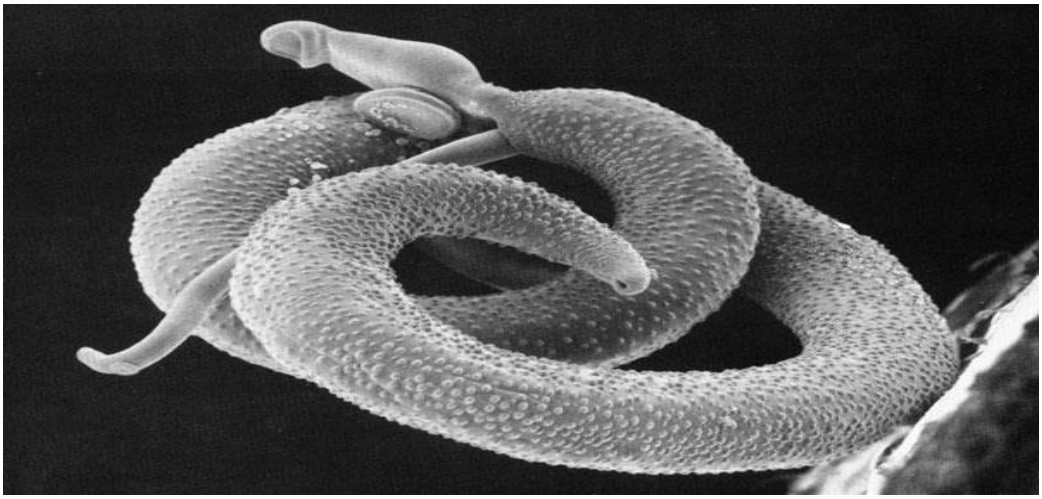


Figure 2.5 Paired adult *S. haematobium* worms (Image adapted from *U.S. Centre for Disease Control and Prevention* 2008) *(including cover page).*

Permanently interlocked the darker female lies within the gynaecophoric canal of the larger male worm. Schistosomes are dioecious and can measure up to 10-20mm in length and 0.5-1.0mm in width. Both sexes have 2 suckers, a ventral and anterior sucker. The male has a deep ventral groove known as the gynaecophoric canal, in which the female lies permanently, facilitating also copulation. The gut of the female appears darker because it is filled with deposits of haematin (breakdown products of haemoglobin) (Rumnajek, 1987).

Table 2.4 The comparison of Schistosoma species.

Item	*S. haematobium*	*S. japonicum*	*S. mansoni*
Adult worm			
Location in host	Vesicle plexus	Mesenteric veins	Mesenteric veins
Female			
Length (mm)	16-26	20-30	10-20
Mature egg			
Shape	Ovoid	Round	Ovoid
Size (µm)	62x150	60x100	61x140
Eggs/day/female	20-300	3,500	100-300

Table adapted from WHO, 2002.
The table shows a comparison of the adult and egg stages between the 3 major bilharziasis species that effect humans

The adult worms of each sex have single openings at their anterior end, which serves as both a mouth and an anus. Around this single gut opening is the oral sucker and situated further back the ventral sucker, which the worms use mainly as an appendage for hanging on to the venous epithelium of the host and as a form of movement for the worm pair. Adult schistosome worms ingest blood cells to make use of the haemoglobin as their primary source of amino acid for growth and development. They employ haemoglobinase to break down the haemoglobin. Free amino acids and small molecules including glucose, purines and pyrimidines, are transported across the tegument via transtegumentary absorption. The tegumental fold of the male worm enclosing both the female and the male seems to also assist in pumping blood into their anterior ends. It is through transtegumentary absorption from the male worm where it seems the female derives much of her nutrition. Adult schistosomes derive their energy largely through the breakdown of glucose and glycogen owing to their nature of being facultative anaerobes (Rumnajek, 1987).

In an *S. haematobium* infection, intertwined together, the adult worms migrate against the blood stream of the portal circulation to the smaller venules of the veins of the pelvic plexuses into the blood vessels of the bladder or intestine (Table 2.4) (Kusel, 1970). The adult schistosomes survive by absorbing amino acids and large amounts of glucose from the host, equivalent of its dry weight (Rollinson *et al*., 1987). The female, after raising her posterior sucker, lays her eggs which then dissolve their way into the bladder and are excreted in the urine. The miracidium (being the free swimming ciliated first larva of the digenetic trematode) hatches from the eggs in water and then in turn seek out and penetrates a suitable snail, the intermediate host being *Bulinus (Physopsis) africanus*, thereby infecting it (no. 3 Figure 2.4). It was Harley (1864) who speculated that there was an intermediate mollusc host, which was not accepted until Miyairi and Suzuki (1913) (Cox, 2002) confirmed this in their studies on *S. japonicum*.

This observation was later confirmed by others, with the snail hosts of *S. mansoni* and *S. haematobium* soon identified. Subsequent to finding the correct snail host and infecting it, the parasite forms a sporocyst (mother sporocyst) at the site of penetration that produces daughter sporocysts that migrate to the snail's hepatopancreatic region. There the parasite asexually produces cercariae (which are the free swimming larva form in which the parasitic fluke passes from the mollusc intermediate host to another host being the human), which in turn are released into water (no.5 Figure 2.4) (Kusel, 1970). The freshwater snails may remain infected for several months, releasing cercariae daily (Despommier *et al*., 2005).

2.3 Pathological characteristics of bilharziasis infections

2.3.1. The stage of infection during which the schistosomes are puncturing the skin

The clinical symptoms of bilharziasis are limited to the five stages of development of the parasite (Gear *et al*., 1966). Cercarial dermatitis (known as swimmer's itch) is caused by penetration of cercariae of schistosomes into human skin, which may provoke an acute inflammatory response. The diagnosis is difficult and treatment is usually not essential. In other endemic areas, several 'non-human' species, including *S. mattheei* and *S. bovis* exist; however, cercariae from bovine schistosomes are not able to complete their life cycle in man; they can often penetrate human skin and transform into schistosomula. They persist for up to 10 days post infection and give rise to an immediate allergic-type hypersensitivity reaction (Taylor, 1987).

2.3.2 The stage of migration

For the duration of this stage, which is 1 – 4 weeks after cercarial infection, symptoms such as fatigue, fever and chills, muscle aches, sometimes associated with rigors, frequent urticaria and diarrhea start to present themselves. These manifestations occur when the parasites are migrating through the blood and lymph systems of their host. Irritation of the lungs resulting in coughs, and seldom coughs containing slight blood, enlargement of the liver and spleen causing, discomfort and tenderness in the region of the liver as the disease prolongs into its' further stages (Gear *et al*., 1966; Gryseels *et al.,* 2006).

2.3.3 The stage of egg-laying

In urinary bilharziasis (*S. haematobium*) the earliest symptom of this stage is often indistinguishable, presenting with lower abdominal pains associated with pain or irritation on micturition followed by haematuria. The urine tends to be bright red and the haematuria can prolong for several years since the eggs of the *S. haematobium* must first pass through the bladder wall before exiting with the urine. Sub-acute appendicitis is also known to occur from this stage of bilharziasis, due to the frequent assault of the adult worms and their eggs to the tissues of the appendix and large intestine. In *S. mansoni* infections almost 50% of all eggs produced end up in the liver (Despommier *et al*., 2005) and those that finish up in the lumen of the small intestine end up in fecal mass, explaining the commonest apparent symptom being diarrhoea. It might be so slight as not to alert the patient or may be so severe as to simulate acute dysentery. In chronic situations conditions such as portal hypertension and splenomegaly often arise. Occasionally the worms of both *S. haematobium* and *S. mansoni* derail in their migration and may reach other points in the tributaries of the portal system or of the pelvic plexuses and veins. When a vital organ such as the brain or spinal cord is involved like the granulomatous lesions around ectopic eggs in the spinal cord from *S. mansoni* and *S. haematobium* infections, the sequelae may be serious (Gear *et al*., 1966).

2.3.4 Stage of cicatrization

Although adult schistosomes are presumed to not cause significant pathological damage to their host, their eggs however can cause intense immunopathologic responses (Despommier *et al*., 2005). Trapped eggs usually lodged in the mesenteric veins or washed back to the liver secrete antigens that elicit vigorous immune responses. Fibrous tissue produced around the worms and their eggs are consequential of the inflammatory manifestations seen in bilharziasis. In *S. haematobium,* egg deposition occurs in the bladder leading to the restriction of the ureter from the fibrosis of the base of the bladder resulting in an obstructed uropathy, and may be followed by hydronephrosis and chronic nephritis (Gryseels *et al.,* 2006). This stage of severity is one of the rarest causes of mortality directly attributable to bilharziasis. Enlargement of the

liver and spleen in both *S. haematobium* and *S. mansoni* infections may result in hepatic bilharziasis a common cause of esophageal varices (Gear *et al*., 1966).

2.3.5 Stage of malignant change

Chronic urinary bilharziasis is often epidemiologically associated with squamous bladder cancer. In South Africa, the most common characteristic of urinary bilharziasis morbidity is the preceding development of cancer of the bladder, which is a form of cancer often encountered in individuals indigenous to these endemic regions (Berry, 1966; Hodder *et al.,* 2000). Carcinoma of the intestine, liver and uterus associated with bilharziasis lesions have been noted and observed, however, these observations appear to be relatively rare (Gear *et al*., 1966; Mostafa *et al*., 1999). The influence of bilharziasis as the source of cancer of the liver and bladder, due to the intensity of the infection or worm burden still remains a vexed question with no definitive reports, since the mechanisms involved in the predisposition of the condition are not well understood and various etiological factors may contribute to the progression of the malignant state (Mostafa *et al*., 1999). This however, occurs more often in South African populations exposed to more highly endemic zones than anywhere else. Research also suggests that environmental factors other than bilharziasis may contribute to the development of this condition i.e. tobacco and exposure to chemicals (Gryseels *et al.,* 2006). King *et al.* (2008) stated that "in the developing world, parasitic infections such as schistosomiasis are common, recurrent and long-lasting health problems that represent an ongoing inflammatory challenge and a significant health threat to the populations who are at continuing daily risk for infection". It has been noted when regarding the fundamental aspects of the host-parasite relationship, that research on *S. haematobium* is still in its formative years compared to *S. mansoni* and *S. japonicum* (Rollinson, 2009), which underscores the need and importance for expanded research on the former species and on its effect on the host.

2.4 Diagnosis of bilharziasis

2.4.1 Parasitological methods

It is unquestionable that there is a dire need for an accurate bilharziasis diagnostic tool that is both sensitive and specific, as it is essentially important that it will help the clinician to provide an adequate management control for the disease to the patient (Peeling *et al*., 2006). Laboratory diagnosis of bilharziasis is usually performed by microscopical detection of eggs in urine (*S. haematobium*) or stool (*S. mansoni* and *S. japonicum*) (Wang *et al*., 2004), or by immunological methods (antibody or antigen detection). Microscopic diagnosis for bilharziasis is the established gold standard tool for the detection and confirmation of an active bilharziasis infection. However in some endemic areas, expert microscopic diagnosis is often not available, its inability to detect low infection intensities may cause delays in the diagnostic treatment in clinical individuals suspected to have bilharzias, and more often not even available. The traditional direct method carried-out for urinary bilharziasis detection is the standard filtration method that involves the detection and quantification of *S. haematobium* eggs in a 10 ml urine sample that generally should be obtained between midday (10:00-14:00) to correspond with the diurnal egg output peak (Mott *et al*., 1982). This common microscope-based method may be inexpensive and straightforward; however, the intensity of the infection and the considerable day-to-day fluctuations in egg output has an effect on the sensitivity of microscopic examinations. Additionally this parasitological method is relatively time consuming and it requires a well trained laboratory technician (Hamilton *et al*., 1998; Doenhoff *et al*., 2004).

For urine investigation purposes urine dip-sticks are also used as analytical means. Midstream urine is usually collected and the dip-stick test being administered within 2 hours after collection. The dipstick tests for blood (microhaematuria), ketones, glucose, pH, bilirubin, urobilinogen and protein. The tests are commonly used in multiple combination strips i.e., five tests on one strip to seven tests on one strip. Dipstick urinalysis is convenient, but false-positive and false-negative results can occur. Specific gravity provides a reliable assessment of the patient's hydration status, while microhaematuria has a range

of causes, from benign to life threatening. Table 2.5 provides a summary on urinalysis performed on *S. haematobium* infected patients.

Table 2.5 *S. haematobium* infection diagnosis based on the detection of eggs in the urine.

Urine collection	The recommended time when urine should be collected is between 11:00 and 14:00 for egg peak output. The sediment should be sampled after the urine has been allowed to stand for 30 min allowing any eggs present to precipitate to the bottom. Several specimens taken on consecutive days should be examined.
Urine analysis	This includes looking for microhaematuria and proteinuria, for which a urine - dipstick is usually employed for detection. An egg count is also done microscopically to estimate the severity of infection.
Egg count	Infection is defined according to the number of eggs per 10 ml of urine: ▪ <100 = light infection ▪ 100–400 = moderate infection ▪ >400 = severe infection

Derived from Bichler *et al*., 2006.

2.4.2 Immunological methods

An immunological diagnostic approach of bilharziasis was proposed when direct parasitological approaches fell short on sensitivity (Van Lieshout *et al*., 2000). Immunodiagnostics of bilharziasis is based on antigen and antibody detection, although they may require a better equipped laboratory than the direct microscopy technique. Another alternative to the existing microscopy-based method is the detection of two antigens the CAA and the CCA which are known to be released in the circulating bloodstream of the infected individuals (Despommier *et al*., 2005). A urinary circulating cathodic antigen cassette (CCA) test kit is available for the detection of these antigens released by the schistosoma parasite. Collected midstream urine is normally used to detect these antigens released largely by the adult worms residing in the host, a positive result on the CCA test kit

is indicative of an active bilharziasis infection (Van Lieshout *et al*., 2000). Although immunological methods tend to yield a higher sensitivity particular for antibody detection, they lack specificity and the ability to differentiate between light and heavy bilharziasis infections. Due to the inflection of the host's immune system, it is possible for the host to indicate separate IgG, IgM, IgA and IgE antibody response, or even a combination of these isotypes (Utzinger *et al*., 2005). On the other hand usually 14% of patients might not even respond with any antibody formation. Some of the commonly used methodologies are based on detection of antibodies directed against the soluble egg antigen (SEA) or as mentioned earlier the circulating cathodic antigen (CCA) (Van Lieshout *et al*., 2000). Depending on the methodology used and the duration of infection in the host, the sensitivity of current antibody assays are not optimal (ranging from 65% to 85%) with the specificity ranging from moderate to high (43% to 87%) (Utzinger *et al*., 2010). "Ultimately, our improved understanding of the full range of schistosomiasis-related disease will provide the basis for an optimal design of the next generation of parasite control" (King *et al*., 2008). Metabolomics may be one of the approaches to reach this goal, and is presently applied in various infectious disorders, including bilharziasis.

2.5 Metabolomics approaches to investigate disease perturbations

2.5.1 Metabolomics

Prior to the present progressive growth in metabolomics technology, molecular biology flourished and was often expressed as the era of the "The Central Dogma: DNA-encodes-RNA-encodes-protein." This formed the basis of much of the biological research for the last 50 years of the previous century. However, this period of research was missing a piece in the sector of biochemistry involving qualitative metabolite analysis (Harrigan *et al*., 2003). A review by Griffin (2006) informs of how, "the term metabolomics (and the related term metabonomics) was coined at the end of the 1990s, to describe the development of approaches which aim to measure all the metabolites that are present within a cell, tissue or organism during a genetic modification or physiological stimulus" (Oliver *et al*., 1998; Nicholson *et al*., 1999). In the 2000s, the field emerged and became known as

metabolomics, with various subfields known by multiple names (metabolite profiling, metabolite fingerprinting even metabonomics, a term often used to indicate metabolomics focused on humans). Metabolomics complements data derived from other more established omics-technologies (genomics, transcriptomics and proteomics) to assist in providing a systems approach to the study of human health and disease (van der Greef, 2004). Metabolomics being the comprehensive analysis of all or a large number of cellular metabolites, has become a rapidly developing field in biomedical science that combines the application of separation and spectroscopic techniques with multivariate statistical analysis in studies of the molecular composition of biofluids, cells, tissues and organisms. (Nicholson *et al*., 1999; Raamsdonk *et al*., 2001). Unlike earlier, more selective analytical methods, metabolomics today offers analytical instruments that can simultaneously quantitate thousands of low molecular weight substances present in a biological sample of interest and sophisticated mathematical tools that can find a molecular signal amongst thousands of pieces of data (Kell, 2004). Table 2.6 gives an indication of a few theoretical aspects that can be identified using a metabolomic approach, using the neurobiological field as illustration.

Table 2.6 Items that can be measured in metabolomics experiments

Biochemistry	**Theoretically**
Metabolites small molecules	Biochemical constitutes
Pathways (e.g. purines catabolites)	Excretion products
Interactive pathways (amino acid metabolism)	Precursor products
Compound classes (e.g. lipids)	Balances (e.g. redox systems)
Conceptually linked systems (e.g. antioxidants)	Collection depots
	Flux
	Snapshot view of biochemistry
	Intergraded signal of genome and environment
	Short and long term status
	Temporal image
	Sub-threshold changes (e.g. toxicology, nutrition)

Table adapted from Metabolomics: concepts and potential neuroscience applications Kristal *et al*., 2007.

2.5.2 Metabolite profiling

Metabolite profiling is the measurement and quantification of all or selective levels of a metabolite set in biological samples, using directed sample preparation and instrumentation to isolate and analyze the compounds of interest (Dunn *et al.,* 2005). On the level of an undergraduate textbook, Garrett and Grisham (2005) define the metabolism as chemical reactions that convert nutrient bio-molecules like lipids, carbohydrates and proteins to release energy and in addition synthesise or degrade molecules present in all organisms. Metabolites are small organic molecules which are thus important modulators, substrates, by-products and building blocks of many different biological processes. Metabolic profiling accordingly aims to give at least theoretically the most temporally up-to-date view (snapshot) into the current status of the biological system (Shulaev, 2006). Therefore, the metabolic state of an organism, as defined by its metabolic profile, can disclose information on the genetic, physiological, functional or pathological status of the system. This leads to essential applications in the research analysis of fundamental biological processes, in addition the molecular diagnosis and prognosis of disease, functional genomics, drug metabolism and toxicity (Dunn *et al.,* 2005). Present metabolomics research is implemented in a way which that it incorporates three fundamental elements: design, data collection and data analysis, which will now be discussed.

1. **<u>Metabolomics experimental design</u>**

   The most important aspect when approaching a metabolomics study is the formulation of a well-structured experimental design, also referred to as the metabolomics work-flow. All steps from sampling human tissue, urine, plasma or other samples to final data acquisition need to be carefully planned (often unique for specific investigations) and meticulously executed in order to reduce sources of error which could contribute to large variability in samples. Such variability may result in a decreased likelihood of detecting inherited, development, behavioural disease factors or even biomarkers. A well planned experimental design could avoid this. A typical metabolomics experimental work-flow/design is shown in Figure 2.6 and includes the following generic aspects: sample collection, data pre-processing, analysis and interpretation of metabolomic data.

The key aspects in a metabolomic study also include three steps; metabolite isolation, metabolite detection and metabolomic data analysis (Goodacre *et al*., 2007).

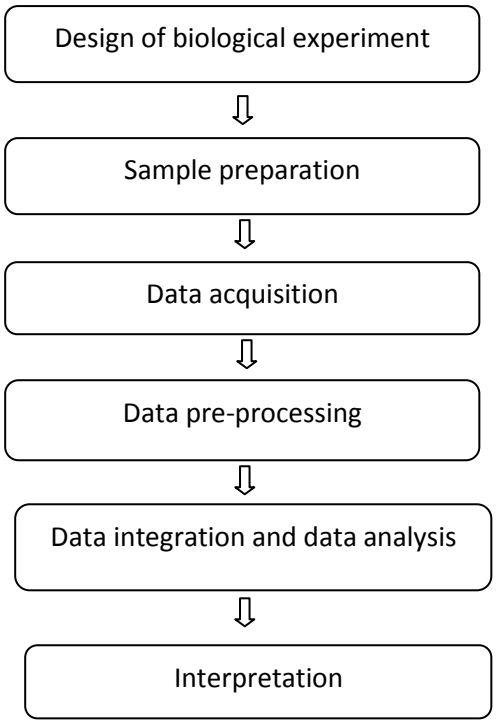


Figure 2.6 The generalized workflow for the design of a metabolomics experiment.
The linear generalized workflow of the experimental design was taken from Goodacre *et al.,* 2007 and is described in detail in that publication.

2. <u>Sample collection and storage</u>

Urine is a convenient biofluid for global metabolite profiling, and is obtainable by the most non-invasive procedure. However, for the generation of useful data, it must be carefully collected and stored (Lauridsen *et al*., 2007). It is essential to note that failure to collect and store samples correctly at the beginning of an investigation can lead to invalid results, e.g. due to microorganisms that may multiply in the samples, resulting in highly distorted metabolomics data or due to disintegration of unstable or volatile metabolites. When taking into account the design of a metabolomics study it may be noteworthy to consider the time of sampling (Maher *et al.,* 2007); for example, in the interest of the disease bilharziasis, it would be more productive to collect midstream

urine between 10h -14h which is the period of most egg output. The next matter to be addressed would be the proper storage of the urine samples. It has been recommended in various literature reviews that the best storage of urine samples from short-term to long-term storage is between 4°C to -80°C, and will be discussed in further detail in Chapter 3 Section 3.2.1.

3. Data generation

As mentioned, the term metabolome refers to the total metabolite complement of a biofluid, cell, tissue or organism. Although there currently are no investigative techniques capable of delivering an extensive measurement of the complete metabolome, there are numerous technologies that can approach this ideal (Nicholson *et al*., 2002). The primary requirement for an extraction protocol is that it must be comprehensive and the extract must represent as large the number of the cellular metabolites as possible (Want *et al.,* 2010). Dunn *et al*. (2011) reveals various important requirements that can be listed for a successful metabolomics experiment. It should be (a) non-selective, in that it is not directed towards a particular compound, as it then becomes targeted metabolomics; (b) quantitative and precise, as a result presenting data which can be meaningfully modelled by statistical procedures; (c) susceptible to an extensive range of metabolite concentrations; and (d) capable in resolving the complex mixture into separate signals from each metabolite. Though not completely, these requirements are fulfilled by hyphenated mass spectrometric techniques, such as GC-MS, LS-MS and CE-MS. NMR is a different, but also most important technology for metabolomics investigations. The use of GC-MS is the analytical method used in this investigation, and will be described in more detail in Chapter 4 section 4.2.2.

4. Data analysis

Most metabolomics applications are based on sample analysis using NMR or MS identification techniques. Once the data on metabolic features have been extracted from controls and affected subjects, classification of subgroups within the samples are determined and identification of biomarkers, which are metabolomic features that

discriminate between the sample classes, is carried out. Before the application of multivariate statistical analysis, data might need some scaling, centering and transformation, to compensate for large dynamic ranges within the data set that might distort or give bias analytical results due to unwanted variation (van den Berg *et al*., 2006). Multivariate statistics include amongst others, principal component analysis (PCA), partial least square analysis, discriminant analysis (PLSDA), hierarchical clustering, and correlation networks, that are available to make sense of the metabolomic data obtained (Ebbels *et al*., 2003).

2.6 Applications of metabolomics in infectious diseases

In this section a number of important infectious diseases and how metabolomics has been used as a tool in their analysis, are described, followed by a more extensive survey of metabolomics investigations on bilharziasis.

2.6.1 The three most important infectious diseases

Metabolomics has become progressively used in the attempt to investigate some of the approximately 20 infectious diseases that affect more than 1 billion people worldwide (WHO, 2012). An important aspect includes the untargeted identification of potential biomarkers, especially of early infection. Metabolites are the apparent candidates for biomarker screening and this is because of their holistic representation of the downstream effects of enzyme catalysis and other biotransformations (Want *et al*., 2007). Unlike the neglected tropical diseases, including bilharziasis, HIV/AIDS, TB and malaria have been the most focused on due to their most important global health burden. Some metabolomics observations will now be briefly discussed as an introduction to this approach, used in infectious disorders.

- HIV/AIDS

Meyer and co-workers (Hewer *et al*., 2006; Hatting *et al.,* 2009), presently at the University of Pretoria in South Africa, described in the Journal of Pharmaceutical and Biomedical Analysis the first metabolomics investigation of HIV infection, using ^{1}H-NMR (proton NMR)

spectroscopy based metabolomic analysis to distinguish between HIV-1 positive/AIDS patients on ARV treatment, to HIV-1 negative individuals. In that study the basis was to use ^{1}H-NMR spectroscopy metabolomics to explore whether they could develop an extensive application, in monitoring of the biological side-effects of antiretrovirals used for the treatment of HIV/AIDS, considering that the potential applications of metabolomics to the identification and study of human diseases is vast. In conclusion, that study indicated that a distinction based on metabolic profiles and chemometric analysis is possible and that the preliminary findings suggested that the use of NMR metabolomics in the monitoring of HIV/AIDS patients on antiretroviral therapy could be expanded to the study of antiretroviral side-effects and the correlation of metabolic disorder. More recently, Williams and co-workers (Williams *et al*., 2011) reported on qualitative urine organic acid profiles of HIV-infected individuals not on antiretroviral treatment. This study incorporated the application of GC-MS metabolomics to evaluate the organic acid profiles in urine of asymptomatic HIV infected individuals compared to uninfected controls. It was previously known that HIV acts directly on the regulation pathways associated with apoptosis (cell death) and that the mitochondria plays a central role in this type of cell death (Pinti *et al.,* 2010). Several organic acids are well-established diagnostic biomarkers of mitochondrial dysfunction (reviewed by Hoffman and Fey, 2005). Therefore, Williams and co-workers chose the analysis of an organic acid metabolome to investigate a possible progressive disruption of mitochondrial structure and function during an HIV infection, using a metabolomics approach. The scope of that study included using a multifaceted analytical bioinformatics procedure, which enabled them to link 10 metabolites, whose occurrence was significantly different between controls and HIV-infected individuals, to disrupted mitochondrial metabolism. These metabolites were significantly different between controls and HIV-infected individuals, and directed to changes in lipid metabolism and oxidative stress, all of which are apparent abnormalities caused by an HIV infection. This study illustrated that an MS metabolomics study is capable of capturing biomarkers of mitochondrial dysfunction which could probably be developed into indicators of an HIV infection. These markers have the potential to define the asymptomatic stage of an HIV infection and can be developed into a method used in the surveillance of more advanced stages of the disease and potentially the response of infected individuals to antiretroviral treatment (Williams *et al.,*

2011). This would, however, need an extensive cohort study of HIV infected individuals to assess the practicality of such an application.

- TB

In a review by Parida and Kaufmann *et al*. (2010), designated "the quest for biomarkers in tuberculosis", they mention that one of the drawbacks in TB research is the lack of reliable biological biomarkers to predict toxicity and efficiency early in the advanced stages of infection. They present a framework of biomarker discoveries for TB in the section of drug and vaccine development using emerging global omics platforms, proposing that metabolomics should be amongst them. In a pioneering study that Loots and co-workers of the North-West University (Potchefstroom campus), conducted in collaboration with colleagues at Stellenbosch University in South Africa (SA), they were able to indicate interesting results using sputum samples analysed from newly diagnosed pulmonary TB individuals, latently infected, and clinically healthy individuals from an endemic region of SA (Olivier and Loots, 2011). In the diagnosis sector they indicate the shortcomings of recent and available TB-diagnostic approaches and that no current tests meet all the specifications of sensitivity, specificity, safety and training simplicity. However, with metabolomics which not only considers the metabolome of the infected organism, it also measures the changes in the host's metabolism due to infection. Therefore, unlike the current gold standard bacterial culture diagnostic method, metabolomics is well on its way to providing new biomarkers with a holistic view of the intra-host changes during TB infection, active disease and treatment. It might also provide potentially less invasive TB diagnostic procedures which are comparatively quick. They thus concluded that metabolomics is able to reveal biosignatures that can distinguish states of a TB infection, which could be used to monitor patients during therapy and to predict potential candidates that do not respond to the treatment leading to effective management of these groups with other drug regimens (Oliver *et al.,* 2011).

- Malaria

Malaria, caused by *Plasmodium* species, is a life-threatening infectious disease in humans, and its prevalence seems to escalate in Southern Africa according to information on malaria control in Africa (Grover-Kopec *et al.,* 2006). The genomics of the malaria parasite is well-studied, but a gap exists in the link between genetic information and annotations for putative associated proteins, but the usefulness of *in silico* data mining was clearly illustrated by the identification of a large number of proteases associated with the *P. falciparum* genome (Wu *et al.,* 2006). Birkholtz and colleagues (Birkholtz *et al.,* 2007), also from the University of Pretoria in South Africa, recently published a review of three omics-technologies (transcriptomics, proteomics and interactomics) to evaluate potential contributions that such a functional genomics approach can make to diminish the gap mentioned above, as well as to direct towards discovery of new targets and therapeutics to combat this parasite. Their review indicated the various challenges that *P. falciparum* investigations pose, like difficulties in obtaining sufficient quantities of recombinant expressed protein for further investigations. They thus also advocate *in silico* data mining investigations, although they recognized existing limitations in this approach (Birkholtz *et al*., 2006). However, metabolomics is another omics-technology that has potential to contribute to systems-level insights on the functional responses on host-parasite interactions. Olszewski *et al.* (2009) thus used metabolomics to view these interactions in *P. falciparum* infection. With the use of MS–based metabolomic analysis of the proliferation of the *P. falciparum* parasite in red blood cell cultures, they were able to reveal a general modulation of metabolites varying in phase with the developmental cycle. Among these was extracellular arginine, which was not converted to ornithine by the parasite. They concluded that the systemic arginine depletion by the parasites might be some form of factor in human malarial hypoargininemia associated with cerebral malaria pathogenesis (Olszewski *et al.,* 2009). According to these authors, their metabolomics approach "*provide a powerful tool to further investigate clinically important biochemical pathways in greater detail and elucidate the metabolomics response to physiologically relevant perturbations*" due to *P falciparum* infection. This was substantiated by the recent report from Basant and co-workers (Basant *et al.,* 2010) who used a metabolomics approach to reveal alterations in urine, serum and stool of mice infected by *P. berghei.* Their results suggests that the

infection in these experimental animals leads to impairment of glycolysis, lipid metabolism, as well as of the metabolism of tryptophan and uracil. Also notable from that study was the distinct sexual dimorphism noticed in the responses to the malaria infection, which might be of significance in understanding the pathophysiology of malaria infection.

A few observations and important conclusions can be made on these metabolomics studies on HIV, TB and malaria, in so far that they have a bearing on investigations on bilharziasis infection:

(1) Metabolomics investigations on infectious diseases are still in a pioneering phase, despite massive support from instances like the Bill Gates and other Foundations to support research on the three major devastating diseases discussed above.

(2) The initial results from the metabolomics investigations resulted in new knowledge on these diseases, an indication that a number of biomarkers (designated as a biosignature) may jointly be considered as important indicators for the characterization of these diseases, and potentially as indicators to monitor responses to their treatment. However, no definitive biomarker or biosignature has yet been revealed for infectious diseases by metabolomics investigations, despite the promising results from this technology.

(3) The pioneering results are mostly outcomes of case-control studies. The requirements to develop biochemical markers to become practical and useful instruments in clinical settings are theoretically feasible, but most probably will be complex to develop (Turnbull, 2011), as the translation of research findings into clinical practice is not straightforward (Hu, 2011). This will be discussed in further detail in Chapter 6. Notwithstanding these reservations, the results discussed above clearly indicate that a metabolomics analysis of infectious disorders produces insight into markers with potential as instruments in clinics dealing with these infectious disorders, and clearly indicates the potential for deepening the understanding of infection biology and disease states, especially with regard to investigations on the category of neglected tropical diseases (Saric *et al*., 2008).

2.6.2 Metabolomics investigations of bilharziasis

2.6.2.1 Experimental animal models

A mouse model was described in the 1990s to examine three antischistosomal drugs, praziquantel, oxamniquine and oltipraz for their ability to reverse the disturbances in carbohydrate metabolism induced by *S. mansoni* infection (Ahmed and Gad, 1995). Using conventional biochemical methods, infected mice were screened every 2 weeks for 16 weeks for their body and liver weights to assess the activities of liver enzymes involved in glycolysis (pyruvate kinase and phosphofructokinase), the tricarboxylic acid (citrate synthase), glycogenolysis (glycogen phosphorylase) and the hexose monophosphate shunt (glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase). These results led to the conclusion that glycolysis is largely stimulated in the livers of infected mice at the expense of other metabolic pathways of glucose utilization. Administration of the drugs tested caused normalization of the measured enzyme activities and support the selection of praziquantel as a drug of choice for treatment of *S. mansoni* infection.

The observations on the enzyme activities substantiated the changes in some metabolites observed in the first metabolomics investigation on the effect of infection causing bilharziasis (Wang *et al.,* 2004). Since then, further metabolomics investigations have been reported on bilharziasis: most used experimental animal models and one specifically focused on humans infected with *S. mansoni,* including the effect of treatment. An overview of six of these investigations will now be presented. The nomenclature used for the experimental animal models are in the form of "animal"-"bilharzia species"-"metabolomics technology". These six investigations will be used as an important basis for the formulation of the aims of the present investigation.

- **<u>The mouse-*S. mansoni*-NMR model (2004)</u>**

In 2004 Wang and co-workers published the first metabolomics investigation on bilharziasis, using mice infected with *S. mansoni* in an attempt for biomarker identification (Wang *et al.,* 2004). Using this mouse model they presented the characterization of a parasitic infection through metabolic profiling using ^{1}H-NMR spectroscopy and multivariate techniques for

data analysis. This study also aspired to enhance the current understanding of a host's metabolic response to a parasitic infection, which is a promising approach for biomarker identification.

In their study design, Wang *et al.* (2004) used 20 female mice (~20 gm each at the date of purchase) of which 10 were used for the control group and the other 10 were infected subcutaneously with 80 *S. mansoni* cercariae. At day 49 after infection they collected urine samples (between 0.7 and 1.5 ml) from each individual mouse, placed in an individual metabolic cage with a collection plastic tube of 10 ml in an ice bath at 0^{o}C. The necessary precautionary measures were taken to ensure minimum contamination of these samples with fecal material or water fed to the animals during the collection period. The same procedure was repeated at day 56 postinfection. The collected urine was transferred into 1.5 ml tubes and stored at -80°C.

For analytical purposes the urine samples were prepared and inspected using ^{1}H NMR. Data reduction and pattern recognition were applied on the ^{1}H NMR spectra. The ^{1}H NMR spectra of a urine sample obtained from an uninfected control mouse and a mouse with a 49 day-old infection showed various metabolite differences. These results included dominance of certain metabolites such as acetate, alanine, citrate, 2-oxoglutarate, p-cresol glucuronide, succinate and trimethylamine. Through visual inspection of the spectra there was a clear indication of differences in the overall composition between urines acquired from the control and *S. mansoni* infected mice, which included elevated levels of lactate and trimethylamine, but depletion of alanine, 2-oxoisocaporate, 2-oxosiovalerate, butyrate, citrate and isovalerate in urine of *S. mansoni* infected mice. To further aid in the identification of a significant difference in metabolites between the infected and uninfected mice additional analysis was performed on the obtained data, which included multivariate PCA and PLS-DA. These analyses showed a clear separation between the infected and uninfected mice which was largely due to the increase of urinary excretion of creatine, tryptophan and pyruvate, and decreased levels of hippurate, a range of short-chain fatty acids including butyrate and propionate.

From their findings Wang and collegues (Wang *et al.,* 2004) were able to categorize the metabolic consequences of an *S. mansoni* infection, which firstly consisted of reduced levels of the tricarboxylic acid cycle (TCA) intermediates, such as citrate, succinate and 2-oxoglutarate in urine samples. A urinary depletion of citrate has an impact on calcium complexation, which may lead to precipitation of calcium salts in the kidney, causing failure. This led to an implied alteration in the utilisation of renal carbohydrates and renal failure. A stimulated glycolysis was also suggested due to the increased levels of pyruvate which was due to the inhibition of pyruvate dehydrogenase. Depleted levels of 2-oxoisovalerate originating from valine and 2-oxoisocaproate a derivative of leucine, suggested a disturbance of the amino acid metabolism. This they linked to impaired liver function due to the serious hepatic effects of a *S. mansoni* infection. Finally, they observed urinary metabolites, like trimethylamine, phenylacetylglycine and p-cresol, which indicated perturbations in the gut microbiota of the infected mice.

- **The Hamster-*S. japonicum*-NMR model (2005)**

Following the pioneering study by Wang and co-workers (Wang *et al.,* 2004), the next experimental model study was the infection of the Syrian hamster by *S. japonicum* (Wang *et al*., 2005). The scope of this investigation was to extend the metabolite profiling of a bilharziasis infection to another host-parasite model like the *S. japonicum* infection in the hamster, with the intent to investigate if "*the response to a particular parasitic infection is host-specific, or whether different hosts respond to different parasites of the same genus (Schistosoma) in a similar way*". Similar to the previous study of *S. mansoni* infection in mouse models, 20 Syrian hamsters approximately 8 weeks old were used to inspect the infection due to *S. japonicum.* Ten hamsters were infected with the *S. japonicum* parasite while the remaining ten hamsters were used as controls. Their urine was collected 36 days post-infection for both the infected and control hamsters. It was collected in small Eppendorf tubes labelled with unique identification numbers and stored in a freezer at -70°C. For preparation purposes of hamster urine samples and subsequent ^{1}H NMR spectroscopic analysis, they adhered to the same protocol used in the study of *S. mansoni* infection in mice. ^{1}H NMR spectra of urine samples obtained from uninfected control

hamsters and infected hamsters showed that metabolites such as acetate, alanine, citrate, creatine, formate, glycerol, hippurate, succinate and taurine were amongst the many metabolites detected in the ^{1}H NMR urine spectra of 36 day infected and control hamsters. Glycerol was also among the metabolites detected by ^{1}H NMR in the hamster urine, which, however, had not been previously reported and was also not found in the urine obtained from mice in their previous work (Wang *et al.,* 2004). Noticeable differences that were observed visually in the overall composition between urine samples obtained from control hamsters compared to hamsters with a *S. japonicum* infection included: elevated levels of acetate, p-cresol glucuronide, lactate and phenylacetylglycine along with depressed levels of citrate, hippurate and succinate. Multivariate analysis such as PLS-DA was carried out to further optimize the visualization of the metabolic effects between infected and control hamster. Based on the VIPs, the metabolite profile, *S. japonicum* infection was characterized by increased urinary excretion of acetate, p-cresol glucuronide, phenylacetylglycine and pyruvate, accompanied by decreased urinary excretion of citrate, dimethylamine, hippurate, proprionate and succinate. These markers exhibited the largest influence on differentiating urine samples from uninfected controls and *S. japonicum* infected hamsters.

From these results Wang *et al.* (2005) noticed that the results of a *S. japonicum* infection using a Syrian hamster model depicted similar results to that of an *S. mansoni* infection in a mouse model. The two models shared many metabolic features like the reduction of TCA cycle intermediates i.e. citrate and succinate. The elevation of urinary excretion of pyruvate due to the inhibition of pyruvate dehydrogenase and the marked increase in enzyme activities involved in carbohydrate metabolism, such as phosphofructokinase and pyruvate kinase. Amongst these were also the changes in various microbiotal-related metabolites including p-cresol glucuronide, phenylacetylglycine, trimethylamine and hippurate. However, along with the similarities of the two studies (S*. mansoni*-mouse and *S. japonicum*-hamster) there were also differences in the urinary metabolites observed in the two animal models. For instance, there were reductions of short chain fatty acid concentrations such as acetate, butyrate and proprionate, which were observed in urine acquired from *S. mansoni*-infected mice (Wang *et al*., 2004) whereas an increased level of acetate and a reduction of proprionate were observed in the *S. japonicum*-infected hamsters. They concluded that

these discrepancies suggested that infection-induced alterations in the short chain fatty acids depended on the type of schistosome species or even the animal host either inhibiting the developing or consumption of these short-chained fatty acids. Also another distinguished difference between the two models were metabolites indicative of an impaired liver function, which included reduced levels of 2-oxoisovalerate and 2-oxoisocaporate and taurine observed in urine obtained from *S. mansoni* infected mice (Wang *et al*., 2004), but not in the hamsters infected with *S. japonicum.*

- **The mouse-*S. mansoni*-CE-MS model (2008)**

Since the description of schistosome infection, using a mouse model (Wang *et al.,* 2004) and the hamster model (Wang *et al.,* 2005), there have been further uses of metabolomics to characterize a schistosome infection using a mouse mode. Garcia-Pérez *et al.* (2008) used CE-MS for the metabolic fingerprint of *S. mansoni* infection, also using the mouse model. This technique was used to establish the capabilities of CE-MS as a diagnostic tool for this parasitic disease, and to assess if this approach could provide a cost-effective and information-rich complementary approach to ^{1}H NMR for metabolomic studies. CE represents a high resolution separation technique which possesses capabilities of producing highly efficient separations of assorted mixtures of analytes present in scarce sample volumes, with minimal or no sample preparation. In this study (Garcia-Pérez *et al.,* 2008), they infected 10 female mice with *S. mansoni* cercariae. However, unlike the procedures used by Wang *et al.* (2004), the mice were infected by a transcutaneous route by submerging them in a tank containing 1cm aquarium water containing *S. mansoni* cercariae. Individual urine samples were collected from the 10 infected mice and the 10 uninfected mice which served as controls, 2 days per week during 8 weeks. Unlike ^{1}H NMR, the urine samples of the CE were supplemented with 20µl of purified water, mixed with buffer (pH=9.5) and injected directly into the CE apparatus.

Multivariate analysis was performed on the data to identify those compounds that contributed to the differences between the infected and uninfected mice. The compounds

that were elevated in the urine of the infected mice were tryptophan, creatine, benzoic acid, uric acid and glycolic acid, while levels of hippuric and citric acids and urea were reduced relative to the control mice. The increased levels of tryptophan and creatine levels and the decrease in hippurate levels were in agreement with the study of Wang *et al*. (2004) using the NMR profiling approach. Alterations of urinary hippuric acid have been associated to microbial-derived metabolites however, when taking into consideration the decreased levels of hippuric acid and also the increased levels of benzoic acid in this study, it is more likely that these changes observed here might be associated with an altered liver metabolism (Garcia- Pérez *et al*., 2008).

- **The mouse-*S. japonicum*-NMR model (2010)**

Unlike the previous investigations on an *S. mansoni* mice model based on a well-established late-stage schistosomal infection, the study by Wu *et al*. (2010) "the metabolomic changes reveal the development of schistosomiasis in mice", is based on investigating the time-course of metabolomic changes in the urine and blood plasma of the *S. japonicum* mouse model. In this investigation the main objectives were to define the host's metabolomic response to infection at the early stages and their dynamic changes during the disease progression. To achieve this 60 female pathogen free mice about 8 weeks old weighing 20 ±2g were purchased, and housed in groups of 5 in plastic cages under environmentally controlled conditions. After 3 weeks of acclimatization half of the mice (n = 30) were infected with 80± *S. japonicum* cercariae, each through a shaved abdominal skin. The rest of the mice (the other half – n=30) served as controls. Plasma and urine samples were collected one day before and after infection for five weeks on a weekly basis. Sample collection was carried out between 08:30 - 11:30 in order to avoid potential metabolic variations due to diurnal cycle. Blood samples (70~80 µl) were collected from the orbital venous plexus of the mice, centrifuged and the supernatant (~30 µl) was transferred into 0.5 ml Eppendorf tubes which were immediately immersed in liquid nitrogen, and stored at -80°C. The urine (50 to~400 µl) samples however, were collected in empty plastic boxes by gently massaging the abdomen of the mice, and transferred into Eppendorf tubes where

they were stored at -80°C. For ^{1}H NMR spectroscopy analysis both urine and plasma samples were prepared.

Multivariate analysis such as principal component analysis (PCA) was employed by using mean-centred NMR data to identify general trends and outliers. In addition a more supervised pattern recognition method, the orthogonal-projection to latent structure discriminant analysis (O-PLS-DA), was used as a validation model, which was crucial for the interpretation of the data. The clinical biochemistry of the serum was measured using an automatic biochemistry analyser, of which independent *t*-tests were conducted. Clinical serum chemistry data, of the control and *S. japonicum* infected mice at week 5 post-infection results, indicated that an *S. japonicum* infection led to significant increases in the activities of alanine aminotransferase, aspartate aminotransferase and globulin with decreases in the levels of albumin, alkaline phosphatase, and triglycerides. Various metabolites such as glucose, lipoproteins, citrate, creatinine and a range of amino acids were detected in the plasma from both control and infected mice. The visual inspection of plasma ^{1}H NMR spectra revealed that the infected samples contained higher levels of *N*-acetyl-glycoproteins along with lower levels of lipids, citrate and alanine than in the controls. The ^{1}H NMR spectra of urine from the infected mice, on the other hand, indicated apparent elevated levels of 2-keto-isocaproate, 2-keto-3-methyl-valarate, pyruvate, 4-cresol glucuronide, 3-ureidopropionate, dimethylamine, trimethylamine and phenylacetylglycine. Also amongst these metabolites were alleviated levels of adipate, taurine, and hippurate as compared to the controls.

For further interpretation of the obtained metabolite data and their differences they employed additional multivariate data analysis. The PCA trajectory of both plasma and urine samples demonstrated that there was a coexistence between the metabolic profiles of the mice with the time course of *S. japonicum* infection and disease progression. In order to identify the metabolites associated with the observed separation, they further compared the metabolic profiles obtained from the infected and the corresponding uninfected-control mice for all matched time points. The plasma and urine samples collected at week 5 post-infection were further separated into two sub-groups. These subgroups were related to the number of worm burden in the mice, ranging from light infection to heavy infection. In order to monitor the variation of metabolites at altered infection levels in plasma and urine

samples, they performed additional O-PLS-DA comparisons between the control groups and the lightly-infected, and the control group and the heavily-infected groups. The blood plasma samples showed significant elevation of lipoproteins and keto-bodies, including D-3-hydroxybutyrate and acetone after infection for 3 and 4 weeks. They documented noticeable changes when comparing the infected mice to their matched controls such as, the metabolic changes of plasma obtained from the lightly infected mice at week 5 post-infection were related to those obtained from the mice at week 4 post-infection, while additional metabolite changes were observed for the heavily infected mice at week 5 post infection. These changes included elevated levels of *N*-acetyl-glycoproteins and lower levels of glucose citrate and choline metabolites. The infected mice showed obvious urinary metabolic changes at week 3 post-infection with significant elevation of 2-keto-3-methyl-valarate, 2-keto-isovalarate, 3-ureidopropionate and some gut microbiota related metabolites, including trimethylamine-*N*-oxide and phenylacetylglycine, together with alleviation of adipate and Krebs cycle intermediates like succinate and citrate as compared to their controls.

The advancement of the disease led to, elevated levels of 4-cresol glucuronide, pyruvate, malonate, glycine, and reduced levels of acetate, taurine and hippurate which were observed in the urine of the infected mice. Observations of the urinary metabolic profiles obtained from the lightly-infected mice at week 5 post-infection on the other hand were characterised by the increase of pyruvate, glycine, 3-ureidopropionate, including microbiota related metabolites such as 4-cresol glucuronide and trimethylamine. When inspecting the heavily infected group, additional elevation of 2-keto-3-methyl-valarate, creatine together with alleviation of hippurate, 2-(4-hydroxyphenyl) propanoic acid and citrate were noted.

PLS models were constructed using Pareto-scaled NMR data of plasma and urine obtained at week 5 post-infection and plotted against the worm burden. From this they found significant correlations between the worm burden and the metabolic changes in plasma and urine samples. The reduced levels of glucose, pyruvate, and increased levels of lysine and *N*-acetyl-glycoprotein that were observed in plasma appeared to be closely associated with the worm burden. In the urine samples however, the worm burden was associated with elevated levels of phenylacetylglycine, 3-ureidopropionate, creatine, 4-cresol glucuronide and alleviated levels of hippurate, 2-(4-hydroxyphenyl) propanoic acid and adipate. In their

investigation they illustrated the alterations of relative concentrations of typical metabolites as a function of the infection period and worm load. It then became evident that *S. japonicum* infection causes a steady increase in the concentrations of 3-ureidopropionate, phenylacetylglycine and pyruvate together with a decrease in citrate. Starting from week 3 post-infection, the elevation of 3-ureidopropionate appears to be positively correlated with infection duration and worm burden. A clear transformation of 3-ureidopropionate occurs at week 4 post-infection along with gut microbiota related metabolites, pyruvate and citrate. From these results they deduce that metabolic changes in plasma and urine samples are all closely related with the existence and severity of a *S. japonicum* infection as indicated by the levels of worm burden.

Since liver injury is one of the manifestations of an *S. japonicum* infection, from these results it was confirmed that the occurrence of liver injuries at week 5 post-infection was present. One of the metabolic consequences of a liver injury is the disturbance of amino acid metabolism, which was noted with high levels of alanine, asparagine, creatine, glutamine and glycine. Another common consequence of liver injury that is significant is stimulated glycolysis, which is apparent by a marked reduction in levels of plasma glucose, liver glucose and glycogen along with the accumulation of liver and urinary pyruvate following 5 weeks infection. Similar to previous observations for *S. mansoni* infection, the *S. japonicum* infection investigated in this research led to a suppression of the TCA cycle with alleviation of plasma citrate and urinary citrate, 2-oxoglutarate and succinate (Wang *et al*., 2004; 2005). The elevations of the keto acids observed in this investigation indicated that *S. japonicum* infection promoted ketogenesis consequential from the degradations of the branched-chain amino acids. Furthermore, it appears that the infection-induced changes in gut microbiota associated metabolites such as phenylacetylglycine, hippurate, trimethylamine and dimethylamine signify that schistosome infection also disturbed the gut microbial ecology. This was largely similar to the effects of infection by *S. mansoni* (Wang *et al*., 2004; Wang *et al*., 2005).

In the course of this investigation elevated urinary 3-ureidopropionate was found in mice infected with *S. japonicum* which correlated with the disease progression and worm burden. Recently in another study this very metabolite was found in the urine sample of mice 53 days post-infection by *S. mansoni* with the combination of NMR capillary electrophoresis

methods (Garcia-Perez *et al.,* 2010). The investigation of mice infected with *S. japonicum*, *however*, detected urinary 3-ureidopropionate in mice at the third week infection, which was four weeks earlier than the previous investigation. Wu *et al*., (2010) concludes that urinary 3-ureidopropionate could be a potential biomarker for early diagnosis of schistosome infection and it should be further justified for certification as an early diagnosis biomarker in the schistosome infected humans and other animals.

The important contribution from this research investigation is that it comprehensively described the time-course changes in metabolic responses in a mouse model and *S. japonicum* infection and the particular emphasis on the possibility of early detection of such an infection (Wu *et al*., 2010).

- **<u>The hamster-co-infection-NMR model (2010)</u>**

In low socio-economic human populations suffering from bilharziasis, co-infection is generally expected to occur. In this study by Wu, Holmes and collegues (Wu *et al*., 2010), the metabolic alterations in the hamster co-infected with *S. japonicum* and *Necator americanus*, a common occurrence in sub-Saharan Africa, as well as parts of South America and southeast Asia (Keiser *et al*., 2002), were investigated using an NMR-based metabolic profiling technique, combined with multivariate statistical analysis. Twenty male hamsters were used in this study, of which half (*n*=10) were co-infected with 250 infective *N. americanus* and 100 *S. japonicum* cercariae via shaved abdominal skin, respectively. The other half (*n*=10) represented the control group. Blood and urine samples were collected at seven time points and at weekly intervals until week five. Urine samples were transferred into Eppendorf tubes and stored in a freezer at -80°C, while blood samples were drawn from the retro-orbital plexus of each hamster via a capillary tube and transferred into Eppendorf tubes. After centrifugation for 10 min the serum was transferred into Eppendorf tubes and kept at -80°C for succeeding ^{1}H NMR analysis.

The serum and urine samples were prepared and analysed, respectively, on a ^{1}H NMR of which their spectra were normalised prior to importing for pattern recognition analysis. The metabolites identified in the ^{1}H NMR spectra of urine included a range of aliphatic organic

acids such as citrate, succinate, 3-hydroxybutyrate, butyrate, formate and acetate. Amongst them were also aromatic metabolites such as 4-ethylphenol, hippurate, 4-cresol glucuronide, phenylacetylglycine and a range of amines like dimethylglycine. The reported metabolites found in the sera of hamsters were mostly amino acids, carbohydrate metabolism-related metabolites like pyruvate and citrate and lipid related metabolites such as acetoacetate. Multivariate PCA was performed between urine spectra obtained from *N. americanus* and *S. japonicum* co-infected hamsters and control non-infected hamsters to provide an overview of the data set and their source of separation. The results of this study concentrated on the effects of the co-infection on the qualitative disturbance of the metabolic profile, it did not try to make any reference regarding the comparative severity of the observed metabolic changes, as a single infection of *S. japonicum* in a hamster had already been described by Wang *et al.,* (2005). The resulting data of co-infection of *N. africanus* and *S. japonicum* indicated depletion of amino acids in the sera of co-infected hamsters. It has been well documented from the previous studies that bilharziasis causes liver injury resulting in disturbance of amino acids. These amino acids are capable of producing keto-acids like 2-ketoisovalerate and 2-ketoisocaporate via amino-transferases, which can be used as an alternative energy source. The marked reduction in levels of glucose and TCA intermediates such as citrate and succinate, in the co-infected hamsters, were also some of this study's findings. Hypoglycaemia is reported to occur in hamsters infected with *N. americanus* (Wang *et al*., 2009) due to marked catabolism and reduction of food intake after the infection. In addition, hypertriglyceridemia which was observed in another research associated with hookworm infection was reported to be caused by decreased lipolytic activity (Mukerjee *et al.,* 1990) and hyperlipidemia was shown by an increase in very low density lipoproteins and low density lipoproteins. This was also observed in this study on the co-infected hamsters.

A series of altered gut microbiotal-related metabolites in the co-infected hamsters such as reduced concentrations of hippurate, trimethylamine-*N*-oxide, 3-hydroxyphenylpropionic acid along with elevated levels of 4-cresol glucuronide and phenylacetylglycine were likewise reported. These altered gut microbiotal related metabolites appeared to be common to all helminths infections studied so far (Wang *et al*., 2004; Wang *et al*., 2009). Metabolic profiles in hamsters co-infected with two parasites and the consequence of the

combined metabolic effects were demonstrated and reported in this study along with responses to the respective single infections (Wu *et al*., 2009).

2.6.2.2 A human model for metabolomics of bilharziasis (2011)

Balog and co-workers (Balog *et al*., 2011) reported the first NMR-base metabonomics investigation of *S mansoni* infection in humans. While metabolomic studies employing mouse and hamster models have revealed metabolic consequences of a parasitic bilharzias infection, very little is known about the metabolic profile of infections in humans, which is the basis of this thesis. In comparison to the animal investigations, the study of Balog *et al.* (2011) is part of a multidisciplinary project that investigates the influence of treatment strategies on *Schistosoma* re-infection rates and *Schistosoma*-related pathology. The study's design entailed a well-characterised cohort of 447 individuals which were selected using a stratified random selection, balanced for sex and age, from a rural area in Uganda near Lake Victoria with a high prevalence of *S. mansoni* infections. Only individuals of 7 years and older were included in the cohort study. Unlike animal models, a study of an infection in humans is more complex due to variable intensities and differences in the time course following the initial infection. Compared to animal models, experiments which are conducted under a controlled environment and the high similarity between the individual animals, in human model experiments, it is crucial to give attention to inter- and intra-individual biological variability. These include different food patterns, co-morbidities due to other infections, and differences in age and gender.

The study was based on the collection of urine samples in an endemic area for *S. mansoni* at different time points following treatment for bilharziasis with praziquantel. The first sample was obtained before treatment was administered to the participants. The second sample was collected 24 hours after the administration of treatment. Two weeks later sample three was collected and a second dose of treatment was administered. Twenty-four hours following treatment the fourth sample was obtained, resulting in the final and fifth sample collected six weeks thereafter. The urine samples were kept cold at 4°C immediately after collection and kept for long term storage at -80°C. After thawing, urine samples were centrifuged for the removal of any solid components and prepared for ^{1}H NMR analysis.

After data processing of ^{1}H NMR spectra and data selection, two types of multivariate pattern recognition were performed, namely children-based PLS-DA and young adults-based PLS-DA. From the children based PLS-DA two class models were built, non-infected *vs.* heavily infected individuals. The PLS-DA model was built using 93 urine spectra, which showed a clear discrimination between heavy and non-infected subjects. The responsible metabolite for this discrimination was 2-C-methylerythritol, a compound of non-mammalian origin which was identified as a major discriminator between the classes.

Analysis of the adult group was also based on the model built for a non-infection *vs*. heavily infected individuals. Metabolites that were involved in the separation of the *S. mansoni* infected individuals and the non-infected controls included reduced concentrations of hippurate and increased phenylacetylglutamine (PAG). This decrease in hippurate levels is in agreement with the data reported for *S. mansoni* infected mice. Elevated levels of phenylacetylglutamine, a human metabolite corresponding to the murine metabolite phenylacetylglycine, were also in agreement with the data reported for *Schistosoma*-infected animals. Using supervised multivariate statistical analysis i.e. PLS-DA, Balog and collegues (Balog *et al.,* 2011) were thus able to discriminate infected from uninfected individuals in the two age groups (children and adults). The potential *S. mansoni* biological perturbations were found to be primarily linked to changes in gut microflora such as depleted levels of dimethylamine and hippurate and increased levels of PAG and trimethylamine. Energy metabolism alterations such as 2-oxoglutarate, succinate and fumarate contributed to the differentiation between infected and uninfected *S. mansoni* individuals. Liver function metabolites were also reported with increased levels of creatine, guanidine-acetate and trimethylamine-*N*-oxide.

These findings of the *S. mansoni* infection human model using NMR resembled data from earlier studies on *S. mansoni* infection in experimental animals and so provided indications for the existence of a response specification for this infection (Balog *et al*., 2011).

2.6.3 Complexity of the disease and common metabolic findings

The organic acids detected in the urine from hosts (mice, hamsters and humans) infected by *Schistosoma* species (*mansoni and japonicum)* as detected by NMR and CE-MS are summarised in Table 2.7. A total of 30 organic acids were reported from these six investigations, and these are listed in alphabetical order in Table 2.7. The last column in Table 2.7 indicates the commonality with which the organic acids were observed and reported in the six investigations. The main findings, summarized in this column, are:

(1) In most of the investigations the organic acids observed, are the outcome of a qualitative assessment only, without a univariate analysis to indicate the statistical significance of the substances listed. This places a clear limitation on a comparative assessment of the findings summarized in Table 2.7.

(2) Thirteen (43.3%) of the organic acids listed, were reported in only one of the six investigations. Of these substances, 6 were observed in the mouse-*S. japonicum* – NMR model, 4 in the mouse-*S. mansoni* - CE-MS model and 3 in the hamster-co-infection-NMR model, respectively. From these observations it cannot be conclusively stated that these differences between the three groups are related to the technology used (NMR and CE-MS), the experimental animal of the models (mice and hamsters) to the mode of infection (*S. mansoni, S. japonicum* or co-infection) or even to the technical expertise of the researchers associated with the investigation teams.

(3) The concentrations of two (6.6%) of the organic acids listed, were reported to be reduced in the urine of most of the experimental models. These are citric acid and succinic acid, both derived from the TCA cycle. Various explanations have been proposed to account for this observation, as listed by Wang *et al.* (2005):

    a. An indication of a perturbation in mitochondrial function due to physiological and/or toxicological stress.
    b. Renal damage is another alternative due to decreased citrate that may trigger precipitation of calcium salts in the kidney which would compromise renal function.

c. Increased urinary pyruvate observed in three of the models may be due to inhibition of pyruvate dehydrogenase which would lead to a decrease in citric acid availability for functioning of the TCA cycle.

Without more detailed clinical and biochemical information, the choice between these alternatives would be speculative.

(4) The presence of increased urinary hippuric acid in all 6 investigations is an indication of an altered detoxification system in the liver.

(5) The presence and increase in 4-cresol glucoronide, as well as phenylacetylglycine, give an indication that the gut biotica are also affected by the *Schistosoma* infection.

In the final analysis it is clear that the consequences of this infectious disease result in a complex pathological profile, despite some common characteristics between the various investigations, as outlined above. Additional research is clearly required, as well as information on the metabolite profile of *S. haematobium* infection, that has not been investigated in any of the metabolomics models described up to now.

Table 2.7 Metabolites obtained in reported bilharzia model studies

Reported organic acid biomarkers identified by metabolomics analysis for various experimental models on bilharziasis infection	Mouse (*S. mansoni*) Wang et al., 2004 (NMR)	Hamster (*S. japonicum*) Wang et al., 2006 (NMR)	Mouse (*S. mansoni*) Garzia-Pères et al., 2008 (CE-MS)	Mouse (*S. japonicum*) Wu et al., 2010 (NMR)	Hamster (Co infection) Wu et al., 2010 (NMR)	Human (*S. mansoni*) Balog et al., 2011 (NMR)	Assessment
Acetic acid	↓	↑	n.r.	n.r.	n.r.	↕	~
Adipic acid	n.r.	n.r.	n.r.	↓	n.r.	n.r.	One
Benzoic acid	n.r.	n.r.	↑	n.r.	n.r.	n.r.	One
Butyric acid	↓	n.r.	n.r.	n.r.	↑	n.r.	~
Citrate	↓	↓	↓	↓	↓	↕	↓: +++++
4-Cresol glucuronide	↑	↑	n.r.	↑	n.r.	n.r.	↑: +++
4-Ethylphenol	n.r.	n.r.	n.r.	n.r.	↓	n.r.	One
Fumaric acid	n.r.	n.r.	n.r.	↓	n.r.	↓	↓ : ++
Glycerol	n.r.	(↑)	↑	n.r.	n.r.	n.r.	~
Glycolic acid	n.r.	n.r.	↑	n.r.	n.r.	n.r.	One
Hippuric acid	↓	↓	↓	↓	↓	↓	↓: ++++++
3-D-Hydroxybutyric acid	↓	n.r.	n.r.	n.r.	n.r.	~	~
2-hydroxyisobuterate	n.r.	n.r.	n.r.	↓	n.r.	n.r.	One
2-(4-hydroxyphenyl)propionic acid	n.r.	n.r.	n.r.	↓	n.r.	n.r.	One
3-hydroxyphenylpropionic acid	n.r.	n.r.	n.r.	n.r.	↓	n.r.	One
4-hydroxyphenylpropionic acid	n.r.	n.r.	n.r.	n.r.	↓	n.r.	One
Indoylsulphate	n.r.	n.r.	n.r.	↑	n.r.	n.r.	One
2-Ketoglutaric acid	↓	n.r.	n.r.	~	n.r.	↓	↓: ++
2-Ketoisocaproic acid	↓	n.r.	n.r.	~	n.r.	↕	~
2-Ketoisovaleric acid	↓	n.r.	n.r.	↑	n.r.	n.r.	~
2-keto-3-methyl-valeric acid	n.r.	n.r.	n.r.	↑	n.r.	n.r.	One
Lactic acid		(↓)	n.r.	n.r.	n.r.	n.r.	~
Malonic acid	↓	n.r.	n.r.	↑	n.r.	n.r.	~
Phenylacetylglutamine	n.r.	n.r.	n.r.	n.r.	n.r.	↑	One
Phenylacetylglycine	↑	↑	n.r.	n.r.	↑	↑	↑: ++++
Propionic acid	↓	(↓)	n.r.	n.r.	n.r.	n.r.	~
Pyruvate	↑	↑	n.r.	↑	n.r.	n.r.	↑: ++
Succinic acid	↓	↓	n.r.	(↓)	↓	↓	↓: ++++
Urea	n.r.	n.r.	↓	n.r.	n.r.	n.r.	One

3-Ureidopropionale	n.r.	n.r.	n.r.	↑	n.r.	n.r.	One
Uric acid	n.r.	n.r.	↑	n.r.	n.r.	n.r.	One

↓ decreased

↑ increased

↕ increased and decreased

n.r. not reported

~ : Reported after early infection, but not reported for 5 weeks post-infection

2.7 Research question and aims of the investigation

Metabolomic studies can be conducted by practising the following approaches: "(i) targeted or non-targeted analytical approach, (ii) metabolite profiling and (iii) metabolic fingerprinting" (Vladimir *et al*., 2006; Halket *et al*., 2005). In this metabolomic investigation we investigate the metabolic effects of a *S. haematobium* infection on a human host, and we present a characterization of the parasitic infection by metabolic profiling, employing gas chromatography (GC-MS) spectroscopy and multivariate pattern recognition techniques. Our experimental design and investigation are based on addressing the following basic question:

What information can be revealed on humans infected by *S. haematobium* through an untargeted metabolomics investigation of the urinary organic acids?

Motivation

1. No metabolomics study of *S. haematobium* has been reported this far.

When this dissertation work started no research had been reported on the biological perturbations of a metabolic investigation of a *S. haematobium* infection in humans. Balog *et al*. 2011 explored the metabonomic investigation of a human *S. mansoni* infection. Although their investigation is the first of its kind to reveal the metabolic consequences of a parasitic bilharzia infection in humans, it was based on evaluating the influence of treatment strategies on *Schistosoma* re-infection rates and *Schistosoma*-related pathology. This dissertation, based on the metabolomic investigation of humans infected with *S. haematobium*, aims at determining the organic acid profiles of individuals infected and uninfected with *S. haematobium* through an untargeted GC-MS approach, with the potential of finding a biological perturbation that can give insight to the schistosome infection.

2. Organic acids are important indications of this infection.

As mentioned in Section 2.2 of this Chapter, schistosomes are facultative anaerobes, deriving their energy primarily via the degradation of glucose and glycogen (Despommier *et al*., 2005). It has been revealed in the 6 models illustrated in Section 2.7.2.1 that a bilharziasis infection causes an alteration in the expression of the hosts liver enzymes associated with the Krebs cycle, fatty acid cycle, urea cycle, amino acid metabolism and catabolism and also disturbances of the gut microflora (Wang *et al*., 2004; Wang *et al*., 2005; Garcia-Pérez *et al*., 2008; Wu *et al*., 2009; Wu et al., 2010; Balog *et al.,* 2011). Organic acids comprise key metabolites of virtually all pathways of the intermediary metabolism, as well as for many exogenous compounds. Therefore, comprehensive qualitative analysis of organic acids in body fluids has the potential of yielding information on the physiological and pathophysiological status of different metabolic pathways, as well as their interrelationships, which might ultimately lead to biomarker discovery.

3. The investigation on *S. haematobium* infection requires an inductive approach.

An untargeted *S. haematobium* infection investigation in humans requires an unprejudiced experimental design, designated as an inductive mode of investigation (Kell, 2006). Such a design should account for the unregulated confounding effects in the human model/host i.e., differences in diet, environmental effects, differences in age and gender along with co-morbidities with other infections. *S. haematobium* potentially co-exists with *S. mansoni* sharing the same habitat causing a co-infection between the two parasites (see Figure 2.3) (Meurs *et al*., 2012). Also the investigation has to be equipped with an efficient detection technology that is able to account for the effects of the bilharzia parasite on the host, such as the parasites metabolic excretions in the host as well as the host's reaction to the bilharziasis infection.

4. The need for a further diagnostic instrument on bilharziasis infection and treatment

Bilharziasis is diagnosed by compiling an egg screening count in urine or feces of an infected individual using a microscope, as described earlier. The parasitological examination is known as the "gold standard" even though by the time that the eggs are detected in the biological samples, the infection most likely has caused extensive damage to its host. Severity of the infection is determined by the number of eggs in 10 ml urine; however, various problems arise in this known diagnostic technique.

In consultation with officials at the National Health Laboratory Services (NHLS) of Elim and Malamulele Hospitals in Limpopo, diagnosis is hampered by several issues:

I. Not all infected individuals excrete eggs during their infection period, some individuals indicate asymptomatic characteristics.

II. Egg excretion varies during the day influencing the infection category of the infected individuals, with midday (11-12am) known to be the peak output of egg excretion (Wolmarans *et al*., 2006).

III. There is also a known possibility of having more than one type of bilharziasis infection considering *S. mansoni, S. haematobium and S. mattheei* (see Figure 2.3) might potentially occupy the same habitat (Ross *et al*., 1994; Meurs *et al*., 2012).

With the use of metabolomics, a metabolic profile of both the infected individual and uninfected individual can help alleviate these diagnostic issues. This can help to identify and investigate clear bilharziasis *S. haematobium* induced perturbations and also uncover possible biomarkers which can be applied to develop new and sensitive diagnostic techniques.

Based on the research question and its motivations, the following aims for this study were thus formulated:

1. Aim 1: Assemble a good set of samples for standardization of the metabolomics technology and for the metabolomics study of humans infected by *S. haematobium*, as well as from comparative controls (see Chapter 3).

2. Metabolic profiling requires a high degree of reliability of the metabolomics data, which is the basis for Aim 2: Propose an approach for an assessment of the repeatability of the generation of metabolomics data matrix required for a bioinformatic analysis to obtain the metabolite profile (see Chapter 4).
3. Aim 3: Determine the metabolic profile of organic acids from humans infected by *S. haematobium* and give a possible biological/pathophysiological interpretation of their profile in comparison to 6 other models of bilharziasis (see Chapter5).

The outcomes of these three aims will be discussed in Chapter 6, as well as proposals for future investigations, based on these outcomes, as well as other existing knowledge in the field of this investigation.

CHAPTER 3

EXPERIMENTAL DESIGN,

MATERIALS AND GENERAL METHODS

3.1 Introduction

Metabolites are intermediates and end products of cellular processes, and their levels may reflect the responses of biological systems, at the systems level (Tyagi *et al*., 2010). Metabolite profiling is a high throughput analytical method for relative quantification of a selected number of metabolites from biological samples such as urine, blood tissues or other biological samples (Shulaev, 2006). Human urine is composed of many classes of compounds and molecules of key metabolites of virtually all pathways of intermediary metabolism, these include organic acids, amino acids, purines, pyrimidines, sugars, sugar alcohols and other compounds at a range of concentrations (Kuhara *et al*., 2005), and was the biofluid of choice for the present study. The quantitative analysis of metabolite levels and their variation in urine has the potential of yielding information that can offer insight into the physiological and pathophysiological condition such as growth, disease development, treatment toxicity, diurnal variation and nutrition on the metabolome and its pathway along with their interrelationships (Nicholson *et al*., 2002).

In order to attain a good quality metabolite profile it is primarily an important aspect when conducting a metabolomics study to have a well structured pre-arranged experimental design. The experiments should produce a non-biased identification and quantification of all metabolites in a biological system, with sample preparation that does not exclude informative metabolites, and the selectivity and sensitivity of the analytical technique should be high (Dunn *et al*., 2005; Goodacre *et al*., 2007).

The number of metabolites which exceeds the number of samples by far, posed a major challenge in metabolomics data analysis. Selection of proper data pre-treatment is, therefore, an essential step in the analysis of this data as it may fundamentally affect the identification of metabolites (Van den Berg *et al*., 2006). The quality of data processing is a vital requirement to enable suitable interpretation of the data, as the primary interest in

metabolomics experiments is to discriminate between *interesting* biological variations from *obscure* sources of variability (Katajamaa *et al*., 2007). Results too optimistic may occur, due to methods that over-fit the data, making rigorous validation a compulsory component of the design of a metabolomics experiment (Westerhuis *et al*., 2008). It should also be noted that two intervening aspects influence the ultimate outcome of metabolomics investigations and should be included in the design of a metabolomics experiment: It is best practice (1) to involve the bioinformatics experts from the experimental design stage to the end in the technique, which allows for the identification of possible biases and improve the validity of the outcomes (Hu *et al*., 2005), and (2) to place equivalent emphasis on the quality of each stage of the process, as each stage presupposes the validity of the preceding stage (Van Batenburg *et al.,* 2007).

Against this background, the primary aim presented in this Chapter is to present the approach followed: " *to assemble a good set of samples for standardization of the metabolomics technology and for the metabolomics study of humans infected by S. haematobium as well as from comparative controls*" (see Chapter 2.7).

3.2 Metabolic profiling of organic acids

The metabolic profiling of urinary organic acids by GC-MS used in this investigation is adopted from a standard operational procedure for the diagnosis of organic acidemias. Following the discovery of isovaleric acidemia in 1966 (Tanaka *et al*., 1966) several organic acidemias, in which organic acids accumulate in the urine, have been discovered by the use of the GC-MS (Fernandes *et al*., 2006). Human urine can thus provide the necessary evidence of an endogenous perturbation, e.g. an inborn error of metabolism, but can also reflect other physiological or pathological perturbations due to exogenous perturbations. Due to its high chromatographic performance, sensitivity and specific identification and quantification of many metabolites, GC-MS is indispensable for the chemical diagnosis of such perturbations. For GC-MS analysis, urinary organic acids are extracted with ethyl ether and ethyl acetate under acidic conditions with or without adding sodium chloride and then dehydrated with sodium sulphate and evaporated to dryness; the residues are derivatized to increase their volatility and, therefore, their suitability for GC-MS analyses (Dalgiesh *et*

al., 1966). In a metabolomics investigation, a comparison is typically made on the differences between a control group and a group suffering from a perturbation, which is bilharziasis due to *S. haematobium* infections in humans in this investigation.

3.3 Experimental design

The broad outline of the experimental design is shown in Figure 3.1. It gives the experimental plan to be implemented in an attempt to answer the proposed research question and aims. The experimental design was arranged to generate two sections: the left side of this flow diagram indicates the procedures to be used to assess the repeatability in the generation of metabolomics data (Chapter 4); the right side indicates the parallel procedure to be followed for analysis of the clinical urine samples (Chapter 5). Each of these sections will be methodically explained and dealt with in detail in the specified chapters.

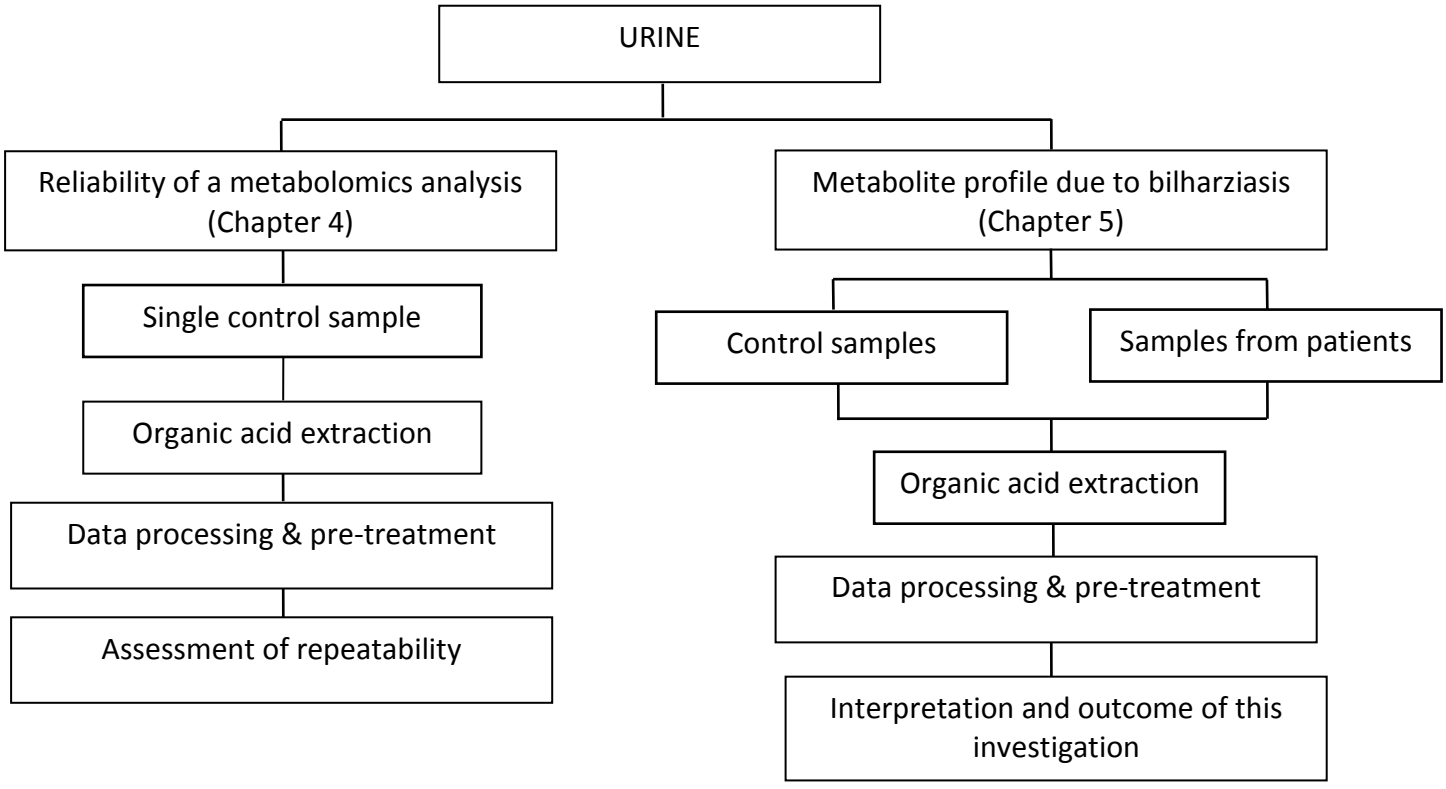


Figure 3.1 Schematic representation of the experimental design of a metabolomics experiment following data generation and analysis towards identification of possible biomarkers.

According to a proposal by Want *et al*. (2010) on profiling procedures for using urine, there are certain key issues to consider when generating GC-MS metabolite profiling experiments. These include:

(1) Sample collection and storage

(2) Chemical modifications of the samples, as applicable

(3) Separation techniques and instrumentation selection

(4) The make-up of test mixtures used

(5) The nature and quantity of the quality control samples and any 'blank' samples

(6) The sequence in which the samples will be ran and how they will be carried out, i.e., if run order is going to be randomised

(7) The number and type of replicates to be analysed

(8) The totality of the run length

(9) The amount of all the samples and batch size

The focus of this Chapter will be on experimental subjects and sample collection, some general methods, while most of the other aspects will be addressed in Chapters 4 and 5.

3.4 Reagents & materials

This section contains information of all the material and apparatus used during this study (see Table 3.1), the companies where the products were purchased and the products' catalogue/CAS number.

Table 3.1 Reagents and apparatus used in the study

Reagents	**Supplier/Manufacturer**	**Cat. or CAS #**
Hydrochloric acid (HCl)	Merck	100319
4-Phenyl butyric acid (mw 164.20)	Fluka	78243
Ethyl acetate HPLC grade/distilled	Sigma	494518
Diethyl ether HPLC grade/distilled	Sigma	309966
Sodium sulphate (Na_sSO_4) anhydrous	Merck	106649
Bis (trimethylsilyl)-trifluoracetamid (BSTFA)	Sigma	T 1506
Trimethylchlorosilane (TMCS)	Sigma	H T 4252
Pyridine	Merck	51 124060LC
Hexane	Sigma	52767
CCA diagnostic test strips (kit)	Rapid Medical Diagnostics	Not provided
UriCheck 9 labstix®	Rapid Medical Diagnostics	Not provided
Laboratory Apparatus		
Kimax culture tubes: Large 16 x 125 mm	Lasec	GIMK 45066A16125
Kimax culture tubes: Large 13 x 100 mm	Lasec	GIMK 45066A13100
Distiller Pasteur pipettes	Merck	612 1702
Roto-torque rotor	Labotech	6700
Graduated pipettes: 10 – 100 µl	Merck	3111000149
Graduated pipettes: 100 – 1000 µl	Merck	311100165
Centrifuge		
Heating block	Pierce	18840
Evaporating adaptor	Pierce	18817
Nitrogen	Pierce	18785
Hamilton syringes 100 µl and 250 µl	Separations	80391 and 80366
Agilent sample vials	Separations	11090500
Agilent sample inserts	Separations	09151819
Agilent sample caps	Separations	06090357
Capillary GC-MS column: DB-1MS (30m x 250µm x 0.25µm)	Agilent	
Agilent 7890A Gas chromatograph	Agilent	7890A
Agilent 5975B XL MSD mass spectrometer	Agilent	5975B XL

(detector) with auto sampler Agilent 7683B injector		
Data analysis software	Automated mass spectral deconvolution and identification system (AMDIS v2.66)	

3.4.1 Reagent preparation

- Internal standard; **3-Phenyl butyric acid** 26.25 mg, dissolve in a few drops NaOH and then to 50ml distilled H_2O
- 5M HCl; 50 ml 32% conc. HCl diluted to 100ml H_2O

3.4.2 Storage and stability

- 5M HCl and other stable reagents stored at room temperature.
- Internal standard stored in fridge (4°C).
- Bis (trimethylsilyl)-trifluoracetamid (BSTFA) & Trimethylchlorosilane (TMCS).
  Store in fridge after new container is opened, but must be at room temperature before opening to dispense and must be at room temperature for use. Use glass pipette to dispense.
- Na_2SO_4 must be desiccated during moist weather.

3.5 Samples

3.5.1 Ethical aspects

This section deals with the aim to obtain samples directed towards the two other aims of this investigation. As this investigation deals with analysis of biofluids from humans, the required ethical procedures were followed prior to sample collection. The metabolic experiments were done according to the ethical code of conduct of the Potchefstroom Laboratory for Inborn Errors of Metabolism (PLIEM, see www.pliem.co.za) of the Centre for

Human Metabonomics of the NWU, as approved by the NWU. The project was ethically also approved as part of the bilharzia research programme of the Unit for Environmental Research and Development of the North-West University. In addition, the Ethics Committee of the Elim Hospital also approved the project and permission from the Makuleke tribal authority was gained to collect the urine samples at the clinics of the National Health Laboratory Services (NHLS) at Elim and Malamulele in the former Venda (Vembe and Livubu regions of Limpopo Province). No informed consent was requested for this study as all samples obtained from the clinic and PLIM were anonymised, except for age, gender and ethnicity.

Three categories of samples were required for this investigation: (1) Samples for repeatability study, (2) samples for the investigation of the bilharziasis infection and (3) quality control (QC) samples.

3.5.2 Sample collection and storage

1. Repeatability of GC-MS metabolomics analysis of urine samples.

One urine sample of approximately ± 20 ml was obtained from a single African[2] male volunteer and used in this part of the investigation. The sample was retained at room temperature and an aliquot of about ± 0.6 ml transferred to a plastic tube for creatinine determination. The remaining part of urine sample was frozen at -40°C and kept until the next morning for the organic acid analysis. This urine was chemically analysed in an experiment to estimate the repeatability of GC-MS metabolome analysis and will be presented in further detail in Chapter 4.

2. Bilharziasis urine samples

Urine samples were collected in the Limpopo province in the former Venda (Vembe and Livubu regions), from the clinics of the National Health Laboratory Service (NHLS) in

[2] **Note:** Samples for the investigation of bilharziasis (patients and controls) were all obtained from African individuals.

Malamulele and Elim Hospitals. The health officials of the NHLS at the Elim and Malamulele Hospitals were eager to cooperate in this project by supplying all required clinical material.

Urine samples were collected over a period of 4 months; 2 months collection in Elim (February and March 2011) and 2 months collection in Malamulele Hospital (April and May 2011). All samples were referred by clinicians to the clinic due to the presence of haematuria, which is the primary indication of *S. haematobium* infection. Samples of approximately ± 20 ml were collected into sterile polypropylene screw-top vessels (147 samples in total collected). Samples were frozen and stored at -20°C in a transportable freezer until transfer in the freezer to the laboratory at NWU. Samples were stored at -80°C for a long storage periods at the NWU. Samples were collected from individuals with characteristics as shown in Table 3.2

Table 3.2 Characteristics of donors of the urine samples

Sample type	**Ethnicity**	**Gender**	**Age**
Repeatability Experiment			
Normal mid-day urine sample	Black	Male	25 yrs
Number of samples in total 1			
Metabolomics of bilharziasis experiment			
Bilharzia Positive & Negative	Black	Male and Female	5 – 34 yrs
PLIEM samples (Bilharzia negative)	Black	Male and Female	4 – 38 yrs
Number of samples and participants in total 147			

All samples showed haematuria, and were classified into two batches, based on the following criteria:

- Bilharzia Positives (BP): tested by the NHLS in the course of a microscopic egg count. Eggs found present in the urine sample was the decisive factor for our patient group.
- Bilharzia Negatives (BN): tested by the NHLS in the course of a microscopic egg count. No eggs were found in the urine samples of the control group.

A second control group was also introduced, and this served as an "uncontaminated" control from participants who did not stay in an endemic area but whose urine samples were analyzed at the PLIEM at the NWU. These urine samples (abbreviated as MC) are known not to be infected with bilharziasis; however they are not "normal", as they were obtained from individuals with clinical symptoms of a suspected metabolic disorder, but

who were found not to have any of the known metabolic diseases. Their selection criteria fit the conditions of those selected in the Limpopo province (Table 3.2).

3. Quality control

A quality control (QC) sample (Want *et al*., 2010; Dunn *et al*., 2011) was prepared by mixing equal volumes (3ml) from each of the 147 samples (bilharzia positive, bilharzia negative and PLIEM controls) as they were being aliquoted for analysis (Chapter 5). This pooled urine was used to provide a representative mean sample containing all the analytes that might be encountered during the analysis of organic acids. The QC samples were evenly distributed over all the batches and were extracted, derivatized and analysed at the same time as the individual study samples, as part of the total sequence order as shown in the relevant section of Chapter 5.

3.5.3 Creatinine determination

For accurate quantitative measurement of the urinary metabolites it is essential to standardize the metabolite concentration by normalizing it to the concentration of a reference metabolite, because of the differences in the concentration of urine. To obtain quantitative values of the metabolites relative to the total creatinine in urine, the correct total creatinine value should be determined. The creatinine values of the experimental participants, both patients and controls were determined by Pathologists: Drs Du Buisson, Kramer, Swart, Bouwer Inc. /Ing.

The volume of urine and derivatisation reagent used is based on the urinary creatinine value of the sample, according to the following guidelines:

- Creatinine values higher than 8.8 mol use 0.5 ml urine.
- Creatinine values lower than 8.8 mol, but higher than 0.44 mol, use 1 ml urine.
- Creatinine values lower than 0.44 mol, but higher than 0.18 mol, use 2 ml urine.
- Creatinine values lower than 0.18 mol, use 3 ml urine.

The creatinine values of the urine sample are used (1) to direct the volume of the sample used for the organic acid extraction, as well as (2) for relative quantification of urinary organic acids, if necessary.

3.6 Methods

3.6.1 Organic acid analysis

The method of organic acid extraction analysis used in the study, is that practised on a daily basis in the PLIEM and described in Reinecke *et al.,* 2011.

To perform the organic acid analysis the urine sample was firstly thawed and if required sonicated to ensure that all the precipitated material became dissolved or suspended. The sample was then transferred into a Kimax test tube where internal standard (4-phenylbutyric acid) was added to a final concentration of 180 mmol/mol creatinine. This ensured a fairly constant ratio between the urinary organic acids, including the internal standard. Phenyl butyric acid was used as the internal standard due to its absence in normal urine and in known pathological conditions, and since it elutes almost in the middle of the organic acid profile and co-elutes with very few other organic acids. The urine samples were acidified by addition of 6 drops 5N HCl to adjust the pH of the urine sample to lower than approximately two.

Six ml of ethylacetate was added to each sample and the mixture shaken on a rotary wheel for 30 minutes. The mixture was then centrifuged for 3 minutes at 3 500 rpm and the upper aqueous ethylacetate organic acid phase was transferred into a clean Kimax test tube. Three ml of diethylether was then added to the aqueous (lower) phase and again shaken for a further 10 minutes. The sample then was again centrifuged where the upper aqueous phase was added to the previous upper ethylacetate organic phase. The ethylacetate/diethylether mixture was then dried with a small amount of sodium sulphate (two spatulas added), centrifuged and transferred into a clean small Kimax test tube (10 ml). The organic solvent was then evaporated to complete dryness under nitrogen at 40°C for ± 45 min.

3.6.1.1 Derivatisation

Once completely dry BSTFA:TMCS was added, used as the derivatisation reagent. Derivatisation was done at 70°C for 45 minutes in a sand bath or heating block. The derivatised mixture was then transferred to a 1,5 ml vial for GC-MS analysis. This approach ensures a fairly constant concentration of organic acids in the derivatisation mixture which contributes to the repeatability of the analysis.

For the analysis of metabolites using the gas chromatography (GC), particularly when interfaced with the mass spectrometry GC-MS which is a preferred technique for the separation of volatile compounds, organic acids must first be converted to derivatives that they are thermally stable, chemically static, and volatile at temperatures of up to about 300°C (Kuhara *et al.,* 2005). Trimethylsilylation has been the method of choice for most organic acid profiling experiments. Silylating agents, including trimethylcholorosilane (TMCS) and *tert*-butyldimethylsilyl (TBS/TBDMS) donors, stabilize a wide variety of compounds for volatilisation during GC. Among silylation reagents, BSTFA is very versatile and is preferred for organic acid analysis because of its reactivity, volatility, excellent solvent properties, and availability in pure form, it also causes less chromatographic interference (Knapp *et al*., 1979). TMCS increases the reactivity of BSTFA and when combined together, BSTFA is easily able to fully derivatize most compounds in 10 min at 60°C making them more volatile, generally less polar and more thermally stable (Kuhara *et al.,* 2005).

3.6.1.2 Gas chromatography – mass spectrometry conditions

The derivatized organic acids were analyzed with an Agilent GC-MS system model 7890A-5975B GC-MSD equipped with a DB-1MS capillary column (30 m x 0.25 mm x 0.25 µm) from Agilent Technologies, Santa Clara, CA, USA. The column is responsible for fractionation of the various metabolites. The samples (1 µl) were injected in splitless mode at a temperature of 280°C in the injection port. The temperature program for the port with the capillary column started at 60°C for 2 min, increasing at 4°C/min to 120°C, and then increased at 6°C/min to 285°C, and kept for 2 min at that temperature. The carrier gas was helium (17.73 psi) and electron impact ionization was applied at 70 eV. Mass Spectral acquisition was performed in scan mode for the detection of all components separated by the GC analysis.

3.7 Deconvolution, peak identification & quantification

In this section we describe how the data is pre-treated and pre-processed before it is statistically analysed.

3.7.1 Identification and quantification of the metabolites

Data quantification was performed using an automated mass spectral deconvolution and identification system (AMDIS software version 2.66), linked to NIST (National Institute of Standards and Technology) Mass spectra search program for the NIST/EPA/NIH Mass Spectral Library containing a library with organic acids. The analytical setting of the AMDIS software was as follows:

- Minimum match factor – 80%
- Type of analysis – Use internal standard for retention index (RI)
    - Component width of 32
    - Adjacent peak subtraction – one
    - Resolution – medium
    - Sensitivity – medium
    - Shape requirements – medium

The software (see Figure 3.2) extracts pure components mass spectra from the GC-MS data files as “raw data” and then uses these spectra to identify the compounds in the chromatogram via a reference library.

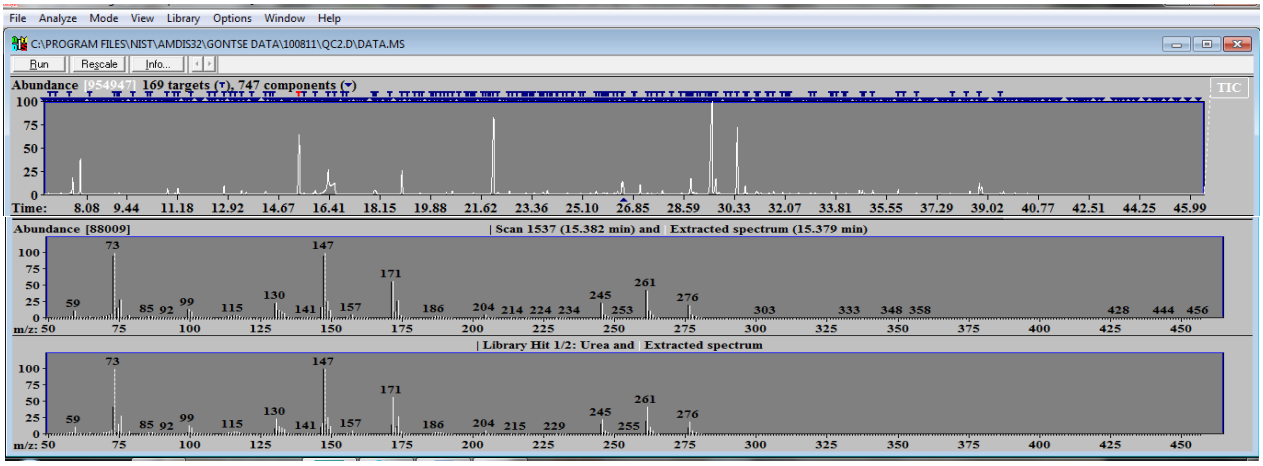


Figure 3.2 Spectral imaging of a chromatogram and mass spectra from AMDIS. Used to identify and extract pure components of mass spectra from the GC-MS via a reference library.

By “raw” one refers to the first level of data that is produced by the instrumentation. In metabolomics, this is usually composed of chromatograms, which is, in this instance, the GC profile. The first hits of identified compounds and integrated area of peaks were generated into a report file and exported to Microsoft Excel©. Peaks with uncertain identification are marked with a question mark by the AMDIS software against the name of the most likely organic acid in the library. These peaks were manually inspected and compared to mass spectra and retention times of the pure compound for identification. The concentrations of identified organic acids were determined according to the internal standard using the formula (Chen *et al*., 2009);

Concentration of the organic acid (µmol/l) = Area (compound)/Area (IS) x Conc. IS.

Concentration can be reported and calculated in either mg/g (*Area org acid / Area IS* 262.5*) or mmol/mol (*Area org acid / Area IS* 180*) creatinine; however, in this experimental investigation it is reported as mmol/mol creatinine.

3.7.2 Rapid medical diagnostics CCA test

The commonly employed direct parasitological method for diagnosis of urinary schistosomiasis is the standard urine filtration method that involves the detection and quantification of *S. haematobium* eggs in 10ml filtrate of a urine specimen. For purposes of quality and accuracy on the urine diagnosis an immunodiagnostic bilharziasis test has also been commercialized and proposed to overcome some of the limitations with parasitological methods such as the one described above. Immunological approaches are based on the detection of antibodies or antigens. This test was used as one of the final assessments of the infection by bilharziasis of the cases described in Chapter 5.

The urine-CCA (Circulating Cathodic Antigen) cassette test is an analysis of antigens released by the parasite and secreted in the urine of its host which is present in all *Schistosoma* species, including animal species. The urine-CCA cassette test is for the qualitative probable detection of an active bilharziasis infection. Intensity of infections differs with differences in location. A broad-spectrum of medium to high level infections with *S. haematobium* can be diagnosed using the urine CCA strip. A positive CCA test result on randomly collected

midstream urine indicates an active bilharziasis infection but is, however, not species specific.

3.7.2.1 Test principle

After applying the urine sample to the assay, the CCA antigen that may be present in the sample, binds to the labelled monoclonal antibody immobilized on the sample membrane. The solution then runs over the strip where the antigen-antibody complex attaches to another monoclonal antibody immobilized at the test line. A pink-colored line develops as a result indicative to a positive infection. The second line is a procedural control, which should always show up to make sure the test works correctly. The intensity of the line is qualitatively related to the intensity of the infection.

3.7.2.2 Specimen collection and preparation

The urine samples used here were those collected in Elim and Malamulele hospitals. This test was performed only as a verification test on the samples obtained from the NHLS laboratory in the above mentioned hospitals. No preparation was done to the samples – they were thawed and tested. The protocols and storage instructions of the CCA test kit described below are those that were supplied with the purchased test kit.

3.7.2.3 Kit components

Each kit contains the following components in sufficient quantities to perform the number of tests indicated on the package label:

- 25 x Test Cassettes each individually packaged
- 1 x Instructions for use
- 1 x 3ml bottle of buffer
- 25 x urine collection devices (plastic pipettes)

3.7.2.4 Precautions

1. Keep storage boxes dry.
2. Do not reuse test cassettes.

3. Do not use test cassettes if foil pouch is punctured or damaged.
4. Never pipette by mouth or allow reagents or patients sample to come into contact with skin.
5. Optimal results will be obtained by strict adherence to this protocol. Reagents must be added carefully to maintain precision and accuracy.
6. Performing the assay outside the prescribed time and temperature ranges may produce invalid results. Assays not falling within the established time and temperature ranges must be repeated.
7. The components in this kit have been quality control tests as a master lot unit. Do not mix components from different lot numbers. Do not mix with components from other manufacturers.
8. Care should be exercised to protect the reagents in this kit from contamination. Do not use if there is evidence of microbial contamination or precipitation. Biological contamination of dispensing equipment, containers or reagents can lead to false results.
9. Do not heat-inactivate samples.
10. All human urine products should be handled as potentially infectious material.
11. Waste disposal. Testing materials should be disposed of in accordance with local, state and/or federal regulations.

3.7.2.5 Assay procedure

Note: Ensure all reagents are equilibrated to room temperature (20-25°C) before commencing the assay.

1. Remove the test cassette and collection device from their pouches just prior to use.
2. Squeeze the pipette bulb and insert the tip into the urine sample.
3. Allow the sample to fill up by gently releasing the bulb.
4. Transfer 1 drop of urine to the circular well of the test cassette by gently squeezing the bulb.
5. Allow the sample to absorb entirely into the specimen pad within the circular well.
6. Hold the buffer bottle vertically and 1 cm above the circular well.

7. Add 1 drop buffer.
8. Read the result exactly **20 minutes** after adding buffer to the test cassette.
9. Any results read outside **25 minutes** should be considered **invalid** and must be repeated.
10. The **blue** control line must turn **pink**. If the control line stays blue the test should be considered **invalid**.
11. Any line in the test area should be considered positive.

Typical results obtained with the CCA-method are shown in Figure 3.2, and were used as the directive for the assessment of *S. haematobium* infection presented in Chapter 5.

Positive

	**Control band turns pink.** **A band is present in the test T area.** The test is positive for Bilharziasis

Negative

	**Control band turns pink.** **No test T band present**. Demonstrates the test was performed correctly but no Bilharziasis antigens were detected.

Invalid

	**Control line stays blue.** Only a pink control line should be considered positive. The test is invalid and should be repeated
	**A test line with no control line.** A pink control line must be present

Figure 3.3 Interpretation of the CCA results

3.7.3 Dipsticks

Dipstick urine analysis was performed using UriCheck 9 labstix® on the bilharziasis patient and control samples obtained from the NHLS and control samples obtained from PLIEM; this was performed to determine the microhaematuria homogeneity of the samples. The results obtained from these dipsticks are, however, not presented in this dissertation. A urine

dipstick consists of a white plastic strip with absorbent microfiber cellulose pads attached to it. Each pad contains the dried reagents needed for a specific test. When performing the test dips the strip is placed into the urine, and let to sit for a specified amount of time (±6 seconds). The colour changes are compared to a standard chart (Chernecky *et al*., 2001).

Although there are more sensitive methods of measuring the levels of bilirubin, protein, glucose, ketones, micro-haematuria and urobilinogen (which are the main substances detected using the dipstick) in urine, the dipstick test method permits detection of a lower concentration of the respective substances. In this investigation the main purpose of utilizing the dipstick was to determine micro-haematuria in both patient and control samples. A brief description of the most common use of the dry reagent dipstick tests follows.

pH: A combination of pH indicators (methyl red and bromthymol blue) react with hydrogen ions (H^+) to produce a colour change over a pH range of 5.0 to 8.5. pH measurements are useful in determining metabolic or respiratory disturbances in acid-base balance. For example, kidney disease often results in retention of H^+ (reduced acid excretion).

Protein: Based upon a phenomenon called the "protein error of indicators," this test uses a pH indicator, such as tetrabromophenol blue, that changes colour (at constant pH) when albumin is present in the urine. Albumin is important in determining the presence of glomerular damage.

Glucose (sugar): The glucose test is used to monitor persons with diabetes. When blood glucose levels rise above 160 mg/dL, the glucose will be detected in urine. Consequently, glycosuria (glucose in the urine) may be the first indicator that diabetes or another hyperglycemic condition is present.

Ketones: Ketones are compounds resulting from the breakdown of fatty acids in the body. These ketones are produced in excess in disorders of carbohydrate metabolism, especially Type 1 diabetes mellitus.

Blood: Red cells and haemoglobin may enter the urine from the kidney or lower urinary tract. Testing for blood in the urine detects abnormal levels of either red cells or

haemoglobin, which may be caused by excessive red cell destruction, glomerular disease, kidney or urinary tract infection, malignancy, or urinary tract injury.

Bilirubin: Bilirubin is a breakdown product of haemoglobin. Most of the bilirubin produced in humans is conjugated by the liver and excreted into the bile, but a very small amount of conjugated bilirubin is reabsorbed and reaches the general circulation to be excreted in the urine. The normal level of urinary bilirubin is below the detection limit of the test. Bilirubin in the urine is derived from the liver, and a positive test indicates hepatic disease or hepatobiliary obstruction.

Urobilinogen: Increased urinary urobilinogen occurs in prehepatic jaundice (hemolytic anemia), hepatitis, and other forms of hepatic necrosis that impair the circulation of blood in the liver and surrounding organs.

Leukocytes: The presence of white blood cells in the urine usually signifies a urinary tract infection, such as cystitis, or renal disease, such as pyelonephritis or glomerulonephritis.

3.8 Statistical analysis

The statistical methods concerning the repeatability study (see Chapter 4) are described in Section 4.4.3. That section contains traditional methods of measuring repeatability, as well as a new qualitative method. The statistical methods used to analyse the data of the bilharziasis experiment (see Chapter 5) involve univariate, as well as multivariate methods which are described in Section 5.2 2.2.

CHAPTER 4

GENERATING METABOLOMICS DATA

4.1 Introduction

Metabolomics aims to achieve unbiased qualitative and quantitative metabolite information from acquired biological samples (Fiehn *et al*., 2000; Bino *et al*., 2004). "In metabolomics experiments, a snapshot of the metabolome is obtained that reflects the cellular state, or phenotype, under the experimental conditions studied "(van der Werf *et al*., 2005). Metabolomic experiments are usually conducted according to an experimental design which chiefly resembles the main components of the generally accepted metabolomics workflow. In order to interpret and make use of the obtained metabolomics data, sources of quantitative variation such as data pre-treatment and data processing ought to be accurately described. Reliability of the instrumentation must be assessed with technical errors of the analytical technique maintained to a minimum in order to differentiate the variance in data from the natural biological inter-individual and intra individual variation among the study groups that are subjected to a certain experimental design. Metabolic profiling thus requires a high degree of reliability of the metabolomics data, which is the basis for the aim addressed in this Chapter: "*To propose an approach for an assessment of the repeatability of the generation of metabolomics data matrix required for a bioinformatic analysis to obtain the metabolite profile*" (see Chapter 2.7).

Keun *et al.* (2002) declares that "the use of -omics technologies evolves from essentially qualitative measurements, and it thus becomes crucial to assess the reliability and repeatability of data generated from these new technologies." In research the implementation of these technologies is fundamental for any real world, furthermore, the reproducibility and robustness of these technologies and techniques will in addition present answers to fundamental questions, such as whether or not signature profiles of chemicals and other stressors can be confidently defined (Keun *et al*., 2002). In any analytical technique, high repeatability can increasingly improve the quantitative accuracy and

sensitivity of the technique and, in addition by decreasing the number of repeats required for any given task increasing sample throughput.

In this chapter we discuss the steps in a metabolomic investigation (Section 4.2), as well as the importance of repeatability, reproducibility, robustness of analytical techniques and quality control procedures for metabolomic projects (Section 4.3). In addition, a repeatability experiment is conducted and analysed using traditional methods and a new qualitative method (Section 4.4 and 4.5). The definitions of repeatability, reproducibility, and robustness are given in Section 4.3.

4.2 Steps in metabolomics investigations

This section proposes to describe how metabolomics data is acquired, handled and analysed.

4.2.1 Sample handling

For metabolomic investigations, sample storage and stability for human biofluids with regards to research are matters that have been addressed concerning NMR and GC-MS based approaches. Through various studies, Lauridsen *et al.* (2007) indicated that urine samples should be stored at or below minus 25°C and recommended not to stay at 4°C for prolonged periods of time. Maher *et al.* (2007) acknowledged similar results using NMR based metabolomics. Short term storage, however, of human urine samples at 4°C was revealed to be stable for up to at least 48 hours using GC-MS profiling, meaning that samples for this type of work can be considered stable (Gika *et al*., 2008). The effects of freeze thaw cycles on the metabolic profiles have been investigated using NMR, LC-MS and GC-MS, where it is recommended to keep freeze – thaw cycles to a minimum to minimise confounding factors in urinary metabolomics analyses.

4.2.2 Data preparation

The common greatest source of variability is the stability of the compounds to be analyzed whether derivatised or not. Since not all samples can be run in a single analytical batch, metabolomics investigations must incorporate analytical techniques that have a well derivatisation system which can produce considerably stable derivatives, since derivatised samples unavoidably reside in the auto-sampler for some time before their injection, especially if the sample batch to be analysed is of a considerable size (Villas-Bôas *et al.,* 2005). Failure to stabilise the sample derivatives in a time period of 24 hours can result in large standard deviations which may be observed in a repeatability analysis i.e. analysing the same sample more than once. Also derivatisation causes the sample to be more volatile and so; losses by evaporation is an additional source of variation for samples held in the auto-sampler for too long (Villas-Bôas *et al.,* 2005).

GC-MS machines produce chromatograms that are potentially information-rich with data volumes that can reach up to 3.24 MB in size and mass spectral based information with ion concentrations of about 0.5s long scans obtained over a 1200s chromatographic run time. "This raw form of data is not suitable for data analysis and needs to be distinguished, to extract useful information from the large capacity of raw data" (Taylor *et al*., 2002).

Krull *et al*., (1999) explained that validation of (analytical) processes is essential not only for commerce but also is as equally important for the academic laboratory. Validation means "to establish the soundness of" and so a valid method needs to be applicable and also be able to provide a clear description of the function for which it is indented to be used. Various research publications are unable to make a clear differentiation between method development and method validation; method development explains the procedures which have been taken to progress to a course that led to a specific investigative result and ultimately to a protocol. Whereas method validation would entail that an assured outcome is always gained if distinct processes are applied to a specified problem. The method should, therefore, be repeatable and reproducible (Fiehn *et al.,* 2007).

Dasgupta *et al.* (1998) and Birkemeyer *et al.* (2003) noted that sample preparation frequently supplies adequate reproducibility for the examination of metabolites for metabolome analysis, although is it highly dependent on the biological make-up of the sample, as well as on the protocols used for extraction, and its concentration. Therefore, internal standardization is essential to minimize the sample variation. Addition of an internal standard to any analytical investigation should be done as early as possible in the analytical process; i.e. if derivatisation is acquired to stabilise the sample, then the internal standard should also be derivatised and undergo the same analytical procedures as the investigated sample, also the internal standard should not exhibit significantly higher or lower boiling points than those of the analytes, and should also not hinder or obstruct the metabolites found in the sample (Villas-Bôas *et al.,* 2005).

4.2.3 Data pre-treatment and data analysis

Before biological insights acquired from metabolomics based studies can be interpreted, they must first be extracted from raw data (Katajamaa *et al*., 2007). Raw non-deconvoluted GC-MS metabolomic datasets are characteristically large, complex and often include a multitude of data files (Carroll *et al.,* 2010), and similar to other omic data sets they are largely underdetermined, which means the data contains numerously more variables than the sum of the samples of which the variable-linked data contains convoluted signals for hundreds to thousands analytes. Complex reliable and interpretable mathematical algorithms and models have consequently been developed for the identification and quantification of signals corresponding to biologically relevant analytes and to achieve a quantitative representation of the metabolic effect(s) related with the experimental factor(s) of interest (Katajamaa *et al*., 2007). This data extraction procedure can be further explained in a number of general steps (Carroll *et al.,* 2010):

(1) Exposure of analytically useful signal features[3] such as peaks of potential metabolites.

[3] The term "**feature**" is mostly used for an item in the original GC-profile that may be a metabolite, a new unknown chemical substance or background noise. **Variables** are features excluding background noise, with or without identification of a biological name. Variables for which a known identification could be achieved, are designated as **metabolites**.

(2) Classification of biologically important signal features by comparing them against a reference library of identified signal characteristics for biological analytes.
(3) Allocation of quantitative signal measurements to each identified biological analyte.
(4) Construction of a data matrix of dimension *n* x *p* where *n* is the number of samples and *p* is the number of metabolites (features). Some form of data normalization against internal standard(s) and/or biological sample mass/volume/amount, is then applied and the data is scaled or transformed.
(5) Implementation of statistical and exploratory data analysis techniques to determine the effect(s) of experimental factors on metabolite levels.
(6) Interpretation of experimental metabolite-level changes in the context of previous understanding about the metabolic system under investigation.

The widespread lack of equipment for carrying out the above data-analysis steps promptly and effortlessly has been one of the greatest challenges weighing down the establishment of metabolomics as a conventional technique (Carroll *et al*., 2010).

4.3 Quality control

Consider the following definitions of repeatability, reproducibility and robustness.

- **Repeatability** is the variation in measurements generated by a single person (the analyst) or instrument (the instrument) on the same item (the sample) and under the same conditions (the standard operational procedure). A measurement may be said to be repeatable when this variation is smaller than some agreed limit. According to the guidelines for evaluating and expressing the uncertainty of NIST measurement results, repeatability conditions thus include:

  - The same measurement procedure
  - The same observer
  - The same measuring instrument, used under the same conditions
  - The same location
  - Repetition within a short period of time.

Hence, experimental work and data are repeatable when we can repeat the experiments in our own experimental setup and obtain consistent results.

- **Reproducibility** is one of the main requirements of the scientific method, and refers to the ability of a test or experiment to be accurately reproduced, or replicated, by someone else working independently, but following the same experimental procedure. It includes the possibility for experimental results to be recreated, ability to be duplicated and the ability to be copied. We mean that others can repeat our experiments and get results consistent with ours.

- **Robustness** is the ability of being able to withstand stresses, pressures, or changes in procedure or circumstance with regard to the experimental procedures and/or analytical equipment. A system, organism or design may be said to be "robust" if it is capable of coping well with variations (sometimes unpredictable variations) in its operating environment with minimal damage, alteration or loss of functionality.

NMR spectroscopy is notoriously recognized to be a relatively stable and reproducible analytical platform (Dumas *et al*., 2006), yet there still is relatively few publications based on the reproducibility of GC-MS based methods (Gika *et al.,* 2007). It therefore vital for these mass spectrometry based studies using such techniques to demonstrate the capability for high quality metabolomics data so that they can be able to provide sound biological insight into the process of development, disease and the attempt of finding possible biomarkers.

In metabolomic investigations, the reduction and control of unwanted sources of variation is important. Such variation can be broadly recapitulated as analytical and biological variation (Maher *et al*., 2007). These include the following illustrations (1) inter-subjective variation, (2) sample collection, storage and stability, (3) sample pre-treatment, including metabolite extraction, and derivatisation prior to analysis, (4) instrument variation and stability. The most frequent form of variation in data resulting from a metabolomics experiment is the result of the sum of the induced biological variation i.e. sample collection and storage, and also including the total uninduced biological variation such as diet (Shurubor *et al*., 2005).

Seemingly countless research papers emphasize on the details of a specific instrument i.e. the type of NMR or mass spectrometer based technique used for data attainment, when in fact this should be the least important aspect than information on how the quality of measurements was guaranteed (Fiehn *et al*., 2007). Regardless of the type of analytical instrument used, there should be practices that are available to ensure and establish long-term data precision. With analytical techniques it's been acknowledged that certain instruments do go through a systematic drift in sensitivity over certain time period; while others may lose specificity for certain compounds without obstructing other metabolites, certain instrumentation may merely lack robustness, producing highly oscillatory and barely manageable measurements (Fiehn *et al*., 2007). Another source of variation is noticed in chromatography – MS systems, the sample inevitably interacts directly with the instrument i.e. with the coating in the GC column and this results in modified changes in measured metabolic feature response over time, both in terms of chromatography and MS.

The scale and timing of signal attenuation is also not consistent across all measured metabolic features and once more it is dependent on the type of biofluids measured. In our laboratory the GC-MS system is used for a wide variety of research applications (i.e. general purpose analyses), ranging from diagnostic metabolomics research in infectious diseases (bilharziasis, HIV and several variants of TB) to analysis of detoxification profiles and screening methods for diagnosis of inherited metabolic disorders, including some of these for commercial purposes. Therefore our GC-MS system has a potentially greater number of sources of contamination that can be introduced through the carry-over effect of column bleeding. For this reason it is a necessary requirement that QC (quality control) samples are periodically analysed throughout an analytical process in order to provide robust quality assurance for each metabolomic feature detected.

During data processing, data conditioning algorithms can use the QC responses as the basis to assess the quality of the data, remove peaks with poor repeatability, and correct the signal attenuation and concatenate batch data together after data acquisition and before statistical analysis (Zelena *et al*., 2009; Stangster *et al* 2006; van de Kloet *et al*., 2009). As noted by Sanchez *et al.* (2010),"a full and independent repeat of a biological experiment is required to assess the robustness and repeatability of results from metabolic studies."

In the current study only one researcher was involved. Hence, we will not consider reproducibility but instead only focus on repeatability. Also, given the scope of the project, no measurement of robustness was considered. The methodology of this chapter will then be further applied to the QC samples included in the metabolomic investigation concerning the bilharziasis samples, as presented in Chapter 5.

4.4 Measuring repeatability

In this section an outline is presented of the experimental aspects, analytical method, as well as the statistical methods required to investigate repeatability. The results of the repeatability experiment are discussed in Section 4.5.

4.4.1 Generation of the data for evaluation of repeatability

Sample collection and preparation

The urine sample used for this study was obtained from an adult African male volunteer, 25 years of age (see Table 3.2). A urine sample of approximately 20 ml was collected into a sterile polypropylene screw-top vessel of which 0.6 ml was transferred into a plastic 5 ml tube for creatinine determination, while the remaining urine was kept frozen at -20°C for later use. Five aliquots of ±3 ml each aliquot were used for the analysis of their organic acid content, done by the same analyst[4].

[4] **Notes:**

(1) The analyst referred to here is Gontse Moutloatse, the student presenting this MSc dissertation.

(2) As indicated in Chapter 4.5, the repeatability experiment was done twice. The first result indicated the outcome of the first experience of the analyst in doing a metabolomics experiment. From this experience more stringent conditions were defined for the second analysis as the results of the first experiment indicated that the repeatability experiment was not satisfactory.

Sample analysis

The protocol, described in Chapter 3.6, was used for the isolation of the organic acids in the five aliquots used for the assessment of repeatability in the generation of a metabolomics data set. The analysis of the batch of five extracted and derivatized organic acid samples was performed on an Agilent 7890A gas chromatograph coupled to an Agilent 5975B XL MSD mass spectrometer (the detector), according to the procedures described in Chapter 3.

Data acquisition and pre-processing

The data from GC-MS analysis were identified and deconvoluted using the AMDIS (Automated Mass Spectral Deconvolution and Identification System) spectral deconvolution software package version 2.66 (Stein *et al*., 1999), as described in Chapter 3. The output from AMDIS was in the *. elu and *. fin files of the reference sample. The resulting AMDIS reports were exported to *.txt files and meta-data were created with the following information: name of the compound, retention time and area of the peak. The area of each peak was used to calculate the relative concentration for each respective compound using, the internal standard (i.e. data normalization) via the following formula:

Relative concentration (mg/L) = area of peak / area of IS * 262.5.

For multivariate techniques to work, and hence to be able to compare samples with large data bases it is a general prerequisite that the generated data table/matrix must be assembled where each sample is characterized by the same number of variables and each of these variables is represented across all observations. Furthermore, it is also imperative that a variable in one sample has the same biological significance and represents the same metabolite for all other samples otherwise; the entire concept of comparing samples based on the systematic changes in metabolite patterns fail (Jonsson *et al*., 2005).

The resulting data matrix (created from the normalised data) has the form:

,

where

- 1,..., n cases. For purposes of this repeatability study the cases represent exactly the same repeat urine samples.
- 1,..., p features; i.e. metabolites like succinic, hippuric etc., expressed as mmol per mole creatinine. (NOTE: data normalization relative to the IS resulted in the IS being a constant value and was thus removed from our repeatability assessment).

4.4.2 Statistical methods

In this section a description is presented on the statistical methods used to measure repeatability. These were two methods, of which the first is based on a more traditional methodology using the coefficient of variation (CV) and then introduce a new qualitative method based on kernel density estimates, developed by Dr. Gerhard Koekemoer of the Statistical Consultation services of the NWU, and co-study leader for this dissertation.[5]

Coefficient of variation

The coefficient of variation is defined as: $CV = s/\bar{x} * 100$, where $\bar{x}$ is the sample mean and *s* is the sample standard deviation. The CV is commonly used to quantify the measurement error; however, this value tends to become large for features with small concentrations. Van Batenburg *et al.* (2011) also suggested that the measurement error is not constant as its value depends on the concentration level of the feature and that measurement errors may be correlated between features. They defined two figures of merit, namely (a) the measurement error distribution based on a model proposed by

[5] As the description of the statistical methods, as well as the algorithms used for their implementation falls, in the field of bioinformatics, the technical aspects of this section were largely done by Dr. Koekemoer.

Rocke and Lorenzato *et al.* (1995) and (b) the measurement error correlation. The model requires injection of blanks in the analytical phase, and according to Rocke *et al.* (1995) still produces a CV that trends to infinity as the concentrations approach zero. In addition, the measurement error calculated approximates the relative standard deviation (RSD), where RSD = CV/100 (van Batenburg *et al.,* 2011).

Hence the CV for each feature (variable) was firstly calculated, after which these values were used to construct the distribution of the CV. Dunn *et.al*. (2011) reported that the FDA (United States of America, Food and Drug Administration) (FDA, 2001) allows a CV of 30% when the objective is biomarker discovery, but stressed that the researcher should aim at the best achievable repeatability.

Qualitative measurement error model

Van Batenburg *et al.* (2011) noted that a quantitative figure of merit to express repeatability could not be derived for multivariate analyses. Here we therefore present a new **qualitative method** that is not affected by small concentrations. The method proposed involves density estimation which is defined next. Let $X_1,....X_n$ be data that is independently and identically distributed (*i.i.d*) with an unknown density function *f*. A kernel density estimator for *f*(*x*) is given by

$$\hat{f}_h(x) = \sum_{i-1}^{n} K((x - X_i)/h)/(nh)$$

Where $K(\bullet)$ is the so-called kernel function and *h* is the bandwidth or window width, Silverman *et al*. (2005). In the current application we used the standard normal density function as kernel and the method proposed by Sheather and Jones (1991) to select a data-driven bandwidth.

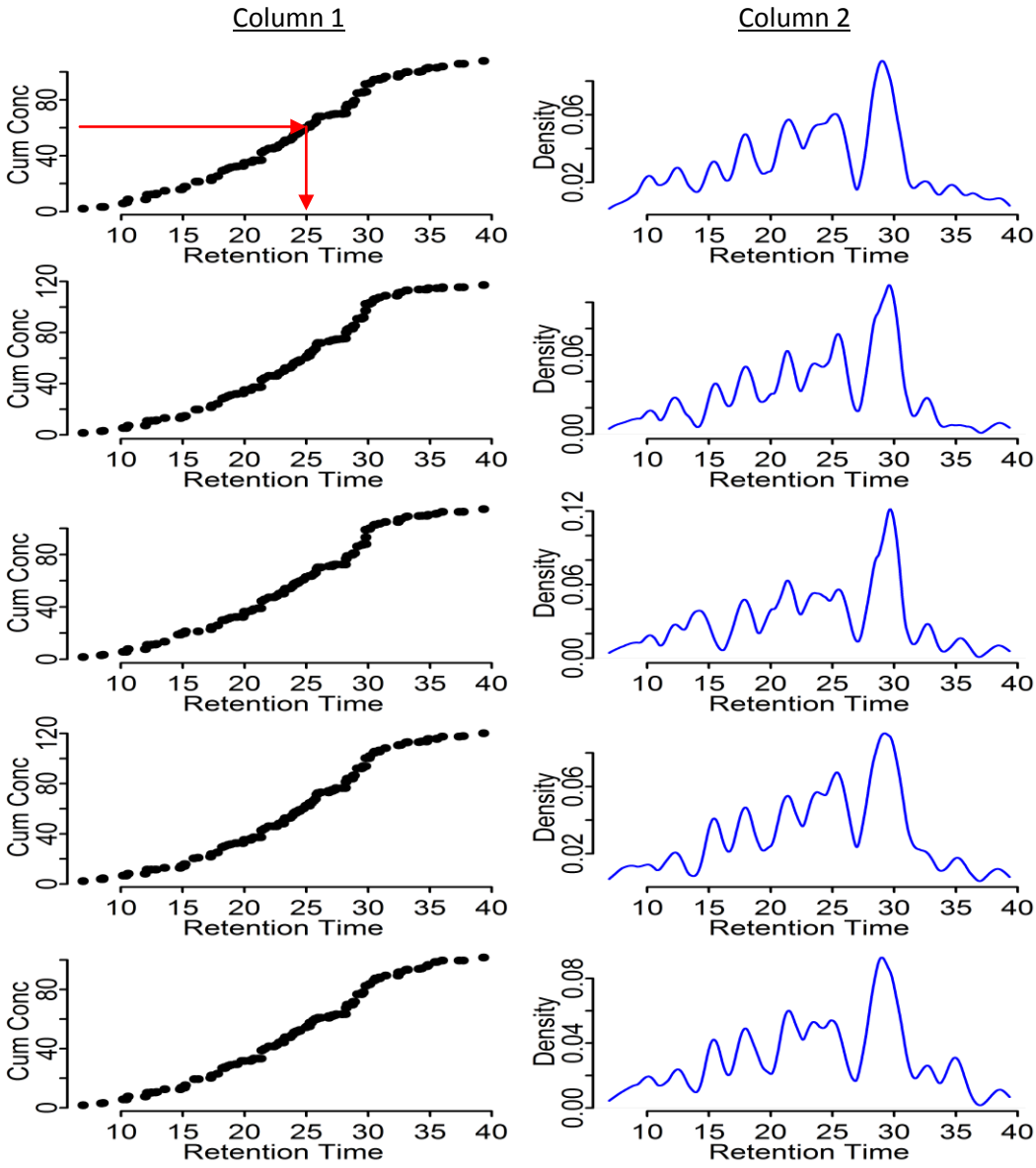


Figure 4.1 Retention time versus cumulative concentration scatter plots (column one). Density estimates for retention time (column two).

The qualitative method takes as input both the retention time and the log-scaled normalized concentrations. The log-scaled concentrations of each sample are then accumulated in order of increasing retention time. These accumulations are performed for each sample individually. Consider the first column of Figure 4.1 as an example where the cumulative concentrations vs. the retention times are plotted for five samples. If the experiment is repeatable, then the cumulative concentrations should all be similar when overlapped. Since these graphs are cumulative it is visually simpler to identify non-

repeatability when we consider density estimates. To accomplish this we used the following procedure;

- We set up a finely spread grid from the smallest concentration to the total accumulated concentration for each sample. As example, consider Figure 4.1 where this grid will be constructed on the y-axis of each sample.

- We then, interpolate retention time values using the finely spaced grid as input. This is illustrated visually (for one grid point) using red arrows in the first row, column one of Figure 4.1. The result is that we map the bivariate data (i.e., retention time and concentration) to univariate data (i.e., retention time), in such a way that a higher density of retention times can be excepted in areas where large concentrations are observed, and a lower density of retention times in areas with less machine activity, or where lower concentrations were recorded.

- Kernel density estimates as defined earlier in this section are then applied to the interpolated retention times of each sample. This is illustrated in the second column of Figure 4.1

- For a repeatable experiment an overlay of these density estimates should show little variance.

4.5 Results

The repeatability experiment as described in Section 4.4.1 was repeated twice (see Note 2 in Chapter 4.4.1). The second repeat was performed under much more stringent conditions, as the results indicate that the first repeatability experiment was not satisfactory. In this section results are presented from both the first and second experiment concurrently. In addition, a data reduction filter was applied to separate the data into 'data used' and 'data not used' categories. The 'data not used' category was used to classify and isolate the less

important data (i.e. peaks/features/compounds considered as electronically generated noise, contaminants from carry-over of previous analyses, artifactual peaks, etc.). The 'data used' category thus contains the data classified as more important/consistent, and will be the data used in the analytical assessment of repeatability. The criterion for retaining data into the 'data used' category is variables with a retention time CV of less than 10%.

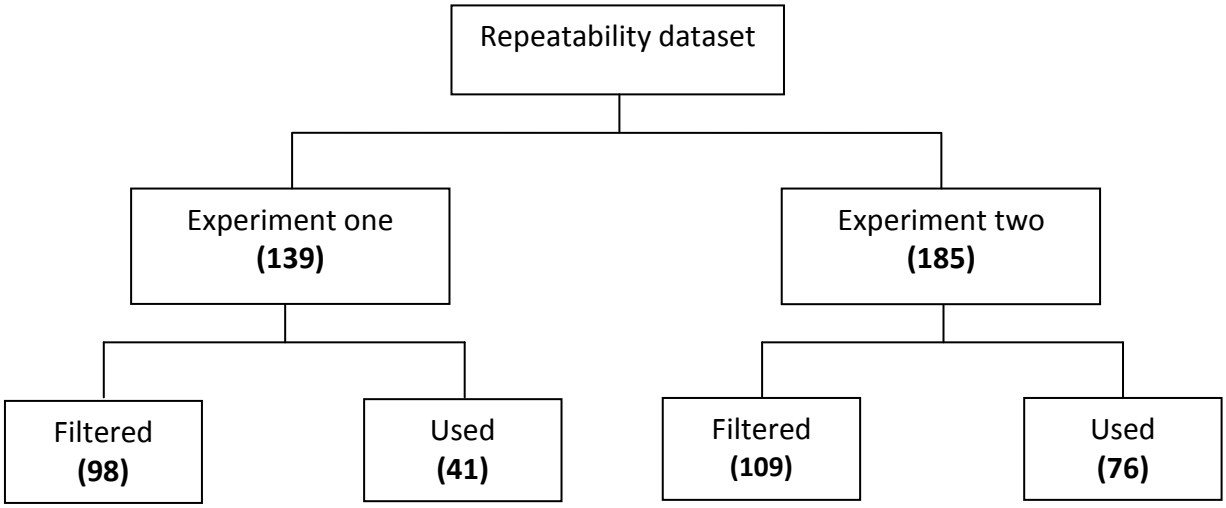


Figure 4.2 Data reduction with number of features given in brackets.

The number of features that was analysed in both experiments is shown in Figure 4.2. The data (using all the features extracted) is presented graphically in Figure 4.3 using a 3D scatter plot with droplines.

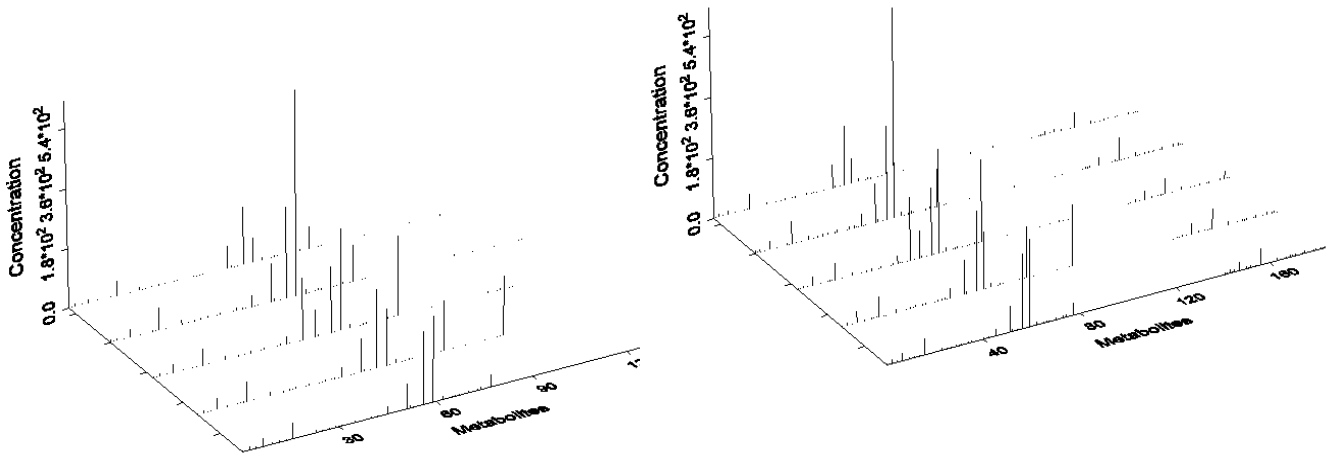


Figure 4.3 Data matrix represented graphically through 3D scatter plots with dropline: Left panel, data for repeatability experiment one and the right panel are the data from experiment two.

Next, the results for the first and second experiment are presented, using the features quantified (see Figure 4.2). Table 4.1 shows the results obtained from the first (row one) and second (row two) repeatability experiment, using 139 and 185 features for experiment one and two, respectively. Row two of Table 4.1 includes results of the second repeatability experiment which were performed under much stringent conditions. These conditions (see Fiehn *et al.* 2007), included

- Cleanliness of the glassware used for the experiment which was self-verified to ensure that the soap was phosphate free and that the glassware was properly cleaned, rinsed and dried.
- Fresh reagents were prepared for conducting the experiment such as the internal standard (4-phenylbutyric acid).
- The nozzle heads of the nitrogen gas drier were rinsed with hexane for removals of any particles from the previous usage.
- The column of the GC-MS machine was checked that it was not dirty.

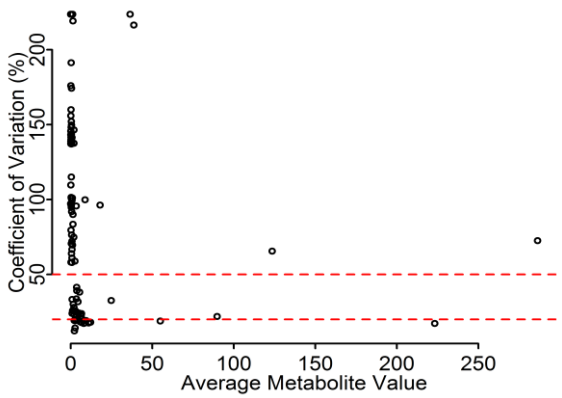
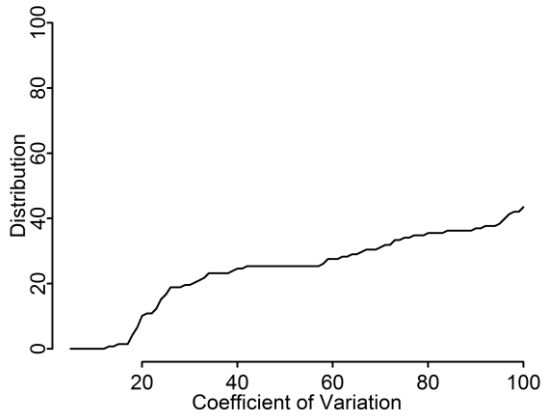
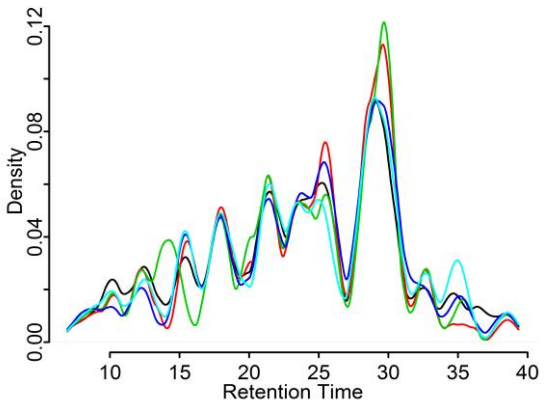
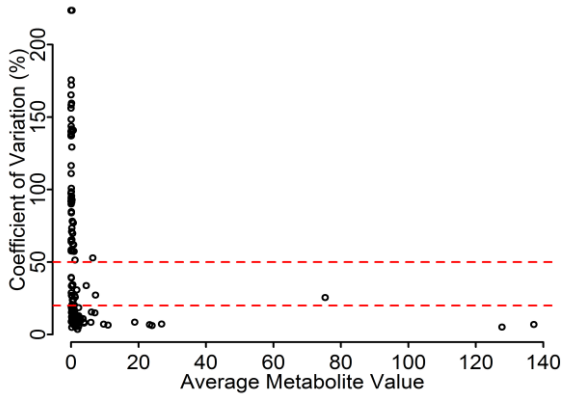
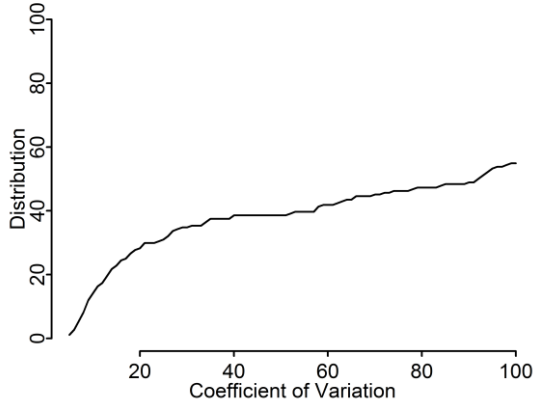
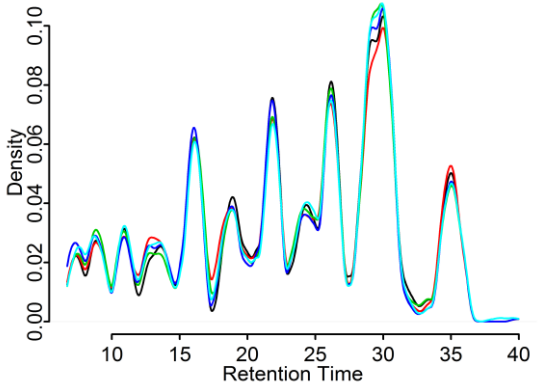
In column A of Table 4.1 the average metabolite value is plotted against the coefficient of variation for both experiment one and two. The dotted red line placed at 20% and 50% can be used as a reference line. It is clear that a large number of features with low concentrations is measured, hence; large CV values are observed. The qualitative method is not influenced by these small concentrations.

Column B consists of graphs of the empirical function coefficient of variation (CV) for both experiments (row 1 and row 2). The CVs for all the variables were used as input. From the CV distributions, (see Table 4.1, column B) it is concluded that experiment two seems to contain more repeatable features at a given CV level, than experiment one.

Column C encompasses the kernel density estimates which are our qualitative results of repeatability. The overlapping of these estimates should show minimal variation when overlapped if the experiment is repeatable; hence, these estimates can be used to visually inspect repeatability. In column C row one the kernel density estimate results of experiment one when overlapped do not visually appear to overlap. This suggests that experiment one was not repeatable. However, the density estimate results of row two column C converge to almost one line.

The results where data reduction was applied to the repeatability data sets of both experiment one and two are summarized in Table 4.2. As shown in Figure 4.2 after the filtering of the data the resulting data matrices of experiment one contained 41 variables from 139 and experiment two contained 76 variables from 185 variables. The CVs and distribution were calculated and the use of qualitative density estimation was applied to assess repeatability.

Table 4.1 Repeatability results. CV distribution, CV plots and density methods

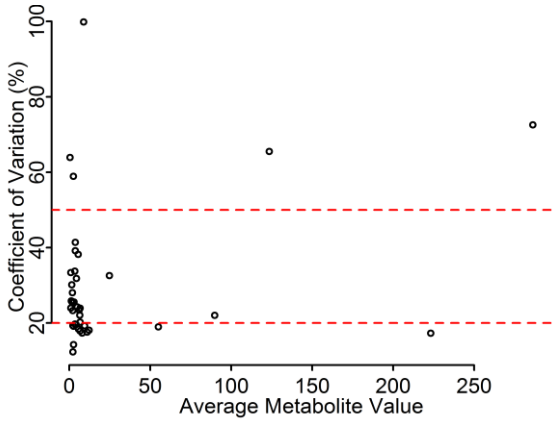
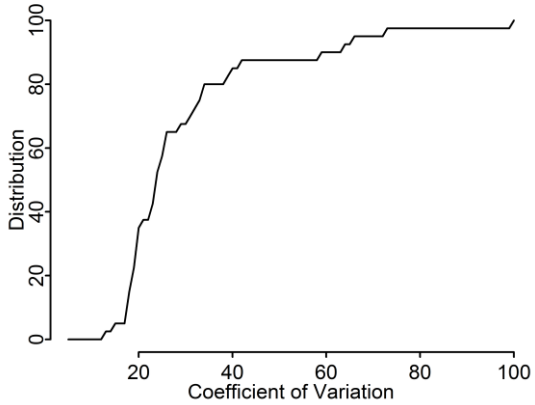
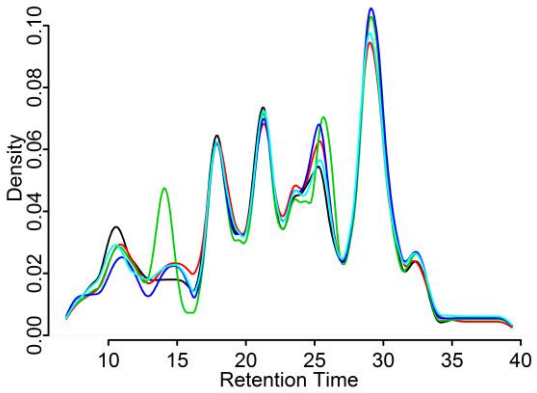
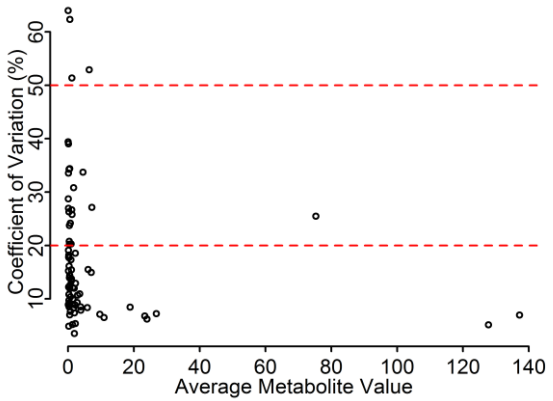
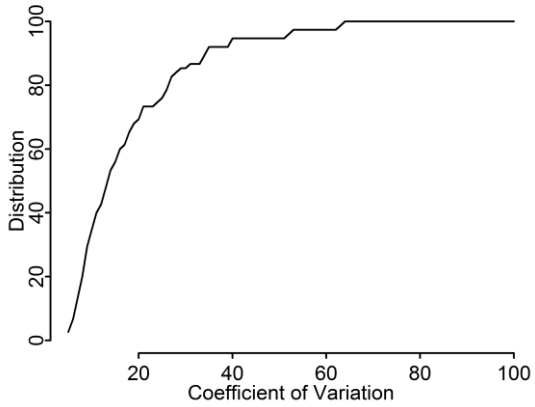
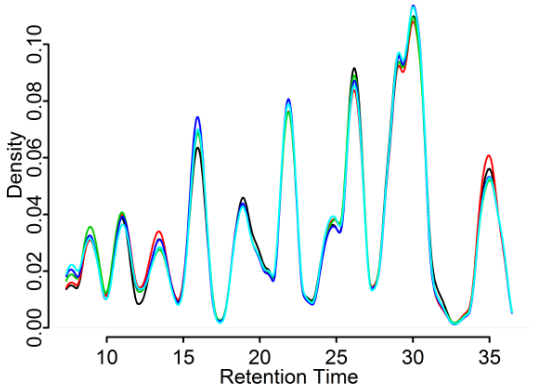
Experiment	Column A: CV distribution	Column B: CV plots	Column C: Density methods
1			
2			

From the CV scatter plots in column A, it should be clear that a large number of metabolites with small concentrations are still measured. However, when compared to column A of Table 4.2 it is evident that most of a large number of the small concentration features with large CV values have been filtered out. From the CV distributions, a clear improvement is also noticeable if only features with a repeatable retention time are considered. Thus, although the number of features considered is smaller, more of these variables are repeatable relative to the total number of features considered, as the filtering scheme proposed seems to filter out most of the non-repeatable features.

It is also important to note that the results of both the CV scatter plots and the CV distributions indicate that the second experiment produced more reliable features. Lastly, the qualitative density method presented in column C indicated plausible repeatability when only 41 features with repeatable retention times were considered, with the exception of certain features measured with retention times ranging from 10 - 16 minutes. However, the 76 features retained from experiment two indicate repeatability with very little variance in the density estimates.

From these results it is concluded that if all features are considered, only the results from experiment two signify repeatability. However, when the features were filtered the remaining 41 features from experiment one also suggest to be repeatable.

Table 4.2 Repeatability results with filtered data sets. CV distributions, CV plots and density methods.

Experiment	Column A: CV distribution	Column B: CV plots	Column C: Density methods
1			
2			

4.6 Discussion

Many factors are involved in conducting a high-quality metabolomics study, the most important of which is a good experimental design and execution. A typical experiment involves collecting biological material, extracting and then measuring potentially hundreds of metabolites simultaneously by GC-MS, spectral processing and multivariate statistical analysis (Goodacre *et al*., 2007). The principal objective is usually to discover metabolites that differentiate between two or more classes of samples, being the bilharziasis positive (patients) and negative (controls) in this investigation. Consequently, it is essential to understand and minimise the spectrum-wide metabolic variation arising from technical sources, as well as inter-individual metabolic variation within each class. Without this, the interpretation of results can be unclear, perplexing and, in the worst case, false. This repeatability chapter as presented here was to establish if the analyst is able to comprehend the analytical basis needed for a good experimental study, that the hands-on ability to actually do the experimental work is up to standard, and that the analyst acquired the ability to handle the software to convert the raw data into biological data (the metabolites under investigation). In this regard, Lu *et al* (2008) asks the vexed question: "How does the analyst pick out the correct deconvoluted spectra from those results without previous knowledge, even though the spectra of metabolites are detected and deconvoluted correctly?" When comparing the trial experiment Table 4.1 experiment one to Table 4.1 experiment two, the results suggests that the analyst acquired the technical skills for repeatability (Aim 2 of this investigation). In the subsequent Chapter the analysis of the bilharziasis samples will thus be presented, including the use of the quality control samples in the context of repeatability.

CHAPTER 5

METABOLITE PROFILE DUE TO *S. HAEMATOBIUM* INFECTION IN HUMANS

5.1 Introduction

The ultimate objective of a metabolomics experimental study is to quantify all of the metabolites in a cellular system, with the intended aim being to understand phenotypic variation, to be able to assemble comprehensive data and models of the cellular organization along with their biochemical function (van der Greef *et al*., 2004). Currently this is unattainable, given the complexity and diversity of biological systems, the lack of simple automated analytical strategies and challenges, such as the chemical complexity and heterogeneity of the metabolites, the dynamic range of the measuring techniques, the throughput of the measurements and the validation of the extraction protocols (Harrigan *et al*., 2003). Given this, van der Greef *et al*. (1986) deemed it of importance to apply appropriate data pre-processing and proper data normalization techniques which are crucial for metabolic profiling and pattern recognition. With this said it should be noted that metabolites differ strikingly in their physical properties and concentrations, with molecules ranging from very polar to very apolar, and with molecular weights ranging from 50D to 1400D, including lipids to sugars and peptides (Milburn, 2006). Given the complexity and diversity of the metabolome a single method or technology wouldn't be able to provide the performance necessary for a comprehensive metabolomic study. However, metabolites are chemical entities and can be analysed by the standard tools of chemical analysis such as molecular spectroscopy and MS (Harrigan *et al*., 2003). The resolution, sensitivity and selectivity of these technologies can be enhanced or modified by coupling them to a gas chromatograph, such as we used in the research for this dissertation (see Chapter 3, Section 3.5.2.2). We thus now address the third aim defined for the present investigation: "To determine the metabolic profile of organic acids from humans infected by *S. haematobium* and give a possible biological/pathophysiological interpretation of their profile in comparison with 6 other models of bilharziasis" (see Chapter 2.7).

5.1.1 Disease biomarkers

The region of importance in metabolomics research which holds great potential is in the discovery of biomarkers (Harrigan *et al*., 2003). Metabolomic investigations that are fashioned to discover differences between a group of individuals from their matched controls are able to obtain metabolic biomarkers associated with diseases and/or drug induced changes (Milburn, 2006). Ultimately one of the objectives of metabolomic research is to be able to specify metabolites that can discriminate classes of samples obtained from different cellular conditions i.e. from patients and controls, by being absent, present, or differentially present either in high or low concentrations. By description, a biomarker, or biological marker, is a substance that can be used as an indicator of a biological state. It is a characteristic used in various scientific fields that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. Harrigan *et al*. (2003) cautions that the identification and use of biomarkers should be practiced in a way that acknowledges that a biomarker discovered in early stages of a disease is most likely going to differ from the biomarker found in the later stages of disease progression. This gives an opportunity to use metabolic biomarker discovery to investigate both the prognostic and the diagnostic aspects of a disease (Harrigan *et al*., 2003). This approach is recognized by the actuality that biochemicals have already been used for diagnostic purposes and are still being routinely used in modern medicine e.g. for glucose in diabetes and cholesterol in cardio-vascular diseases (Milburn, 2006). Furthermore, biochemistry being a well established field in science allows for the observed changes characterised by altered, or stressed physiological states discovered in metabolites by metabolomics to be comprehended.

5.1.2 Metabolite detection

Metabolomic approaches should take into consideration that metabolite identification can become a complex procedure in metabolomics and does not offer 100% coverage (Dunn *et al*., 2001). This is owing to the fact that the human metabolome has not been entirely characterized and identified and the libraries and databases of experimental data applied

for the identification of the human metabolome have not yet been completed to reflect all known metabolites. In addition, the human metabolome is a collection of endogenous metabolites, those produced by the organism like metabolic end products, and exogenous metabolites i.e. drugs and food components along with metabolites from the gut micro flora. Furthermore, there currently are chromatographic peaks that are not identified in metabolomics data sets for chromatography MS acquired data (Want *et al*., 2006).

5.2 Methods

An organic acid analysis protocol was used to analyse the collected bilharziasis urine samples. The detailed protocol is described in Chapter 3, Section 3.5.2.

5.2.1 Samples

Samples were collected from the NHLS clinics in Limpopo which contained 73 bilharziasis positive samples and 49 bilharziasis negative samples along with 24 control samples collected from the Centre of human metabolomics at Potchefstroom Laboratory for Inborn Errors of Metabolism (PLIEM) (For further detail, see Chapter 3.5.2). Aliquots of the samples were stored at -80°C for long term storage till the time of analysis. In this chapter we annotated the bilharziasis positive samples by BP, the bilharziasis negative samples by BN, and the NWU control samples by MC.

5.2.1.1 QC samples

A common observation in a GC-MS metabolic analysis has been that for the first few injections of the sample, unrepresentative results have been recorded (Gika *et al.,* 2007). This is mainly because of small changes in the chromatographic retention time or the signal intensity of the apparatus, commonly known as a machine drift (Zelena, *et al*., 2009). To counteract for this it is good practice to run at least five pooled quality control (QC) samples at the beginning of the run and use the data derived from these samples to demonstrate the systems' analytical suitability. Therefore, the QC samples' pre-run serves as a repeatability

quality assessment on the machine. For this investigation the QC samples were prepared by mixing equal volumes from each of the bilharziasis positive (BP), bilharziasis negative (BN) and NWU metabolic laboratory controls (MC) (see Chapter 3.5.2). This pooled urine was then further used as a representative "standard" sample containing all the analytes that could potentially be encountered during the analysis. A total of 90 QC samples was prepared for quality control purposes. However, one QC sample was excluded due to handling error; hence, only 89 samples were used.

5.2.1.2 Batch composition and run order

Each of the samples were analysed by GC-MS in an untargeted analysis, but not all of the samples were run or could be run in a single analytical batch because of issues ranging from instrument availability to routine maintenance procedures. The experiment was a construction of 15 sample batches and each batch consisted of 16 samples that contained: a blank (a sample vial containing only hexane and no sample contents), 6 QC samples, 5 BN samples and 5 BP samples, randomly ordered as outlined below. The BN and BP samples were subjected to a designed run order since the samples were run in a time sequence, sometimes experiencing gradual changes in the instrument's sensitivity over time. This might produce unreliable results due to contamination from remnants on the MS column. To rectify the problem these samples could be carefully randomized so that all the experimental groups would be affected to the same extent. This meant designing a run order that was orthogonal to the samples to minimise machine bias, as outlined by Gika *et al*. (2007), Want *et al*. (2010) and Dunn *et al*. (2011). In this experiment our designed run order was the following:

,

With [B] being the blank which is a sample vial that contained only hexane and no sample contents, [Q] was a randomly selected quality control sample, [C] was the control sample which can either be a bilharziasis negative or a NWU metabolic laboratory sample (allocated to the run order at random), and [P] was a randomly selected bilharziasis positive patient sample. This run order represented a single batch and including the blank it consisted of 17 vials which ran on the GC-MS for 17h. Each sample thus ran for 60min on the GC-MS. Blank

runs were carried out at the beginning of each run to minimise and evaluate an eventual chromatographic carry over effect.

The QC samples in each batch were also to be used to filter out features, prior to the bilharziasis data analysis. The repeatability of the selected features using the QC samples was then investigated using the methods as described in Chapter 4 (see Section 4.4.3), and a principal component analysis done, which is a multivariate method described in Section 5.2.2.

5.2.2 Statistical methods

Metabolomic investigations are notorious for generating large amounts of raw data, of which the challenge of handling, processing and analysing this data can become apparently clear and time consuming. Therefore researchers require the implementation of specialized arithmetical, statistical and bioinformatic skills (Mehrotra *et al*., 2006). Metabolomics has unique bioinformatics needs in addition to others common in microarray or proteomics data. In this section we describe the statistical methods used to analyse the quality of the QC samples (Section 5.2.2.1), and the methods to analyse the data of the bilharziasis experiment (Section 5.2.2.2).

5.2.2.1 Repeatability methods

The QC samples were used to test the repeatability of the data generated and to filter out features to be excluded from the bilharziasis study. The repeatability method presented is the qualitative measurement of error model as described in more detail in Chapter 4 (see Section 4.4.3).

5.2.2.2 Statistical methods for the bilharziasis study

The metabolomics data generated contained a large number of zeros, which complicates the statistical analysis. These zeros have been replaced by randomly generated values from a Beta (0.1, 1) distribution that is bounded between zero and the detection limit, prior to data normalization. These generated random numbers are extremely small, and do not alter the covariance structure amongst the features. Hence, a different set of randomly values does not alter the statistical results and conclusions reached.

To ensure that all features contribute to the multivariate analysis, data transformation is required since the scales of the features measured differ drastically. The following transformations were applied to the data prior to analysis: The shifted-log transformation, a non-parametric transformation function as described by Koekemoer *et al*., (2008) and two transformations in which the data were replaced by ranks. The first ranking method replaced any tied ranked values with the average rank, whilst the second method replaced tied ranked values with the minimum rank. It should be noted that data transformation has a large effect on the eventual data analysis. Thus, based on the group separations obtained through these different transformation methods, as well as the biological interpretation of the respective outcomes, we selected the output of the rank-replacement transformation (tied ranks replaced with minimum rank), to present in this thesis. These ranked features were then scaled according to the formula $Y = X / \sqrt{\sum X^2/(n-1)}$, where X refers to any one of the features used and *n* is the number of samples.

Following this scaling, the data were also been centred. Centering converts all the concentrations to fluctuate around zero instead of around the mean of the metabolite concentrations. It is, therefore, used to focus on the fluctuating parts of the data (Jackson *et al*., 1991; Bro *et al*., 2003) and leaves only the relevant variation (being the variation between the samples) for analysis. Subsequently, the univariate and multivariate statistical methods (illustrated in Figure 5.1) will be described.

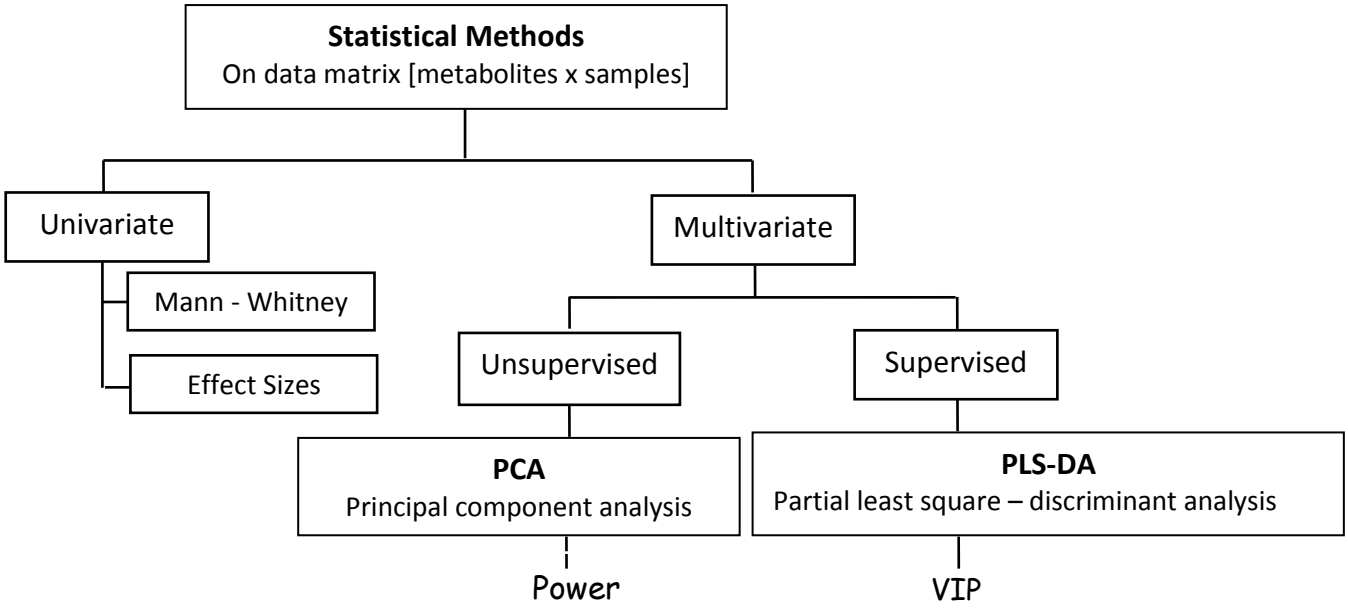


Figure 5.1 Work-flow of statistical methods

The data was subsequently analysed using univariate as well as multivariate techniques. The univariate methods involve, the non-parametric Mann-Whitney test (based on the untransformed data) and the calculation of effect sizes (based on the transformed data). The Student's *t*-test and Mann-Whitney test can be used to identify statistical differences between samples of distinct classes (Wiener *et al*., 2004), whereas the effect sizes are used to ascertain the practical significance of a statically significant result. Note that the student t-test based on the original data is not reported in this study, since normality assumptions could not be confirmed. The Cohen (1988) effect size used in this thesis is given by $d = \left| \overline{X}_1 - \overline{X}_2 \right| / S_{\max}$, where $\overline{X}_1$ and $\overline{X}_2$ is the group means and $S_{\max}$ is the maximum of the group standard deviations. The value of d should be interpreted as follows:

$d < 0.2$: small effect; $d \approx 0.5$: medium effect; $d > 0.8$:large effect.

In this thesis the effect sizes based on the transformed data have been used as a class discrepancy measure and utilised to identify potential markers. These univariate procedures are easily understandable but their applicability when used in their own is rather limited when dealing with thousands of highly correlated variables. For this reason the use of multivariate methods are also incorporated and discussed next.

The objective of the multivariate methods used in this thesis, is to investigate the existence of a natural separation amongst the experimental groups, and then to find a model that can quantify this separation. These methods are therefore usually divided into two groups namely, unsupervised and supervised. Since a large number of variables are measured in metabolomics studies, multivariate methods involving dimension reduction are commonly used.

Unsupervised methods attempt to analyse a set of observations without calculating or examining any associated outcome. Unsupervised data analysis uses procedures that attempt to find the natural separation of patterns to facilitate the understanding of the relationship between the samples, without using any prior knowledge to guide the analysis, and to highlight the variables responsible for these relationships. One such an unsupervised pattern recognition method frequently used in metabolomics experiments is principal component analysis (PCA). PCA is primarily concerned with the transformation of a large set

of related variables into a new, smaller set of uncorrelated variables (Joliffe, 1986). The new variables are termed latent variables, or principal components (PCs). The PCs attempt to express the maximum variation of the original data. Each principal component may be thought of as an axis in multidimensional space, and each object can then be characterised by how far away it lies to a particular axis, this coordinate is often called the object or the sample score. The principal components will be presented in terms of score-plots, which illustrate the score that a sample receives on the latent variables (i.e. the principal components). The percentage of variance explained by the principal component model will also be presented as a summary fit-statistic. Lastly, identification of important variables was performed by calculating the modelling power of each variable. According to (Brereton 2003) a modelling power of greater than 0.5 signifies an important variable in the PCA projection.

Supervised data analysis considers each object with respect to an observed response and includes regression and classification problems depending on the output type under consideration, i.e. a numerical value in the first case and class label in the second (Wold *et al*., 1984). In this thesis we consider the classification problem and use partial least squares discriminant analysis (PLS-DA) as statistical methods to build a discriminant model. The variables important in projection (VIPs) in the PLS-DA were calculated and used to rank the features in order of discrimination contribution. Features with a VIP value greater than one (Chong and Jun, 2005), were selected as potential markers. Supervised multivariate analysis tends to over-fit the data. Therefore, to assess the quality of the fitted model some form of validation is required (Westerhuis *et al*., 2008). Fit statistics of the PLS-DA models were reported as the percentage variance explained for the metabolites (R^2X), the percentage variance for the group membership of the patients and controls (R^2Y), and the predictive R^2Y values (commonly called Q^2). The Q^2 value gives an indication whether the model over-fitted. A value closer to 1 is preferred for these statistics. In addition, the specificity and sensitivity estimates using only the final marker list of the PLS-DA approach were evaluated by cross-validation. For this, a data set was constructed which included only the important metabolites that were identified. Next, a PLS-DA model was constructed for this data set and an appropriate cut-off point was determined by calculating the Youdin index (Fluss *et al.*, 2005). Then, we let $\mathbf{P_{BN}}$ and $\mathbf{P_{BP}}$ be the observed occurrence probabilities of a

bilharziasis negative and a bilharziasis positive patient and let α be the fraction of cases to be removed in the cross-validation. Subsequently, 10 000 unique stratified samples of size n_{CV} = [α·n], with n = total number of negatives and patients, were selected from the data, stratified according to the observed occurrence probabilities, that is, $n_{CV} = n_C + n_P$, where n_C and n_P are the sample sizes for the withheld controls and patients, respectively. For each of the 10 000 samples the n_{CV} cases were withheld, a PLS-DA model was built using the remaining cases and the group membership of the withheld cases was predicted. For this, the sensitivity and specificity, as well as the percentage of misclassified cases were recorded. Lastly, the standard deviation and the average of the recorded information were calculated over the 10 000 samples and reported as cross-validated estimates of sensitivity, specificity and percentage of misclassifications, as well as the respective values for α and the cut-off point.

All statistical analysis was performed using the R statistical package version 2.13.0.

5.3 Results

A detailed schematic work flow-diagram of the design, organic acid analysis and statistical analysis of the bilharziasis study is presented below in Figure 5.2.

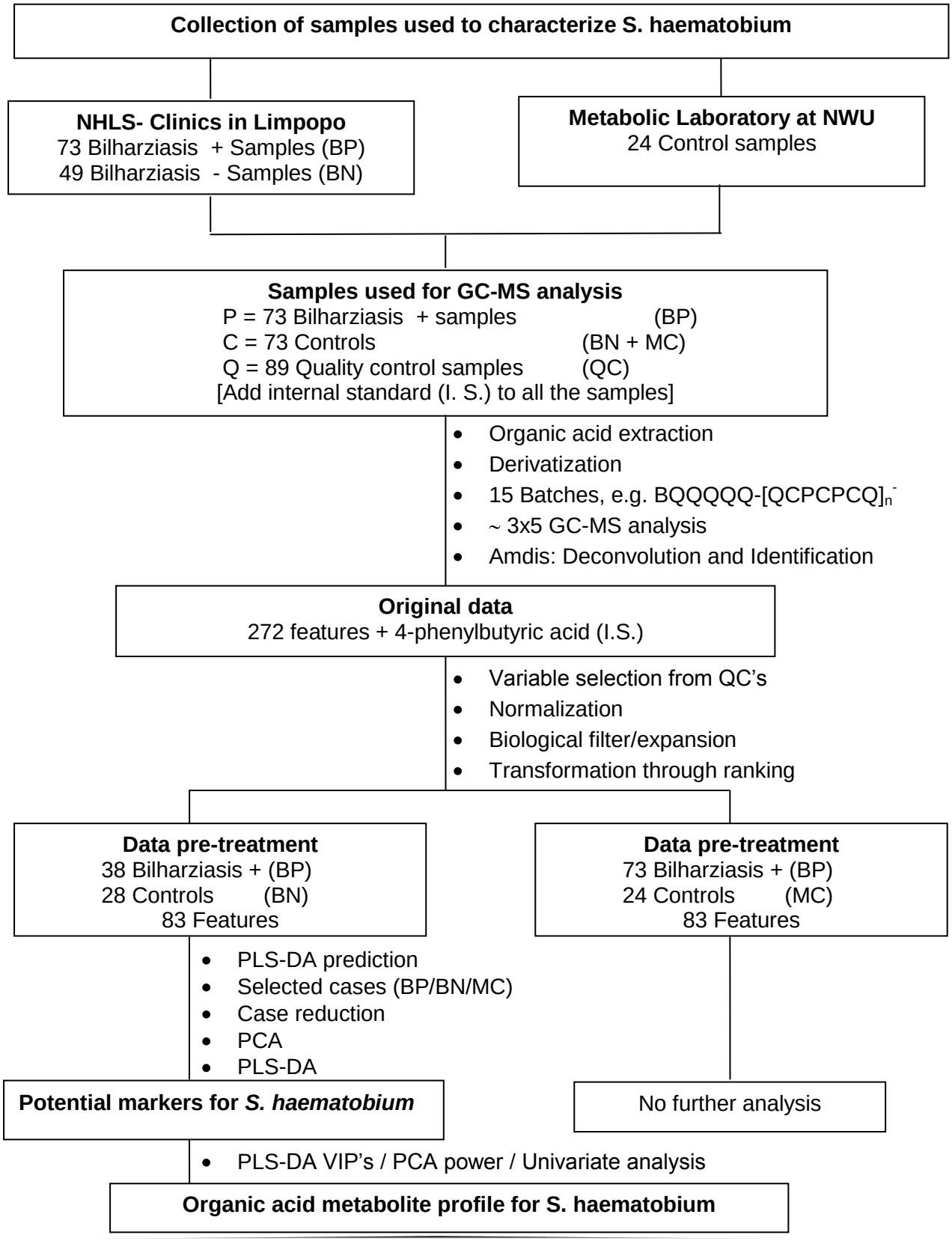


Figure 5.2 Metabolomics investigation work-flow.

Following the organic acid extraction methods (see Section 3.5.2), and the run order design as described in Section 5.2.1.2, a data matrix was compiled consisting of 73 bilharziasis positive, 49 bilharziasis negative and 24 NWU metabolic controls. The biographical information of these samples is described in Chapter 3, Section 3.4.2. A total of 273 (including 4-phenyl butyric acid as an internal standard) features were measured. An additional data matrix consisting of the 89 QC samples were also constructed for quality control purposes. All data matrices were normalised using the internal standard through the regression approach (Redestig *et al*., 2009; Sysi-Aho *et al*., 2007). The results in this section are divided into two parts; first an analysis of the 89 QC samples are presented to illustrate feature selection and to verify the quality of the selected features. Secondly, the analysis of the bilharziasis experiment based on these selected features.

5.3.1 Quality assurance

The QC data matrix was used for two purposes namely, to reduce the number of features investigated and to inspect the quality of the final list of variables that will be analysed. The reduction of variables is summarised in Figure 5.3 and was performed as follows. For each of the 273 initial features the percentage of samples containing the specific feature was calculated.

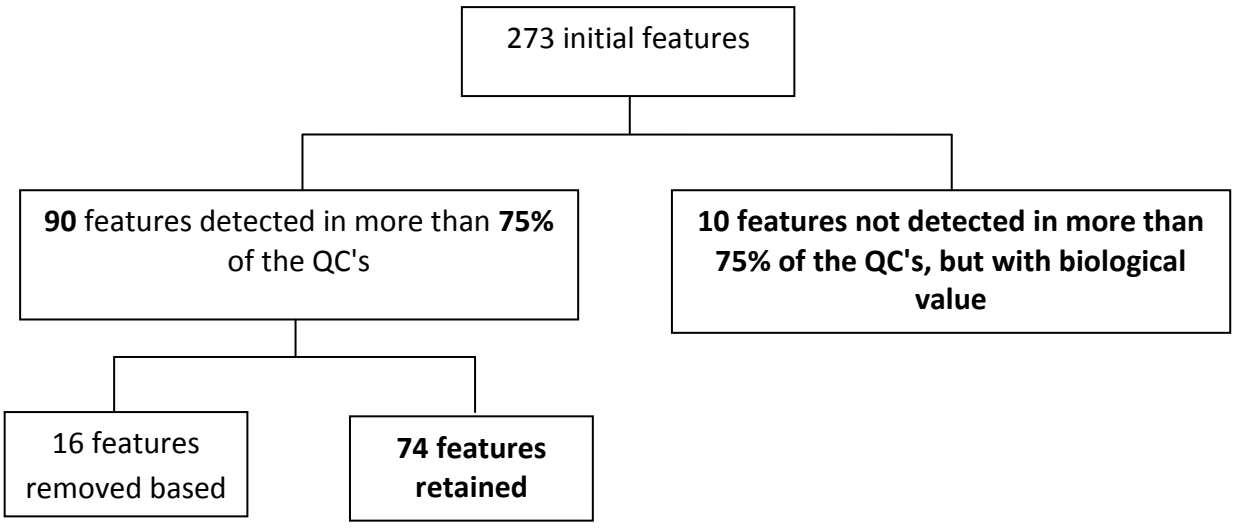


Figure 5.3 Feature selection diagram.

These percentages were used to reduce the number of features, by removing those features detected in less than 25% of the QC samples. This feature reduction step was adapted from Dunn *et al*., (2011), who proposed to remove all features detected in less than 50% of the QC samples. The number of features was reduced from 273 to 90 using this approach. These features, now designated as metabolites, were then further reduced to 74 based on biological interpretation. In addition, 10 features, removed due to the 25% rule, and thus not included in the 74 were then added back, based on their biological importance in literature on bilharziasis infection. Hence, the final data matrix used in the bilharziasis experiment consisted of 84 features (including internal standard).

The quality (repeatability) of the 83 (Internal Standard removed) selected features were subsequently investigated, using the qualitative method as described in Chapter 4 (see Section 4.4.3) applied to the QC samples. The log transformation had been applied to the data prior to analysis. The resulting density estimates are reported for each batch and are presented in Figure 5.4, where the batch number is given in the top right corner of each cell in the Figure. Analogous to Section 4.4.3 a filter has been applied to reduce the number of features based on the consistency of the retention time. In the first three columns of Figure 5.4 the density estimate results are presented for a 20 % filter, i.e., only 47 of the 83 features were used to construct the density estimates as their retention times had a coefficient of variation of less than 20%. From this analysis, we conclude that the intra-batch variation is acceptable (small variation amongst density estimates) for most batches with the exception of batch 15. In addition, the density peaks are consistent amongst the different batches.

The analysis of the bilharziasis data will be performed using the 83 selected features; hence, we also present the non-filtered (83 features) repeatability results in columns 4, 5 and 6 of Figure 5.4. From these graphs it is clear that the density peaks amongst the batches are again consistent, suggesting that the experiment was successfully reproduced for each batch. However, the intra-batch variation of the density estimates was larger when compared to the results where the 20% filter was applied. It should be noted that this data set contained the features added back, based upon their importance in bilharziasis literature. These added features might be responsible for the larger intra-batch variation

observed. In general we consider the quality assessment of the QC samples as satisfactory and subsequently present the results of the bilharziasis study.

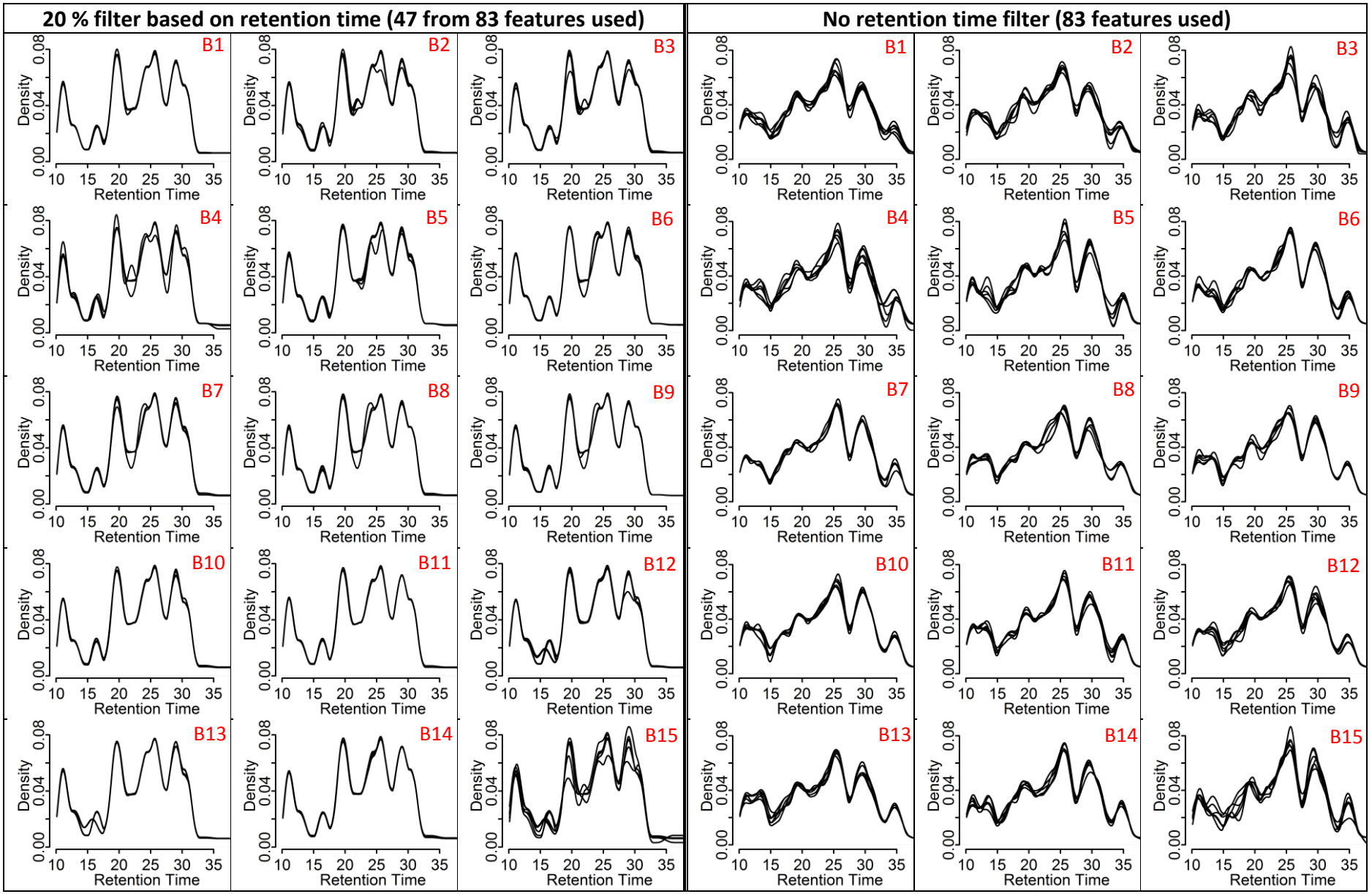


Figure 5.4 Qualitative quality assessment of the QC samples

5.3.2 Further evaluation of the data generated in the bilharziasis experiment

This section is divided into three parts. First we investigate whether a batch effect is present in the data. Secondly, selection of a control group is discussed and lastly, the results of the analysis concerning the bilharziasis experiment are presented.

The reduced normalized data set consisting of the bilharziasis cases and the 83 selected features were transformed using the rank transformation as described in Section 5.2.2. This transformation was applied in order to make the scales more comparable, and therefore to reveal the biological information encapsulated in the data.

5.3.2.1Batch effect

Prior to commencing, we investigate whether a batch effect exists for the bilharziasis data. Given the large number of batches analysed (15) and the importance of the bilharziasis positive group, we only present the batch effect results using this group. Figure 5.5 contains a 3D (left column) and a 2D (right column) PCA score plot, where the batch number has been used as a labelling character.

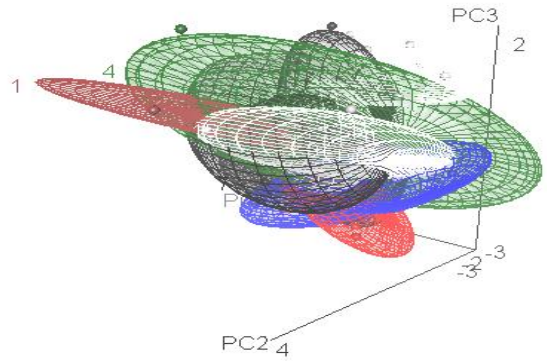
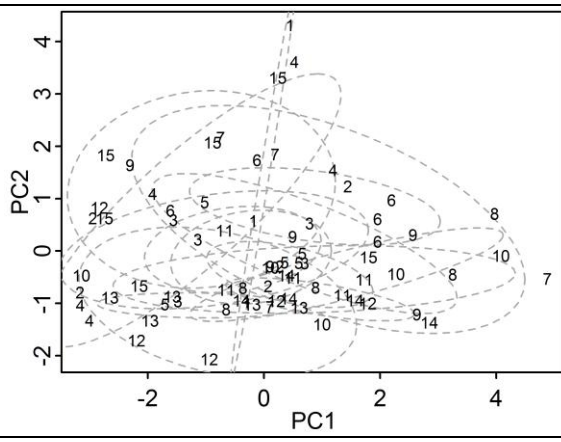
PCA 3D	PCA 2D
	

Figure 5.5 PCA scores plots investigating batch effect.

It is clear from Figure 5.5 that there is no batch effect present in the data since a complete overlap of all the batches is observed. We now precede with the analysis of the bilharziasis data by firstly selecting appropriate controls.

5.3.2.2 Selection of the control group

A PCA was performed using the NWU controls (MC) and the bilharziasis positive (BP) samples. The two and three dimensional scores plots of this analysis are presented in the first row of Figure 5.6. From these score plots we observe no separation between these two groups, and further investigate whether the bilharziasis negative (BN) controls would differ from the NWU controls (MC). The score plots of this PCA analysis are presented in the second row of Figure 5.6. In this case no natural separation was observed.

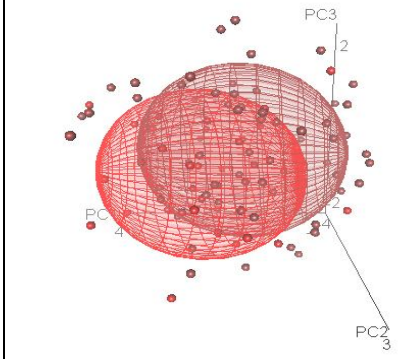
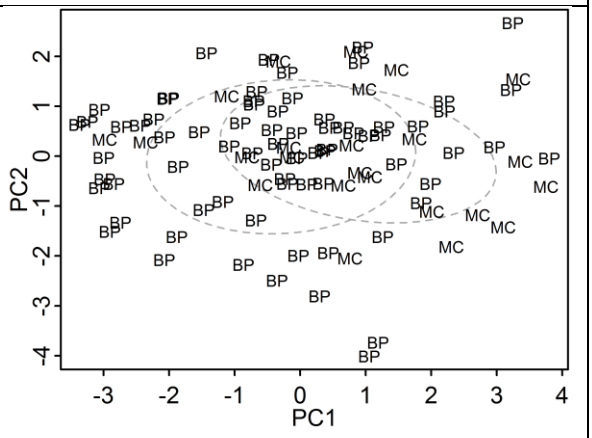
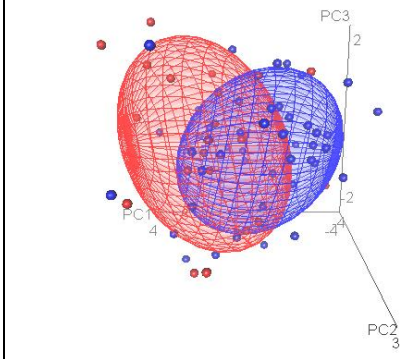
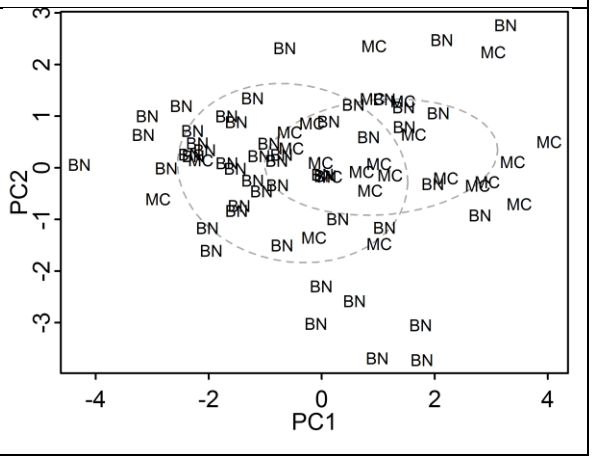
Group	PCA 3D and 2D scores plot	
BP vs. MC		
BN vs. MC		

Figure 5.6 PCA score-plots to assist in selecting a control group. Legend: NWU controls (MC) – red; Bilharziasis positive (BP) – brown; Bilharziasis negative (BN) – blue.

Hence, we excluded the NWU control samples from any further statistical analysis, and only focus on the 73 bilharziasis positive and 49 bilharziasis negative samples. The cumulative percentage of variance explained for the PCA's performed above is given in Table 5.1.

Table 5.1 PCA fit statistics

Groups	PCA number of components	Cumulative % of variance explained
BP vs MC	2	23.79%
	3	29.22%
BN vs MC	2	25.57%
	3	31.63%

Note that the cumulative percentage of variance explained by the PCAs is small. This is common for an analysis that contains a large number of features (such as metabolomic investigations), since the more features introduced, the larger the total variance. We now proceed to analyse the data sets of the BP and BN groups.

5.3.3 Bilharziasis positive versus bilharziasis negative selection

After data transformation and centering, a PCA analysis was performed using the BP and BN cases, of which the score plots are presented in the first two columns of Figure 5.7 and the summary fit statistics of the PCA analysis in Table 5.2.

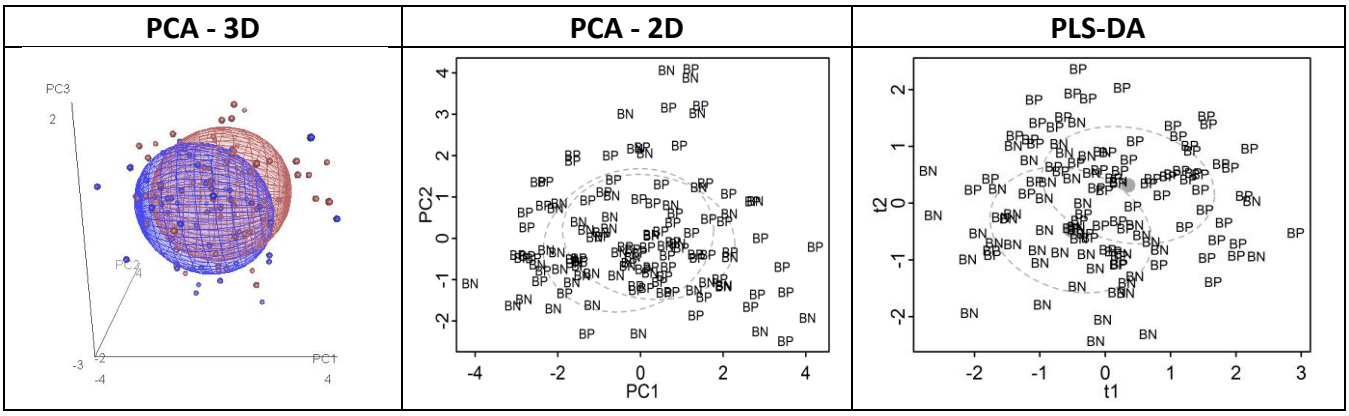


Figure 5.7 PCA score plots and PLS-DA score plot. Plots created from data containing 83 features, 73 BP cases & 49 BN cases. Legend: Bilharziasis positive (BP) – brown & Bilharziasis negative (BN) – blue.

Table 5.2 PCA and PLS-DA fit statistics of 83 features 73 BP cases & 49 BN cases.

PCA		PLS-DA			
Number of component extracted	**Cumulative % of variance explained**	**Number of component extracted**	**Cumulative R^2X**	**Cumulative R^2Y**	**Q^2**
2	24.07%	1	0.120	0.136	-0.099
3	29.59%	2	0.187	0.295	-0.359

No natural separation between the bilharziasis positive and negative groups was observed from the 2D, as well as the 3D scores plots. For any metabolomic investigation it is important to observe such a separation, without which the identification of potential biomarkers is cumbersome. On closer inspection of the PCA scores plots we noted that not all cases intersect. Therefore, we decided to follow an approach to reduce the samples, and to investigate only those cases for which a separation was plausible. This approach was followed by Reinecke *et al*. (2011). The effect is that the underlying population is reduced to a smaller target population from which we can derive bilharziasis markers. In the present investigation the following approach was used to reduce these cases: Firstly, a PLS-DA model was fitted to the data using all the cases. The resulting PLS-DA scores plots are shown in Figure 5.7 (column three) and the summary fit statistics is presented in Table 5.2. From these summary values we then deduced that the PLS-DA model did not fit extremely well with low R^2X, R^2Y values and negative Q^2 values. Note however, that the purpose of this model is case elimination and, therefore, more of a means to an end then an end in itself. Also, although the Q^2 values suggest that only one PLS-DA component should be extracted, we extracted two components, based on the PLS-DA score plot (see Figure 5.7), for case elimination. From the PLS-DA scores plot it is clear that a large number of cases do contain information concerning the metabolic profile of a bilharziasis infected individual. Using the PLS-DA model, we aimed to remove samples from the study for which the model could not predict the group membership correctly. These were cases for which the metabolic profile of the bilharziasis positive individuals resembled that of the bilharziasis negative individual and vice - versa. The PLS-DA group predicted values for both the BP and BN cases are presented in Figure 5.8.

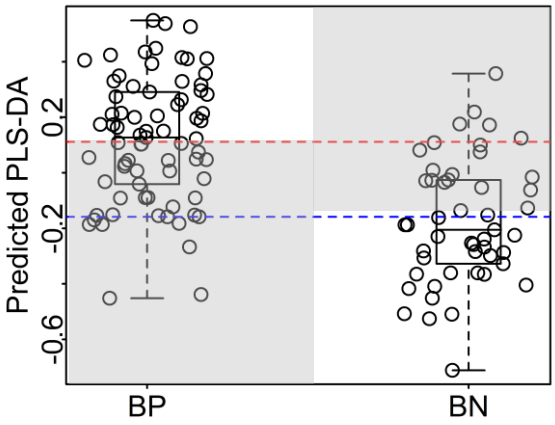


Figure 5.8 Box plot representation of the PLS-DA predicted values.

It is clear that a large number of wrong class predictions will result from the PLS-DA model. The 10th percentile value of the BP predicted values (indicated by the blue dotted line), as well as the 90th percentile of the BN predicted values (indicated by the red dotted line), was used to reduce the cases. Any BN case beyond the 10th BP percentile (blue dotted line), was removed. Similarly any BP case below the 90th BN percentile (red dotted line), was removed. These cases are located in the grey shaded areas in Figure 5.8. This case reduction reduced the number of BP samples from 73 to 38 and the BN samples from 49 to 28. The normalized data of these 66 cases were then extracted, transformed, centred and submitted to a PCA. The scores plots of this analysis are presented in the first two columns of Figure 5.9 and the summary fit statistics of the PCA analysis in Table 5.3.

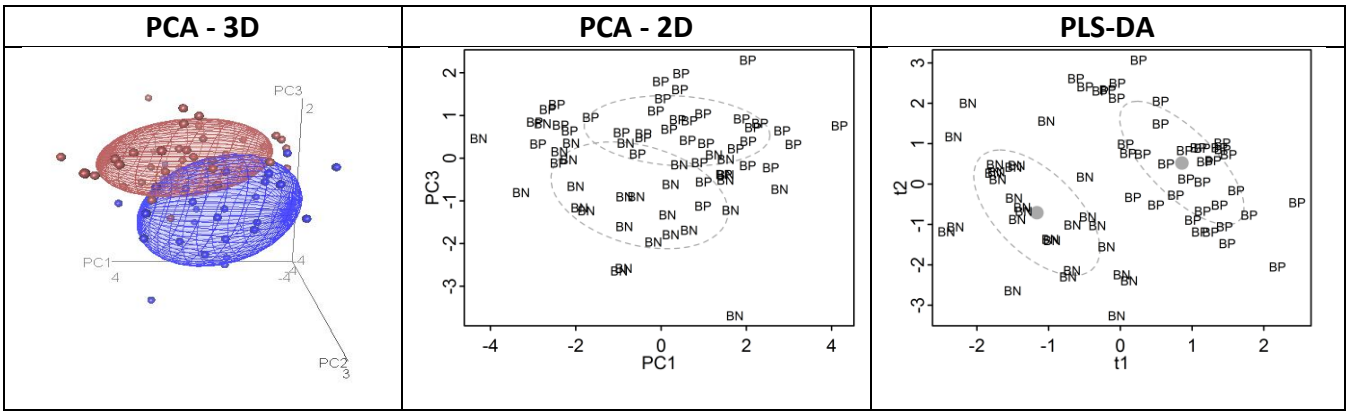


Figure 5.9 PCA score plots and PLS-DA plot. Plots created from cases swept containing 83 features 38 BP cases & 28 BN cases. Legend: Bilharziasis positive (BP) – brown & Bilharziasis negative (BN) – blue.

...Table 5.3 PCA and PLS-DA fit statistics of 83 features 38 BP cases & 28 BN cases.

PCA		PLS-DA			
Number of component extracted	**Cumulative % of variance explained**	**Number of component extracted**	**Cumulative R^2X**	**Cumulative R^2Y**	**Q^2**
2	27.09%	1	0.104302	0.656939	0.488265
3	33.49%	2	0.219957	0.842729	-0.31533

From these graphs we observed a natural separation between the reduced BP and BN groups. However, the separation is not mutually exclusive. When comparing the 3D scores plots of the PCA analysis based on all cases (see Figure 5.7) with the 3D PCA scores plot of the reduced cases (see Figure 5.9), the separation is evident. For the reduced cases PCA, principle component three is more informative concerning the separation; hence we present PC1 versus PC3 in Figure 5.9. The cumulative percentages of variance explained (see Table 5.3) by the PCA model are slightly larger when compared to the PCA performed on all the cases (see Table 5.2). Concerning the PLS-DA model, a clear separation is visible (see Figure 5.9) and a positive Q^2 value of 0.49 is observed for the model where one component was extracted. This suggests that the prediction capability of the PLS-DA model improved; nevertheless we extracted two components for biomarker identification, again based on the location of the cases in the PLS-DA score plot (see Figure 5.9). We could now proceed to biomarker identification for the targeted population. This was accomplished by performing the Mann Whitney U test (based on the untransformed data), calculating the effect sizes (based on the transformed data), evaluation of PCA power calculations, as well as the PLS-DA VIP values.

5.3.4 Metabolite profile of the selected controls and bilharziasis infected cases

All of the above mentioned statistics were then utilised to identify features responsible for the separation of the reduced BP and BN groups (see the final section of Figure 5.2). The identified features, as well as some of the descriptive statistics, all calculated from the untransformed data, are presented in Table5.4.

The features identified are categorized into four main groups namely; energy metabolism, impaired liver function, exogenous origin and other. Twenty features have been identified in this study as potential markers for bilharziasis infection which are displayed using bold font. In addition, the statistics of markers found from other studies are also presented (not highlighted in bold). Distributional comparisons between the bilharziasis positive and negative groups can be derived from the box-plots (constructed from the transformed data) presented in Figure 5.10. From these box-plots it is evident that all of the 20 identified features contain some form of discriminatory value. However, all these variables should be jointly used to discriminate between the two classes. To test the quality of the discriminatory power of the 20 identified features, we subsequently performed a PCA, and PLS-DA (including a cross-validation of the model) based only on these 20 features. The results of this validation are presented in Appendix 1 and are briefly summarised in Section 5.4.1.

Table 5.4 Final table on biomarkers of bilharziasis perturbations

Metabolite	Mean G1 ($x10^{-6}$)	Sd G1 ($x10^{-6}$)	Mean G2 ($x10^{-6}$)	Sd G2 ($x10^{-6}$)	mw p-value	d (effect sizes)	VIP
1. Energy metabolism:							
1.1 Carbohydrate metabolism – End products of glycolysis							
Pyruvic acid	N/D						
Lactic acid	277975.2	658649.9	325806	703234.5	0.490208	0.171	0.853
2-Hydroxybutyric acid	6083.251	23682.05	7948.145	29468.66	0.207399	0.318	0.994
1.2 Intermediates of the tricarboxylic acid cycle							
Citric acid	79340406	4.82E+08	1771163	1768592	0.286586	0.260	0.555
2-Ketoglutaric acid	3.27E+08	2.02E+09	5419.392	15752.31	0.031103	0.549	1.065
Succinic acid	116815.3	265299.3	347579.5	451091.6	0.000551	0.913	1.883
Fumaric acid	14313.4	61407.11	4269.728	11743.65	0.872063	0.042	0.571
2. Impaired liver function:							
2.1 Perturbation of amino acid metabolism							
2-Keto-3-Methyl valeric acid	147.336	467.8406	56.95901	178.4444	0.002873	0.769	1.443
2-Ethylhydracrylic acid	76752.87	311103.5	6683.079	11367.6	0.01655	0.609	1.170
2-Hydroxyisocaproic acid	20621.34	106707.4	5579.918	22369.3	0.811606	0.059	0.150
2.2 Perturbation of fatty acid metabolism							
Methylsuccinic acid	36483.27	124860.9	13316.57	49008.78	0.003433	0.751	1.572
Pimelic acid	1948.74	5049.065	42782.52	186120.3	0.006508	0.674	1.352
Adipic acid	151023.1	550957.4	13918.26	13372.29	0.831654	0.053	0.926

3. Exogenous origin: Disturbance of the gut microbiota / composition of the diet / environmental influences							
3-3-Hydroxyphenyl-3-hydroxypropionic acid	561355.3	748239.6	173526.9	473054.2	6.10E-07	1.299	2.314
Hydroquinone (4-hydroxyphenol)	1.22E+08	7.54E+08	382.9808	1516.942	0.000245	0.942	1.768
3-Hydroxybenzoic acid	146969.4	788798.5	80198.7	396438.3	0.01655	0.606	1.211
2-Hydroxyphenylacetic acid	27633.14	100337.1	4234.148	11091.6	0.002185	0.382	1.546
2-5-Dihydroxybenzoic acid	35825	123340.9	2327.778	5915.08	0.014271	0.605	1.262
p-Hydroxyphenyllactic acid	51324.94	188740	10537.23	21028.25	0.002185	0.781	1.487
Phenylacetylglutamine	2.19E+11	1.35E+12	380717.3	1811323	0.35604	0.224	0.472
p-Hydroxymandelic acid	30313.78	97606.02	3596.101	4707.11	0.000289	0.884	1.736
3-Methoxy-4-hydroxybenzoic acid	145011.6	730370.8	5276.234	13268.45	0.027168	0.549	1.090
Glycolic acid	641196.1	1272155	210891.4	259159.8	0.045906	0.484	1.144
4.Other:							
4.1 Endogenous/exogenous Phase II detoxification							
X-glucoronide	18448.47	78588.58	405.1709	1432.561	0.013238	0.617	1.479
N-Phenylacetylglycine	84.95997	523.5712	52.95171	277.1375	0.005285	0.692	1.471
m-Hydroxyhippuric acid	1.31E+08	8.08E+08	97775.77	404038.4	0.003915	0.749	1.398
Hippuric acid	2.87E+10	1.77E+11	2011537	2922774	0.994861	0.003	0.326
4.2 Adrenalin metabolism							
3-Methoxy-4-hydroxyphenylacetic acid	107554.1	190171.4	40710.19	65305.39	0.00013	1.002	1.874
Vanillylmandelic acid	75530.45	112475.6	69867.23	215853.9	0.003915	0.703	1.438
3-4-Dihydroxyphenylglycol	376826.1	2318253	227.1552	458.0313	0.055264	0.487	0.945

Note: 3-Methoxy-4-hydroxyphenylacetic.acid = homovanillic acid; 3-methoxy-4-hydroxyglycolic acid = vanillylmandelic acid

3-4-Dihydroxyphenylglycol = 3-4-Dihydroxyphenylethyleneglycol = DOPEG

G1 = Bilharziasis positive G2 = Bilharziasis negative

Sd = Standard deviation

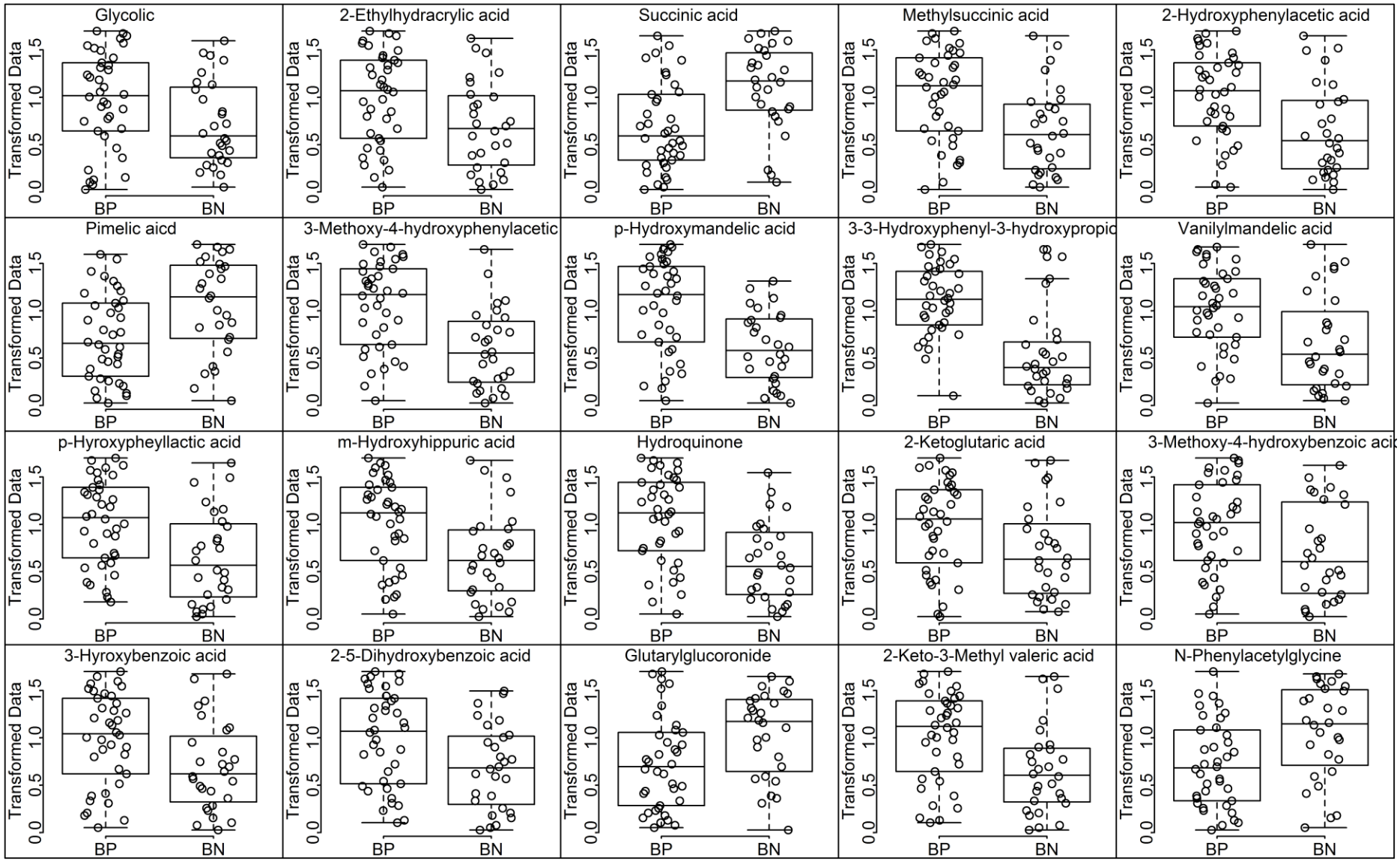


Figure 5.10 Box-plots of the transformed data for the 20 identified features.

5.4 Discussion

5.4.1 Validity of the indentified markers

We successfully derived a list of metabolic indicators (the metabolomics profile in Table 5.4) for the *S. haematobium* infection. The validation of the PLS-DA models fitted and therefore the discriminatory value of the identified variables are presented in Appendix 1. The following aspects are relevant:

(1) We decided on a number of PLS-DA components=2 (based on score plots) whilst the Q^2 value suggested the extraction of only one PLS-DA component. From Appendix 1 we concluded that the choice of two could be further motivated again, based on the PLS-DA scores plots (see Figure 5.13), as well as the high cross-validated estimates of sensitivity and specificity (see Table 5.6).

(2) In the whole study we never obtained a complete separation (see PCA scores plots), however, the PLS-DA model showed that the list of variables identified in this study, did contain discriminatory values, which could be used to predict group membership with a high degree of accuracy (see Table 5.6).

We can, therefore, now proceed to the biological interpretation of these features.

5.4.2 Assessment of the bilharziasis infected group by the immunological test.

The urinary bilharziasis samples of our current investigation were microscopically tested by the Limpopo NHLS (National Health Laboratory Services) where their infection was confirmed. From these samples 73 were identified as bilharziasis positive samples. We then further re-tested the same samples using the immunological CCA (Circulating Cathodic Antigen) test. The re-testing was done due to the popular knowledge that the gold standard of microscopic egg counts for bilharziasis diagnosis is not the most sensitive method. The principle of the test is based on the assay of antigens that may be present in the sample to bind to a colloidal carbon conjugate of anti-CCA monoclonal antibodies using a

nitrocellulose strip of the sample (van Dam *et al*., 2004). The general measured sensitivity of the CCA test is at a high from (76.9% - 100%), whereas its specificity ranges from moderate to high at levels of (43.4% - 87%) (Utzinger *et al*., 2010).

BP (Bilharziasis Positive)	No. of cases	% CCA
Original cases	73	74%
Cases used after case sweep	38	68%

The 73 samples that were collected and positively tested for bilharziasis by the NHLS laboratory were re-tested with the CCA test kit, of which 74% of those samples yielded a positive test result. From these samples we performed a further case reduction (see Section 5.3.3, Figure 5.8) where the 73 cases were reduced to 38 samples. The selected 38 samples from the 73 samples yielded a positive percentage of 68, which is still relatively high. We further worked on these samples where we proceeded for biomarker discovery using the targeted population (see Table 5.4).

5.4.3 General biological significance of the markers.

Understanding the dynamic responses of the host to bilharziasis on a systems level is one of the key importances to providing insight into the mechanisms underlying the disease's progression. Here we extended our metabolic profiling from urine, employing (i) a different host-parasite model (humans versus mice/hamster) and (ii) a different bilharziasis affecting organism (*S. haematobium*) in order to investigate the origin of the observed changes in urine and to relate the changes in the urinary metabolic profile (see Table 5.4) to potential mechanisms of pathology. The 20 feature markers that were identified provided us with information as to which features were responsible for the separation between the control (BN) and the infected group (BP). The concentration variation of the different features responsible for the separation between these groups is identified and provided in mean value in Table 5.4. The most notable observations will subsequently be discussed. It should be noted, however, that it will not be attempted to account for changes observed in every

single of the 20 metabolites, given the general agreement of all metabolomics studies on *Schistosomal* infection that the host-parasite interaction is extremely complex and influenced by several endogenous and exogenous factors. However, important perspectives on the broader significance of such infections will be presented in Chapter 6.

1. Energy metabolism

It was well established some time ago that *S. japonicum* is dependent on respiration (aerobic) as its source of energy, while *S. mansoni* mostly use glycolysis (anaerobic) for its energy requirements (Despommier *et al*., 2005). We could not, however, substantiate an effect of *S. haematobium* infections on glycolysis in our experimental population, as pyruvate is not normally detectable in organic urinary analysis, and the lactic acid, as well as 2-hydroxybutyric acid levels did not differ significantly between the controls and the patients (Table 5.4).

The elevated urinary excretion of pyruvate noticed in the mouse model of *S. mansoni* infection is consistent with an inhibition of pyruvate dehydrogenase and insufficient levels of acetyl-CoA for the Krebs cycle, as shown by depletion of citric, 2-ketoglutaric and succinic acids (Wang *et al.,* 2004). Despite its dependence on respiration for energy production, mice infected with *S. Japonicum* likewise led to suppression of the Krebs cycle, as indicated by alleviation of plasma citric acid and urinary citric, 2-ketoglutaric and succinic acids. In fact, decreased urinary succinic acid was also the most consistent observation from all metabolomics studies on *Schistosoma* infection described in Chapter 2 (Table 2.7). The decrease in succinic acid was also substantiated for the *S. haematobium* infected cases, as shown in Table 5.4. It thus seems reasonable to conclude that impaired energy metabolism is a consequence of all forms of *Schistosomas* infection in humans, apparently independent of the age of those affected (Balog *et al.,* 2011).

2. Impaired liver function

One of the consequences of infection with bilharziasis is liver injury, which is indicated by disturbance of amino acid and fatty acid catabolism. In our investigation there were increased levels of urinary 2-keto-3-methylvaleric acid and 2-ethylhydracrylic acid. 2-Keto-3-methyl valerate is a metabolite of isoleucine in man, produced by cytosolic branched

chain aminotransferase, whereupon it is further degraded by branched chain keto acid dehydrogenase. A deficiency of the branched chain keto acid dehydrogenase complex results in an accumulation of the branched-chain amino acids and the corresponding -keto- and -hydroxy acids in blood and urine. The urinary increase of 2-keto-3-methyl valerate in *S. haematobium* infected individuals in this investigation seems to support apparent indications of elevations in the keto acids promoting ketogenesis resulting from the degradation of the branched–chain amino acids also observed in mice infected with *S. japonicum* (Wu *et al*., 2010), and similar in the infection of *S. mansoni* in mice (Wang *et al*., 2004). Urinary excretion of 2-ethylhydracrylic acid is variably increased in defects of isoleucine oxidation at distal steps in the catabolic pathway and is diminished when proximal steps of the oxidative pathway are blocked, as in branched-chain keto-acid decarboxylase deficiency ('maple-syrup-urine' disease). There was a significant increase in the urinary 2-ethylhydracrylic acid and methylsuccinic acid of our *S. haematobium* infected individuals as compared to the controls. Normal human urine contains small amounts (less than 4 mg/g of creatinine) of 2-ethylhydracrylic acid (Mamer *et al*., 1976) and low concentrations of methylsuccinic acid is a normal constituent seen in human urine. Both increase, however, in defective isoleucine catabolism. We finally also encountered increases in 2-hydroxyisocaproic acid in some of the patients (not generally significant due to great variability), which likewise originated from the metabolism of the branched-chain amino acids and is also regarded as an indicator of impaired liver function due to *S. mansoni* infection (Wang *et al.,* 2004).

There has been commonality in previous investigations indicating the ability of schistosomes of taking up phospholipids and triacylglycerols from the host. The investigation on *S. mansoni*-infected mice revealed reductions in the concentrations of short chain fatty acids such as acetate, butyrate and proprionate (Wang *et al*., 2004) whilst increased levels of lipids were observed in the sera of the co-infected *S. japonicum* hamsters at week 5 post-infection (Wu *et al*., 2009) and in the urine of single infected *S. japonicum* hamsters (Wang *et al*., 2005). In our investigation, however, a reduced level of pimelic acid and no short chain fatty acids were noted (Table 5.4). The latter are volatile substances which are not normally detectable in urinary organic acid analysis. The investigation by Wang *et al.* (2005) suggests that such conflicting results could also be due to infection-induced changes in the

short chain fatty acids that depend on the type of schistosome species and/or the type of host. In conclusion, Wang *et al.* (2004) regard periportal hepatic fibrosis as the most important clinical complication of impaired liver function due to chronic schistosoma infection, which can be related to accumulation of collagen in the liver resulting in the long-term disturbances of amino acid and fatty acid degradation.

3. Disturbance of the gut microbiota

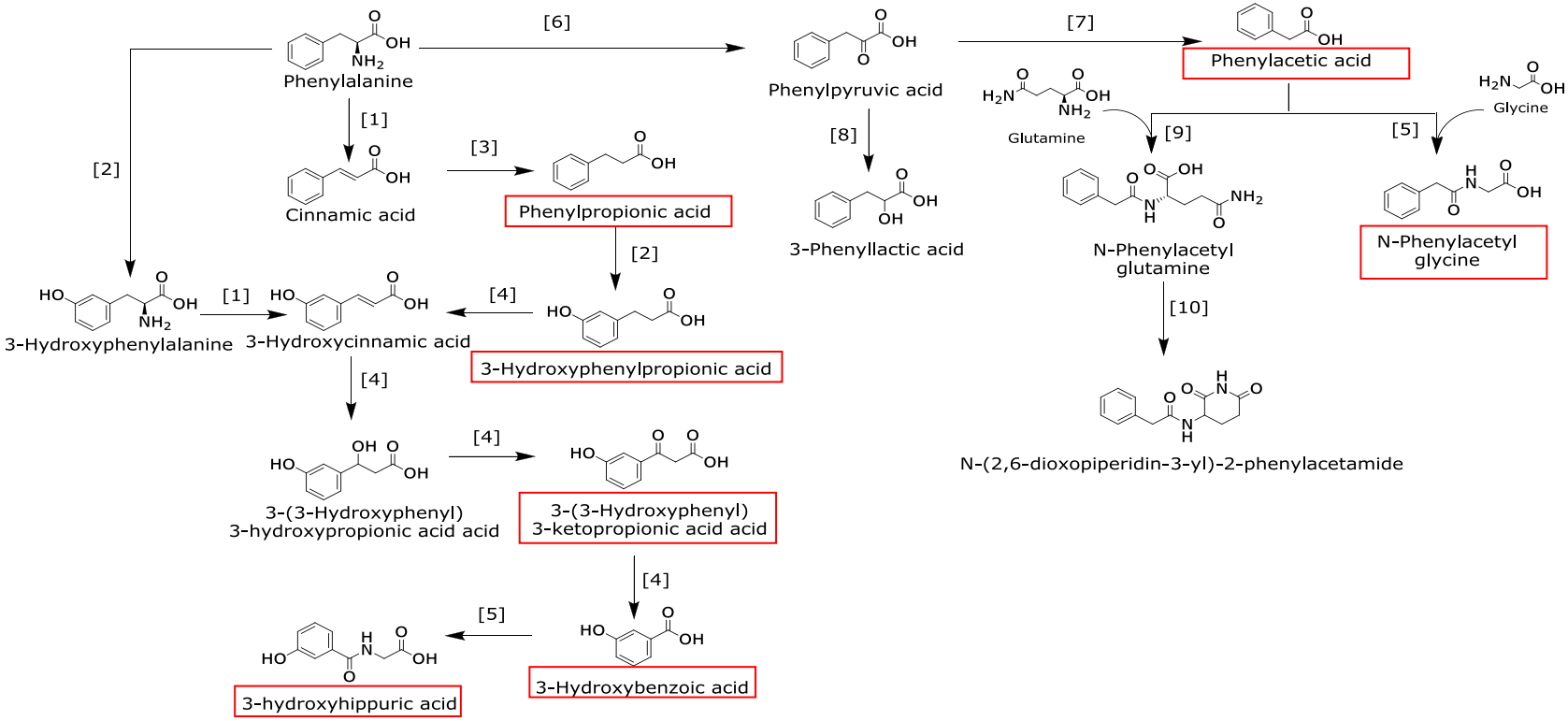
Research suggests that the relationship between the gut flora and humans is not merely commensal, a non-harmful coexistence, but rather a symbiotic relationship. The microorganisms perform a host of useful functions, such as fermenting unused energy substrates, training the immune system, preventing growth of harmful pathogenic bacteria, regulating the development of the gut, producing vitamins for the host, e.g. biotin and vitamin K, and producing hormones to direct the host to store fats (Zheng *et al*., 2011). From the literature study of the six mentioned investigations of bilharziasis (see Chapter 2, Section 2.7.2), each investigation gives an indication of the effect of bilharziasis on the relationship of the host-microbial equilibrium, each indicating disturbances of metabolites associated with the gut-microbiota. These are reflected by elevated levels of urinary constituents like trimethylamine, phenylacetylglycine and p-cresol glucuronide obtained from *S. mansoni*-infected mice (Wang *et al*., 2004; Garcia-Pérez *et al*., 2008), *S. japonicum*-infected mice (Wang *et al*., 2005; Wu *et al*., 2010), and co-infected *S. japonicum*- infected hamsters (Wu *et al*., 2009). Similar views were also linked to the reduction of hippurate and 3-hydroxyphenylpropionic acid, suggesting that such effects is a universal consequence of bilharziasis, as well as co-infection with *N. americanus* (Wu *et al*., 2010).

We have found several metabolites derived from the aromatic amino acids (phenylalanine and tyrosine) to be significantly influenced by the *S. haematobium* infection (Table 5.4). In this regard, Wu *et al.* (2010) noted: "*It is known ... that phenylalanine and tyrosine are initially converted to phenylacetic acid and hydroxyphenylacetic acid, respectively, by intestinal bacteria."* These and other metabolites formed by the gut microbiotica become absorbed by the host, and "*once absorbed ... are conjugated ... via phase II of detoxification in the liver",* producing the various conjugates observed due to parasitic infections, as is also included in Table 5.4. In addition to the disturbance of the gut we observed highly elevated

levels of p-hydroxymandelic acid. Since this metabolite comes from the sub-class of phenols it is absorbed and biotransformed by the colon's microflora. There it is known to exhibit anti-fungal qualities, in our investigation due to the influx of the *S. haematobium*-infection worm burden in our human subjects; its' over activity suggests an attempt to restore the symbiotic microbiota environment.

Based on these general observations, we propose a model of metabolic pathways originating from phenylalanine to account for its catabolites observed in the present study (see Figure 5.11). The proposed phenylalanine pathway seems to be a significant result within this experiment and it is through this model we propose to explain better the perturbations that we observed in our investigation.

Figure 5.11 Proposed metabolic pathways of products formed from phenylalanine through degradation and detoxification reactions



List of enzymes

[1]	Phenylalanine ammonia-lyase,	[6]	Tyrosine transaminase
[2]	Unknown	[7]	Pyruvate dehydrogenase
[3]	2-Enoate reductase	[8]	Lactate dehydrogenase
[4]	ß-Oxidation	[9]	Phenylacetylglutamine synthase
[5]	Glycine-N-acylase	[10]	Spontaneous

Through the results of the metabolomics approach used within this investigation we were able to observe a difference between the metabolites, marked in red boxes in Figure 5.11, of the infected individuals (bilharziasis positive) and the controls (bilharziasis negative). From this model our results correlates with the following:

- Phenylalanine has been proposed to be converted to phenylpropionic acid by certain species of *Clostridium* (gut microbiota), with cinnamic acid being the possible intermediate in that conversion. In a study of Moore *et al*. (2002), they reported that the sedimentary bacterium *Streptomyces maritimus* produces benzoyl-CoA from phenylalanine involving a phenylalanine ammonia lyase-mediated conversion of phenylalanine to cinnamic acid (see Figure 5.11). We observed significantly increased urinary levels of phenylpropionic acid – a compound known to be produced by anaerobic gut flora – in our bilharziasis infected individuals as compared to our controls.
- 3-hydroxyphenylpropionic acid (HPPA) was also detected in the urine of our infected subjects. This metabolite is associated with specific types of intestinal bacterial activity and it is known to be a major product of several *Clostridium* species (Lord *et al*., 2008). An increase in HPPA is consequential of an increased intestinal microflora metabolism of dietary catechin and caffeine in the diet. The observation of this metabolite in our infected case-subjects thus again indicates a disturbance of the gut microbiota.

  Hydroxycinnamic acid was not detected in the organic acid profiles of individuals infected and uninfected with *S. haematobium* of our current study. This metabolite is most commonly found linked to a quinic acid moiety in fruits and foods.

- 3-(3-hydroxyphenyl)-3-hydroxypropanoic acid (HPHPA) however**,** a by-product of hydroxycinnamic (Figure 5.11) was observed. HPHPA contained the highest urinary concentration in our investigation and on the biomarker list (Table 5.4) of human subjects infected with *S. haematobium*. HPHPA is an organic acid detected in human urine and its presence is ascribed to nutritional sources, e.g. dietary phenylalanine. However, the source of this metabolite is also known to be derived from multiple species of anaerobic bacteria of the *Clostridium* genus (Shaw, 2010). Phenols and polyphenols are absorbed after biotransformation by the colon microflora; the main products of the colon are benzoic acid, phenylacetic acid, phenylpropionic acid and their

derivatives, and 3-(3-hydroxyphenyl)-propionic acid (Williamson *et al*., 2010). The release of such elevated concentrations of HPHPA implies that there is an increased disturbance of the normal microbial ecosystem in the presence of *S. haematobium* worms.

The appearance of 3-hydroxybenzoic acid and its derivative 3-hydroxyhippuric acid in urine is known to be a consequence of unabsorbed phenols like phenylalanine and tyrosine. When this occurs, the ingested polyphenols, not absorbed or excreted in the bile, reach the colon where they are extensively metabolized by two central locations, the microflora and the liver, into various aromatic acids.

- We observed both 3-hydroxybenzoic acid and 3-hydroxyhippuric acid in the urine of our *S. haematobium* infected individuals in significant amounts as compared to the controls, as indicated in the model above Figure 5.11. The presence of the metabolites could be indicative of impaired intestinal absorption due to the existence of the infection.

  Furthermore, 2-5-dihydroxybnzoic acid was observed in increased levels in the urine of *S. haematobium* infected human subjects. This metabolite, not presented in our model, is found on our biomarker list Table 5.4. 2-5-dihydroxybenzoic acid is identified in having a broad spectrum of biological activities such as anti-inflammatory, antirheumatic and antioxidant, which could be activated due to the presence of the *S. haematobium* worm and its disturbance of the microbiota-mammalian equilibrium.

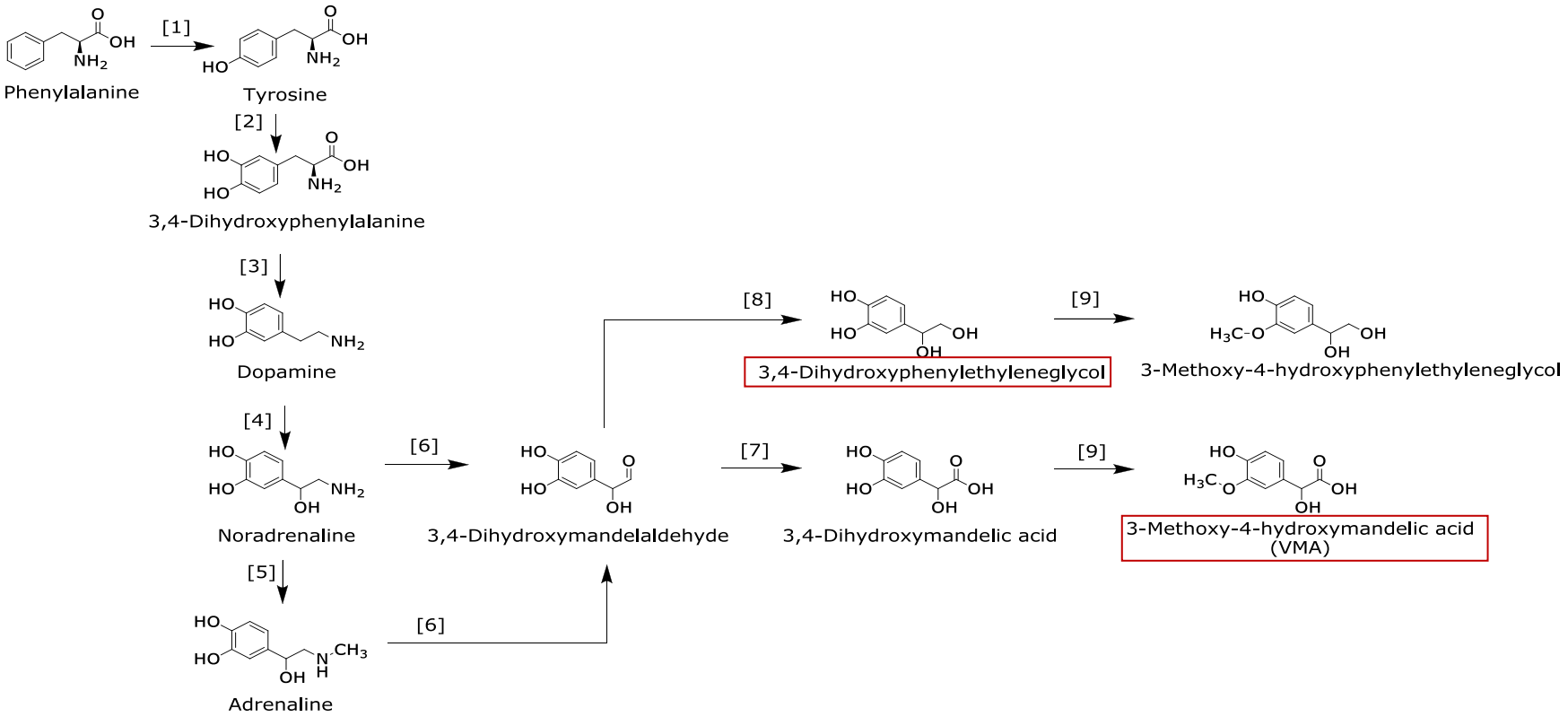
- Phenylacetic acid, another derivative of phenylalanine, with phenylpyruvic acid as the intermediate metabolite, was detected in the urine of our case-subjects infected with *S. haematobium* (Figure 5.11). Phenylacetic acid is the production of abundant endogenous compounds formed from dietary sources together with several other acidic urinary metabolites by the gut microbiota. We also observed N-phenylacetylglycine, a conjugate of phenylacetic acid and glycine. N-phenylacetylglycine was observed in increased levels in the urine of our infected individuals when compared to the controls.
- Phenylacetic acid ingested by humans is usually excreted in urine as phenylacetylglutamine. This metabolite, however, was not observed in our organic acid extraction due to the ammonia molecule attached to the compound.

Since the scope of this study is based on determining the organic acid profiles of individuals infected and uninfected with *S. haematobium* through an untargeted GC-MS approach, with potential of finding a biological perturbation that can give insight to the schistosome infection (Chapter 2, Section 2.7), we can conclude that the metabolites observed through the use of a metabolomics approach could be interpreted according to the proposed model given above (Figure 5.11).

4. Catecholamine metabolism

It has been reported by previous research that the underlying molecular markers of a schistosome infection are found primarily in changes in gut microflora, energy metabolism and liver function. We have in this current research also shown explanatory results corresponding with these findings (Section 5.4.3). We however, also present for the first time, metabolic perturbations in the catecholamine metabolism. It remains to be established, however, whether this is unique to a *S. haematobium* infection. Based on our novel findings we present a model (Figure 5.12), based on a section of the known metabolism of catecholamines from the phenylalanine origin, to further explain the results obtained from the metabolomic approach used in this investigation.

Figure 5.12 Markers of phenylalanine, tyrosine and catecholamine metabolism.



List of Enzymes

[1] Phenylalanine hydroxylase
[2] Tyrosine hydroxylase
[3] DOPA decarboxylase
[4] Dopamine hydroxylase
[5] Phenylethanolamine N-methyltransferase
[6] Monoamine oxidase
[7] Aldehyde dehydrogenase
[8] Aldehyde reductase
[9] Catechol-O-methyltransferase

Tyrosine, derived from phenylalanine through phenylalanine hydroxylase, or through the diet, is the precursor of several well-known catecholamines, as shown in Figure 5.12. They are called catecholamines because they contain a catechol or 1,2-dihydroxybenzene group, and are derived from the amino acid tyrosine. The presence of metabolites from these pathways has not been reported in previous studies on *schistosomal* infections, but was detected to be significantly present in the urine of *S. haematobium* infected individuals (Table 5.4). In our current study we observed an accumulation of urinary catecholamine metabolites specifically: 3-methoxy-4-hydroxyphenylacetic acid (homovanillic acid), 3-methoxy-4-hydroxyglycolic acid (vanillylmandelic acid) and 3-4-dihydroxyphenylgycol (DOPEG) in higher concentration levels in our *S. haematobium* infected individuals than there were in our controls.

- One of the first steps in catecholamine metabolism is the conversion of tyrosine to 3-4-hydroxyphenylalanine (L-DOPA) by the enzyme tyrosine hydroxylase (see Figure 5.12). This is the rate-limiting step in the biosynthesis of the catecholamines dopamine, norepinephrine and epinephrine (van der Heyden *et al*., 2003). After binding to its receptor(s) and reuptake into neurons, dopamine is inactivated to its metabolite homovanillic acid (Wevers *et al*., 1999).

- Homovanillic acid, one of the catecholamine metabolites from our biomarker list (see Table 5.4), presented with the second highest urinary concentration in our biomarker list of our *S. haematobium* infected subjects. Homovanillic acid, a major catecholamine metabolite of dopamine occurring in human biofluid, has been used as a peripheral measure of dopaminergic activity in the central nervous system (Sumiyoshi *et al*., 1998).
- Increased levels of vanillylmandelic acid were also observed in urine samples of our S. *haematobium* infected case-subjects. Vanillylmandelic acid, the major end-product of norepinephrine and epinephrine (also known as adrenalin) metabolism, is produced almost exclusively from the removal and metabolism by the liver of catecholamines and their metabolites that circulate in the bloodstream (Eisenhofer *et al*., 1996). It is produced through intermediary metabolites such as 3-4-dihydroxymandelic acid (Figure 5.12). The substantial production of vanillylmandelic from circulating 3-4-dihydroxyphenylglycol and 3-methoxy-4-hydroxyphenylglycol, most of which is derived

from neuronal norepinephrine metabolism, explains why vanillylmandelic acid is a relatively insensitive marker for pheochromocytoma compared with the precursors norepinephrine and epinephrine.

- An increase in vanillylmandelic in our *S. haematobium* infected subjects could be consequential of an over stimulation in the release of norepinephrine and epinephrine which could be for a number of reasons, such as: an increase in the release of norepinephrine from the kidneys (norepinephrine is one of the hormones produced by the adrenal glands, which are found on top of the kidneys), a hormone released into the blood during times of physical stress, such as infection and fever and emotional stress.
- In addition to the neurotransmitters that were observed in the urine of the S. haematobium infected human individuals, was 3-4-dihydroxyphenylglycol (DOPEG) which was observed in elevated concentrations when compared to the controls. DOPEG is a normal norepinephrine metabolite present in CSF, plasma and urine in humans. The norepinephrine is metabolized within the nerves, is converted to DOPEG and much of the DOPEG formed in the nerves is further metabolized by catechol-O-methyltransferase to 3-methoxy-4-hydroxyphenolglycol (MOPEG) in extraneuronal cells (Eisenhofe *et al*., 1994).

The discovery of catecholamines in a *S. haematobium* infection in humans, when compared to the results obtained by research presented of the 6 models (Chapter 2, Section 2.6.2) of a *S. mansoni* or *S. japonicum* infection is new. This observation might be a result of species specific or even due to a specification on the species – host result of which further investigation needs to be employed to be able to know the exact link.

5. Diet or other environmental influences

Urinary glycolic acid was observed in increased levels in our human subjects infected with *S. haematobium* and on our perturbation marker list. This metabolite is of exogenous origin and is known well as the fruit acid. It is derived from sugar cane and it can, therefore, be considered a natural product. The increase of glycolic acid in the urine of our investigation subjects is of a dietary effect. In addition, increases of 3-methoxy-4-hydroxybenzoic acids

(vanillinic acid) were observed in the GC-MS spectra of our BP subjects. Vanillinic acid is a phenolic acid found in some forms of vanilla and many other plant extracts. It is the intermediate product in the two-step bioconversion of ferulic acid to vanillin (Lesage-Meessen *et al*., 2006). Vanillic acid is a metabolic byproduct of caffeic acid and is often found in the urine of humans who have consumed coffee, chocolate, tea and vanilla-flavoured confectionary.

From these observed changes in urinary metabolic perturbation markers, it would appear that the results obtained from this investigation coincide with the rationale that a three-way interaction may exist between the host, the parasite, and the resident host microflora (Nicholson *et al*., 2004). A further discussion of these, and other related aspects from two investigations, will subsequently be presented in Chapter 6.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

6.1 Introduction

It was noted in Chapter 1 that "bilharziasis is one of the most prevalent parasitic infections in the world, and continues to be a global public health concern in the developing world. The main infectious disease-causing species are *S. haematobium, S. mansoni* and *S. japonicum*. According to the recent weekly epidemiological record of the World Health Organisation (WHO), bilharziasis is endemic in 76 countries and territories (WHO, 2011), and it is estimated that more people are infected by the *S. haematobium* species than with the other two combined" (Rinaldi *et al.,* 2011). However, the fundamental aspects of *S. mansoni* and *S. japonicum* are well researched and described, whereas insights on *S. haematobium* lag well behind (Rinaldi *et al.*, 2011). The same applies to animal, as well as human models describing the consequences of *S. haematobium* infection. In fact, this dissertation presents the first investigation on the profile of individuals who show infection by *S. haematobium,* coming from a free-living population from a rural area (African children and adults from the Limpopo Province of South Africa, with an age distribution of 5 to 34 years of age, including male and female subjects). Against this background, the following research question was proposed for the present investigation: "What information can be revealed on humans infected by *S. haematobium* through an untargeted metabolomics investigation of the urinary organic acids?" (Chapter 2.7). It was also noted in Chapter 2.7 that the approach to answer this research question would entail an untargeted investigation of a sub-section of the human metabolome (for which the organic acids were selected) from individuals infected by *S. haematobium,* which in essence required an unprejudiced experimental design, designated as an inductive mode of investigation (Kell, 2004). The outcomes of this investigation with regard to the research question are: (1) The untargeted metabolomics investigation revealed new and potentially important information on humans infected by *S. haematobium* (see discussion on Aim 3 below), but (2) also indicated the

limitations of such an experimental approach, most of which were known or anticipated before the commenced investigation, but became emphasized through the present study.

The following Sections propose to focus on aspects related to the three aims cited in Chapter 2 and also an attempt to make clear recommendations linked to each aim based on the results obtained.

6.2. Sampling for an investigation on *S. haematobium* infection

Aim 1: "Assemble a good set of samples for standardization of the metabolomics technology and for the metabolomics study of humans infected by *S. haematobium*, as well as from comparative controls" (Chapter 2.7). This aim was successfully accomplished:

Ethical clearance was obtained from the Elim Hospital and also from the Makuleke tribal authority where permission was granted to collect the urine samples at the clinics of the National Health Laboratory Services (NHLS) at Elim and Malamulele in the former Venda (Vembe and Livubu regions of Limpopo Province) region. The project was also ethically approved as part of the bilharziasis research programme of the Unit for Environmental Research and Development of the North-West University. The collection of samples and controls from the NHLS clinics both in Malamulele and Elim occurred in four months (two months of sample collection in each clinic), and additional controls samples were collected from the Potchefstroom Laboratory for Inborn Errors of Metabolism (PLIEM). The samples were obtained from 147 randomly selected individuals with a balance of sex and age. The samples were collected and stored according to protocol as mentioned in Section 3.5.2 at the North-West University. The presence of infection was recorded by the NHLS laboratory technicians through microscopic detection of schistosome eggs in urine samples. The urine samples finally selected for the metabolomics analysis were further screened for infection through utilizing the CCA test kit.

However, several limitations were linked to the cohort of samples:

1. The collection of urine was done at Limpopo district, where *S. haematobium* is known to be endemic. The collection of urine samples at the villages surrounding

Elim and Malamulele in Limpopo was of a sensitive nature, as permission had to first be granted by the Makuleke tribal authorities. The NHLS clinics of these two communities were more than happy to get involved and with their help we were able to obtain urine samples. The collection of urine was achieved through the deliverance of patient urine samples sent to the clinic, which limited the formula of a well-designed case–control study which could have better benefitted this investigation.

2. In Chapter 2, Section 2.6.1 we present an outline of 6 models where metabolomics was used to investigate bilharziasis. In those investigations rodent models were used. In comparison to these animal studies our study is based on urine sample collection of *S. haematobium* infection of humans in an endemic area and this makes our investigation more complex, due to the variation in infection intensities and differences in the time course of the infection, unlike in the mice models where all the case-mice are infected at the same time with the same number of cercariae. In addition, the urine samples collected were from free-living individuals; this presents a cohort study with inter- and intra-individual biological variability amongst our investigation group. Differences in the nature and status of nutrition were not reported. The participants could also have been co-infected with other diseases and parasitic infections, all of which are features influencing the metabolite profile and specificity of the markers. The infection being investigated *S. haematobium* might include co-inhabitation with *S. mansoni* (see Figure 2.3) which can cause confounding effects to the infected individuals (Meurs *et al*., 2012).

3. Haematuria was one of the diagnostic markers for an *S. haematobium* infection. Dipstick tests were then performed on the collected urine samples, patients and controls, to test for micro-haematuria. This was done to make sure that the samples were homologous and that if a separation was found between the two groups, bilharziasis infected and bilharziasis non-infected, it was not due to blood in the urine. However, there are other infections within the urinary tract that also may cause haematuria.

4. The urinary-CCA test performed on the collected samples is a presumptive detection of bilharziasis in individuals with clinical signs and symptoms consistent with an active bilharziasis infection. However, the test may provide false negative results known to occur in a low parasitic level of infection i.e., individuals in the first four weeks of infection. Therefore, the sensitivity of the test varies with the intensity of the infection.

Recommendations on improvement for sampling in a future metabolomics investigation on *S. haematobium* infection.

I. Create a well planned experimental design, which minimises sample variability during sample collection and storage. It is of most importance with any analytical investigation that the methods utilized for the identification and quantification of metabolites presented in complex biological matrixes should be repeatable and accurate within the limits of quantification; therefore, the urine samples should be used with a high level of confidence.

II. Maintain the collection of samples to a single endemic area i.e., a single village in a certain district, as this will minimise location-induced variability due to environmental differences. Also collect the samples at the same seasonal time-frame i.e., when the release of the cercariae is at its peak – this is usually when the daily average maximum temperatures reach 30°C or higher. The samples of our current investigation were collected over four months and this could have induced a time variation amongst our samples.

III. Get an informed and written consent from the participants to participate in the investigation and to provide personal information on an internationally standardized questionnaire (Lengeler *et al*., 2002) to support the overview of the severity of the bilharziasis infection in the area of choice for sample collection. Note: The questionnaire was developed for diagnosing *S. haematobium* at a community and individual

level in Nigeria. The questionnaire proved to be informative, giving a predictive range of the infection rate based on its questions and most of all it was low-cost.

IV. Involve the participation of physicians or specialists in infective diseases to obtain a clear clinical profile of the participants in the cohort study, i.e., what other diseases or infections they might have, if they are under any treatment. Have a clear and defined homogeneity between the control and experimental groups to make them as comparable as possible i.e., gender, age, ethnicity, diet, lifestyle and other possible confounding influences.

V. Strategies should be employed with regard to the selection process of participants and samples i.e., pre-check the urine samples for micro-haematuria to make sure they are homogenous and re-analyse the selected samples with the CCA test to verify what infection category they fall into before performing any metabolomics quantification analysis.

6.3. Repeatability in the generation of a metabolomics data matrix

Van Batenburg *et al.* (2011) claimed: "*A single value for repeatability as a figure of merit for such data does not suffice."* and "*Because of all these* (analytical) *issues discussed* (*above*)*, the repeatability will not do as a figure of merit for a comprehensive metabolomics data set."* However, metabolic profiling requires a high degree of reliability of the metabolomics data, which is the basis for **Aim 2**: "Propose an approach for an assessment of the repeatability of the generation of metabolomics data matrix required for a bioinformatic analysis to obtain the metabolite profile" (Chapter 2.7).

With regard to this aim (Chapter 4), a description is presented on the statistical methods that were used to measure repeatability. The **first** of these is based on a more conventional methodology using the coefficient of variation (CV) and the **second**, introduces a new qualitative method, based on kernel density estimates, developed by Dr. Gerhard Koekemoer of the Statistical Consultation services of the NWU, and co-study leader for this

dissertation. As shown in Chapter 4 (Section 4.3), the aim of repeatability was successfully achieved.

The qualitative method has since been improved to include lower and upper bounds within which the observed density estimates must lie if the experiment is repeatable (Mason *et.al*., 2012 *in preparation*). The details of the improvement are outside the scope of the current dissertation, however, we present a graph (see Figure 6.1 B), which can be compared to the repeatability density estimates presented in Chapter 4 (see Table 4.2, row 2). For comparison purposes, we present this graph again in Figure 6.1 (A).

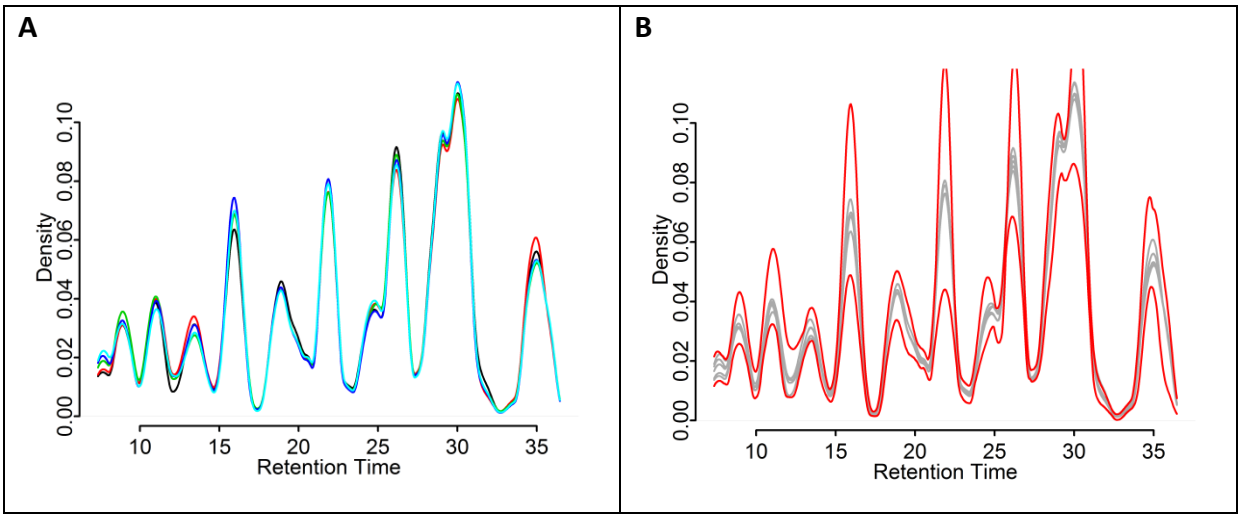


Figure 6.1 Method for assessing repeatability: (A) kernel density estimates which are our qualitative results of repeatability **(B)** kernel density estimates within lower and upper bounds-red lines.

The density lines in Figure 6.1 (A) are shown as the gray lines in Figure 6.1 (B). From Figure 6.1 B we conclude that the observed density estimates (gray lines) are all bounded within the lower and upper red lines, which defines a repeatable experiment. From these results of the experimental work done in Chapter 4 and Chapter 5 with the utilization of the QCs which represent repeatability, it is substantiated that the investigational work conducted in this dissertation is repeatable and accurate within the limits of quantification.

6.4. Biomarkers and the pathophysiological profile of *S. haematobium* infection

Definition and requirements for a biomarker:

Biomarkers (biological markers) are widely used today in Chemical Pathology and Clinical Chemistry practice. Biomarkers are defined as "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention" (Frank *et al*., 2003; Biomarkers Definitions Working Group, 2001).

In medicine biomarkers can be utilized to inspect the function of an organ or the status of health. For example, rubidium chloride is a traceable substance used as a radioactive isotope introduced inside the body to evaluate perfusion of heart muscle. It can also be a substance whose detection indicates a particular disease state, for example, the increase in blood glucose levels could indicate diabetes or the presence of an antibody may indicate an infection. More specifically, a biomarker indicates a change in expression or state of a biomolecule that correlates with the risk or progression of a disease, or with the susceptibility of the disease to a given treatment.

In clinical trials, the biochemical biomarkers used are those which are obtained from body fluids that are easily available to the primary phase researchers. In metabolomics however, the primary objective for biomarker discovery is to characterize metabolic differences between altered, stressed or otherwise abnormal physiological states, with a single biomarker defined from a metabolomics investigation the exception rather than the rule, e.g. sarcosine for prostate cancer (Baum *et al*., 2010). A biosignature on the other hand is a profile of combined biomarkers (Parida *et al*. 2010).

Biomarkers for infection by schistosoma species

Van der Greef *et al*. (2003) pronounced that the identification of a biomarker "*can be refined by recognizing that early markers of a disease will differ from markers found in later stages of disease progression*" (see Chapter 5, Section 5.1.1). In this regard the study compiled by Wu *et al*. (2010) documented elevations of urinary 3-ureidopropionate in mice infected with *S. japonicum.* The presence of the same metabolite had been recorded in the

urine samples of mice 53 days after infection by *S. mansoni* (Garcia-Perez *et al*., 2010). Furthermore, in the study by Wu *et al.* (2010) elevations of 3-ureidopropionate was recorded in mice at the third week of *S. japonicum* infection, four weeks earlier than the previous investigation. From this analysis they suggested that urinary 3-ureidopropionate could be a potential biomarker for early diagnosis of schistosome infection. Ongoing they postulated that elevated levels of urinary 3-ureidopropionate could suggest that the schistosomal infection causes reduced activities of the β-ureidopropionase leading to a disturbed uracil metabolism. This metabolite was not observed in the current study presented in this dissertation. It was concluded however that the presence and specificity of 3-ureidopropionate warrants further investigation in experimental studies using animal models and / or a cohort of humans infected by schistosomas.

The study presented in this dissertation is related to that of Balog *et al*. (2011) who investigated a cohort of humans infected by *S. mansoni* (Chapter 2, Section 2.6.2.2). In contrast to my approach, they recommend not to build a "mechanistic" pathway-based interpretation but to consider the metabolic perturbations observed in the urine as a signature response of the host on the infection. (See my proposal on possible pathways on products formed from phenylalanine through degradation and detoxification reactions, (Chapter 5 and Figure 5.11)). The important biomarkers that they emphasize to distinguish between controls and *S. mansoni* individuals are linked to gut microflora, energy metabolism and liver function. This is in accordance with my discussion in Chapter 5 (compare Table 5.4). They then noted that their proposal raised two basic questions "*(i) does the signature represent the morbidity or the infection status of the host and (ii) how specific is the signature for S. mansoni?"* These leading questions are indicative of the importance of metabolomic studies on bilharziasis, which I will address in the discussion on Aim 3. Nonetheless, there is still a dire need of a metabolite profile or metabolic signature that is specific for these infectious diseases. Taking into account factors like difference of diet, environmental exposure, geographical location and co-infections amongst the hosts' the specificity of the presented metabolic signatures still remains to be evaluated. This brings me to the third aim of this dissertation.

Aim 3: "To determine the metabolic profile of organic acids from humans infected by *S. haematobium* and give a possible biological/pathophysiological interpretation of their profile in comparison to 6 other models of bilharziasis" (see Chapter 2.7).

In this investigation we discovered that the metabolite profile of *S. haematobium* infection in main lines seems to be similar to the infection of the other two species (*S. mansoni* and *S. japonicum*). The scope of this investigation as mentioned above is to determine a metabolic signature between individuals infected and uninfected with *S. haematobium* bilharziasis; however, when comparing the results obtained from our metabolomic investigation to the summary of the 6 models presented in Chapter 2, Table 2.7 the same commonalties became evident. For example, quite common with the bilharziasis infection is the reduced levels of the TCA intermediates. The *S. japonicum* infection in mice led to the suppression of TCA cycle with alleviation of plasma citrate and urinary citrate, 2-oxoglutarate and succinate (Wu *et al*., 2010) which was similar to previous observations for *S. mansoni* infection (Wang *et al.,* 2004; Wang *et al*., 2005). In our investigation our findings of the depletion of the TCA cycle included 2-ketoglutaric acid (= 2-oxoglutaric acid) and succinic acid in the urine of the infected BP individuals as compared to the controls. However, unlike the other mice models infected with *S. mansoni* which indicated alleviation in both plasma citrate and urinary citrate, there was an increase in citric acid and fumaric acid in the urine of our *S. haematobium* human infected individuals.

- Citric acid (citrate) is a weak acid that is formed in the tricarboxylic acid cycle or that may be introduced with diet. *S. haematobium* is found almost exclusively in the venus plexus that drains the urinary bladder, where the eggs of the *S. haematobium* must first transverse the wall of the bladder before exiting with the urine. These conditions could be consequences of the acidic increase in the host's urine leading to hypocitraturia causing the urine to crystallize, an indication of a urinary tract infection, and a possible underlying cause of the microhaematuria which is seen in most of the infected BP urine samples. A high level of urinary citrate excretion is a common tool in the differential diagnosis of kidney stones, or even renal tubular acidosis. We can deduce that the increase in urinary citric acid in humans infected with *S. haematobium,* unlike its decrease in

S. mansoni mice models (Wang *et al*., 2004) is due to the response of the host to the worm load and the infection itself.

- Table 2.7 represents a summary of the results obtained by each model in consideration to certain metabolites. With regard to succinic acid it is noticed that it decreases in all the six investigative models where it was observed. Our findings also indicated that an infection with *S. haematobium* causes a decrease in succinic acid. Succinic acid is an important substrate in the functionality of the mitochondria where it is oxidized by an enzyme complex called succinate dehydrogenase into fumaric acid. The concentrations of succinic acid and fumaric acid differed between the infected and uninfected individuals, which could be linked to mitochondrial dysfunction.
- The interpretation of changes in 2-ketoglutaric acid is likewise not straightforward. Changes in 2-ketoglutatric acid may be influenced by metabolites preceding it in the TCA cycle, like citric acid (discussed above) and aconitic acid, which was not significantly altered in our investigation.

The surveillance of the above mentioned results in this investigation and that of the 6-models (Table 2.7) suggests that changes in the intermediates of the TCA cycle are a universal consequence of a bilharziasis infection, although the levels of changes are not always similar.

In the present study we observed fundamental metabolites excreted in urine of our case-subjects infected with *S. haematobium* corresponding with the metabolic pathway model proposed for phenylalanine and tyrosine (see Chapter 5, Figure 5.11 and Figure 5.12). The detection of these biological markers from the phenylalanine – tyrosine proposed pathway was of noteworthy significance as these results included the novel observation of catecholamine metabolites not previously reported in the current literature.

- Phenylacetylglycine is an acyl glycine – acyl glycines are normally minor metabolites of fatty acids. However, the excretion of certain acyl glycines is increased in several inborn errors of metabolism. In certain cases the measurement of these metabolites in body fluids can be used to diagnose

disorders associated with mitochondrial fatty acid beta-oxidation. Phenylacetylglycine a glycine conjugate of phenylacetic acid was observed in increased levels in the urine of our investigative bilharziasis case-subjects. This metabolite was also observed in previous bilharziasis research (Table 2.7) in elevated levels (Wang *et al* 2004; Wang *et al*., 2005; Wu *et al*., 2010). Additionally research conducted by Nicholls *et al*. (2000) documented elevated urinary phenylacetylglycine in animals known to be exhibiting abnormal phospholipid accumulation in many tissues and suggest that with further research phenylacetylglycine may thus be useful as a surrogate biomarker for phospholipidosis (Dalaney *et al*., 2004; Nicholls *et al*., 2000). In our current investigation, however, the elevation of phenylacetylglycine in the urine of our investigative S. haematobium infected subjects could be a result of a gut microbial infection.

- Remarkably high levels of 3-(3-hydroxyphenyl)-3-hydroxypropanoic acid (HPHPA) (see Table 5.4), a decarboxylated metabolite of phenylalanine through intermediate metabolites (see Section 5.4.3 Figure 5.11) were also observed in the urine of our *S. haematobium*-infected individuals. This metabolite was identified as the one metabolite with the highest VIP value of 2.31 on our biomarker list (Table 5.4). Interestingly to HPHPA is that previous research carried out by Shaw *et al.* (2010), reported on identifying 3-(3-hydroxyphenyl)-3-hydroxypropionic acid which was observed in elevated levels in urine samples of children with autism (a disorder of neural development with impaired social behaviour), as compared to their age and sex appropriate controls and also in an adult with recurring diarrhea caused by a *Clostridium difficile* infection (Shaw, 2010).
- Additionally Lis *et al*. (1976) reported elevated urinary levels of 4-hydroxyhippuric acids in autistic subjects and the immediate pre-conjugation precursor of that compound, 4-hydroxybenzoic acid, which could potentially be derived from either phenylalanine or tyrosine. In the current study we observed m-hydroxyhippuric acid (3-Hydroxyhippuric acid) in a human model of subjects infected with *S. haematobium* (see Table 5.4

- Vanillylmandelic acid, a catecholamine metabolite discovered in the urine of our infected subjects (see Figure 5.12 marked in a red box), is a metabolite of the amino acid tyrosine produced by *Clostridia* species. It has been postulated that *Clostridia* overgrowth in the gut may interfere with the body's production and metabolism of important catecholamines such as dopamine, norepinephrine, and productions of tyrosine, which if elevated could lead to autism and depression (Shaw *et al*., 2010). It has also been reported that the gut bacterial production of p-cresol is consequential to autism, *Clostridium difficile*, being one of the notable p-cresol producer, is known to have inhibitory effects on dopamine β-hydroxylase effecting dopamine, an important neurotransmitter.
- Homovanillic acid (3-methoxy-4-hydrophenylacetic acid) presented also as one of the neuro-metabolites that we discovered in elevated concentrations in the urine of bilharziasis infected individuals. Its increase is possibly associated with the disturbance of the symbiotic relationship between the gut-microbiota and the mammalian host. Since neuropathological diseases are thought to be connected to the gut, research has brought evidential findings of a microbial-mammalian co-metabolism association, implying that there is a bidirectional neurohomoral communication system that incorporates brain and gastrointestinal functions, a term known as the gut-brain axis (Holmes *et al*., 2011). Research conducted by Socha *et al.,* 2010 proposed that homovanillic could be a potential biomarker for autism.

Therefore, there is seemingly some sort of high degree of metabolic proximity suggesting that there might be a connection between autism, an abnormal gut microbial metabolism of aromatic amino acids and a *S. haematobium* infection due to the worm burden. However, the existence and the exact nature of any link between these three possibilities remain to be clarified.

With regard to the above mentioned results, obtained through the metabolomics approach used in this investigation, the limitation we are faced with is the question of specificity of our findings and the difficulty to evaluate them on the basis of the available information

(compare also the two questions raised by Balog *et al*., 2011 above). To overcome such limitations the following criteria have to be met.

(1) A well established experimental design must be produced to obtain samples which comply with well-defined criteria like the impact of factors such as: geographical location, differences in ethnicity, age, gender, diet and co-morbidities. If these factors are not dealt with then the metabolomic signatures discovered for bilharziasis, such as the ones we have presented above, can never be specific biomarkers for bilharziasis.

(2) It has also come to our attention that there are other infectious diseases which show comparative symptoms, e.g.

- **HIV-infection**: in a research conduct by Williams *et al.* (2011) they observed changes in the same metabolites of the TCA cycle (succinic acid and fumaric acid which both increased in their study) as those that were observed in the current investigation (Section 5.4.3). An indicator related to gut disturbances (indole-3 acetic acid) was also found in the *S. japonicum* mice model (Wu *et al*., 2010).

- **Malaria**: Basant *et al.* (2010) observed the presence of phenylacetylglycine and lactate in malaria infections, both of which were observed in the current study as well as in the infection of *S. mansoni* and *S. japonicum* (Wang *et al.,* 2004; Wu *et al*., 2010; Balog *et al*., 2011). Phenylacetylglycine levels were increased in all three investigations including the current one of a *S. haematobium* infection. Lactate was observed in elevated levels in both the *S. mansoni* and *S. japonicum*; however, in the current investigation lactic acid levels did not differ significantly between the controls and the patients. The investigation on malaria also saw ureidopropionate, which was proposed by Wu *et al*. (2010) to be an early biomarker for *S. japonicum* infection.

- ***E. caproni* infection**, which is also a neglected tropical disease and which is caused by food-borne trematodiasis (Saric *et al.,* 2008), observed hippurate, phenylacetylglycine and succinate in urine, as well as trimethylamine and

trimethylamine-N-oxide observed in NMR analysis of urine from bilharziasis studies on mice and hamsters.

- *K. pneumonia* and other bacteria are associated with **sepsis** (Khodakova and Beloborodova, 2007). Phenyllactic acid, p-hydroxyphenylacetic acid and p-hydroxyphenyllactic acid were reported to be present in sepsis cases, while a marked increase of p-hydroxyphenyllactic acid was observed in the current investigation.

- In **Crohn's diseases** (Inflammatory bowel disease), Jansson *et al.* (2009), observed several metabolites from tyrosine and phenylalanine. In the current investigation these included; phenylpropionic acid, 3-hydroxyphenylpropionic acid, 3-(3-hydroxyphenyl)-3-hydroxypropanoic acid, homovanillic acid and vanillylmandelic acid (see Section 5.4.3, Figure 5.11 and 5.12). An interruption of these aromatic phenolic acids could be an outcome of a disturbance of the gut microflora or impaired intestinal resorption.

All these examples are related to the specificity of schistosomas infection, as was also expressed by the basic questions formulated by Balog *et al*. (2011). Dysbiosis may be the underlying perturbation for some of these findings (see Lord *et al.,* 2008, Sekirov *et al.,* 2010). Clarification of these issues can only be achieved by further investigations of these infectious diseases.

In conclusion of the outcomes of Aim 3, the metabolomic analysis of urinary samples of humans infected with *S. haematobium* in an inductive nontargeted study proved to be valuable. Through this approach I was able to present and substantiate metabolic signatures associated with a bilharziasis infection and present possible physiological interpretation and novel information improving on the existing knowledge of a *S. haematobium* infection. The metabolic signatures that were observed in this dissertation as a result of a *S. haematobium* infection will most probably be different at various stages of infection and therefore represent more an indication of the infection status rather than a

particular state of morbidity. Famously the "golden-standard" method in diagnosing bilharziasis still remains the detection of eggs in urine and feces under a light microscope, or through an immunological approach by detecting soluble antigens secreted from the hatching-eggs by means of the antigen detecting test, e.g. CCA. These diagnostic tests however, provide broad specificity. The microscopy-based method is specific; it is simple and cheap, however, its limitation to day-to-day egg output fluctuations leading to failure to detect infections of low intensities maintain. Therefore, the development of early and accurate diagnostic methods are mandatory and in urgent need.

6.5. Towards clinically applicable information on *S. haematobium* infection

Translation research is by definition "a way of thinking about and conducting scientific research to make the results of research applicable to the population under study and is practiced in the natural and biological, behavioural, and social sciences" (Wikipedia, 2012). Translational research strives to transform scientific discoveries arising from laboratory, clinical, or population studies into clinical applications to reduce morbidity, and mortality (Woolf *et al*., 2008). Metabolomics are often also referred to as "translation research", meaning the translation of metabolomics information into practical applications. However, as stressed by Hu (2011), this "is not a straightforward process." Nevertheless, the challenge to use metabolomics as an approach to deepen the understanding of epidemiological diseases, like the neglected tropical diseases, requires more and improved experimental investigations.

The investigation of metabolic profiling to enhance our understanding of the biological responses to parasitic infections has been accelerating at a remarkable pace. It holds promise as a basis for novel diagnostic tests with high sensitivity and specificity and for improved disease surveillance. "It is thus increasingly important for researchers, manufacturers, regulators, and clinicians to critically appraise the value of new biomarkers as they emerge as candidates for further investigation and possible clinical application (Morrow and de Lemos 2007)."

Benchmarks for the assessment of the applicability of biomarker findings

Since the use of high-throughput screening of metabolites has led researchers to the discovery of various metabolic signatures that have the potential to be developed into robust biomarkers that can be applied in the clinic, the following three fundamental questions, formulated by Morrow and de Lemos (2007) must then be addressed in all metabolomics investigations aimed towards practical applications (Mamas *et al*., 2010):

1. **"*Can the clinician measure it?*"**

   In general a biomarker is a substance that can be utilised medically as an indicator of a biological state. Therefore, in order for the discovered biomarker to be used practically and regularly in clinics it is required that its metabolomic analytical technique should be easily available, allow for accurate quantification of the metabolite with optimum capabilities for high-throughput, hold a quick turn-around time and must be inexpensive. Since the discovered metabolomic biomarkers could be utilised in general public places like hospitals and clinics, then the evaluation of new biomarker properties and its assay(s) require that it undergo stringent and thorough examinations of its pre-analytical and analytical performance which includes sample handling, measurement conditions and inspection of sample type.

2. **"*Does it add new information?*"**

   The influence of the association between the novel biomarker and its outcome or the disease of interest, its consistency and the degree to which can be an improvement on the already existing established methods, by either contributing to them or by replacing them, is a fundamental criterion with regards to the practice of the candidate biomarker and its potential value to the clinic. For this reason, with biomarkers external validation is a critical step for its path toward clinical integration.

3. ***"Will it help the clinician to manage patients?"***

   Biomarkers have the ability of providing a variety of possible clinical applications that may enhance the care of patients such as: they can reflect the entire spectrum of diseases showing changes from early disease states to terminal stages; diagnostic assessment of an acute or chronic clinical syndrome; provide risk assessments of patients with the clinical syndrome; selection of an appropriate treatment intervention and monitoring the response to treatment. The above mentioned potential uses of biomarkers merit deliberation in the evaluation of a new biomarker's value and therefore, steady progresses in the discovery of new biomarkers and development towards more sophisticated clinical application offers promising possibilities in enhancing the care of patients.

Proposals for future studies

According to the WHO (2012), the target of controlling, eliminating and eradicating neglected tropical diseases has gathered significant momentum over recent years. However, until the current state of safe water provision, sanitation and hygiene improves, many neglected tropical diseases and other communicable diseases will not be eliminated and certainly not eradicated (WHO, 2012).

The current approaches for schistosomiasis control are those that have been recommended by the Ministry of Health (MoH) and the WHO of which they are similar in various important aspects: *"(i) emphasis on preventive measures such as health education, safe-water supply, and sanitation; (ii) snail control only as an auxiliary measure; (iii) close interaction among the national, regional, and local levels of health organs; and (iv) local availability of praziquantel for the cases diagnosed through the health service*s" (World Health Assembly, 2001).

From the metabolomics investigation it is clear that an infection with *S. haematobium* induces changes in the concentration of a range of metabolites in urine. To be useful as an "authentic" biomarker, the metabolic candidate must be reproducible, robust, specific and preferably easy to measure (Bonassi *et al*., 2001). In the current investigation of *S.*

haematobium, to assess the specificity of the biomarkers indentified (Table 5.4) for potential diagnosis of infection, the obtained metabolic signatures were compared to altered metabolic patterns associated with other parasite-rodent models. This was the basis of our third aim.

For future studies I finally recommend:

(1) Extending the approach of the present investigation to other host-parasite models, e.g. hamster or mice infected with *S. haematobium* and applying complementary metabolic profiling methods such as gas chromatography time-of-flight (GC-TOF) analysis or ultra performance liquid chromatography (UPLC) in combination with mass spectrometry (MS) to help confirm the specificity of the metabolic perturbations associated with an *S. haematobium* infection.

(2) Improving the case-control approach as indicated above and expand it to a cohort-control study of humans infected with *S. haematobium* with an improved case selection and still applying high-throughput screening analytical methods.

(3) For future studies I propose a model (Figure 6.3), based on a flow chart of Yoshida *et al.* (2012). The proposed future study is based on the discovery of the small scaled cohort study where repeatability was tested, the intra-variance of the metabolites investigated by including QCs and by eliminating samples presenting with the same profiles as the infected group (MC).

The future study must have a well established experimental design. Better sample collection, must be practised from a larger scaled cohort study and a reputable metabolome analysis system that is both robust and repeatable should be employed.

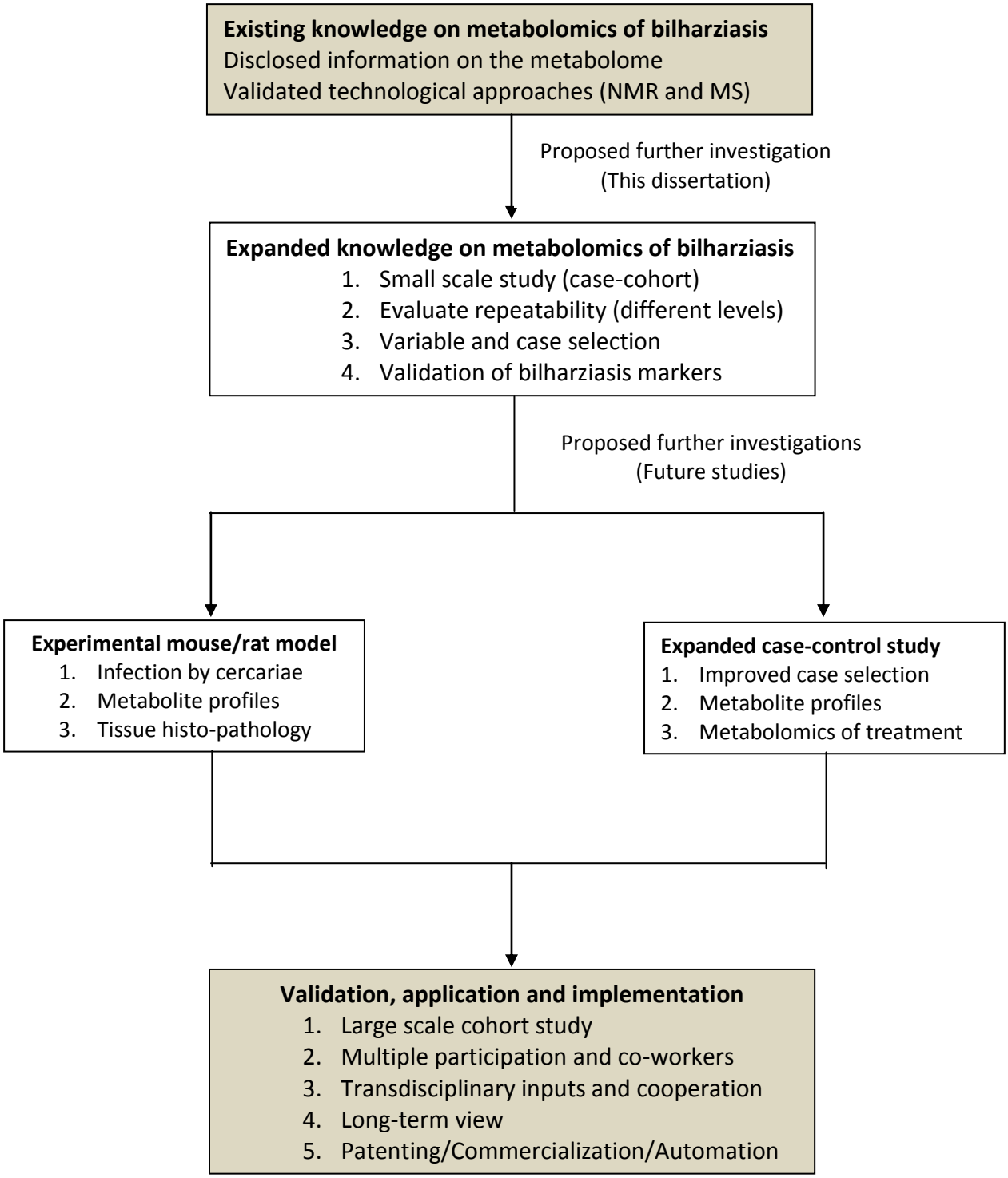


Figure 6.3 Proposed future studies flow chart.

REFERENCES

Ahmed, S. A. & Gad, M. Z. (1995). Effect of schistosomal infection and its treatment on some key enzymes of glucose metabolism in mice livers. *Arzneimittel-Forschung,* 45(12):1324-1328.

Allen, J., Davey, H. M. & Broadhurst, D. (2003). High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature Biotechnology*, 21:692-696

Appleton, C. C. (1977). The influence of above-optimal constant temperatures South African *Biomphalaria pfeifferi* (Mollusca: Planorbidae). *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 71:140–143.

Balog, C. I., Meissner, A., Göraler, S., Bladergroen. M. R., Vennervald, B. J., Mayboroda, O. A. & Deelder, A. M. (2011). Metabonomic investigation of human *Schistosoma mansoni* infection. *Molecular BioSystems*, 7(5):1473-80.

Basant, A. Rege, M. Sharma, A. & Sonawat, H. (2010). Alterations in urine, serum and brain metabolomic profiles exhibit sexual dimorphism during malaria disease progression. *Malaria Journal*, 9:1-10

Baum, C. E., Price, D. K. & Figg, W. D. (2010). Sarcosine as a potential prostate cancer biomarker and therapeutic target. *Cancer Biology & Therapy,* 9(5):341-342.

Berry, A. (1966). A cytopathological and histopathological study of bilharziasis of the female genital tract. *Journal of Pathology and Bacteriology*, 91:325-338.

Bichler, K. H., Savatovsky, I., Members of the Urinary Tract Infection (UTI) Working Group of the Guidelines Office of the European Association of Urology (EAU): Naber, K. G, Bischop, M. C., Bjerklund-Johansen, T. E., Botto, H., Cek, M., Grabe, M., Lobel, B., Redorta, J. P & Tenke, P. (2006). "EAU guidelines for the management of urogenital schistosomiasis." *European urology,* 49(6):998-1003.

Bilharz, T. (1853). A study on human helminthography, derived from information by letter from Dr. Bilharz in Cairo, along with remarks by Prof. Th. V. Siebold in Breslau. *Zertschiriftfür Wissenschartliche Zoologie*, 4:53-71.

Bino, R., Hall, R., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B. J., Mendes, P., Roessner-Tunali, U., Beale MH, Trethewey, R. N., Lange, B. M., Wurtele, E. S.& Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*, 9:418-425.

Biomarkers Definitions Working Group (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*, 69(3):89–95.

Birkemeyer, C., Kolasa, A. & Kopa, J. (2003). Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. *Journal of Chromatography A*, 993:89-102.

Birkholz, L. van Brummelen, A. C., Clark, K., Niemand, J. Maréchal, E. Llinás, M. & Louw, A. I. (2007). Exploring functional genomics for drug target and therapeutics discovery in *Plasmodia*. *Acta Tropica,* 105:113-123.

Bonassi, S. Neri, R. & Puntoni, R. (2001). Validation of biomarkers as early predictors of disease. *Mutation Research,* 480-481:349-358.

Brereton, R. G. (2003). Chemometrics – data analysis for the laboratory and chemical plant. John Wiley & Sons Ltd, West Sussex, England, 246.

Bro, R. & Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, 17:16-33.

Carroll, A., Badger, R. M. & Millar, A. H. (2010). The metabolome express project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. *BMC Bioinformatics*, 11(376)1-13.

Chen, J., Meng, C., Marayan, S. B., Luan, W. & Bennett, M. J. (2009). The use of deconvolution reporting software and backflush improves the speed and accuracy of data processing for urinary organic acid analysis. *Clinica Chimica Acta*, 405:53-59.

Chernecky, C. C. & Berger, B. J. (2001). Laboratory Tests and Diagnostic Procedures, 3rd ed. Philadelphia, PA: W. B. Saunders Company.

Chong, I. G. & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems,* 78(1-2):103-112.

Claudino, W. M., Quattrone, A., Biganzoli, L., Pestrin, M., Bertini, I. & Di Leo, A. (2007). Metabolomics: available results, current research projects in breast cancer, and future applications. Journal of Clinical Oncology, 25:2840–2846.

Cohen, J. (1988). Statistical power analysis for behavioural sciences. Second edition. Hillsdale. New Jersey: Erlbaum.

Cox, F. E. G. (2002). History of Human Parasitology. *Clinical Microbiology Reviews*, 5(4):595-612.

Crompton, D. W. (1999). "How Much Helminthiasis Is There in the World?" *Journal of Parasitology,* 85:397–403.

Dalgliesh, C. E., Horning, E. C., Horning, M. G., Knox, K. L. & Yarger, K. (1966). A gas-liquid-chromatographic procedure for separating a wide range of metabolites occurring in urine or tissue extracts. *Biochemical Journal,* 101:792-810.

Dasgupta, A. (1998). Gas chromatography-mass spectrometric identification and quantification of aniline after extraction from serum and derivatization with 2,2,2 trichloroethyl chloroformate, a novel derivative. *Journal of Chromatography B*, 716:354–358.

Delaney, J., Neville, W. A., Swain, A., Miles, A., Leonard, M. S. & Waterfield, C. J. (2004). Phenylacetylglycine, a putative biomarker of phospholipidosis: its origins and relevance to phospholipid accumulation using amiodarone treated rats as a model. *Biomarkers*. 9(3):271-290.

Despommier, D. D., Gwadz, R. W., Hoetz, P. J. & Knirsch, C. A. (2005). Parasitic Diseases 5th ed. LLC, New York: Apple Trees.

Doenhoff, M. J., Chiodini, P. L. & Hamilton, J. V. (2004). Specific and sensitive diagnosis of schistosome infection: can it be done with antibodies? *Trends in Parasitology*, 20:35–39.

Doumenge, J., Mott, K. E., Cheung, C., Villenave, D., Chapuis, O., Perrin, M. F., Reaud-Thomas, G. (1987). Centre d'Etude de Geographie Tropicale:WHO Atlas of the Global Distribution of Schistosomiasis. Bordeaux: Presses Universitaires de Bordeaux.

Dumas, M. E., Maibaum, E. C., Teague, C., Ueshima, H., Zhou, B., Lindon, J. C., Nicholson, J. K., Stamler, J., Elliott, P., Chan, Q. & Holmes, E. (2006). Assessment of Analytical Reproducibility of ^{1}H NMR Spectroscopy Based Metabonomics for Large-Scale Epidemiological Research: the INTERMAP Study. *Analytical Chemistry,* 78:2199-2208.

Dunn, W. B., Bailey, N. J. C. & Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *Analyst,* 130:606–625.

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B. & Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 6(7):1060-1083.

Ebbels, T. M. D., Keun, H. C., Antti, H., Bollard, M. E., Beckonert, O., Holmes, E., Lindon J. C. & Nicholson, J. K. (2003). Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Analytica Chimica Acta,* 490:265-276.

Eisenhofer, G. (1994). Plasma normetanephrine for examination of extraneuronal uptake and metabolism of noradrenalin in rats. *Naunyn-Schmiedeberg's Archives of Pharmacology,* 349:259–269.

Eisenhofer, G., Aneman, A., Hooper, D., Rundqvist, B., & Friberg, P. (1996). Mesenteric organ production, hepatic metabolism and renal elimination of norepinephrine and its metabolites in humans. *Journal of Neurochemistry.* 66:1565–1573.

Eltawil, K. M. & Plassmann, M. (2011). Theodor Maximillian Bilharz (1825-1862): the discoverer of Schistosomiasis. *International Journal of Parasitology Research*, 3(1):17-25.

FDA. (2001). *Guidance for Industry, Bioanalytical Method Validation*. Food and Drug Administration, Centre for Drug Evaluation and Research (CDER).

Fernandes, J., Saudubray, J. M., Van den Berghe, G. & Walter, J.H. (2006). Inborn Metabolic Diseases. Fourth Revised Edition. Medizin: Springer Verlag.

Fiehn, O. (2002). Metabolomics–the link between genotypes and phenotypes. *Plant Molecular Biology,* 48:155–171.

Fiehn, O. (2007). Validated high quality automated metabolome analysis of *Arabidopsis thaliana* leaf disks. (*In* Nikolau, B. J. & Wurtele, E. S. eds. Concepts in plant metabolomics. Springer).

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R. N. & Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nature Biotechnology,* 18:1157-1161

Fluss, R., Faraggi, D. & Reiser, B. (2005). Estimation of the youden index and its associated cutoff point. *Biometrical Journal,* 47(4):458-472.

Frank, R. & Hargreaves, R. (2003). Clinical biomarkers in drug discovery and development. *Nature Reviews*. 2:566-580.

Garcia-Pérez, I., Couto, A. A., Angulo, S., Li, J. V., Utzinger, J., Ebbels, T. M., Legido-Quigley, C., Nicholson, J. K., Holmes, E. & Barbas, C. (2010). Bidirectional correlation of NMR and capillary electrophoresis fingerprints: a new approach to investigating *Schistosoma mansoni* infection in a mouse model. *Analytical Chemistry*, 82:203-210.

García-Pérez, I., Whitfield, P., Bartlett, A., Angulo, S., Legido-Quigley, C., Hanna-Brown, M. & Barbas, C. (2008). Metabolic fingerprinting of *Schistosoma mansoni* infection in mice urine with capillary electrophoresis. *Electrophoresis,* 29(15):3201-3206.

Garrett, R. H. & Grisham, C. M. (2005). Biochemistry. 3rd ed. Singapore: Thomson.

Gear, J. H. S. & Wolsteholme, B. (1966). Bilharzia in South Africa with special reference to the laboratory diagnosis, South African Institute for Medical Research, Johannesburg. *South African Medical Journal*: 877-880.

Gear, J. H. S., Pitchford, R.J. & van Eeden, J. A. (1980). Atlas of Bilharzia. Johannesburg, South Africa Institute for Medical Research. *South African Medical Journal,* 48 maps.

Gerritsen, T., Walker, A. R. P., De Meillon, B. & Yeo, R. M. (1953). *Transactions of Royal Society of Tropical Medicine and Hygiene*, 47:137.

Gika, H. G., Theodoridis, G. A, Wingate, J. E. & Wilson, I. D. (2007). "Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine". *Journal of Proteome Research,* 6(8):3291-3303.

Gika, H. G., Theodoridis, G. A., & Wilson, I. D. (2008). Liquid chromatography and ultra-performance liquid chromatography– mass spectrometry fingerprinting of human urine—Sample stability under different handling and storage conditions for metabonomics studies. *Journal of Chromatography A*, 1189:314–322.

Goodacre, R. (2010). An overview of what else but metabolism. *Metabolomics*, 6:1-2.

Goodacre, R., Broadhurst, D., Smilde, A. K., Kristal, B. S., Baker, J. D., Beger, R., Bessant, C., Connor, S., Capuani, G., Craig, A., Ebbels, T., Kell, D. B., Manetti, C., Newton, J., Paternostro, G., Somorjai, R., Sjo¨stro¨m, M., Trygg, J. & Wulfert, F. (2007). Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, 3:231–241.

Griffin, J. L. (2006). The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philosophical Transactions of the Royal Society B,* 361:147-161.

Grover-Kopec, E. K., Blumenthal, M. B., Ceccato, P., Dinku, T., Omumbo, J. A. & Connor, S. J. (2006). "Web-based climate information resources for malaria control in Africa." *Malaria journal,* 5(1):38.

Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. (2006). Human schistosomiasis. *Lancet*, 368:1106-1118.

Hagan, P., Blumenthal, U. J., Dunn, D., Simpson, A. J. G. & Wilkins, H. A. (1991). Human IgE, IgG4 and resistance to reinfection with Schistosoma haematobium. *Nature*, 349:243-245.

Halket, J. M., Waterman, D., Przyborowska, A. M., Patel, R. K., Fraser, P. D. & Bramley, P. M. (2005). Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *Journal of Experimental Botany*, 56:219–243.

Hamilton, J. V., Klinkert, M. & Doenhoff, M. J. (1998). Diagnosis of schistosomiasis: anti- body detection, with notes on parasitological and antigen detection methods. *Parasitology*, 117:41–57.

Harrigan, G., Goodacre, R. (2003). Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis. Boston: Kluwer Academy.

Hattingh, Z., Walsh, C., Veldman, F. J., & Bester, C. J. (2009). Metabolic profile of HIV-infected woman in Mangaung, Free State. *South African Journal of Clinical Nutrition,* 22:23-28.

Hewer, R., Vorster, J., Steffens, F. E. & Meyer, D. (2006). Applying biofluid ^{1}H NMR-based metabonomic techniques to distinguish between HIV-1 positive/AIDS patients on antiretroviral treatment and HIV-1 negative individuals. *Journal of Pharmaceutical and Biomedical Analysis*, 41:1442-1446.

Hodder, S. L., Mahmoud, A. A. F., Sorenson, K., Weiner, D, M., Stein, R. I., Ouma, J. H., Koech, D. & King, C. H. (2000). Predisposition of to urinary tract epithelial metaplasia in *Schistosoma haematobium* infection. *The American Journal of Tropical Medicine and Hygiene,* 63(3, 4):133–138.

Hoffman, G. F. & Feyh, P. (2005). Organic acid analysis. (*In*: Blau, N., Duran, M. & Blaskovics, M. E. eds. Physician's guide to the laboratory diagnosis of metabolic diseases. Revised 2nd ed. Heidelberg, Springer-Verlag, p. 27-44).

Holmes, E., Li, J. V., Athanasiou, T. Ashrafian, H. & Nicholson, J. K. (2011). Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends in Microbiology*, 19(7):349-359.

Hotez, O., Ottesen, E., Alan, F. & Molyneux, D. (2006). The neglected tropical diseases: The ancient affliction of stigma and poverty and the prospects for their control and elimination. *Hot topics in infection and immunity in children*, Springer, New York: 23-33.

Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D. & Savioli, L. (2007). Control of Neglected Tropical Diseases. *The New England and Journal of Medicine*, 357(10):1018-1027.

Hu, F. B. (2011). Metabolic profiling of diabetes: from black-box epidemiology to system epidemiology. *Clinical Chemistry*, 57(9):1224-1226.

Hu, J., Coombes, K. R., Morris, J. S. & Baggerly, K. A. (2005). The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. *Briefings in functional genomics and proteomics*, 3:322-331.

Jackson, J. E. (1991). A user's guide to principal components John Wiley & Sons, Inc.

Jamison, D. T., Breman, J. G., R.Measham, A., Alleyne, G., Claeson, M., Evans, D. B., Jha, P., Mills, A. & Musgrove, P. (2006). *Disease Control Priorities in Developing Countries*, 2nd ed. New York: Oxford University Press.

Jansson, J. Willing, B., Lucio, M. Fekete, A. Dicksved, J., Halfvarson, J., Tysk, C. & Schmitt-Kopplin, P. (2009). Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE,* 4(7):1-10.

Joliffe, A. (1986). Principal Components Analysis. New York: Springer.

Jonsson, P., Bruce, S. J., Moritz, T., Trygg, J., Sjöström, M., Plumb, R., Granger, J., Maibaum, E., Nicholson, J. K., Holmes, E. & Antti, H. (2005). Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst*, 130:701-707.

Katajamaa, M. & Oresic, M. (2007). Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A*, 1158:318-328.

Keiser, J., N'Goran, E. K., Singer, B. H., Lengeler, C., Tanner, M. & Utzinger, J. (2002). Association between *Schistosoma mansoni* and hookworm infections among schoolchildren in Côte d'Ivoire. *Acta Tropica*, 84(1):31-41.

Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. *Current Opinion in Microbiology*, 7:296-307.

Kell, D. B. (2006). Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discovery Today*, 11(23-24):1085-1092.

Keun, C. H., Ebbels, T. M. D., Antti, H., Bollard, M. E., Beckonert, O., Schlotterbeck, G., Senn, H., Niederhauser, U., Holmes, E., Lindon, J. C. & Nicholson, J. K. (2002). Analytical reproducibility in ^{1}H NMR- based metabolomic urinalysis. *Chemical Research in Toxicology,* 15:1380-1386.

Khodakova, A. & Beloborodova, N. (2007). Microbial metabolites in the blood of patients with sepsis. *Critical Care*, 11(4):5.

King, C. H. (2009). Toward the Elimination of Schistosomiasis. *The New England Journal of Medicine*, 360(2):106-109.

King, C. H. & Dangerfield-Cha, M. (2008). The unacknowledged impact of chronic schistosomiasis. *Chronic Illness*, 4:65-79.

King, C. H., Dickman, K. & Tisch, D. J. (2005). Reassessment of the cost of chronic helmintic infection: a meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet*, 365:1561-1569.

Knapp, D. R. (1979). Handbook of analytical derivatization reactions. New York: Wiley.

Koekemoer, G. & Swanepoel, J. W. H. (2008). A semi-parametric method for transforming data to normality. *Statistics and Computing,* 18:241-257.

Kristal, B. S., Kaddurah-Daouk, R., Beal, M. F. & Matson, W. R. (2007). Metabolomics: Concepts and potential neuroscience applications. (*In*. Gibson, G. E. and Dienel, G., Eds. Handbook of Neurochemistry and Molecular Biology, 3rd Edition. Brain energetics from genes to metabolites to cells: Integration of molecular and cellular processes, in press 5:889-912).

Krull, I. S. & Swartz, M. (1999). Analytical method development and validation for the academic researcher. *Analytical Letters,* 32:1067-1080.

Kuhara, T. (2005). Metabolome profiling of human urine with capillary gas chromatography/mass spectrometry. (*In* Tomita, M. & Nishioka, T., ed. Metabolomics the frontier of systems biology. Tokyo: Springer-Verlag).

Kusel, J. R. (1970). Studies on the structure and hatching of the eggs of *Schistosoma mansoni*. *Parasitology*, *60*:79–88.

Lauridsen, M., Hansen, S. H., Jaroszewski, J. W., & Cornett, C. (2007). Human urine as test material in H-1 NMR-based metabonomics: Recommendations for sample preparation and storage. *Analytical Chemistry*, 79:1181–1186.

Lengeler, C., Utzinger, J. & Tanner, M. (2002). Questionnaires for rapid screening for Schistosomiasis in sub-Saharan Africa. *Bulletin of the World Health Organization*, 80(3):235-242.

Lesage-Meessen, L., Delattre, M., Haon, M., Thibault, J. F., Ceccaldi, B. C., Brunerie, P. & Asther, M. (1996). A two-step bioconversion process for vanillin production from ferulic acid combining *Aspergillus niger* and *Pycnoporus cinnabarinus*. *Journal of Biotechnology*, 50:107-113.

Lis, A. W., McLaughlin, I., McLaughlin, R. K., Lis, E. W. & Stubbs, E. G. (1976). Profiles of ultraviolet-absorbing components of urine from autistic children, as obtained by high-resolution ion-exchange chromatography. *Clinical Chemistry*, 22:1528–1532.

Lord, S. R. & Bralley, A. J. (2008). Clinical applications of urinary organic acids part 2: dysbiosis markers. *Alternative Medicine Review*, 13(4)292-306.

Lu, H., Dunn, W. B., Shen, H., Kell, D. B. & Liang, Y. (2008). Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. *Trends in Analytical Chemistry*, 27(3):215-227.

Maher, A. D., Zirah, S. F. M., Holmes, E., & Nicholson, J. K. (2007). Experimental and analytical variation in human urine in H-1 NMR spectroscopy-based metabolic phenotyping studies. *Analytical Chemistry*, 79:5204–5211.

Mamas, M., Dunn, W. B., Neyses, L. & Goodacre, R. (2010). The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Archives of Toxicology,* 85:5-17.

Mamer, O. A., Tjoa, S. S., Scriver, C. R. & Klassen, G. A. (1976). Demonstration of a New Mammalian Isoleucine Catabolic Pathway Yielding an R Series of Metabolites. *The Journal of Biochemistry*, 160:417-426.

Marshall, A. (2004). Health Biotechnology Innovation in Developing Countries: Open secrets. *Nature Biotechnology*, 22:1.

Mason, S. W., Van Furth, M., Solomons, R., Reinecke, C., Van Reeneu, M. & Koekemoer, G. (2012). Assessment of analytical repeatability and reproducibility of mass spectrometry-based metabolomics data from tuberculosis infected cerebrospinal fluid. (In preparation).

Mehrotra, B. & Mendes, P. (2006). Bioinformatics approaches to integrate metabolomics and other systems biology data. (*In*: Saito, K, Dixon R. A, Willmitzer, L. eds. Plant Metabolomics; Vol. 57. Berlin, Heidelberg: Springer-Verlag).

Meurs, L., Mbow, M., Vereecken, K., Menten, J. Mboup, S. & Polman, K. (2012). Epidemiology of mixed Schistosoma mansoni and Schistosoma haematobium: infections in northern Senegal. *International Journal for Parasitology,* 42:305–311.

Milburn, M. (2006). Metabolomics for biomarker discovery. *Innovations in Pharmaceutical Technology,* 21:40-42.

Miyairi, K., & Suzuki, M. (1913). On the development of *Schistosoma japonicum. Tokyo Iji Shinshi Gokyo Medical Journal*, 1836:1–5.

Moodley, I., Kleinschmidt, I., Sharp, B., Craig, M. & Appleton C. (2003). Temperature-suitability maps for schistosomiasis in South Africa. *Annals of Tropical Medicine and Parasitology,* 97(6):617-627.

Moore, B. S., Hertweck, C., Hopke, J. N, Izumikawa, M. J., Kalaitzis, A., Nilsen, G., O'Hare, T., Piel, J., Shipley, P. R., Xiang, L., Austin, M. B. & Noel, J. P. (2002). Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. *Journal of Natural Products*, 65:1956–1962.

Morrow, D. A. & de Lemos, L. A. (2007). Benchmarks for the assessment of novel cardiovascular biomarkers. *Circulation*, 115:949-952.

Mostafa, H. M., Sheweita, S. A. & O'Connor, P. J. (1999). Relationship between schistosomiasis and bladder cancer. *Clinical Microbiology Reviews*, 12(1):97-111.

Motari, M., Quach, U., Thorsteinsdóttir, H., Martin, D. K., Daar, A. S. & Singer, P. A. (2004). Health Biotechnology Innovation in Developing Countries, South Africa – blazing a trail for African biotechnology. *Nature Biotechnology*, 22:37-41.

Mott, K. E., Baltes, R., Bambagha, J. & Baldassini, B. (1982). Field studies of a reusable polyamide filter for detection of Schistosoma haematobium eggs by urine filtration. *Tropenmed* Parasitol, 33:227–228.

Mqoqi, N. P., Appleton, C. C. & Dye, A. H. (1996). Prevalence and intensity of Schistosoma haematobium in the Port St John's district. *South African Medical Journal*, 86:76–80.

Mukerjee, S., Chander, R., Tekwani, B. L., Gupta, S., Katiyar, J. C., Shukla, O. P. & Kapoor, N. K. (1990). Molecular basis of hyperlipidemia in golden hamsters during experimental infection with *Ancylostoma ceylanicum* (Nematoda:Strongylidae). *International Journal of Parasitology*, 20(2):217-223.

Murray, C. J. L. (1996). Rethinking DALY's. (In: *The Global Burden of Disease.* Murray, C. J. L., and Lopez, A. D., eds. Global burden of disease and injury series, Harvard School of Public Health on behalf of the WHO and World Bank, Harvard University Press).

Nicholls, A. W., Nicholson, J. K., Haselden, J. N., & Waterfield, C. J. (2000). A metabonomic approach to the investigation of drug induced phospholipidosis: a NMR spectroscopy and pattern recognition study. *Biomarkers,* 5:410-423.

Nicholson, J. K. (2006). Global systems biology, personalized medicine and molecular epidemiology. *Molecular Systems Biology*, 2:52.

Nicholson, J. K., Lindon, J. C. & Holmes, E. (1999). Metabonomics : understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29:1181-1189.

Nicholson, J. K.; Holmes, E. & Wilson, I. D. (2004). Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews Microbiology*, 3 (5):431–438.

Nicholson, J.K., Connelly, J., Lindon, J. C., & Holmes, E. (2002). Metabonomics: a platform for studying drug toxicity and gene function. *Nature Reviews Drug Discovery*, 1 (2):153-161

Oliver, I. & Loots, D. T. (2011). An overview of tuberculosis treatments and diagnostics: what role could metabolomics play? *Journal of Cell and Tissue Research*, 11(1):2655-2671.

Oliver, S. G., Winson, M. K. & Kell, D. B. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology*, 16:373–278.

Olszewski, K. L., Morrisey, J. M., Wilinski, D., Burns, J. M., Vaidya, A. B., Rabinowitz, J. D. & Llinás, M. (2009). Host-parasite interactions revealed by *Plasmodium falciparum* metabolomics. *Cell Host & Microbe*, 5:191-199.

Parida, S. K. & Kaufmann, S. H. E. (2010). The quest for biomarkers in tuberculosis. *Drug Discovery Today*, 15(3/4):148-157.

Peeling, R. W., Smith, P. G. & Bossuyt, P. M. (2006). A guide for diagnostic evaluations. *Nature Reviews Microbiology*, 4 (9):2–6.

Pinti, M., Nasi, M., Gibellini, L., Roat, E, De Biasi, S., Bertoncelli, L. & Cossarizza, A. (2010). The role of mitochondria in HIV infection and its treatment. *Journal of Experimental and Clinical Medicine,* 2:145-155.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., Berden, J. A., Brindle, K. M., Kell, D. B., Rowland, J. J., Westerhoff, H. V., van Dam, K. & Oliver, S. G. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*, 19:45-50.

Redestig, H., Fukushima, A., Stenlund, H., Moritz, T., Arita, M., Saito, K. & Kusano, M. (2009). Compensation for systematic cross-contribution improves normalization of mass spectrometry based metabolomics data. *Analytical Chemistry*, 81:7974–7980.

Reinecke, C. J., Koekemoer, G., & van der Westhuizen, F. H., Louw, R., Lindeque, Z. J., Mienie, J. L. & Smuts, I. (2011). Metabolomics of urinary organic acids in respiratory chain deficiencies in children. *Metabolomics*:1-20.

Rinaldi, G., Okatcha, T. I., Popratiloff, A., Ayuk, A. M., Suttiprapa, S., Mann, V. H., Liang, Y. S., Lewis, F. A., Loukas, A. & Brindley, P. J. (2011). Genetic manipulation of *Schistosoma haematobium*, the neglected schistosome. *PLoS Neglected Tropical Diseases*, 5(10):1-10.

Robosky, L. C., Robertson, D. G., Baker, J. D., Rane, S. & Reily, M. D. (2002). In vivo toxicity screening programs using metabonomics. *Combinational Chemistry and High Throughput Screening*, 5(8):651-662

Rocke, M. D. & Lorenzato, S. (1995). A two-component model for measurement error in analytical chemistry. *Technometrics*, 37(2):176-184.

Rollinson, D. (2009). A wake up call for urinary schistosomiasis: reconciling research effort with public health importance. *Parasitology*, 136:1593–1610.

Rollinson, D. & Southgate, V. R. (1987). The genus *Schistosoma*: a taxonomic appraisal. (*In* Rollinson, D. & Simpson, A. J. G., eds, The Biology of Schistosomes. From Genes to Latrines, London, Acadamic Press, p. 1-49).

Ross, G., Bartley, P.B., Sleigh, A.C., Olds G. R., Li, Y., Willliams, G. M. & McCanus, D. P. (2002). Schistosomiasis. *The New England Journal of Medicine,* 346:1212–1220.

Rumnajek, F. D. (1987). Biochemistry and physiology. (*In* Rollinson, D. & Simpson, A. J. D., eds, The Biology of Schistosomes. From Genes to Latrines, London, United Kingdom: Academic Press).

Saathoff, E., Olsen, A., Magnussen P., Kvalsvig, J. D., Becker, W. & Appleton, C. C. (2004). Patterns of Schistosoma haematobium infection, impact of praziquantel treatment and re-infection after treatment in a cohort of schoolchildren from rural KwaZulu-Natal/South Africa. *BMC Infectious Diseases*, 4(40):1-10.

Sanchez, D. H., Szymanski, J., Erban, A., Udvardi, M. K. & Kopka, J. (2010). Mining for robust transcriptional and metabolic responses to long-term salt stress: A case study on the model legume *Lotus japonicus*. *Plant Cell and Environment*, 33:468–480.

Sandbach, F. R. (1976). The history of schistosomiasis research and policy for its control. *Medical History*, 20:259-275.

Saric, J., Li, V.J., Wang, Y., Keiser, J., Bundy, J. G., Holmes, E. & Utzinger, J. (2008). Metabolic profiling for an *Echinostoma caproni* infection in the mouse for biomarker discovery. *PLoS Neglected Tropical Diseases*, 2(7):1-15.

Schutte, C. H. J., Fripp, P. J. & Evans, A. C. (1995). An assessment of the schistosomiasis situation in the Republic of South Africa. *Southern African Journal of Epidemiology and Infection*, 10:37–43.

Sekirov, I., Russel, L. S., Antunes, M. C. L. & Finlay, B. B. (2010). Gut microbiota in health and disease. *Physiological Reviews*, 90:859-904.

Shaw, W. (2010). Increased urinary excretion of a 3-(3-hydroxyphenyl)-3-hydroxypropionic acid (HPHPA), an abnormal phenylalanine metabolite of *Clostridia spp*. in the gastrointestinal tract, in urine samples from patients with autism and schizophrenia. *Nutritional Neuroscience*, 13(3):135-143.

Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society B*, 53(3):683-690.

Shulaev, V. (2006). Metabolomics technology and bioinformatics. *Briefings in Bioinformatics*, 7(2):128-139.

Shurubour, Y. I., Paolucci, U., Krasnikov, B. F., Matson, W. R. & Kristal B. S. (2005). Analytical precision, biological variation, and mathematical normalization in high data density metabolomics. *Metabolomics*, 1:75-85.

Silverman, B. W. & Ramsay, J. O. (2005). Functional Data Analysis. 2nd ed. New York: Springer.

Socha, E., Kałużna-Czaplińska, J & Rynkowski, J. (2010). Determination of homovanillic acid and vanillylmandelic acid in urine of autistic children by gas chromatography/mass spectrometry. *Medical Science Monitor*, 16:445-450.

Solanky, K. S., Bailey, N. J., Beckwith-Hall, B. M., Davis, A., Bingham, S., Holmes, E., Nicholson, J. K. & Cassidy, A. (2003). Application of biofluid 1H nuclear magnetic resonance-based metabonomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. *Analytical Biochemistry*, 323(2):197-204.

Stein, S. E. (1999). An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry,* 10:770-781.

Sumiyoshi, T., Yotsutsuji, T., Kurachi, M., Itoh, H., Kurokawa, K. & Saitoh, O. (1998). Effect of mental stress on plasma homovanillic acid in healthy human subjects. *Neuropsychopharmacology*, 19(1):70-73.

Sysi-Aho, M., Katajamaa, M., Yetukuri, L., & Oresic, M. (2007). Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC Bioinformatics*, 8:93.

Tanaka, K., Budd, M. A., Efront, M. L. & Isselbacher, K. J. (1966). Isovaleric acidemia: A new genetic defect of leucine metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 56:236-242.

Taylor, J., King, R. D., Altmann, T. & Fiehn, O. (2002). Application of metabolomics to plant genotype discrimination using statistics and machine learning. *Bioinformatics*, 18(2):241-248.

Taylor, M. G. (1987). Schistosomes of domestic animals: Schistosoma bovis and other animal forms. (*In*: Soulsby, E.J .L., ed., Immunology, Immunoprophylaxis and Immunotherapy of Parasitic Infections, Boca Raton, FL, CRC Press, p. 49-90).

Turnbull, D. (2011). A new biomarker for mitochondrial disease. *The Lancet Neurology,* 10(9):777-778.

Tyagi, S., Raghvendra, U. S., Kalra, T. & Munjal K. (2010). Applications of metabolomics a systemic study of the unique chemical fingerprints: an overview. *International Journal of Pharmaceutical Sciences Review and Research,* 3(1):83-86.

US Centers for Disease Control and Prevention. (2008). Schistosomiasis. Health Information for Overseas Travel. Elsevier: Atlanta:297300.

Utzinger, J., N'Goran, E. K., Caffrey, C. R. & Keiser, J. (2010). From innovation to application: Social–ecological context, diagnostics, drugs and integrated control of schistosomiasis. *Acta Tropica,* 120 (1):121-137.

Utzinger, J., Zhou, X. N., Chen, M. G. & Bergquist, R. (2005). Conquering schistosomiasis in China: the long march. *Acta Tropica*, 96:69–96.

Van Batenburg, M. F., Coulier, L., van Eeuwijk, F., Smilde, A. K. & Westerhuis, A. J. (2011). New figures of merit for comprehensive functional genomics data: the metabolomics case. *Analytical Chemistry*, 83:3267-3274.

Van Dam, G.J., Wichers, J.H., Falcao Ferreira, T.M., Ghati, D., van Amerongen, A. & Deelder, A.M. (2004). Diagnosis of schistosomiasis by reagent strip test for detection of circulating cathodic antigen. *Journal of Clinical Microbiology*, 42:5458–5461.

Van de Kloet, M. F., Bodeldijk, I., Verheij, R. E., & Jellema, R. H. (2009). Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *Journal of Proteome Research*, 8:5132-5141.

Van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7:142.

Van der Greef, J., Stroobant, P. & van de Heijden, R. (2004). The role of analytical sciences in medical systems biology. *Current Opinion in Chemical Biology*, 8:559-565.

Van der Greef, J., Tas, A. C., Bouwman, J. & ten Noever de Brauw, M. C. (1986). Pattern recognition of complex matrix profiles generated by soft ionization methods. *Advances in Mass Spectrometry*, 10:1227-1228.

Van der Heyden, J. C., Rotteveel, J. J. & Wevers, R. A. (2003). Decreased homovanillic acid concentrations in cerebrospinal fluid in children without a known defect in dopamine metabolism. *The European Journal of Paediatric Neurology*, 7:31–37.

Van der Werf, M. J., de Vlas, S. J., Brooker, S., Looman, C. W., Nagelkerke, N. J., Habbema, J. D. & Engels, D. (2003). "Quantification of Clinical Morbidity Associated with Schistosome Infection in Sub- Saharan Africa." *Acta Tropica,* 86:125–39.

Van der Werf, M. J., Jellema, R. H. & Hankemeier, T. (2005). Microbial Metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *The Journal of Industrial Microbiology and Biotechnology*, 32:234-252.

Van Lieshout, L., Polderman, A. M. & Deelder, A. M. (2000). Immunodiagnosis of schisto- somiasis by determination of the circulating antigens CAA and CCA, in particular in individuals with recent or light infections. *Acta Tropica*, 77: 69–80.

Villas-Bôas, S. G., Mas, S., Akesson, M., Smedsgaard, J. & Nielsen, J. (2005). Mass spectrometry in metabolome analysis. *Mass Spectrometry Reviews*, 24(5):613-646.

Vinayavekhin, N., Homan, E. A. & Saghatelian, A. (2009). Exploring disease through metabolomics. *ACS Chemical Biology*, 5(1):91-103.

Vladimir S. (2006). Metabolomics technology and bioinformatics. *Briefings in Bioinformatics*, 7 (2):128-139.

Wang, Y., Holmes, E., Nicholson, J. K., Cloarec, O., Chollet, J., Tanner, M., Singer, B. H. & Utzinger, J. (2004). Metabonomic investigations in mice infected with *Schistosoma mansoni*: an approach for biomarker identification. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34):12676-12681.

Wang, Y., Utzinger, J., Xiao, S. H., Xue, J., Nicholson, J. K., Tanner, M., Singer, B. H. & Holmes, E. (2005). System level metabolite effects of a *Schistosoma japonicum* infection on the Syrian hamster. *Molecular and Biochemical Parasitology,* 146(1):1-9.

Wang, Y., Xiao, S. H., Xue, J., Singer, B. H., Utzinger, J. & Holmes, E. (2009). Systems metabolic effects of a *Necator americanus* infection in Syrian hamster. *Journal of Proteome Research*, 8(12):5442-5450.

Want, E. J., Nordström, A., Morita, H. & Siuzdak, G. (2007). From exogenous to endogenous: the inevitable imprint of mass spectrometry in metabolomics. *Journal of Proteome Research*, 6:459-468.

Want, E. J., O'Maille, G., Smith, C. A., Brandon, T. R., Uritboonthai, W., Qin, C., Trauger, S. A. & Siuzdak, G. (2006). Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Analytical Chemistry*, 78(3):743-752.

Want, J. E., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, S. R., Shockcor, J., Holmes, E. & Nicholson, J. K. (2010). Global metabolic profiling procedures for urine using UPLC-MS. *Nature Protocols*, 5(6):1005-1018.

Watson, W. (2007). Brick by brick: an informal guide to the history of South Africa. Claremont, South Africa: New Africa Books.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J. J., van Duijnhoven, J. P. M. & van Dorsten, F. A. (2008). Assessment of PLSDA cross validation. *Metabolomics,* 4:81-89.

Wevers, R. A., de Rijk-van Andel, J. F., Bra¨utigam, C., Geurtz, B., van den Heuvel, L. P., Steenbergen-Spanjers, G. C., Smeitink, J. A., Hoffmann, G. F. & Gabreëls, F. J. (1999). A review of biochemical and molecular genetic aspects of tyrosine hydroxylase deficiency including a novel mutation. *Journal of Inherited Metabolic Disease*, 22:364–73.

Wiener, M. C., Sachs, J. R., Deyanova, E. G. & Yates, N. A. (2004). Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Analytical Chemistry*, 76(20):6085-6096.

Wikipedia. (2012). Translational research. http://en.wikipedia.org/wiki/Translational_research Date of access: 27 March. 2012.

Wikoff, W. R., Gangoiti, J. A., Barshop, B. A. & Siuzdak, G. (2007). Metabolomics identifies perturbations in human disorders of propionate metabolism. Clinical Chemistry, 53:2169–2176.

Williams, A., Koekemoer, G., Lindeque, Z., Reinecke, C. J. & Meyer, D. (2011). Qualitative serum organic acid profiles of HIV-infected individuals not on antiretroviral treatment. *Metabolomics*:1-15

Williamson, G. & Clifford, N. M. (2010). Colonic metabolites of berry polyphenols: the missing link to biological activity? *British Journal of Nutrition*, 104:48-66.

Wilson, R. A. (1987). Cercariae to liver worms: development and migration in the mammalian host. (*In* Rollinson, D. & Simpson, A. J. G., ed. The Biology of Schistosomes. From Genes to Latrines, London, Acadamic Press, p. 115-146).

Wolmarans, C.T., De Kock, K.N., Potgieter, A. & Postma, S. (2006). Occurrence of urinary schistosomiasis in the greater Rustenburg area, North West province, an area not typical for endemic bilharzia. *South African journal of science,* 102(5):246-248.

Woolf, S. H. (2008). The meaning of translational research and why it matters. *JAMA*, 299(2):210-213.

World Health Assembly (WHA). (2001). World Health Assembly endorses WHO's strategic priorities, WHA-54.19 Geneva, World Health Organization.

World Health Organization (WHO). (1995). The control of schistosomiasis, second report of the WHO expert committee. Technical Report Series Geneva, World Health Organization.

World Health Organization (WHO). (2002). Prevention and Control of Schistosomiasis and Soil-transmitted Helminthiasis, Report of a WHO Expert Committee. Technical Report Series, No 912 Geneva, World Health Organization.

World Health Organization (WHO). (2011). Weekly epidemiological record, 25 February, 86th year. No. 9 Geneva, World Health Organization: 73-80.

World Health Organization (WHO). (2012). Accelerating work to overcome the global impact of neglected tropical diseases: A roadmap for implementation, Geneva, World Health Organization.

Wu, H., Xue, R., Dong, L., Liu, T., Deng, C., Zeng, H. & Shen, X. (2009) Metabolomic profiling of human urine in hepatocellular carcinoma patients using gas chromatography/mass spectrometry. *Analytica Chimica Acta*, 648:98-104.

Wu, J. F., Holmes, E., Xue, J., Xiao, S. H., Singer, B. H., Tang, H. R., Utzinger, J. & Wang, Y. L. (2009). Metabolic alterations in the hamster co-infected with *Schistosoma japonicum* and Necator americanus. *The International Journal for Parasitology,* 40(6):695-703.

Wu, J., Xu, W., Ming, Z., Dong, H., Tang, H. & Wang, Y. (2010). Metabolic changes reveal the development of Schistosomiasis in mice. *PLoS Neglected Tropical Diseases,* 4(8):1-11.

Wu, Y., Wang, X., Liu, X. & Wang, Y. (2006). Data-mining approaches reveal hidden families of proteases in the genome of malaria parasites. *Genome Biology*, 13:601-616.

Yoshida, M., Hatano, N., Nishiumi, S. Irino, Y., Izumi, Y., Takenawa, T. & Azuma, T. (2012). Diagnosis of gastroenterological diseases by metabolome analysis using gas chromatography-mass spectrometry. *The Journal of Gastroenterology*, 47:9-20.

Zelena, E., Dunn, B. W., Broadhurst, D., Francis-Mclntyre, S., Carroll, K. M., Begley, P., O'Hagan, S., Knowles, J. D., Halsall, A.; HUSERMET Consortium, Wilson, I. D. & Kell, D. B. (2009). Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Analytical Chemistry*, 81:1357-1364.

Zhang, Q., Wang, G., Du, Y., Zhu, L. & Jiye, A. (2007). GC/MS analysis of the rat urine for metabolomic research. *Journal of Chromatography B*, 854:20-25.

Zheng, X., Xie, G., Zhao, A., Zhao, L., Yao, C., Chiu, N. H., Zhou, Z., Bao, Y., Jia, W., Nicholson, J. K. & Jia, W. (2011). The footprints of gut microbial mammalian co-metabolism. *Journal of Proteome Research,* 10:5512–5522.

APPENDIX

Marker validation

The 20 identified features (see Table 5.3) were extracted from the normalized data set, transformed, centred and analysed. The PCA and PLS-DA score plots are given in Figure 5.13, and the summary fit statistics for both the PCA and PLS-DA model are presented in Table 5.5. From the PCA score plots (both 3D and 2D) we still cannot obtain a mutually exclusive separation between the two experimental groups. However, when compared to Figure 5.9 (cases swept, all 83 features) the PCA based on the 20 features produces scores from which the separation is more clearly defined in the direction of the first principal component (which is the component that explains the most variation). The score plots obtained from the PLS-DA model suggest that a mutually exclusive separation can be obtained using this supervised method. From Figure 5.13 it should be noted that a PLS-DA model with two components extracted (t1 and t2), can be used since the separation is not mainly driven by the first component (t1), and the addition of component two (t2) might produce less group misclassifications.

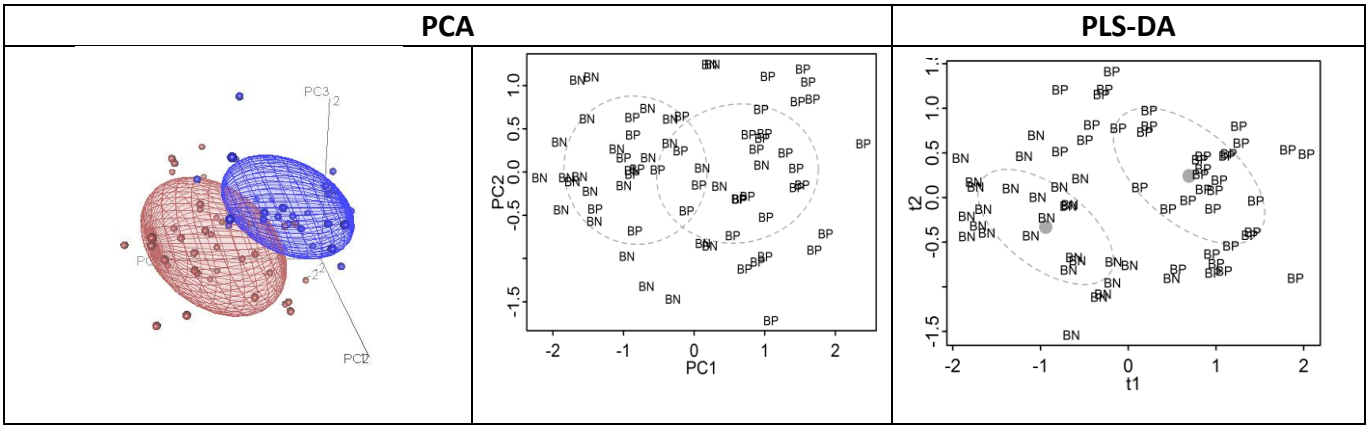


Figure 5.13 PCA score plots and PLS-DA plot. Plots created from cases swept containing 20 features 38 BP cases & 28 BN cases. Legend: Bilharziasis positive (BP) – brown & Bilharziasis negative (BN) – blue

Table 5.5 PCA and PLS-DA fit statistics of 83 features 38 BP cases & 28 BN cases.

PCA		PLS-DA			
Number of component extracted	**Cumulative % of variance explained**	**Number of component extracted**	**Cumulative R^2X**	**Cumulative R^2Y**	**Q^2**
2	37.36%	1	0.261527	0.582451	0.533119
3	46.78%	2	0.359564	0.767921	-0.28571

This might further be justified by the R^2Y value increasing from 0.58 to 0.76 for 1 and 2 PLS-DA components extracted, respectively. However, the predictive R^2Y, i.e., Q^2, suggests that only one component should be extracted with a Q^2 value of 0.53. Regardless of the Q^2 value, we extracted two PLS—DA components and subsequently, subjected this model to a thorough cross validation to investigate the model's prediction capability, for cases unseen by the model. These values will be presented as cross-validated estimates of sensitivity and specificity. If both the sensitivity and specificity are high, we conclude that the PLS-DA model did not over-fit the data, i.e. the mutual exclusive separation obtained in the PLS-DA scores plots (see Figure 5.13) is substantiated and therefore, the selected features do contain sufficient information concerning group separation.

All of the 66 (i.e., 28 BN and 38 BP) cases were initially used to determine a suitable cut-off point, i.e. a point that defines group classification in the PLS-DA model. This cut-off point was determined using the Youdin index (Fluss *et.al*., 2005), and is displayed using a green line in Figure 5.14, which shows the PLS-DA predicted values using box-plots (left figure), as well as the sensitivity (blue dots) and specificity (red dots) for different cut-off points (right figure).

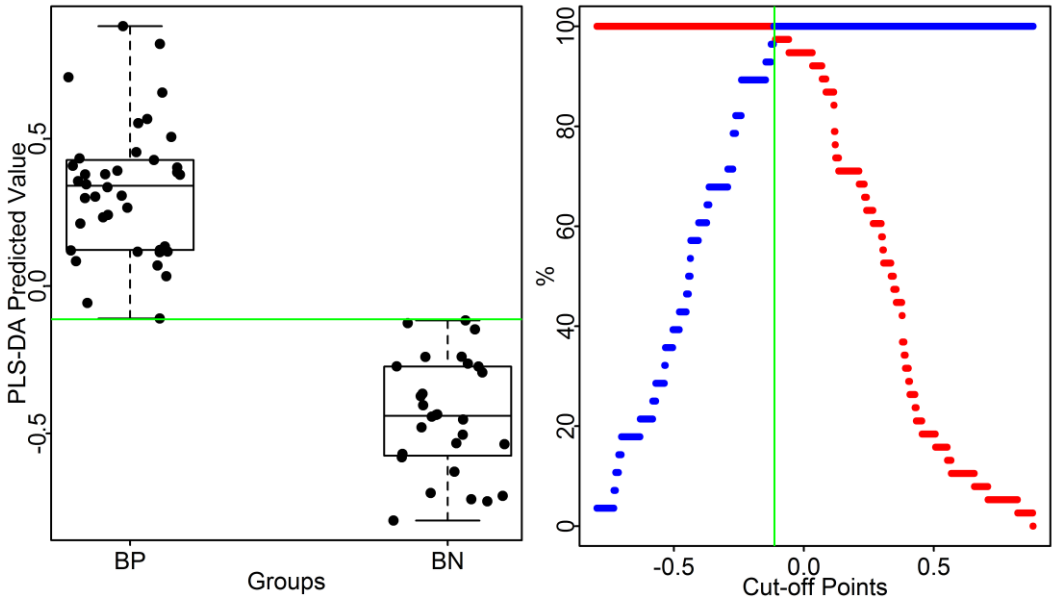


Figure 5.14 Left: **Box-plots of the predicted PLS-DA model for both groups (green line: Youdin cut-off point). Right: Sensitivity (blue dots) and specificity (red dots) with the Youdin selected cut-off point in green.**

From Figure 5.14 we can also conclude that the PLS-DA model is able to predict all BN and BP cases correctly, using the Youdin index as cut-off-point. To investigate whether this exceptional performance can be expected for cases not seen by the model, we now proceed to the cross-validation as described in Section 5.2.2.2.

Thirty percent of the cases were withheld in the cross-validation, which resulted in 8 and 11 of the 28 and 38 bilharziasis negative (BN) and bilharziasis positives (BP) groups, respectively. The normalized data of the remaining cases were then transformed, centred and used to construct the PLS-DA model with two components extracted. The cases withheld were then transformed and centred (using linear interpolation of the untransformed vs. centred transformed values of the cases used). This technique is illustrated in Figure 5.15, using succinic acid as an example. In this graph the transformed and centred values of the 19 randomly selected withheld cases were obtained using linear interpolation and are indicated by the solid black dots.

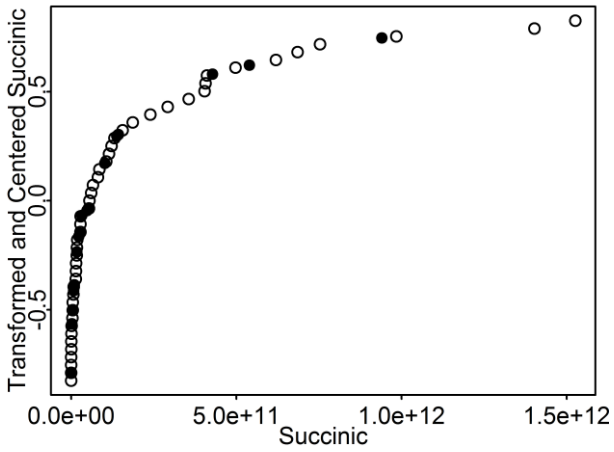


Figure 5.15 Succinic acid untransformed data versus Succinic rank transformed and centred data. (Transparent black dots: cases used to build the PLS-DA model; filled black dots: cases withheld for validation, hence interpolated).

This interpolation was performed for each of the 20 selected features, which were then used as input to the PLS-DA model, to predict the group membership of these cases. Subsequently, the sensitivity, specificity, and the number of misclassifications were recorded. This procedure was performed 10 000 times, each time with different cases removed. Note that if all possible ways of selecting 8 and 11 cases from 28 and 38, BN and BP cases, respectively, were investigated, then a total of 3.74×10^{15} of these samples would have been used. Using the 10 000 estimated values of sensitivity, specificity and percentage of misclassification, we calculated summary statistics and kernel density estimates of the distributions of these values, which are displayed in Table 5.6 and Figure 5.16.

Table 5.6 .Summary statistics concerning sensitivity, specificity and misclassifications

	Sensitivity	**Specificity**	**% Misclassification**
Mean	85.76	95.43	8.64
Standard error	12.55	5.70	5.97

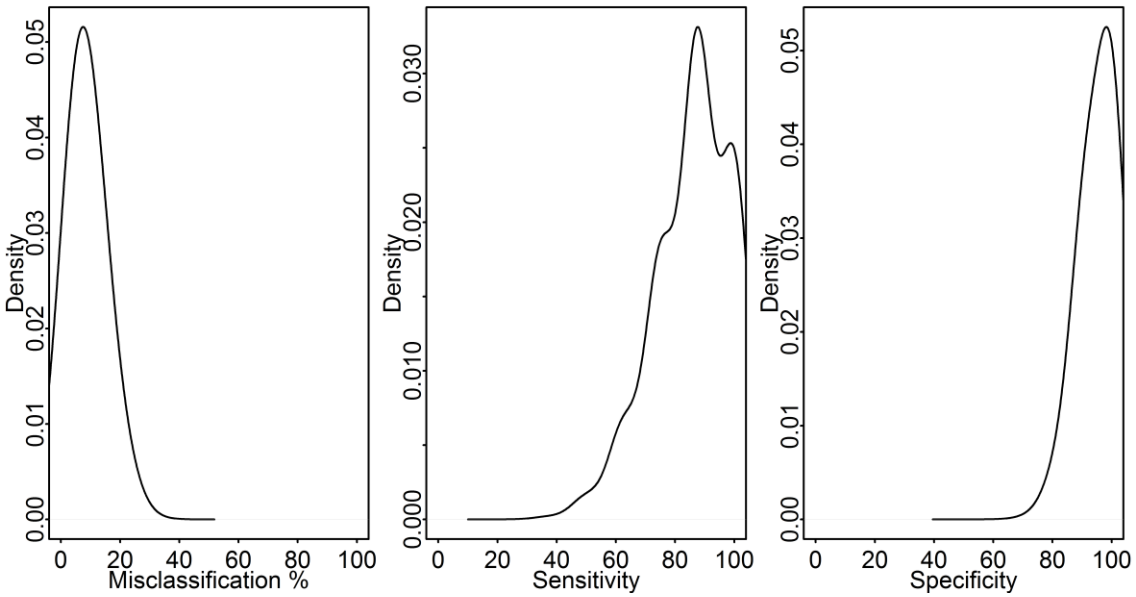


Figure 5.16 Kernel density estimates of the cross-validated percentage of misclassification, sensitivity and specificity.

From Table 5.6 we conclude that the group prediction of the PLS-DA, for unseen data, is high, both when predicting BP and BN cases. Note that the model can more often predict the correct group membership of BN cases than that of the BP group. Also, it should be stressed that these results only apply to the smaller target population, defined by removing cases from the study. In general, we conclude that the features identified in this study, do contain discriminatory value, which can be used to predict group membership with a high degree of accuracy.