FACULTY OF **VETERINARY MEDICINE**
approved by EAEVE

Proteomic analysis of *Taenia* spp. excretion/secretion proteins: the search for *Taenia solium*-specific diagnostic antigens

Bjorn Victor

Proefschrift voorgedragen tot het behalen van de graad van
Doctor in de Diergeneeskundige Wetenschappen, 2014

Promotoren:
Prof. Dr. Pierre Dorny
Dr. Sarah Gabriël
Dr. Magnus Palmblad

Vakgroep Virologie, Parasitologie en Immunologie
Faculteit Diergeneeskunde, Universiteit Gent
Salisburylaan 133, B-9820 Merelbeke

"I may not have gone where I intended to go, but I think I have ended up where I needed to be."

– Douglas Adams, *The Long Dark Tea-Time of the Soul* (1988).

Acknowledgements

A PhD is never a one-man show, so there are a number of people who deserve a moment in the spotlights:

First and foremost, my promoters, who gave me countless opportunities, helped me find additional funding, allowed me to make mistakes, assisted me with the writing of the manuscripts and the thesis and most of all, gave me the chance to prove myself as a PhD candidate:

Pierre, this is my third thesis under your supervision and I truly admire your courage! Your unit is an incredibly fun and stimulating place to work. My sincerest gratitude for your continuous support, advise, kind encouragement and constructive thinking. It is safe to say that I could never have done this without you! Thank you very much!!

Sarah, you had the bad luck of sitting right next to me where you were never safe from the endless stream of questions (not even with the small forest that suddenly appeared between our desks :-). I thank you very much for the many answers, constructive comments, time and advice you gave me throughout this endeavour. Furthermore, I feel I should also give you credit for your efforts to teach me some diplomacy...

Magnus, you introduced me to proteomics, mass spectrometry, scripting and whisky. I learned a great deal from you about proteomics and science in general. I really want to express my gratitude for all your help, good advice, hospitality, patience, effort, and maybe most of all, for putting up with my stubborn way of doing things...

The staff at the ITM: Guy, Leen, Anke, Ellen, Gill, Caroline, Ko, Famke, Julie, Nadia and Nick: thank you for the encouragement and the fantastic times we had over a few beers... Nicolas[1], thank you for the many conversations and laughs on the third floor! Katja and Kirezi[1], thank you for helping me with the writing of the manuscripts. Nadia, extra thanks for your help with my thesis. Daniëlle, thank you for sorting out the financial side of my PhD.

[1]No longer working at the ITM at the time of writing.

The staff and students at the LUMC: André, Yuri, Yassene, Rob, Alex, Linda, Kate, Katia, Simone, Tiziana, Suzanne, Dana, Anton, Hans and Hannah. Thanks you for helping me so many times, explaining everything twice, letting me sleep on your couches, sharing rooms with me during conferences and for the good times in the pub.

The staff at the Veterinary Faculty (Ghent University): Peter and Jimmy, thank you for all the help and advice with the 2D-PAGE analysis. Sien, thanks for bringing the IEF strips to Antwerp. Prof. Vercruysse, thank you very much for letting me work in your lab. Dirk, thank you for the help with my thesis. Mieke, thank you for sorting out the administration in Ghent.

The staff at Department of Medical Protein Research, VIB, Ghent: Lennart and Şule, thank you for the time and effort you put into the spectral comparison work (and I hope it's not over yet!!).

Johan, thank you for your comments and advise on the EST paper.

Nynke, you paved the way for me and this thesis and for that I am truly grateful!

I wish to thank the School of Veterinary Medicine of the University of Zambia and the Universidad National Mayor de San Marcos in Lima (Peru) for their assistance and letting us use their facilities during the production of the ESPs.

I am deeply grateful to my parents for giving me the chance to study and for supporting me in everything I've ever done (no matter how foolish!).

And last but certainly not least, I want to thank Iris. What would I do without you? Without your support, this PhD would simply not exist. Thank you for always being there for me, no matter what, and for taking care of our boys during the many times I wasn't home. And to my boys, Pieter and Jeroen, thank you for always putting a smile on my face!

The research leading to this thesis has received funding from:

- The Research Foundation – Flanders (FWO) (project number: G.0192.10N)
- The European Unions Seventh Framework Program (FP7/2007-2013) under grant agreement no. 221948 (ICONZ).
- The Institute of Tropical Medicine (ITM) Secondary Research Funding (SOFI-A).

Table of Contents

Acknowledgements **v**

List of Figures **xi**

List of Tables **xv**

List of Abbreviations **xvii**

1 General Introduction **1**

2 *Taenia solium*: An Introduction **5**

2.1 *Taenia solium* taeniasis and cysticercosis . . . 5

2.2 Human (neuro)cysticercosis: a dead-end . . . 8

2.3 Public health and economic importance . . . 10

2.4 Eradication, elimination, control and prevention . . . 11

2.5 Diagnosis of *Taenia solium* . . . 15

2.5.1 Cysticercosis . . . 16

2.5.2 Taeniasis . . . 26

3 The Hitchhiker's Guide to Proteomics **31**

3.1 Proteomics: what's in a name? . . . 31

3.2 Tandem mass spectrometry-based proteomics . . . 33

3.2.1 Sample preparation and separation . . . 33

3.2.2 Liquid chromatography and tandem mass spectrometry . . . 36

3.2.3 Data analysis . . . 40

3.3 Helminth proteomics . . . 48

4 Rationale and Objectives **53**

5 Partially Sequenced Organisms, Decoy Searches and False Discovery Rates 55
5.1 Introduction . . . 56
5.2 Materials and methods . . . 57
5.2.1 Databases . . . 57
5.2.2 Sample preparation . . . 57
5.2.3 Liquid chromatography – tandem mass spectrometry . . . 58
5.2.4 Analysis pipeline . . . 58
5.3 Results and discussion . . . 59
5.3.1 Dataset . . . 59
5.3.2 PeptideProphet false discovery rate estimations for composite databases . . . 59
5.3.3 PeptideProphet false discovery rate estimations without decoy databases . . . 61
5.4 Conclusions . . . 61

6 Proteomic Analysis of *Taenia solium* Metacestode Excretion/Secretion Proteins 63
6.1 Introduction . . . 64
6.2 Materials and methods . . . 65
6.2.1 Parasite material and excretion/secretion proteins . . . 65
6.2.2 One-dimensional polyacrylamide gel electrophoresis . . . 65
6.2.3 In-gel trypsin digest . . . 65
6.2.4 Liquid chromatography – tandem mass spectrometry . . . 66
6.2.5 Data analysis . . . 66
6.2.6 Databases . . . 67
6.3 Results and discussion . . . 68
6.3.1 *Taenia solium* metacestodes and *in vitro* excretion/secretion protein production . . . 68
6.3.2 One-dimensional polyacrylamide gel electrophoresis . . . 68
6.3.3 Data analysis . . . 68
6.3.4 Proteins identified in *Taenia solium* excretion/secretion proteins . . . 70
6.3.5 Relative protein abundance . . . 73
6.3.6 *Taenia solium* proteins . . . 73
6.3.7 Proteins matching sequences from other species . . . 74
6.4 Conclusions . . . 75

7 Use of Expressed Sequence Tags for the Identification of *Taenia solium* Metacestode Excretion/Secretion Proteins 77
7.1 Introduction . . . 78

7.2 Materials and methods . . . 78
7.2.1 Generation of the dataset . . . 78
7.2.2 Database design and data analysis . . . 79
7.3 Results and discussion . . . 79
7.3.1 Identified proteins and Gene Ontology annotation . . . 79
7.3.2 Comparison between the two studies . . . 84
7.4 Conclusions . . . 84

8 The Search for Differential Proteins 87
8.1 Introduction . . . 88
8.2 Materials and methods . . . 89
8.2.1 Parasite material and excretion/secretion proteins . . . 89
8.2.2 Two-dimensional polyacrylamide gel electrophoresis . . . 90
8.2.3 One-dimensional polyacrylamide gel electrophoresis and in-gel trypsin digest . . . 90
8.2.4 Liquid chromatography – tandem mass spectrometry . . . 91
8.2.5 Data analysis – spectral comparison . . . 91
8.3 Results and discussion . . . 94
8.3.1 Parasite material and excretion/secretion proteins . . . 94
8.3.2 Two-dimensional polyacrylamide gel electrophoresis . . . 94
8.3.3 One-dimensional polyacrylamide gel electrophoresis . . . 95
8.3.4 Data analysis – spectral comparison . . . 95
8.4 Conclusions . . . 100

9 General Discussion 101
9.1 Introduction . . . 101
9.2 Proteomics bioinformatics . . . 102
9.3 *Taenia solium* protein analysis . . . 103
9.4 Towards a *Taenia solium*-specific antigen detection assay . . . 106
9.5 Conclusions and recommendations . . . 110

10 Bibliography 111

11 Summary 141

12 Samenvatting 145

A Appendix 149
A.1 Supporting information for Chapter 5 . . . 149

List of Figures

1.1 Larval and adult stages of a *Taenia solium* tapeworm. 2
1.2 Countries and areas at risk of cysticercosis, 2011. 3

2.1 Taxonomy of *Taenia solium* in relation to other helminths mentioned in the thesis. 6
2.2 *Taenia solium* life cycle. 7
2.3 Neuroimaging examples of magnetic resonance imaging (MRI) and computed tomography (CT). 16
2.4 The seven diagnostic bands in the enzyme-linked immunoelectrotransfer blot (EITB) recognised by human sera. 19

3.1 The common "omics" sciences complement each other. 32
3.2 Schematic representation of isoelectric focussing (IEF), one-dimensional polyacrylamide gel electrophoresis (1D-PAGE) and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). 34
3.3 Principle of matrix-assisted laser desorption/ionisation (MALDI). 37
3.4 Principle of electrospray ionisation (ESI). 37
3.5 Schematic representation of an ion trap. 39
3.6 Theoretical fragmentation of a doubly charged parent ion. 40
3.7 *De novo* analysis of the collision-induced fragmentation spectrum of the DARWIN peptide. 41
3.8 Spectral library searching approach. 41
3.9 Protein database searching approach. 43
3.10 Global false discovery rate (FDR) estimation by using a concatenated target+decoy database database searching approach. 44

3.11 Global false discovery rate (FDR) estimation by calculating posterior probabilities P for each peptide spectrum match. 45
3.12 Illustration of shared peptides – the protein inference problem. 47
3.13 Two illustrations of the underrepresentation of proteomic studies on helminths and the number of reviewed helminth proteins compared to some of the "big players" in the field e.g. *Homo sapiens*, *Escherichia* and *Saccharomyces*. 49

5.1 Deviations in estimated false discovery rates (Δ_{FDR}) calculated relative to the (median) FDR estimation of the full database at three different probability thresholds. 60
5.2 Deviations in estimated false discovery rates (Δ_{FDR}) calculated relative to the median FDR estimation of the full database at three different probability thresholds for target databases with only canonical sequences or canonical sequences and isoforms without any decoy database. 61

6.1 Two examples of one-dimensional polyacrylamide gel electrophoresis illustrating the similarities between different excretion/secretion protein fractions from cysts obtained from Peruvian pigs and the differences in protein profiles observed between excretion/secretion protein fractions obtained from cysts from Zambian pigs. 69

7.1 Gene Ontology level 2 pie charts displaying the biological processes, the molecular functions and the cellular components of the 297 proteins that were identified in the *Taenia solium* metacestode excretion/secretion proteins. 83

8.1 Overview of the spectral comparison methodology. 92
8.2 Restriction fragment length polymorphism profiles of *Taenia* metacestodes. 94
8.3 Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) profiles of *Taenia solium* and *Taenia hydatigena* excretion/secretion proteins. 95
8.4 One-dimensional polyacrylamide gel electrophoresis profiles of *Taenia* metacestode excretion/secretion proteins. 96
8.5 Filtering of the *Taenia solium* and *Taenia hydatigena* experimental spectra with the spectral comparison methodology. 96

9.1 Overview of the envisaged road towards a species-specific diagnostic assay for *Taenia solium*. 109

A.1 False discovery rate behaviour at three different probability thresholds for composite databases consisting of a target database with only canonical sequences or canonical sequences and isoforms and an equally sized reversed ($n = 1$), scrambled ($n = 10$) or randomised ($n = 10$) decoy database. . . . 150

A.2 False discovery rate behaviour at three different probability thresholds for target databases with only canonical sequences or canonical sequences and isoforms without any decoy database. 151

List of Tables

2.1 Overview of the commonly used diagnostic tools and techniques for the detection of *Taenia solium*, grouped according to the lifecycle stage and the host. 17
2.2 Four diagnostic criteria for neurocysticercosis. 21
2.3 Morphological differences between *Taenia solium*, *Taenia saginata* and *Taenia asiatica*. 27

6.1 The ProteinProphet estimations for the four merged fractions. 69
6.2 Proteins ($n = 76$) identified in *Taenia solium* metacestode excretion/secretion proteins organised by Gene Ontology annotations for biological process and molecular function. 71
6.2 Table 6.2 continued. 72

7.1 Protein groups ($n = 58$) newly identified in *Taenia solium* metacestode excretion/secretion proteins, organised by Gene Ontology annotation information on biological processes and molecular functions. 81
7.1 Table 7.1 continued. 82
7.2 Gene Ontology level 2 annotations identified in this study alongside the ones identified in the previous study. 85

8.1 Peptides ($n = 11$) identified using the spectral comparison method. 98

List of Abbreviations

ΔM mass accuracy

1D-PAGE one-dimensional polyacrylamide gel electrophoresis

2D DIGE two-dimensional difference gel electrophoresis

2D-PAGE two-dimensional polyacrylamide gel electrophoresis

ABC ammonium bicarbonate

ABZ albendazole

AC alternating current

BLAST basic local alignment search tool

BLASTP protein-protein BLAST

cAMP cyclic adenosine monophosphate

cDNA complementary DNA

CID collision-induced dissociation

CLTS community-led total sanitation

CSF cerebrospinal fluid

CT computed tomography

DALY disability adjusted life year

DC direct current

DNA deoxyribonucleic acid

DTT dithiothreitol

E-value expectation value

EITB enzyme-linked immunoelectrotransfer blot

ELISA enzyme-linked immunosorbent assay

EM expectation-maximisation algorithm

ESI electrospray ionisation

EST expressed sequence tag

ETD electron-transfer dissociation

FAO Food and Agriculture Organisation of the United Nations

FDR false discovery rate

FERG Foodborne Disease Burden Epidemiology Reference Group

FTICR Fourier-transform ion cyclotron resonance

g RCF - Rotational centrifugal force

GADPH glyceraldehyde-3-phosphate dehydrogenase.

GI GenInfo Identifier

GO Gene Ontology

GP glycoprotein

HMM hidden Markov models

ICC ion charge control

IEF isoelectric focussing
IL interleukin

kDa kilodalton

LC liquid chromatography
LC-MS/MS liquid chromatography and tandem mass spectrometry
LDS lithium dodecyl sulfate
LLGP lentil lectin purified glycoproteins

m/z mass-to-charge ratio
MALDI matrix-assisted laser desorption/ionisation
MDA mass drug administration
MGF Mascot generic format
MRI magnetic resonance imaging
mRNA messenger ribonucleic acid
MS/MS tandem mass spectrometry
MudPIT multidimensional protein identification technology

NCBI National Centre for Biotechnology Information
NCBInr National Centre for Biotechnology Information non-redundant
NCC neurocysticercosis
NMC number of missed cleavages
NN neural networks
NSP number of sibling peptides
NTT number of tryptic termini

OIE World Organisation for Animal Health
OMSSA open mass spectrometry search algorithm

PCR polymerase chain reaction
pI isoelectric point
PSM peptide spectrum match
PTM posttranslational modifications
PZQ praziquantel

RF radio frequency
RFLP restriction fragment length polymorphism
ROC receiver operating characteristic
RPLC reversed-phase liquid chromatography

S discriminant score (calculated from the search tool scores)
SCG single cysticercus granuloma
SCX strong-cation exchange
SDS sodium dodecyl sulfate
SILAC stable isotopic labelling with amino acids in cell culture

TBLASTN protein-nucleotide 6-frame translation BLAST
TFA trifluoroacetic acid
Th T-helper
TOF time-of-flight
TPP Trans-Proteomic Pipeline

UHPLC ultra-high pressure liquid chromatography

WHO World Health Organisation

YLD healthy life years lost due to living with disability
YLL years of life lost due to premature death

CHAPTER 1

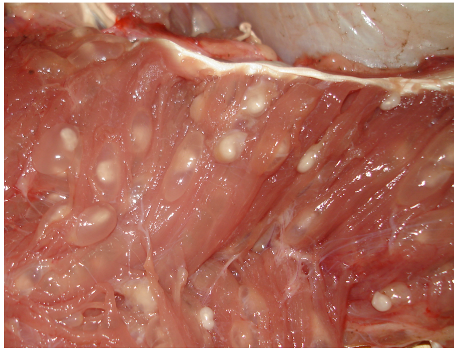
General Introduction

Tapeworms...

Many people have heard of them, but unfortunately, most of the knowledge is based on one of the many urban myths that are still being told as true, like losing weight by infecting yourself with a tapeworm.

The truth is of course far more intriguing: the tapeworm in those myths is actually the adult stage of a fascinating parasite that can belong to a number of different genera e.g. *Diphyllobothrium*, *Hymenolepis* and *Taenia* (Linnæus, 1758). The *Taenia* genus contains many different species (Hoberg *et al.*, 2000; Hoberg, 2006), but only three of them are known to infect humans: *Taenia solium* (the pork tapeworm), *Taenia saginata* (the beef tapeworm) and *Taenia asiatica* (the Asian tapeworm).

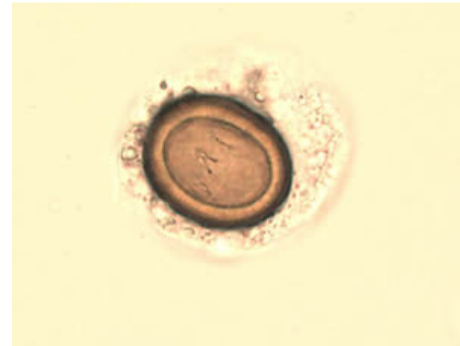
For *T. saginata* and *T. asiatica*, humans only act as the final host (taeniasis), meaning that they harbour the adult tapeworm in the small intestine, which does not pose a serious health threat. However, *T. solium* is of much more importance as humans not only harbour the adult worm (Figures 1.1b and 1.1c), but unfortunately, also act as accidental intermediate hosts. Intermediate hosts (normally pigs, but also humans, dogs, ...) infect themselves when ingesting eggs which contain oncosphere larval stages (Figure 1.1d) and can develop the metacestode larval stages (cysticerci) in the subcutaneous and muscle tissues (cysticercosis; Figure 1.1a), the central nervous system (neurocysticercosis), the eyes (ophthalmic cysticercosis) and a number of other organs like the liver and the lungs.



(a) Viable cysticerci in muscle tissue.

(b) Head (scolex) of an adult worm.

(c) Adult worm segments (proglottids).



(d) Egg with oncosphere inside.

Figure 1.1: Larval (a and d) and adult stages (b and c) of a *Taenia solium* tapeworm. Images: Dr. Sarah Gabriël (a), Bjorn Victor (c), Centres for Disease Control and Prevention Image Library (http://www.dpd.cdc.gov/dpdx/HTML/Image_Library.htm) (b and d).

In humans, neurocysticercosis can cause a number of neurological symptoms (e.g. epilepsy) and even death, while porcine cysticercosis is responsible for serious economic losses, because infected carcasses lose value and are often destroyed.

Most developed countries have eliminated *T. solium* by permanently breaking the transmission cycle through proper human waste management and controlled farm and meat industries. This is unfortunately not the case in many developing countries: *T. solium* is still found in large parts of Africa, Asia and Latin America (Figure 1.2).

Wherever elimination of the parasite is not yet feasible, we need to settle for the next best thing: control of *T. solium*. A number of possible control measures have been suggested, ranging from treatment of tapeworm carriers (either selective or through mass drug administration) and vaccination of pigs to improvement of the pig husbandry system and sanitary conditions and health education. However, the effectiveness of these measures needs to be evaluated and this is where diagnostic tools come in. Of course, diagnostic tools serve other purposes as well e.g. assessment of the prevalence of taeniasis and cysticercosis in a particular region or country (prior to the implementation of control measures) and identification of individual cases of taeniasis/cysticercosis.

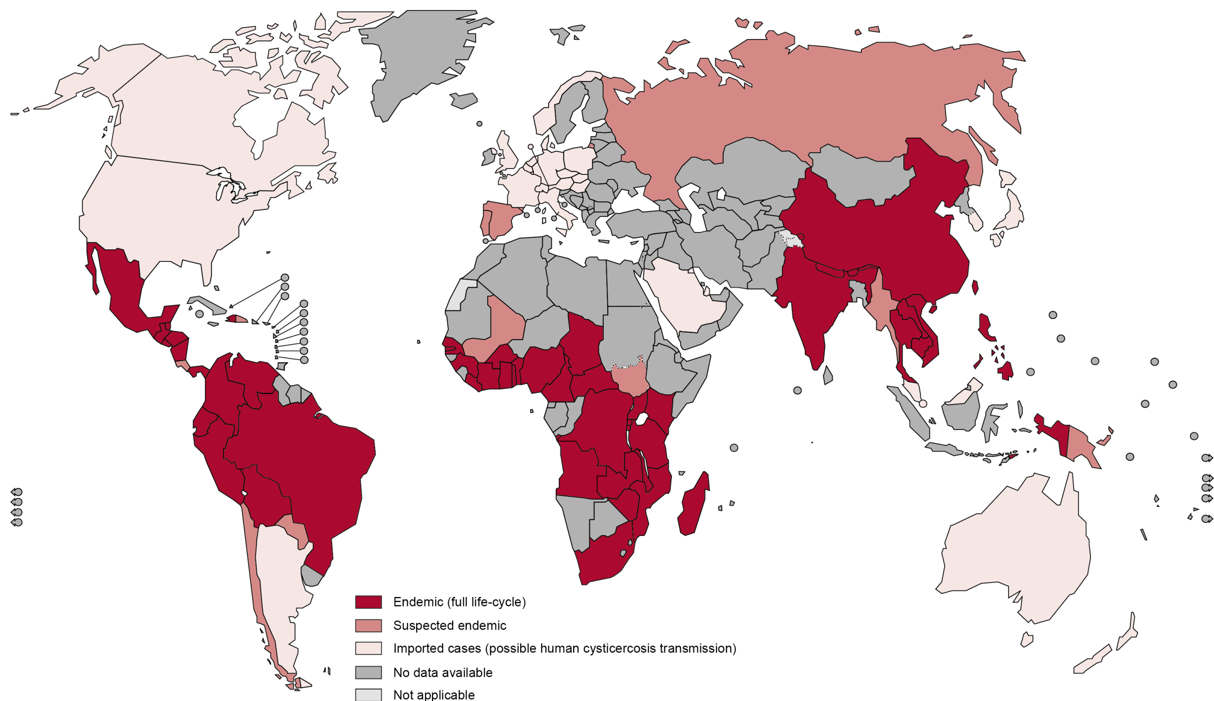


Figure 1.2: Countries and areas at risk of cysticercosis, 2011.
Adapted from: "Global distribution of cysticercosis, 2011" (World Health Organisation) (`http://www.who.int/taeniasis/Global_distribution_cysticercosis_2011.png`).

Depending on the intended purpose (e.g. clinical or epidemiological), the host and the life cycle stage of the parasite, different diagnostic tools are employed. A number of them focus on detecting the parasite directly (e.g. neuroimaging for brain cysts, antigen detection for parasite proteins, meat inspection for cysts in the meat and stool examinations for the detection of eggs and proglottids), while other assays focus on the reaction of the host (e.g. antibody detection). The development of antigen and antibody detection assays for human and porcine cysticercosis as well as taeniasis has received a lot of attention. Antibody detection is generally more sensitive compared to antigen detection, but it fails to distinguish between an active infection, a past infection and contact (without infection). Antigen detection, on the other hand, only detects active infections (viable cysts), but the currently used assays were originally developed to detect circulating antigens from *T. saginata* cysticerci in cattle. From this it follows that their use in humans and pigs is based on cross-reactions with *T. solium* antigens. Although convenient for the detection of human (neuro)cysticercosis, the use of these assays is compromised in pigs, more specifically, in regions where *T. solium*, *T. hydatigena* and/or *T. asiatica* are sympatric.

Making antigen detection more specific would greatly improve its reliability in Latin America and Asia (where *T. solium* can co-occur alongside other *Taenia* species), but to do that, differential proteins (i.e. proteins that are unique for *T. solium*) need to be identified. To achieve this, a detailed analysis of the proteins that are produced by the parasites (called excretion/secretion proteins) is necessary.

The field of proteomics enables this type of protein identification studies by combining the latest advancements in liquid chromatography, (tandem) mass spectrometry and bioinformatics. Mass spectrometry involves the measurement of the mass-to-charge ratio (m/z) of charged peptides in the gas phase. The end result of such an experiment is a dataset with spectra, which are then matched to peptides/proteins in the subsequent data analysis step. Since this thesis involves a number of mass spectrometry experiments and critical thoughts on data analysis, the most relevant concepts and techniques will be introduced in a separate chapter.

Utilising these highly specialised tools to investigate the excretion/secretion proteins of the *T. solium* larval stages should provide valuable insights into their nature and function. Furthermore, comparison of this proteome with that of other *Taenia* species might lead to the identification of one or several unique proteins. These may, in turn, lead to a specific diagnostic assay for *T. solium* porcine cysticercosis.

CHAPTER 2

Taenia solium: An Introduction

2.1 *Taenia solium* taeniasis and cysticercosis

Taenia solium is a tapeworm belonging to the Taeniidae family (Figure 2.1). The adult stage of the worm (Linnæus, 1758) develops in the upper third section of the small intestine of humans, the only final host, where it causes an infection called taeniasis (Figure 2.2). A mature tapeworm usually measures 2-7 m long and consists of a scolex, a short neck and a strobila made up of 700-1,000 proglottids (Flisser, 1994, 2013). The scolex, which is roughly the size of a pinhead (0.6-1 mm), has four suckers and a rostellum which has a double crown of 22 to 32 hooks. The hooks that constitute the inner crown are slightly larger than those present on the outer crown (160-180 µm versus 110-140 µm) (Flisser *et al.*, 2004). The neck is the area of proliferation from which the proglottids of the strobila grow. Tapeworms are hermaphroditic organisms i.e. as the immature proglottids grow further away from the neck, they develop both male and female reproductive organs. Once these organs are fully developed, the proglottids are called mature and they measure about 2.1-2.5 mm long and 2.8-3.5 mm wide. The gravid proglottids, which contain between 50,000 and 60,000 eggs and measure 3.1-10 mm long and 3.8-8.7 mm wide, detach from the distal end of the tapeworm and leave the body with the stools (Flisser, 2013). In general, 2-5 segments are released roughly 2-3 times a week (Flisser, 1994). The eggs are spherical (26-34 µm) and consist of (i) a delicate outer layer, which is often quickly lost, (ii) a thick embryophore (prismatic keratin blocks held together by a cement-like substance, which give the *Taenia* egg its radial appearance), (iii) a thin oncospheral membrane and (iv) the oncosphere or hexacanth embryo itself (Murrell, 2005; Smyth and McManus, 2007). The eggs may be ingested by pigs and humans through contamination of food and water resulting from poor hygienic conditions. Furthermore, pigs display

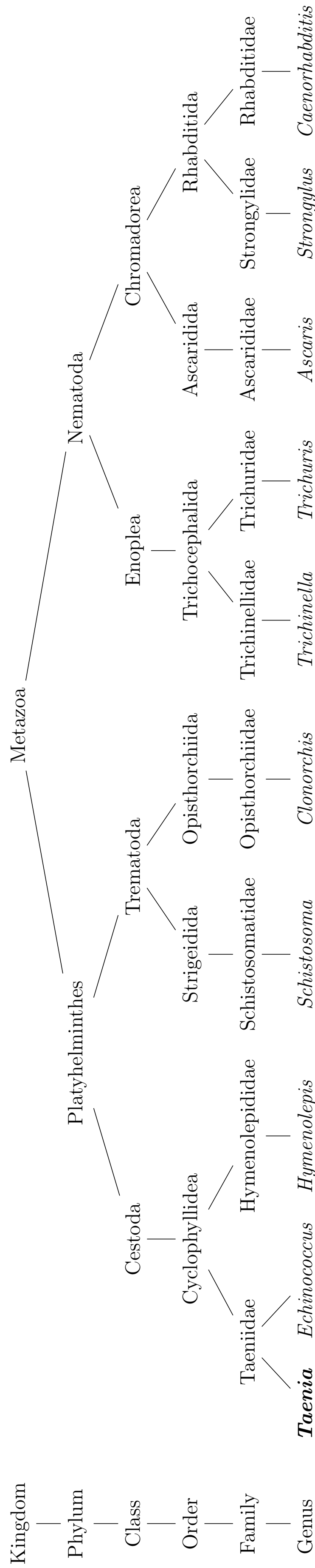


Figure 2.1: Taxonomy of *Taenia solium* in relation to other helminths mentioned in the thesis. Source: `http://www.uniprot.org/taxonomy/`

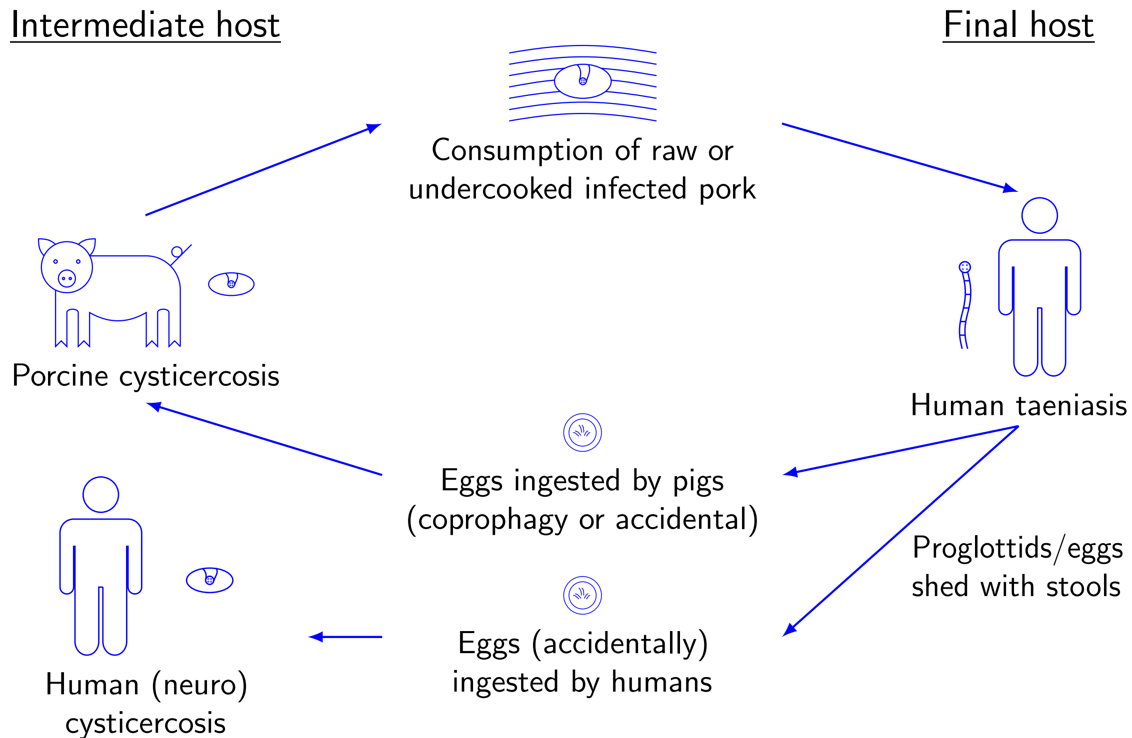


Figure 2.2: *Taenia solium* life cycle.

coprophagic behaviour, which creates further opportunities to get infected. Following ingestion, the digestive system will liberate the oncospheres from their embryophores and oncospheral membranes. Aided by their hooks, the oncospheres will then penetrate and migrate through the intestinal mucosa. Once inside the blood flow, they circulate through the body until they invade the subcutaneous tissues and the skeletal and cardiac muscle tissues. There, they develop into metacestode larval stages (cysticerci or cysts), thereby causing human and porcine cysticercosis. The cysticerci take approximately three months to fully develop, in which time they can grow up to 5-15 mm in size, and stay alive for a least one year (Flisser, 2013). Furthermore, cysticerci can also develop in the central nervous system, causing neurocysticercosis (NCC), the eyes (ophthalmic cysticercosis) and a number of other organs e.g. liver and lungs (Flisser, 1994; García *et al.*, 2003b; Bastos *et al.*, 2007). Once developed, the cysticerci are (usually) separated from the host tissues by a thin collagenous layer and undergo four involutionary stages. The first is the vesicular stage, where the cyst is almost translucent with a visible invaginated scolex. This also is the infectious stage. The colloidal stage begins when the scolex starts to degenerate and the cyst fluid becomes turbid. This stage is followed by the granular stage with further degradation of the entire cyst until finally, the cyst has fully degenerated into a calcified nodule (Escobar, 1983; Coyle and Tanowitz, 2009). The life cycle is completed when humans eat raw or undercooked infected pork. The digestive process liberates the cysticerci from the meat and causes the scolex to evaginate. The scolex will then attach

itself to the wall of the small intestine, after which the tapeworm will grow to maturity over a period of 2-4 months (Yoshino, 1934; Flisser, 2013).

2.2 Human (neuro)cysticercosis: a dead-end ...

Humans are not the common intermediate hosts for *T. solium*, pigs are. Both are infected in the same way, but for the parasite, humans are a dead-end, since the only way to complete the life cycle would be through cannibalism. For humans however, the consequences of cyst development can range from harmless and painless to life-threatening.

Cysticercosis outside the central nervous system causes little to no major symptoms, except in some rare cases where a massive parasite burden can cause muscular pseudohypertrophy and ophthalmic cysticercosis can lead to visual disturbance, proptosis and blindness (Venkataraman and Vijayan, 1983; Chang and Keane, 2001). Subcutaneous nodules as they are often observed in Asia and Africa (but rarely in Latin America) are usually small, movable and painless (García *et al.*, 2003b).

Morbidity and mortality are mostly due to neurocysticercosis (NCC), when the oncospheres have migrated through the blood-brain barrier and have developed into cysticerci in the central nervous system. Clinical manifestations vary from person to person as they are related to the location, the size and the number of cysts and the severity of the host immune response. The latter largely depends on the involutionary stage of the cysticerci. When cysts are in their vesicular stage, there is very little inflammation, since the cysticerci direct the host immune response towards a T-helper 2 (Th2) non-inflammatory response in a number of ways (Hewitson *et al.*, 2009; Moreau and Chauvin, 2010; Peón *et al.*, 2013; Singh *et al.*, 2013a) e.g. by secreting proteases that degrade interleukin (IL) 2 (Baig *et al.*, 2005) and by stimulating immunoglobulin production. These immunoglobulins rarely damage the parasite, but are instead taken up by the cysticerci and degraded as a source of amino acids (White *et al.*, 1997b; White, 2000). Typically, the immune response at this stage is marked by the production of significant amounts of interleukins IL-4 and IL-10 (Abbas *et al.*, 1996; Maizels *et al.*, 2004, 2009). However, as the cysticerci start to degrade i.e. enter the colloidal stage, the Th2 mediated response switches to a pro-inflammatory Th1 response, as is evidenced by an increase in Th1 associated cytokines such as IL-12 and IL-2 (White *et al.*, 1997b; Restrepo *et al.*, 1998).

Regarding the location, intraparenchymal NCC is the most common presentation and usually carries a good prognosis. While some patients with few cysts can even remain free of symptoms, 50-80% of patients with intraparenchymal brain cysts present the most common (and sometimes the only) symptom associated with NCC: epileptic seizures

(Del Brutto *et al.*, 1992a; García *et al.*, 2003b). It is generally accepted that seizures occur mainly because of active inflammation resulting from the death of the cysticerci and subsequent parenchymal irritation (Takayanagui and Odashima, 2006). It is also commonly observed that seizures become less frequent as the cysticerci become calcified and the inflammation becomes less pronounced.

Extraparenchymal NCC is less common, but generally more serious. Symptoms and prognosis depend largely on whether the cysts are located in the convexities of the cerebral hemisphere, the ventricular system, the fissures (especially the Sylvian fissure), the basal cisterns or the spine (García *et al.*, 2005). Cysticerci in the gyri of the cerebral convexities usually have an appearance and clinical presentation similar to intraparenchymal NCC (Salgado *et al.*, 1997), while intraventricular NCC may require surgery, because the cysts can block the flow of cerebrospinal fluid, which in turn may lead to violent headaches, hydrocephalus, intracranial hypertension and sudden death (Lobato *et al.*, 1981; Nash, 2003; Llompart Pou *et al.*, 2005; Torres-Corzo *et al.*, 2006; Takayanagui and Odashima, 2006). Cysticerci that develop in the fissures are often larger than normal and their tumour-like growth is associated with intracranial hypertension and seizures (Del Brutto *et al.*, 1992b; Takayanagui and Odashima, 2006). Perhaps the worst prognosis comes from a rare form of subarachnoid NCC associated with a large translucent cyst without a scolex, which develops in the basal cisterns (Bickerstaff *et al.*, 1952; García *et al.*, 2003b). In some cases, cysts cluster together and resemble a bunch of grapes (racemose NCC). As these cysts grow, they block the flow of cerebrospinal fluid causing intense inflammatory reaction, obstructive hydrocephalus and in severe cases, death (Bickerstaff *et al.*, 1952; Pittella, 1997). Spinal NCC is rare (1% of all NCC cases) with clinical signs usually caused by inflammation and compression of the spinal cord (Colli *et al.*, 2002; White, 2000).

In 2002, García *et al.* published the consensus that was reached by a panel of experts regarding treatment of NCC for each clinical presentation. Additional guidelines and publications by several experts in the field further illustrate the treatment modalities that can be offered to patients (Nash, 2003; Del Brutto *et al.*, 2006; Nash and García, 2011; Del Brutto, 2012). At this point, it is important to note that the treatment of (life threatening) symptoms is always the first priority. Patients suffering from seizures are first given antiepileptic drugs like carbamazepine, and if the aetiology of the seizures is confirmed i.e. intraparenchymal cysticerci are identified, the neurologist may decide to also administer antiparasitic drugs to kill the cysticerci. Currently, two antiparasitic drugs are in use for the treatment of NCC: praziquantel (PZQ) and albendazole (ABZ). ABZ is preferred over PZQ because of the slightly higher efficacy, lower cost and higher availability (Sotelo *et al.*, 1990). That being said, ABZ only kills around 70% of all living cysts and clears the parasites completely (in one round of treatment) in less than 40% of all cases (Botero *et al.*, 1993; Del Brutto *et al.*, 2006; Carpio *et al.*, 2008). The question of whether or not to

treat intraparenchymal NCC still remains somewhat controversial, since the degradation of the cysts due to the use of PZQ or ABZ can cause the inflammation that is responsible for the onset of epileptic seizures and other neurological symptoms. Therefore, corticosteroids like dexamethasone and prednisone are often given alongside the antiparasitic drugs to decrease or prevent the inflammation. Hydrocephalus and intracranial hypertension resulting from extraparenchymal NCC require immediate attention, which in many cases means the surgical removal of cysts that (may) cause the obstruction of the ventricles, shunt placement for hydrocephalus and sometimes removal and/or decompression of large/critical cysts prior to use of cysticidal drugs (García *et al.*, 2002; Nash, 2003; García *et al.*, 2005). However, surgery is difficult for cysts that develop in the basal cisterns, which adds to the grim prognosis of this type of NCC. The optimal duration of cysticidal treatment for this type of lesion is not known, but therapy should likely be sustained for longer than is the case for intraparenchymal NCC (García *et al.*, 2002).

2.3 Public health and economic importance

The *T. solium* taeniasis/cysticercosis complex has a major economic impact and poses a serious threat to public health in large parts of Africa, Asia and Latin America (Phiri *et al.*, 2003; Zoli *et al.*, 2003; Rajshekhar *et al.*, 2003; Flisser *et al.*, 2003; Ito *et al.*, 2004, 2011; Assana *et al.*, 2013).

In order to estimate the health impact (or burden) of (neuro)cysticercosis and a number of other parasitic zoonoses (e.g. echinococcosis, fasciolosis and schistosomiasis) the World Health Organisation's (WHO) preferred metric is the "disability adjusted life year" (DALY) (Torgerson and Macpherson, 2011). DALYs are calculated by summarising the years of life lost due to premature death (YLL) and the healthy life years lost due to living with disability (YLD). Ideally, DALYs should facilitate comparisons between diseases, countries and regions. However, accurate data regarding these parameters is often scarce which, if left uncorrected, can result in an underestimation of the true burden. In regions where the data is available, the total number of DALYs related to neglected tropical diseases (which include neglected zoonotic diseases like cysticercosis) can be on a par with other major infectious diseases like malaria and HIV/AIDS, as is the case in Nepal (Devleesschauwer *et al.*, 2014). To assess the health impact on a global scale, Murray *et al.* (2012) presented the Global Burden of Diseases, Injuries, and Risk Factors Study 2010 DALY calculations for 291 diseases and injuries in 21 regions and estimated that for cysticercosis, 503,000 DALYs were lost (all ages). In a recent publication, Torgerson *et al.* (2014) argued that it is likely to be considerably higher, since there is a lack of clarity over the proportion of epilepsy DALYs that was ascribed to cysticercosis. Torgerson *et al.* further stated that the Foodborne Disease Burden Epidemiology Reference Group (FERG) that was established

by WHO in 2006, will undertake its own DALY calculation.

The economic impact associated with *T. solium* cysticercosis is equally hard to estimate, since accurate information (especially for porcine cysticercosis) is also relatively scarce (Carabin *et al.*, 2005; Schantz, 2006). Nonetheless, the figures that have been published indicate significant economic losses. For example, in West Cameroon, the total annual costs due to *T. solium* cysticercosis were estimated at over EUR 10 million. Losses in pig husbandry accounted for 4.7% of that amount while the remaining 95.3% was taken up by losses directly or indirectly due to human cysticercosis (Praet *et al.*, 2009). Even higher numbers were reported in Eastern Cape Province (South Africa) where the overall cost was estimated to vary from USD 18.6 million to USD 34.2 million.

These figures illustrate the massive impact of the disease and the need to control, eliminate and, if possible, eradicate it. This need was also recognised by the WHO and cysticercosis was included in the "Neglected Zoonotic Disease" list (WHO, 2006) and in the "Global Plan to Combat Neglected Tropical Diseases (2008-2015)" which aims to prevent, control, eliminate and eradicate neglected tropical diseases and zoonoses (WHO, 2007). More recently, the commitment to prevention, control, elimination and eradication programmes was further recognised and confirmed by the World Health Assembly with the adoption of resolution WHA66.12 on May 23, 2013 (WHO, 2013).

2.4 Eradication, elimination, control and prevention

T. solium cysticercosis is considered a potentially eradicable disease (Schantz *et al.*, 1993). Indeed, *T. solium* was successfully eliminated in most European countries by improving hygienic and sanitary conditions, controlled (indoor) pig husbandry and rigorous meat inspection (García *et al.*, 2003b; Schantz, 2006). In developing countries, however, the situation is more complicated. These countries are often endemic regions and lack the necessary financial resources and political, social and economic structures to set up and maintain an elimination program. Nonetheless, even in endemic countries, it is possible to implement a number of measures in an attempt to control the disease (García *et al.*, 2007).

Treatment of human tapeworm carriers is probably the most logical and obvious control measure. The drugs currently used for this are niclosamide and praziquantel (PZQ). Niclosamide became available in 1960 and was the first safe, synthetic taeniacide (Pearson and Hewlett, 1985; Ditzel and Schwartz, 1967). Its efficacy is about 85% with a single oral dose of 2 g (Murrell, 2005), often given in combination with a purgative (Jeri *et al.*, 2004; Rajshekhar, 2004). Later, in 1972, PZQ was introduced and quickly proved very effective against a wide range of cestodes and trematodes (Katz *et al.*, 1979; Andrews *et al.*, 1983).

For human taeniasis, the efficacy is around 95% with one oral dose of 5-10 mg/kg. The use of PZQ does require an extra degree of caution. Unlike niclosamide, PZQ is absorbed from the intestine and is also used for the treatment of human (neuro)cysticercosis (García *et al.*, 2002). Although there is a report of neurological symptoms after a single taeniacidal dose of PZQ (Flisser *et al.*, 1993), the low dose makes it unlikely that this will be a common occurrence. It may however be more relevant in regions where schistosomiasis (treated with PZQ at 40 mg/kg) occurs together with cysticercosis. Furthermore, niclosamide has little to no side effects, while the use of PZQ can result in headaches and nausea (WHO, 1995). On the other hand, PZQ is cheaper[1] and more widely available than niclosamide and is therefore the drug of choice from an economic point of view (Allan *et al.*, 1997; Murrell, 2005; Pawlowski *et al.*, 2005). Another issue to consider is whether to opt for identification and treatment of individual carriers (selective treatment) or for mass treatment of an entire population (mass drug administration or MDA). Individual identification and treatment has the advantage of being more controlled. The obvious downside is the amount of community involvement, resources and field visits necessary for the identification, treatment and follow-up. Interventions like this require blood and/or stool samples from each individual and when found positive, the individual needs to report back (or be tracked down), treated and be available for follow-up (Gilman *et al.*, 2012). The alternative is mass treatment, where the entire community is treated at the same time. This has several advantages compared to the previous method and although it was once claimed that the massive expulsion of proglottids and eggs into the environment could potentially cause an increase in porcine cysticercosis (Keilbach *et al.*, 1989), recent expert opinions tend to disagree, because the proglottids would be expelled anyway and the rest of the tapeworm is not infectious.
Finally, treatment of taeniasis without implementation of other control measure(s) leaves the possibility of reinfection by consumption of infected pork. To prevent this, treatment would have to be repeated at regular intervals for at least as long as the life span of the infected pigs. Therefore, combined treatment of human taeniasis and porcine cysticercosis is a logical next step (Gonzalez, 1997; Gonzalez *et al.*, 2003; García *et al.*, 2006).

Treatment of porcine cysticercosis was initially tried with flubendazole, praziquantel and albendazole, but this was rather impractical due to the need for multiple doses (and a subsequent higher cost) and a number of side effects including death, lethargy, anorexia and the fact that, although dead, the cysts remained intact and made the meat appear unsuitable for human consumption (Telléz-Girón *et al.*, 1981; Flisser *et al.*, 1990; Torres *et al.*, 1992; Gonzalez *et al.*, 1995). This changed with the introduction of oxfendazole. A single dose of 30 mg/kg oxfendazole was shown to kill all cysts (except some in the brain) after 12 weeks (Gonzalez *et al.*, 1996, 1997, 1998). Furthermore, pigs did not get re-infected

[1] Praziquantel costs about USD 0.01 per cure (Pawlowski *et al.*, 2005)

for at least three months post treatment (Gonzalez *et al.*, 2001), but, unfortunately, it also takes 3 to 6 months for the cysts to resolve and the meat to become aesthetically suitable for human consumption (Gonzalez *et al.*, 1998; Sikasunge *et al.*, 2008). The aesthetic aspect aside, there is also a withdrawal period of 17 days before which the meat is not suitable for human consumption because of drug residues (Moreno *et al.*, 2012). The use of oxfendazole in pregnant sows is considered safe at a dose of 13.5 mg/kg (Morgan, 1982). However, the effect of the recommended dose of 30 mg/kg on pregnant sows and the piglets still requires further study. Additional limitations to the use of oxfendazole are the scarce availability in many endemic regions and the fact that the drug is not currently registered for use in pigs (Mkupasi *et al.*, 2013a). Despite these limitations, oxfendazole is generally considered safe, inexpensive, very effective against muscle cysts (although less against brain cysts), easy to administer in suspension and as an added bonus, the drug was found to be highly effective against *Ascaris suum*, strongyles and *Trichuris suis* (Mkupasi *et al.*, 2013a,b). In hyperendemic regions, re-infection is likely to occur. To prevent this from happening, treatment can be combined with vaccination.

Vaccination of pigs was first tried by Molinari *et al.* (1983) with a total extract from *T. solium* cysticerci. The following years, a number of purified fractions of *Taenia* spp. metacestode proteins were evaluated under experimental conditions i.e. with well nourished pigs of similar age, gender and genetic background and with a single infection with limited numbers of eggs, often from the same tapeworm. However, to date, only two vaccines have been successfully used in field trials where circumstances often differ greatly from experimental conditions e.g. in the field, pigs can be genetically (very) heterogeneous, malnourished, stressed and exposed to multiple *T. solium* infections with varying amounts of eggs, possibly from different tapeworms, as well as various other diseases (Sciutto *et al.*, 2007c).
The first vaccine, S3Pvac, is a combination of three synthetic peptides (GK1, KETc1 and KETc12) that were identified by using *Taenia crassiceps* murine cysticercosis as an experimental model (Manoutcharian *et al.*, 1996; Toledo *et al.*, 1999, 2001; de Aluja *et al.*, 2005; Sciutto *et al.*, 2007b,c, 2013). S3Pvac was evaluated under field conditions on two separate occasions in rural Mexico. The first trial showed that S3Pvac reduced the prevalence of cysticercosis in pigs from 15.8% to 7.5% and caused a 97.9% reduction of the total parasite load (Huerta *et al.*, 2001). In a follow-up paper, Sciutto *et al.* (2003) hinted at the possible involvement of the pigs' genetic background when it comes to resistance to naturally transmitted *T. solium* cysticercosis. During the second trial, Sciutto *et al.* (2007b) observed a reduction in porcine cysticercosis from 10% to 4% (with one dose) and to 3% (with two doses). However, statistical analysis revealed no significant difference between the values. This was likely due to the difference in sample sizes. In a recent review, Sciutto *et al.* (2013) commented further on the cysticidal effect of the vaccine on

early established cysticerci in experimentally lightly infected young pigs. The vaccinated pigs harboured 11-84% damaged cysts, while the control animals had 0.4-40% damaged parasites. The importance of this observation lies mostly in the fact that a damaged cyst is less likely to develop into a tapeworm as evidenced by a reduction in *in vitro* cysticercus evagination from 70% to 38% and a reduced *in vivo* transformation into tapeworms when fed to immunologically uncompromised golden hamsters (from 59% to 18%). In contrast, S3Pvac was found to be ineffective in damaging fully developed muscle cysticerci of naturally, heavily infected adult pigs (de Aluja *et al.*, 2011).
The second vaccine, TSOL18, is derived from the *T. solium* oncosphere larval stage and has been successful in experimental vaccine trials (Lightowlers, 2003; Gonzalez *et al.*, 2005a; Lightowlers, 2006) and, more recently, in field studies as well. In a first study conducted by Assana *et al.* (2010) in Cameroun, 2-3 month old piglets were given three vaccinations along with a single oxfendazole treatment (at the time of the second immunisation). At 12 months of age, necropsy revealed a prevalence of 20% in the control group, while no cysts were found in any of the vaccinated animals. A second study in Peru, where a combination of two recombinant antigens, TSOL16 and TSOL18, was given twice, but without any anthelmintic treatment, the number of viable cysts in the vaccinated group was found to be significantly different from the number of viable cysts in the control group (3 versus 33,416, respectively) (Jayashi *et al.*, 2012). Of the three viable cysts, two were found in the brain and since consumption of raw brain is a rare practice in Peru, these cysts would not have posed much risk to human health (Gonzalez *et al.*, 1998). The number of observed degenerated cysts was 80 and 665 for the vaccinated group and the control group, respectively. Despite initial thoughts that vaccination had no effect on established cysts (Lightowlers, 2010), results from the study in Peru suggest that the protection provided by the TSOL16 and TSOL18 antigens led to an increase in the number of non-viable cysticerci in the vaccinated animals. These promising results notwithstanding, a major disadvantage of TSOL18 vaccination is the need to give animals more than one immunisation (often two or three are given). This adds to the cost and limits feasibility, especially in traditional pig husbandry systems (free roaming pigs and no breed control). Further factors to consider when attempting vaccination are the immunological immaturity of neonatal piglets, which may prevent very young piglets from being properly immunised and the possibility that maternally-derived antibodies might interfere with the immunisation (Lightowlers, 2013). Gonzalez *et al.* (1999) showed that passively transferred antibodies can persist for months in piglets born to *T. solium* infected sows. However, these antibodies are likely to be directed against the cysticerci and since TSOL18 is an oncosphere-derived protein and anti-TSOL18 antibodies are almost undetectable in pigs harbouring mature cysticerci, there is little chance that the maternal antibodies will interfere with the immune response to the TSOL18 vaccine. However, this may not be the case if the sow was vaccinated during pregnancy (Lightowlers, 2013).

Health education has also been shown to be effective and generally includes education of the community on *T. solium* transmission, disease prevention, pig-management and sanitation knowledge (Sarti *et al.*, 1997; Ngowi *et al.*, 2008). A slightly different initiative is the community-led total sanitation (CLTS) approach, where the aim is not so much to educate the community about the diseases etc., but to make the communities critically question their own open defecation behaviour (Kar and Chambers, 2008). This is done through a process of social awakening, stimulated by local facilitators who publicly analyse open defecation in order to shock, disgust and shame the community members. The end-goal is to achieve and sustain an "open defecation free" status. This type of interventions come with an initial cost (training, posters, travel, etc.), but in the long run, no further costs are required, since the communities take over. Costs are of course higher if these initiatives are followed-up by or combined with infrastructure improvements such as the construction of latrines and pigpens (Sarti and Rajshekhar, 2003).

As a control measure, carcass inspection at official slaughterhouses is rather ineffective (Gilman *et al.*, 2012). One of the main reasons for this is the lack of official slaughterhouses (and meat inspection) in rural areas. Furthermore, when cysts are found, the carcass is confiscated with little to no compensation for the farmer (Gonzalez *et al.*, 2003). To avoid confiscation, farmers and/or traders often inspect the base of the tongue for the presence of cysticerci to identify the heavily infected animals and slaughter these for their own consumption ("backyard slaughter") or sell them illegally (Praet *et al.*, 2010a). Also, many pigs harbour only a few viable cysts which are easily missed by meat inspection (Sciutto *et al.*, 1998b; Dorny *et al.*, 2004b; Gilman *et al.*, 2012) .

Finally, freezing of meat at −24℃ or −15℃ for 1 or 3 days, respectively (Sotelo *et al.*, 1986) and the use of gamma-radiation (Verster *et al.*, 1976) have also been proposed, but the feasibility of both these methods is limited at best in the poor, rural areas where the disease is mostly occurring. Salting, where 2.5 kg pieces of pork are covered with ordinary salt and pickled for 20 days, has also been suggested, but perhaps the most realistic option to make meat suitable for consumption is heating, since it can be included in the daily food preparation routine. In general, meat is considered safe for consumption when an internal temperature of 80℃ is reached (Murrell, 2005).

2.5 Diagnosis of *Taenia solium*

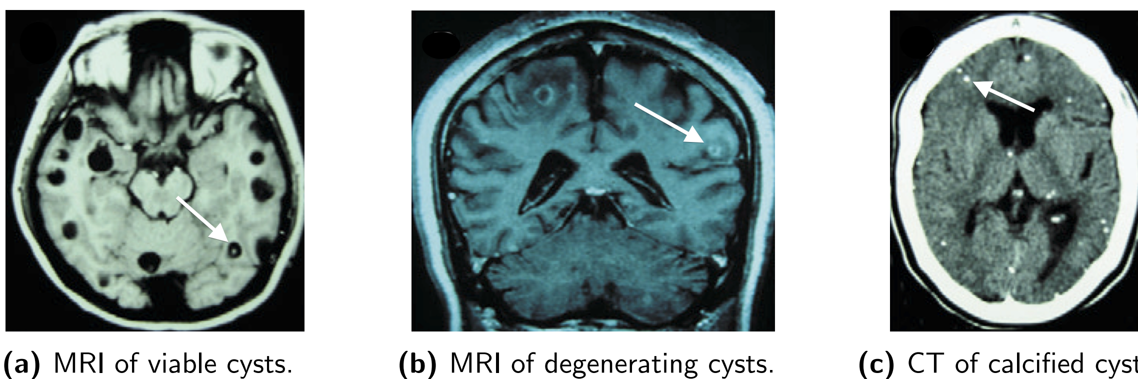
Depending on the host and the life cycle stage of the parasite, different diagnostic tools are currently in use. There is no ideal test for all purposes and deciding which tool to use for which purpose is not trivial. Even in the case where the most optimal tool is obvious, it may still not be feasible to use it due to limited local infrastructure, availability, the

number and nature of the samples and the price tag of the assay.
The intended purpose is the first thing to establish and this can vary from the identification of an individual tapeworm carrier to confirmation of the aetiology of the cysts, the screening of a population (human or porcine) to determine the prevalence of *T. solium* infection in an epidemiological study, the assessment of the efficacy of control measures and the evaluation of a newly developed tool. Furthermore, the "best tool" need not necessarily be one tool. Multiple tools can be combined in support of an individual diagnosis or for prevalence estimations based on Bayesian analysis of imperfect tests (i.e. not 100% sensitive and specific) and in the absence of a "gold standard" (Dorny *et al.*, 2004b). Table 2.1 provides an overview of the tools and techniques that are commonly utilised for the detection of (neuro)cysticercosis and taenisasis. In the following sections, the different tools and their respective advantages and disadvantages will be discussed in more detail.

2.5.1 Cysticercosis

Human (neuro)cysticercosis

Diagnosis of human NCC is ideally done by combining clinical evaluation, neuroimaging techniques and immunological methods. Neuroimaging techniques like computed tomography (CT) or magnetic resonance imaging (MRI) offer information regarding the location, the number and the involutionary stage of the cysts (Figure 2.3).



(a) MRI of viable cysts. **(b)** MRI of degenerating cysts. **(c)** CT of calcified cysts.

Figure 2.3: Neuroimaging examples of magnetic resonance imaging (MRI) (a-b) and computed tomography (CT) (c).
Reprinted from The Lancet, volume 362, number 9383, García, H. H., Gonzalez, A. E., Evans, C. A. W., Gilman, R. H., and Cysticercosis Working Group in Peru., *Taenia solium* cysticercosis, 547–556, Copyright (2003), with permission from Elsevier.

Although MRI is the most accurate technique and can visualise the different stages of the cysts, inflammation around the cysts and even forms of the disease which are difficult to detect with CT (e.g. intraventricular cysticercosis and brainstem cysts), it is scarcely available in endemic regions and comes with a hefty price tag (Suss *et al.*, 1986; García *et al.*, 2003b). CT on the other hand, offers better detection of calcified cysts [a common

Table 2.1: Overview of the commonly used diagnostic tools and techniques for the detection of *Taenia solium*, grouped according to the lifecycle stage and the host.

Human (neuro)cysticercosis: detection of the metacestode larval stage
- Clinical evaluation: seizures, headaches, subcutaneous nodules, patient history (travel/stay in endemic areas), ... - (Neuro)imaging techniques: MRI[1] and CT[2] - Immunological tools: → Antibody detection: immunoblot (EITB)[3] and various ELISAs[4] → Antigen detection: HP10 and B158/B60 ELISA - Molecular tools: PCR[5] (+ enzymatic digestion) (used mainly as confirmatory tool)
Porcine cysticercosis: detection of the metacestode larval stage
- Post-mortem diagnosis: carcass inspection - Ante-mortem diagnosis → Tongue palpation and eyelid examination → Immunological tools: *Antibody detection: immunoblot (EITB) and various ELISAs *Antigen detection: HP10 and B158/B60 ELISA → Molecular tools: PCR (+ enzymatic digestion) (used mainly as confirmatory tool)
Taeniasis: detection of the adult stage
- Stool sample examination: proglottids and/or eggs - Coproantigen detection ELISA on stool samples - Molecular diagnosis (copro-PCR) on stool samples - Antibody detection: immunoblot assay on serum

[1] magnetic resonance imaging
[2] computed tomography
[3] enzyme-linked immunoelectrotransfer blot
[4] enzyme-linked immunosorbent assay
[5] polymerase chain reaction

occurrence in NCC patients from endemic regions, accounting for over 95% of the general population and up to 30% of symptomatic cases (Zea-Vera *et al.*, 2013)], but CT images are rarely characteristic for the disease (García *et al.*, 1994).

Immunological methods should complement the neuroimaging to confirm the aetiology of the cysts i.e help differentiate the *T. solium* lesions from tuberculomas, mycotic granulomas and primary or metastatic brain tumours which may present with similar lesions on neuroimaging (Del Brutto *et al.*, 2001). These methods focus either on (i) detecting parasite antigens or (ii) detecting host antibodies directed against the parasite. The different techniques and their respective advantages/disadvantages have, in recent years, been reviewed in great detail by a number of experts (Dorny *et al.*, 2003, 2004a; Deckers

and Dorny, 2010; Esquivel-Velázquez *et al.*, 2011b; Rodriguez *et al.*, 2012; Giri and Parija, 2012). The next paragraphs will offer a summary of the most important characteristics and considerations relevant to this thesis.

Antibody detection has been in use since the early 1900s (Weinberg, 1909; Moses, 1911) in a wide variety of assays (e.g. complement fixation) (Flisser *et al.*, 1979), but the real breakthrough came in 1971, when Engvall and Perlmann described the enzyme-linked immunosorbent assay (ELISA). This assay enabled the analysis of multiple samples at the same time and was generally better than any other test before it. Initially, the antigens that were used in ELISA were crude soluble whole metacestode extracts or cyst fluids, not necessarily derived from *T. solium*. Due to the fact that the metacestodes were easy to maintain in laboratory conditions, *T. crassiceps* cysts were often used as a source of antigen (Pardini *et al.*, 2002). These crude antigens had moderate sensitivity and rather poor specificity (Schantz and Sarti-Gutierrez, 1989), but fortunately, advancements in protein purification made it possible to develop assays with more refined antigens (Ito *et al.*, 1998; Ferrer *et al.*, 2005a). Progress in purification methods also lead to the development of the enzyme-linked immunoelectrotransfer blot (EITB) (Tsang *et al.*, 1989), a Western blot assay that is currently considered the test of choice for serodiagnosis of cysticercosis (Rodriguez *et al.*, 2012). This assay uses lentil lectin purified glycoproteins (LLGP) from homogenised *T. solium* metacestodes and had an initial sensitivity and specificity of 98% and 100%, respectively. However, later reports indicated that the sensitivity was lower in cases with few or single cysts or when cerebrospinal fluid was used instead of serum (Wilson *et al.*, 1991). Reaction with one or more of the seven diagnostic proteins (at 13, 14, 18, 21, 24, 39-42, and 50 kDa) is considered a positive result (Figure 2.4). The need for parasite material limits the large scale production of the EITB, resulting in the evaluation and use of recombinant proteins and synthetic peptides in the EITB format, the Falcon assay screening test-enzyme-linked immunosorbent assay (FAST-ELISA) format and the QuickELISA™ format (Hancock *et al.*, 2004, 2006, 2003; Greene *et al.*, 2000; Scheel *et al.*, 2005; Bueno *et al.*, 2005; Handali *et al.*, 2004; Lee *et al.*, 2011).

Regardless of the format, antibody detection generally struggles to distinguish an active infection from a past infection or exposure (without infection), as antibodies can linger for up to a year and more after the infection was cleared or treated (García *et al.*, 1997; Ito *et al.*, 1999; García *et al.*, 2001). Despite this characteristic, several assays have been proposed for a stage-specific diagnosis i.e. distinguish between viable and degrading metacestodes. IgG4 antibodies have been linked with active infection (Chung *et al.*, 2002) and were suggested as a marker in NCC follow-up in saliva (Malla *et al.*, 2005). Also, a recombinant 10-kDa metacestode protein, called CyDA, looked promising in the diagnosis of active NCC with no cross reaction with chronic inactive NCC or other parasitic infections (Chung

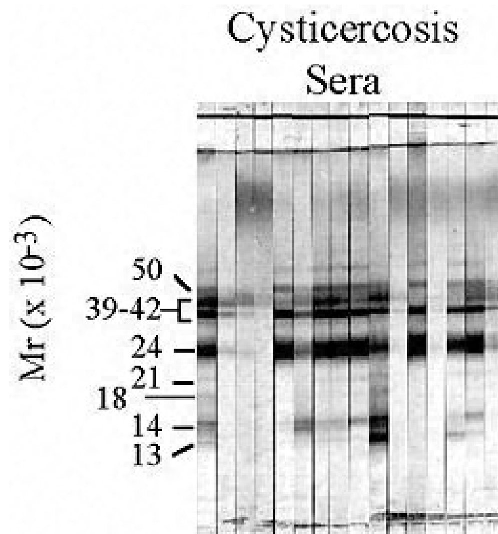


Figure 2.4: The seven diagnostic bands in the enzyme-linked immunoelectrotransfer blot (EITB) recognised by human sera.
Image (adapted): Centres for Disease Control and Prevention Image Library (`http://www.dpd.cdc.gov/dpdx/HTML/Image_Library.htm`).

et al., 1999). Later studies involving deletion mutants revealed that amino acid residues 30-34 (Asn-Met-Thr-Val-Met) reacted only with sera from active cysticercosis cases and that these residues were almost exclusively recognised by IgG4 antibodies (Chung *et al.*, 2002). Recently, sequence alignment analysis located the peptide within a variable region of the 8 kDa antigen family, further emphasising the diagnostic importance of this antigen family (Ferrer *et al.*, 2012).

Notwithstanding promising results, antibody assays for a stage-specific diagnosis are not widely used [e.g. due to major limitations like the need for cerebrospinal fluid (Barcelos *et al.*, 2007)] and require further validation in both clinical and field settings.

A better detection of viable cysticerci can be achieved with antigen detection assays. Since the production of (circulating) antigens is directly related to the viability of the cysts, dead cysts will not register in these assays.

The antigen composition of helminths can be roughy divided into surface, excretory/secretory and somatic antigens. Excretory/secretory proteins (ESPs) are released by the parasite as either "waste" proteins (excretion) or as "functional" molecules (secretion) (Harnett, 2014). Secretory proteins are often identifiable by a specific N-terminal signal peptide although leaderless secretion (a non-conventional/non-classical secretory pathway) has also been described (Kuchler *et al.*, 1997).
ESPs are at the interface of interaction between host and parasite and can represent 8-20% of the proteome of an organism (Nagaraj *et al.*, 2008). Specifically for *T. solium*,

ESPs help the parasite to (i) penetrate the host, (ii) establish and maintain an infection, while being attacked by the host immune system, and (iii) provide for its own nutrition by incorporating metabolites from the host. Perhaps the most important function is the evasion of the host immune system. *T. solium* metacestodes have developed various strategies to deal with both the innate and the adaptive immune system of its host and many of them are shared with other helminths or helminth stages e.g. the production of proteases and protease inhibitors, inhibition of complement formation and the directing of the host immune response towards a Th2 non-inflammatory response. A large number of experts have reviewed these strategies and proteins in great detail e.g. White *et al.* (1992, 1997b); Riganò *et al.* (2001); Sciutto *et al.* (2007a); Siracusano *et al.* (2008a); Maizels *et al.* (2009); Hewitson *et al.* (2009); Siracusano *et al.* (2012) and McSorley *et al.* (2013). Evading the immune system is however not the only function of ESPs. Obtaining nutrients is equally important. Recent evaluation of four cestode genomes (including *T. solium*) has revealed a reduction in overall metabolism capability and an increased ability to absorb nutrients (Tsai *et al.*, 2013). This is evidenced by e.g. a large number of hydrophobic ligand binding proteins found in parasite ESPs (Hancock *et al.*, 2003; Lee *et al.*, 2007). Aside from the above mentioned functions, ESPs are also involved in signalling, defence against oxygen-mediated killing mechanisms and various binding processes (Kim *et al.*, 2000; Guillou *et al.*, 2007; Siles-Lucas *et al.*, 2008). Furthermore, many ESPs can have multiple functions.

Since ESPs are so vital to the survival of the parasites, they have been extensively studied with diagnosis, vaccination and development of antiparasitic drugs in mind (Nagaraj *et al.*, 2008). Furthermore, ESPs are used as an indicator of parasite viability i.e. contrary to antibody levels, ESP levels decrease relatively quickly when the metacestodes enter a degenerative/calcified stage.

Currently, two antigen detection assays are being used. The first one, the HP10 ELISA (Harrison *et al.*, 1989), where HP10 is the circulating antigen detected in this assay, has a sensitivity of 72% in cerebrospinal fluid of NCC confirmed patients (Correa *et al.*, 1989). The second assay, the B158/B60 ELISA (Brandt *et al.*, 1992; Van Kerckhoven *et al.*, 1998; Dorny *et al.*, 2000), where B158 and B60 are the monoclonal antibodies used to capture the circulating antigens, has a reported sensitivity and specificity of 90% and 98%, respectively (Praet *et al.*, 2010b). Both assays were originally designed to detect antigens produced by *T. saginata* metacestodes in cattle. The fact that they are used in the detection of *T. solium* (neuro)cysticercosis is due to a (convenient) cross reaction between *Taenia* species.

For clinical case management, antigen detection has proven to be very valuable in a number of ways. Since a positive test is directly linked to the viability of the cysts, it is a well suited follow-up tool for treatment, with circulating antigen levels decreasing in parallel with parasite resolution on brain imaging (García *et al.*, 2000, 2010). Furthermore, antigen

levels not only correlate with severity of the infection (Zea-Vera *et al.*, 2013), but also with the type of NCC. In the case of intraparenchymal NCC, both antigen assays showed reduced sensitivity, but test results still scaled well with the number of cysts (Rodriguez *et al.*, 2009; Bobes *et al.*, 2006). Extraparenchymal NCC, which carries a poorer prognosis, shows much higher antigen levels, particularly in patients with basal subarachnoid NCC (García *et al.*, 2010). Very high antigen levels can thus indicate extraparenchymal NCC or massive intraparenchymal NCC (Rodriguez *et al.*, 2012).

In 2001, Del Brutto *et al.* presented four diagnostic criteria for neurocysticercosis stratified on the basis of their diagnostic strength. These criteria were revisited in 2012 with notes on recent advances in neuroimaging and immunodiagnostic techniques (Table 2.2) (Del Brutto, 2012).

Table 2.2: Four diagnostic criteria for neurocysticercosis. Interpretation of these criteria permits two degrees of diagnostic certainty (Del Brutto, 2012).

Diagnostic criteria
Absolute
- Histological demonstration of the parasite from biopsy of a brain or spinal cord lesion
- Evidence of cystic lesions showing the scolex on neuroimaging studies
- Direct visualisation of subretinal parasites by fundoscopic examination
Major
- Evidence of lesions highly suggestive of neurocysticercosis on neuroimaging studies
- Positive serum immunoblot for the detection of anticysticercal antibodies
- Resolution of intracranial cystic lesions after therapy with albendazole or praziquantel
- Spontaneous resolution of small single enhancing lesions
Minor
- Evidence of lesions compatible with neurocysticercosis on neuroimaging studies
- Presence of clinical manifestations suggestive of neurocysticercosis
- Positive CSF[1] ELISA[2] for detection of anticysticercal antibodies or cysticercal antigens
- Evidence of cysticercosis outside the central nervous system
Epidemiological
- Individuals coming from or living in an area where cysticercosis is endemic
- History of travel to disease-endemic areas
- Evidence of a household contact with *T. solium* infection
Degrees of diagnostic certainty
Definitive
- Presence of one absolute criterion
- Presence of two major plus one minor and one epidemiological criteria
Probable
- Presence of one major plus two minor criteria
- Presence of one major plus one minor and one epidemiological criteria
- Presence of three minor plus one epidemiological criteria

[1] cerebrospinal fluid
[2] enzyme-linked immunosorbent assay

However, the criteria (still) depend heavily on neuroimaging and the EITB assay, while antigen detection is not thoroughly considered. Recently, the criteria have been questioned in resource-poor settings where imaging is not feasible and the suggestion to include the B158/B60 ELISA as a major criterion was put forward by Gabriël *et al.* (2012). Del Brutto commented on this in his revisited paper by suggesting minor adaptations to the criteria according to the most common patterns of disease expression in a given region. To further illustrate the added value of antigen detection, it was demonstrated that a positive antigen ELISA in serum can predict undetected viable brain metacestodes, thereby selecting a group of high-risk patients who would benefit from more advanced imaging techniques and possibly antiparasitic treatment or even surgery (Zea-Vera *et al.*, 2013). The authors also suggested to investigate whether antigen ELISA could be routinely recommended in patients with a diagnosis of calcified NCC on CT.

In some peculiar cases, NCC presents itself as a single parenchymal brain enhancing lesion (or single cysticercus granuloma; SCG), which accounts for up to 60% of NCC cases in the Indian subcontinent (Rajshekhar, 1991; Kumar Garg *et al.*, 2000; Prabhakaran *et al.*, 2007) and has also been reported in Latin America and a number of non-endemic countries (Del Brutto *et al.*, 2012). Although the diagnosis of neurocysticercosis can be established with a sensitivity of 99.5% and a specificity of 98.9% when those lesions fulfil a rigid set of clinical and radiologic criteria (Rajshekhar and Chandy, 1997), it is generally observed that immunodiagnostic tools (i.e. both antigen and antibody detection assays) show low sensitivity in those cases (Proaño-Narvaez *et al.*, 2002), especially when the cyst is starting to degenerate. To remedy this drop in sensitivity, the six most recognised EITB glycoproteins in Indian patients [50, 38, 24, 18, 14 and 13 kDa according to Prabhakaran *et al.* (2004)] were unfolded with 5 M urea to expose more/other epitopes. In Western blot format, sera from SCG patients that were negative in the "native" EITB, were tested with these unfolded proteins, resulting in an increase in sensitivity from 0% to 46%. When repeating the analysis in the ELISA format, the sensitivity went from 13% to 42% (Prabhakaran *et al.*, 2007).

Overall, antibody detection (with the EITB as best example) has a higher diagnostic sensitivity and specificity than antigen detection, but it tends to overestimate the true prevalence (especially in the endemic regions), due to its inability to discriminate between active and past infections or contact. The use of the EITB as a screening tool to refer seropositive patients to more specialised health centres has been suggested. However, seropositive patients with neurological symptoms will most likely not be treated differently (or more urgently) than seronegative patients with seizures and/or (surgery requiring) intracranial hypertension (García *et al.*, 2012). Therefore, the role of antibody detection is more towards confirmation of the aetiology of the cysts and identifying individuals at

higher risk of disease progression (e.g. early stages of extraparenchymal NCC and patients with multiple viable cysts who will likely encounter problems in later years). Since antigen detection provides insights in the viability of the (brain) cysts and is therefore an important tool in the follow-up of treated cases, it can also be used to support (early) detection of the above mentioned cases, especially because extraparenchymal NCC presents itself with very high levels of circulating antigens.

Porcine cysticercosis

Porcine cysticercosis is diagnosed post-mortem by carcass inspection, or ante-mortem by inspection or palpation of the tongue or by using a variety of immunodiagnostic tools.

A complete necropsy of the carcass (i.e. complete slicing of the muscle tissues and organs in 3-5 mm slices) is the only true “gold standard”. However, this procedure is detrimental to the commercial value of the meat and is very labour intensive. Therefore, it is never used for routine meat inspection. Nonetheless, it does have its use in the validation of (new) immunodiagnostic tools (Dorny *et al.*, 2003). Routine meat inspection at slaughterhouses is limited to a number of incisions in the *T. solium* cysticerci predilection sites e.g. the masseter muscles, triceps brachii muscle, tongue and heart (Dorny *et al.*, 2004b). This method requires trained staff and while being 100% specific, it lacks sensitivity. In two studies in Zambia, one using a Bayesian approach to estimate test characteristics and another to validate those findings, the reported sensitivity of the carcass inspection varied from 22.1% (estimated) to 38.7% (validated) with no detection of light infections (1-100 cysts) (Dorny *et al.*, 2004b). In addition to the low sensitivity, lots of pigs are not slaughtered at the “official” abattoirs, so even heavily infected carcasses can enter the food chain without detection.

Tongue palpation (inspection of the base of the tongue for the presence of nodules) is often used by “middlemen” who purchase pigs in the villages and at the markets and (unofficial) slaughterslabs, since it is quick, requires no infrastructure and does not damage the meat. However, although 100% specific when performed by trained staff, it is even less sensitive than carcass inspection for moderate to heavy infections and also misses all light infections (Dorny *et al.*, 2004b). Eyelid examination has also been suggested and compared with tongue palpation (Singh *et al.*, 2013b). Specificity was determined at 100% for both tongue palpation and eyelid examination, while the sensitivity was determined at 70% and 25%, respectively. Although the 70% sensitivity is in line with a previous report by Gonzalez *et al.* (1990), the available information in both reports suggests that the animals harboured moderate to high infections, hinting at the possibility that the 70% sensitivity of the tongue palpation is applicable to heavier infections only.
Clinical signs of porcine neurocysticercosis (excessive salivation, excessive blinking and tearing, and subconjunctival nodules) have also been proposed as a means to identify

infected pigs (Prasad *et al.*, 2006), however, most authors report no clinical signs at all, except for the presence of subconjunctival nodules in massive infections.

Apart from a complete necropsy, all the methods described above lack sensitivity. Despite this, their use (in slaughterhouses) remains vital as they help keep (heavily) infected carcasses out of the food chain. However, in large-scale epidemiological studies or in assessments of control measures, those tools are impractical and immunological tools are a better choice.

Immunodiagnostic assays offer increased sensitivity and, as is also the case for human (neuro)cysticercosis, the assays can be divided into antibody detection and antigen detection. Antibody detection assays using a wide range of antigens (from crude *Taenia* spp. extracts to purified fractions) have been described (and reviewed) with variable sensitivities and specificities, mainly depending on how severe the infection is (de Aluja *et al.*, 1996; Biondi *et al.*, 1996; Vaz *et al.*, 1996; Sciutto *et al.*, 1998a; Ko and Ng, 1998; Sciutto *et al.*, 1998b; D'Souza and Hafeez, 1999; Pinto *et al.*, 2000; Nunes *et al.*, 2000; Dorny *et al.*, 2003; Sato *et al.*, 2003). In short, there is no "go-to" antibody assay for porcine cysticercosis. That being said, the EITB (Tsang *et al.*, 1989), the assay of choice for human (neuro)cysticercosis, has also been evaluated for the detection of porcine cysticercosis.

In 1991, evaluation of the EITB assay reported 100% sensitivity and specificity in a control group of 137 animals. Additionally, four pigs were experimentally infected and their immune response was quantified by EITB. All four animals harboured 12-74 cysts after full necropsy and showed IgG activity directed mainly towards the 42 kDa and 50 kDa bands (Tsang *et al.*, 1991). A year before, identical results were obtained by Gonzalez *et al.* (1990). Since then, the EITB assay has been used throughout the world on pig serum samples (Sakai *et al.*, 1998; García *et al.*, 1999, 2003a; Rodriguez-Hidalgo *et al.*, 2006; Gomes *et al.*, 2007; Krecek *et al.*, 2008, 2011; Devleesschauwer *et al.*, 2013). Despite the initially promising results, later reports have indicated that pigs can have a seropositive result on EITB while being negative at full necropsy. Various reasons have been proposed with maternal antibodies (Gonzalez *et al.*, 1999) and secondary transmission (Gonzalez *et al.*, 2005b) being amongst the most plausible explanations. Further research into the correlation between antibodies detected with the EITB and the number of cysts found after a full necropsy revealed that there are two important factors to be considered when estimating the actual infection status: (i) the age of the pig (i.e. the older pigs have a higher probability of being infected) and (ii) the number of EITB bands (i.e. the probability of having (many) cysts is higher for pigs that react with 3 or 4+ diagnostic bands than for pigs reacting with only one or two bands) (Gavidia *et al.*, 2013).
A similar age relation was shown in e.g. Mozambique (Pondja *et al.*, 2010) and Peru (García *et al.*, 2003a), while Jayashi *et al.* (2013) made similar observations about the

number of identified EITB bands. Furthermore, their evidence suggests that the more cysts in an animal, the more consistent the EITB results will be. Jayashi *et al.* go on to propose a modification of the EITB cut-off point from one to ≥ 3 bands, based on receiver operating characteristic (ROC) curve analysis. This cut-off point modification changed the EITB characteristics from a sensitivity of 88.9% and a specificity of 48.3% (for one or more bands) to a sensitivity of 77.8% and a specificity of 76.4% (for three or more bands). The ≥ 3 bands cut-off point offered the best trade-off in sensitivity and specificity and more closely resembled the prevalence of porcine cysticercosis determined by necropsy (16.8% versus 57.9% for one band and 32.7% for three or more bands). Coincidentally, the second best cut-off point, four or more bands, resulted in a prevalence of 16.8% (i.e. identical to the one found by necropsy)! However, the individual animals did not match between the two. A further observation was that a relatively high proportion of animals that were negative at necropsy still reacted to ≥ 3 bands in the EITB assay. This could be due to antibodies triggered by exposure and/or aborted infections (the Achilles' heel of any antibody assay), but also to non-specific reactions. Indeed, although the EITB for human NCC has been well studied in that regard, there is a need for thorough specificity studies involving non-*T. solium* (helminth) infections.

In conclusion, the use of the EITB for porcine cysticercosis is possible, though not ideal. The EITB remains a complex, expensive and labour intensive test with potentially unknown cross reactions. Also, its use may require resetting the cut-off point to better suit the conditions (field or experimental) and the objective of the study e.g. ≥ 1 band to "rule out" the disease and ≥ 3 bands to "rule in" the disease.

Antibody detection assays are not the only tool available. The two antigen detection assays that were described for human NCC can also be used to diagnose porcine cysticercosis and here too, they only detect viable cysticerci. The B158/B60 ELISA characteristics were evaluated in Zambia using a Bayesian approach and resulted in 87.6% sensitivity and 94.7% specificity. A follow-up study to validate these findings resulted in 64.5% sensitivity and 91.2% specificity (Dorny *et al.*, 2004b). Although most assays display lower sensitivities and specificities in medium to light infections, the B158/B60 ELISA was able to detect a single viable cyst in the brain of a pig (Nguekam *et al.*, 2003). A later study in South Africa also used a Bayesian approach to evaluate the prevalence [56.7%, close to the one found in Zambia (64.2%) (Dorny *et al.*, 2004b)] and test characteristics of the B158/B60 ELISA as well as the HP10 ELISA (Krecek *et al.*, 2008, 2011). The study revealed a slightly lower sensitivity for the B158/B60 ELISA compared to the HP10 ELISA (63.3% versus 70.4%, respectively) and a specificity of 87.0% for the B158/B60 ELISA versus a 66.1% specificity for the HP10 ELISA.

As is evidenced by the above mentioned specificities, there is an important drawback to using antigen detection in pigs: there is strong evidence of cross reaction with metacestodes of *Taenia hydatigena*, a non-zoonotic *Taenia* species that is transmitted between dogs (the final host) and small ruminants (the intermediate hosts) (Dorny *et al.*, 2003, 2004a). Pigs may also act as intermediate hosts for *T. hydatigena*, although there are regional variations in their role in the life cycle i.e. *T. hydatigena* infections in pigs are rare in Africa, but they are a common observation in Asia and Latin America (Conlan *et al.*, 2012). Furthermore, in many Asian countries [e.g. Nepal, Korea, China, Taiwan, Thailand, Indonesia, Vietnam, Japan, and the Philippines as reported by Eom *et al.* (2009)] the metacestodes of a third *Taenia* species, *T. asiatica*, are found in pig viscera (Eom, 2006; Eom and Rim, 1993) and given the known cross reactions between *T. solium*, *T. saginata* and *T. hydatigena*, it is logical to expect a cross reaction with *T. asiatica* as well.

All this considered, antigen detection is seriously compromised in regions where these parasites coexist in the pig. In Laos, maximum-likelihood adjusted prevalence of *T. solium* and *T. hydatigena* in pigs was 4.2% and 55.9%, illustrating the potential error that can be made in these regions (Conlan *et al.*, 2012).

A number of attempts have been made to address the need for a *T. solium*-specific (antigen detection) assay. Deckers *et al.* (2009) developed nanobodies (camelid-derived single-domain antibody fragments) that did not cross react with *T. hydatigena*, *T. saginata*, *T. crassiceps* or *Trichinella spiralis* (cross reaction with *T. asiatica* could not be assessed). Unfortunately, attempts to detect antigens in serum from naturally infected pigs with ELISA resulted in a high background signal, preventing assessment of the real reactivity. A different approach towards a more sensitive and more specific assay was reported by Diaz-Masmela *et al.* (2013). They used two-dimensional polyacrylamide gel electrophoresis blots in combination with liquid chromatography and tandem mass spectrometry to identify seven antigens that were specific for porcine cysticercosis: tropomyosin 2, alpha-1 tubulin, beta-tubulin 2, annexin B1, small heat-shock protein, 14-3-3 protein, and cAMP-dependent protein kinase. None of these proteins cross reacted with serum pools of pig infected with *Ascaris suum*, cysticerci of *T. hydatigena* and hydatid cysts of *Echinococcus* spp. or with a serum sample from a cow infected with *T. saginata*.
Despite this progress, there is currently no *T. solium*-specific antigen detection assay available.

2.5.2 Taeniasis

Diagnosis of the adult tapeworm in the intestine of its final host, humans, can be done with a number of techniques, ranging from stool examinations to molecular biology.

The first and most obvious way to detect a tapeworm is a stool sample examination. This can be macroscopic or microscopic. The former can detect (gravid) proglottids, while the latter is used to detect eggs and can be preceded by some form of concentration method (e.g. sedimentation and formalin-ether concentration) (Murrell, 2005). These techniques all have low sensitivity due to the intermitted release of eggs/proglottids and the relatively small size of the sample. Specificity of stool examination depends on the tapeworm tissue found in the sample. When eggs are found, specificity is limited to the genus level, since *T. solium* eggs cannot be distinguished from other *Taenia* that are known to infect humans (*T. saginata* and *T. asiatica*). When scolices and gravid or mature proglottids are recovered (e.g. after treatment), it is theoretically possible to distinguish between the three *Taenia* species (Table 2.3) (Murrell, 2005). Unfortunately, even with optimised purges, recovery rates are low: 25-33% and 53-69% for scolices and gravid proglottids, respectively (Jeri *et al.*, 2004).

Table 2.3: Morphological differences between *Taenia solium*, *Taenia saginata* and *Taenia asiatica*.

		Taenia solium	*Taenia saginata*	*Taenia asiatica*
Scolex	Rostellum	Present	Absent	Present
	Hooks	22-32	Absent	Absent
Mature proglottids[1]	No. of testes	375-575	800-1,200	324-1,216
	Ovary	3 lobes	2 lobes	2 lobes
	Vaginal sphincter	Absent	Present	Present
Gravid proglottids[1]	Unilateral uterine branches	7-16	14-32	11-32
	Branching pattern	Dendritic	Dichotomous	Dichotomous
	Expulsion from host	Passively[2] (in groups)	Actively[3] (single)	Actively[3] (single)

[1] Comparison of the internal structures can be facilitated by a simple hematoxylin-eosin staining (Mayta *et al.*, 2000).
[2] Usually with faeces.
[3] Outside defecation.
Adapted from "WHO/FAO/OIE Guidelines for the surveillance, prevention and control of taeniasis/cysticercosis" (Murrell, 2005).

More complex methods like electrophoresis assays using mobility of glucose phosphate isomerase (Le Riche and Sewell, 1977, 1978) or total protein (Bursey *et al.*, 1980) to distinguish between species were evaluated in the late seventies and early eighties. Since these assays often required very fresh material without additives (to preserve enzyme activity), their use was limited.

An ELISA that could detect *Taenia* antigens in faecal material made its appearance in the early nineties (Allan *et al.*, 1990, 1992). This coproantigen detection ELISA uses polyclonal rabbit antibodies directed against whole worm extract of *T. solium* and is able to detect 2.6 times as many confirmed cases of taeniasis compared to a stool examination. Sensitivity was determined at 98% and specificity at 99%, but on the genus level, meaning that the

test is unable to distinguish between *Taenia* species (Allan *et al.*, 1996). To solve this issue, new IgG capture antibodies were produced against *T. solium* whole worm extract while new IgG detecting antibodies were produced against *T. solium* adult excretion/secretion proteins (both in rabbits) (Guezala *et al.*, 2009). The authors calculated 100% specificity and 96% sensitivity for *T. solium* tapeworm carriers. Apart from an increase in sensitivity, coproantigen ELISAs have other advantages as well: they can detect tapeworms in their immature, prepatent stage [even after one week in animal models (Avila *et al.*, 2003)] and can be used as an early indicator of cure or treatment failure (Bustos *et al.*, 2012).

In an ongoing search for increased sensitivity and specificity, researchers started looking into DNA for the development of a diagnostic tool. Small differences in the genomic or mitochondrial DNA were exploited and made into tools to differentiate between *T. solium* and *T. saginata*/*T. asiatica*. Initially, DNA probes were used (Harrison *et al.*, 1990), while later efforts employed (multiplex) polymerase chain reactions (PCR) and PCR followed by enzymatic digestion or single-strand conformation polymorphism (Gasser and Chilton, 1995; Gasser *et al.*, 1999; González *et al.*, 2000b, 2002, 2004; Montero *et al.*, 2003; Rodriguez-Hidalgo *et al.*, 2002). Close scrutiny of *T. solium* mitochondrial DNA enabled the distinction between a genotype restricted to Asia and a genotype in Africa and Latin America (Nakao *et al.*, 2002, 2003), while a range of highly polymorphic markers (such as microsatellites) made it possible to analyse genetic variation at the community level (Campbell *et al.*, 2006). As with coproantigen detection, DNA analysis also allows for the detection of immature worms (Yamasaki *et al.*, 2004) and, although many extraction protocols require pure and clean samples (e.g. proglottids and/or eggs), it is possible to extract DNA from faecal material directly with a lower detection limit of 40 eggs/g faecal matter (Mayta *et al.*, 2008). Additionally, since DNA is the same for all stages of the life cycle, DNA based assays can serve double duty as as confirmation tool for cysticercosis lesions as well (Geysen *et al.*, 2007).

A recent study by Praet *et al.* (2013) in Zambia compared for the first time the performances of stool examination, coproantigen ELISA and real-time copro-PCR and used Bayesian modelling to estimate test characteristics and prevalence. They found high specificities for all three tests: 99.9%, 92% and 99% for stool examination, coproantigen ELISA and copro-PCR, respectively, although the 92% was lower than previously reported. Sensitivities for stool examination, coproantigen ELISA and copro-PCR were 52.5%, 84.5% and 82.7%, respectively. Given the observed genus specificity of the coproantigen ELISA and low sensitivity of the stool examination, Praet *et al.* argued that a more widespread use of copro-PCR would enable better detection of tapeworm carriers, a key factor in *T. solium* control.

The above discussed tests share a dependance on parasite material. Collecting and handling stool samples and/or tapeworm segments not only carries with it the risk of exposure to infective *Taenia* eggs, but also to bacteria, protozoa and other helminth eggs. Furthermore, collection of faecal material is not always "acceptable" for social or cultural reasons, nor is it practical i.e. a container has to be given to the individual for collection of the sample and this person has to return it as well. For those reasons, an immunoblot assay that can detect specific antibodies in serum was developed (Wilkins *et al.*, 1999). The assay is based on the identification of two protein groups in *T. solium* excretion/secretion proteins (32.7 and 37.8 kDa) by specific antibodies in the serum of *T. solium* tapeworm carriers. Sensitivity and specificity were determined at 95% and 100%, respectively. In later modifications, the native proteins were replaced by two baculovirus expressed recombinant antigens to eliminate the need for fresh parasite material (Levine *et al.*, 2004, 2007). A drawback of this assay is that antibodies may persist, even though the patient no longer harbours the adult *Taenia* (Deckers and Dorny, 2010).

CHAPTER 3

3

The Hitchhiker's Guide to Proteomics[1]

3.1 Proteomics: what's in a name?

The term *proteome* was coined in 1994 by Marc Wilkins in a symposium on two-dimensional gel electrophoresis in Siena, Italy, and appeared in print one year later (Wilkins *et al.*, 1995). A proteome was defined as the entire PROTEin complement expressed by a genOME. Soon thereafter, the study of a proteome was referred to as *proteomics*. Nowadays, proteomics is a mature post-genomic research field that not only aims to study the proteins in a given cell, but also their isoforms, modifications, structures and interactions (Martins-de Souza, 2014).

The relationship between DNA, mRNA and proteins is not necessarily straightforward. A significant change in mRNA abundance is not necessarily followed by a corresponding change in protein abundance. Gygi *et al.* (1999b) examined 106 yeast genes with corresponding information on mRNA and protein levels. When calculating a Pearson product moment correlation coefficient with only the low abundant 40 to 95 proteins, correlation between protein levels and mRNA levels was bad (consistently between 0.1 and 0.4). If the 11 most abundant proteins were included, the correlation increased to 0.94. They then estimated the correlation for all yeast proteins to be less than 0.4. Two years earlier, Anderson and Seilhamer (1997) came to similar conclusions when studying mRNA and protein abundances in human liver. Similarly, studies of parasitic protozoa, like *Plasmodium falciparum* showed discrepancies between transcriptomic and proteomic data, albeit with a higher correlation of up to 0.59 (Le Roch *et al.*, 2004).
Only 30-40% of the variance in protein abundance is explained purely by mRNA abun-

[1]Title adapted from the novel "The Hitchhiker's Guide to the Galaxy", written by Douglas Adams and published by Pan Books in 1979.

dance. Mainly it is due to extensive regulatory processes occurring after mRNA is made i.e. post-transcriptional regulation, translation efficiency, and protein half-life and degradation (Varshavsky, 1996; Vogel and Marcotte, 2012).

A further interesting observation in this regard was made by Schrimpf *et al.* (2009) and Laurent *et al.* (2010) and involves that fact that the abundances of orthologous proteins across divers taxa (bacteria: *Escherichia coli*, *Pseudomonas aeruginosa*, fungi: *Saccharomyces cerevisiae*, rice: *Oryza sativa*, fruit fly: *Drosophila melanogaster*, nematode: *Caenorhabditis elegans* and humans) correlate remarkably well. Much better in fact than protein abundance versus transcript abundance within each organism or transcript abundances across organisms. This lead the authors to conclude that there is also an evolutionary component involved and changes in mRNA expression could be offset by opposite changes in translation rate or protein half-life, and vice versa. Over evolutionary time scales, such small changes may accumulate, resulting in appreciable changes of mRNA abundance, whereas protein abundance would remain roughly constant.

Examples like this highlight how many of the common "omics" sciences (like *genomics*, *transcriptomics*, *proteomics* and *metabolomics*) complement each other (Figure 3.1).

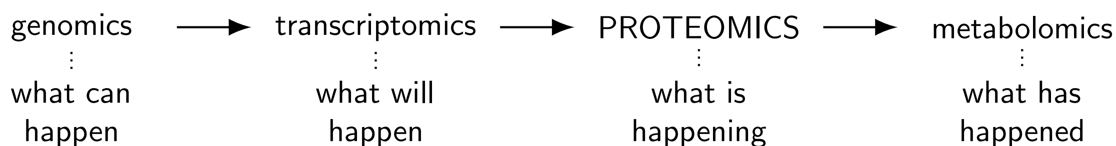


Figure 3.1: The common "omics" sciences complement each other.

In genomics, one studies the genome, which is the complete set of DNA in a cell or organism. Simply put, genomics tries to sequence and assemble the DNA and study the function and structure of the genome. To better understand the functional elements of the genome, researchers started looking at gene transcripts (mRNA). The complete set of transcripts in a cell at a given time, development stage or condition is called the transcriptome. The field of transcriptomics studies these transcripts to determine the transcriptional structure of genes (5' and 3' ends, splicing patterns, ...) and to quantify the changing expression levels of each transcript under various conditions and development stages (Wang *et al.*, 2009b). The translation of mRNA leads towards proteins, proteomes and proteomics. Proteins are often considered the workhorses who carry out the majority of cellular processes. Furthermore, they have received a large amount of attention by researchers looking to develop diagnostic assays, since, for example in the case of *T. solium* cysticercosis, circulating antigens are likely to be easier to detect than circulating DNA (Ramahefarisoa *et al.*, 2010). In order to get the whole story of what happens in a cell, the field of metabolomics studies the substances produced in or by biological/cellular processes.

These metabolites are usually smaller than 1.5 kDa and reflect the sum of all up-stream regulatory events, as well as direct inputs from the external environment.

3.2 Tandem mass spectrometry-based proteomics

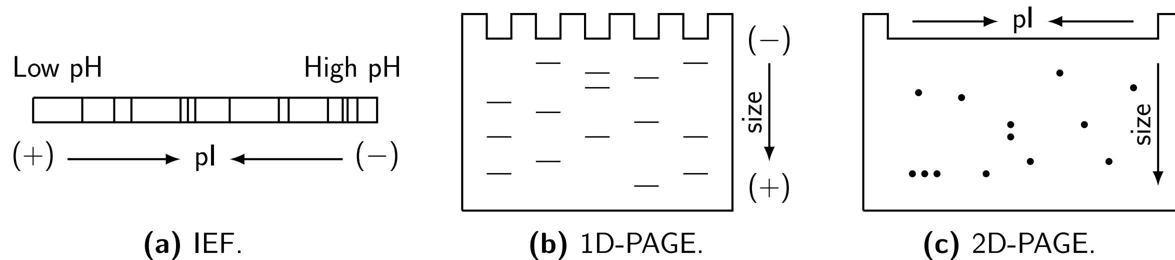
This section describes a generic proteomics workflow. Since it is impossible to cover all the currently available tools, techniques and equipment, the text is limited to those most relevant to this thesis.

3.2.1 Sample preparation and separation

The sample of interest can be just about anything that contains proteins or peptides: eukaryotic tissue and derived extracts, cells and organelles, biological fluids (e.g. plasma or serum, urine, cerebrospinal fluid, saliva, tears, nasal mucus, milk, amniotic fluid, semen or snake venom), prokaryotic organisms, seeds, vegetal matter and plants (Carrette *et al.*, 2006). In cases where endogenous protease activity is to be expected (e.g. after cell lysis), the addition of protease inhibitors is recommended. Further sample preparation often includes concentration steps and general sample clean-up (from centrifugation for the removal of insoluble particles to dedicated clean-up kits to separate proteins from detergents, salts, lipids and nucleic acids).
At this stage, it is of the utmost importance to realise that most if not all of the downstream manipulations and/or analyses operate by the "garbage in, garbage out" principle, meaning that no good results can come from a bad sample preparation (Lidwell *et al.*, 2010).

Currently, no method or instrument is capable of identifying and quantifying the components of a complex sample in a single-step operation (Abdallah *et al.*, 2012). Hence, several separation methods are employed prior to mass spectrometry in order to reduce the complexity of the sample and to increase the identification rate of less abundant proteins.

One of the most common protein separation methods is one-dimensional polyacrylamide gel electrophoresis (1D-PAGE; Figure 3.2b) (Laemmli, 1970) which separates the proteins based on size. Throughout the years, numerous optimisations like neutral pH conditions instead of alkaline, pre-cast gels and ready made reagents have made this a very robust and highly reproducible technique. At around the same time, another separation technique, called isoelectric focussing (IEF; Figure 3.2a), which relied on a separation based on the isoelectric point (pI) of a protein (instead of the size), was also being developed (Svensson, 1961; Vesterberg and Svensson, 1966). A few years later, the two techniques were combined into two-dimensional (2D) PAGE (Figure 3.2c) (MacGillivray and Rickwood, 1974; Klose, 1975; O'Farrell, 1975).



(a) IEF. **(b)** 1D-PAGE. **(c)** 2D-PAGE.

Figure 3.2: Schematic representation of isoelectric focussing (IEF), where proteins are separated based on their isoelectric point (pI; a), one-dimensional polyacrylamide gel electrophoresis (1D-PAGE), where proteins are separated based on their size (b) and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE; c). 2D-PAGE combines IEF as a first dimension of separation and 1D-PAGE as a second dimension.

The advantages of 2D-PAGE are numerous. The high resolving power and large sample loading capacities allow for the simultaneous display of hundreds of proteins on a single gel (Carrette *et al.*, 2006). Furthermore, 2D-PAGE can detect protein isoforms and posttranslational modifications and can demonstrate changes in relative abundance of visualised proteins. These characteristics have made 2D-PAGE a very popular choice for proteomic research (Rabilloud, 2002; Görg *et al.*, 2004; Carrette *et al.*, 2006; Rabilloud and Lelong, 2011).

The benefits of 2D-PAGE notwithstanding, its use has been declining over the past few years. The reasons for this trend lie with the many limitations of the technique: (i) 2D-PAGE tends to identify only the most abundant proteins due to the limited dynamic range of the staining methods compared to the large dynamic range of protein levels in a proteome, (ii) proteins with extreme masses ($<$ 10 kDa or $>$ 150 kDa) and extreme isoelectric point (pI) values (pI $<$ pH 3 or pI $>$ pH 10) are poorly resolved, (iii) poorly soluble proteins (e.g. membrane proteins) are often lost or underrepresented, (iv) different posttranslational modifications (PTMs) can alter the behaviour of a protein during electrophoresis, (v) poor reproducibility, (vi) co-migration of multiple proteins in a single spot, and (vii) the difficulty to automate (Gygi *et al.*, 1999b, 2000; Rabilloud, 2002; Wu and MacCoss, 2002; Lilley *et al.*, 2002; Carrette *et al.*, 2006; Abdallah *et al.*, 2012). The 2D difference gel electrophoresis technique (2D DIGE) addresses some of these shortcomings by enabling the comparison of two different protein samples simultaneously on the same gel (Unlü *et al.*, 1997; Viswanathan *et al.*, 2006; Timms and Cramer, 2008; Minden *et al.*, 2009; Beckett, 2012). Before running the gel, the proteins in each sample are labelled with fluorescent cyanine dyes [Cy3 = sample 1, Cy5 = sample 2 and Cy2 = equal amount of each sample as internal standard as proposed by Alban *et al.* (2003)]. Cyanine dyes have both a high fluorescent extinction coefficient (i.e. they are bright) and they each possess unique absorption and emission spectra (i.e. they each have a different colour). Once labelled, the protein samples can be mixed and separated as a normal 2D-PAGE

sample. During development of the gel, sophisticated imaging equipment scans each dye individually and makes a gel image. Then, the gel images are analysed with appropriate software to detect common proteins (or a change in abundance) as well as proteins unique to any one of the samples. Unfortunately, the software is not without issues and often requires user intervention, especially during spot matching across different gels (Abdallah *et al.*, 2012).

Due to the limitations mentioned above, 2D-PAGE is being phased out by advancements in mass spectrometry, especially for quantitative proteomics. Labelling methods like stable isotopic labelling with amino acids in cell culture (SILAC) (Ong *et al.*, 2002) are designed around the fact that both the labelled and unlabelled peptides have the same chromatographic and ionisation properties, but can be distinguished by their specific mass-shift. However, these labelling methods are often expensive and require large sample volumes and many preparation steps. Therefore, label-free methods have provided a reliable, versatile, and cost-effective alternative to labelled quantitation (Neilson *et al.*, 2011). The specific details about the labelling and label-free methods have been discussed and reviewed by several experts in the field (Gygi *et al.*, 1999a; Ong *et al.*, 2002; Ross *et al.*, 2004; Mann, 2006; Choi *et al.*, 2008a; Lundgren *et al.*, 2010; Neilson *et al.*, 2011). The use of shotgun proteomics combined with multi-dimensional liquid chromatography separation techniques has also provided an alternative to gel-based separation (Wu and MacCoss, 2002; Motoyama and Yates, 2008; Gilmore and Washburn, 2010).
In a technique often referred to as multidimensional protein identification technology (MudPIT), a protein sample is fully digested with a protease and the resulting peptides are separated based on several characteristics before being analysed by the mass spectrometer (Link *et al.*, 1999). This shotgun approach effectively eliminates many of the disadvantages of gel-based separations as well as any loss related to incomplete sample extraction for the gel (Rabilloud, 2002).

Whether or not proteins were separated first, they still need to be processed before they can be analysed with a mass spectrometer. This generally implies (i) the addition of dithiothreitol (DTT) to reduce the cystines, (ii) the addition of iodoacetamide to alkylate the cysteines, (iii) the addition of a protease, commonly trypsin, to digest the proteins (Olsen *et al.*, 2004) and, when using PAGE, (iv), an elution of the resulting peptides from the gel piece (Shevchenko *et al.*, 1996, 2006).

3.2.2 Liquid chromatography and tandem mass spectrometry

Liquid chromatography

Protein sample complexity may have been reduced by a first separation step (e.g. 1D-PAGE), but the subsequent tryptic digest again adds complexity. Therefore, another separation step is required, this time on the peptide level. Frequently, the method of choice is reversed-phase liquid chromatography (RPLC). Peptides are loaded onto a stationary phase made out of C18 chains bound to a silica base material. A gradual increase of organic solvent (e.g. acetonitrile) in the mobile phase will elute the peptides based on the strength of their hydrophobic interaction with the stationary phase. The eluted peptides can either be collected in fractions (off-line) or can be directly loaded onto the ionisation source of the mass spectrometer (on-line). Although RPLC is much used, it is often overlooked when optimising the experimental set-up. Peterson *et al.* (2009) found that a slight decrease (from 5% to 2%) in the concentration of acetonitrile during the desalting step, resulted in a 26% increase in peptide identifications. Logically, most of these peptides were hydrophilic.

When the protein sample is digested without prior separation, the resulting sample complexity often requires multidimensional liquid chromatography before sufficient separation is achieved (Motoyama and Yates, 2008). One of the most widely used combinations is ion exchange chromatography (e.g. strong cation exchange or SCX) and subsequent RPLC (Wolters *et al.*, 2001).

Sample ionisation

In short, mass spectrometry is the measurement of the mass-to-charge ratio (m/z) of gas-phase ions. Hence, peptides need to be ionised and transferred to the gas phase before they can be analysed. The two most used ways to achieve this are matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI).

The MALDI technique (as it was introduced by Karas *et al.* in 1987) uses a nitrogen laser (at 337 nm) to shoot at a matrix/analyte mixture that is fixed on a target plate (Figure 3.3). The ultraviolet energy from the laser is absorbed by the matrix material and transferred to the analytes, causing both to transfer to the gas phase. Ionisation of the analytes is believed to happen during or shortly after the laser pulse as well as during secondary reactions in the expanding plume of desorbed material (Knochenmuss, 2006). Combining (reversed-phase) liquid chromatography and MALDI is possible in an off-line format i.e. the liquid chromatography outlet is not directly coupled to the ionisation source. Instead, the LC eluent has to be mixed with MALDI matrix first and then spotted onto target plates (Pereira *et al.*, 2013). Although not as convenient as a direct coupling (on-line), it does offer a big advantage: preservation of the chromatographic separation on a

MALDI target plate yields as much time as desired to obtain (tandem) mass spectrometry spectra, without risk of missing ions as they elute from the LC column (Zhen *et al.*, 2004).

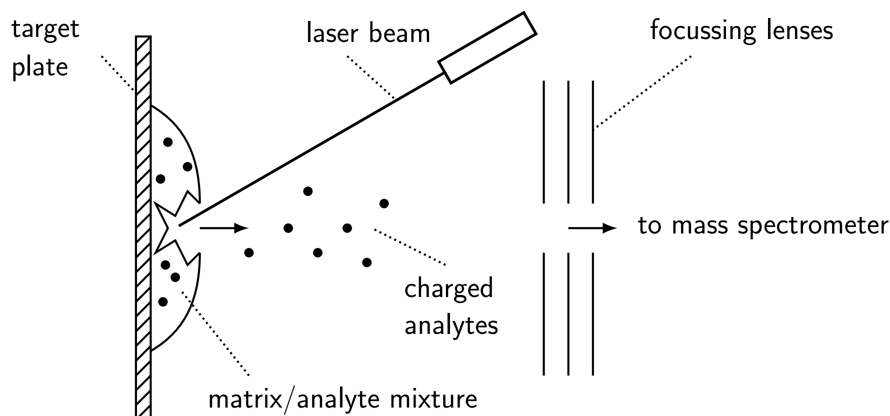


Figure 3.3: Principle of matrix-assisted laser desorption/ionisation (MALDI).

3

ESI also became popular in the 1980s (Yamashita and Fenn, 1984; Fenn *et al.*, 1989) and, like MALDI, it is a soft ionisation method, meaning it does not fragment (larger) molecules. The ESI principle involves a flow of liquid analytes, a thin conducting needle at 3-4 kV and a potential difference between the needle and the inlet of the mass spectrometer (Figure 3.4). Positively charged analytes accumulate at the liquid surface, while negatively

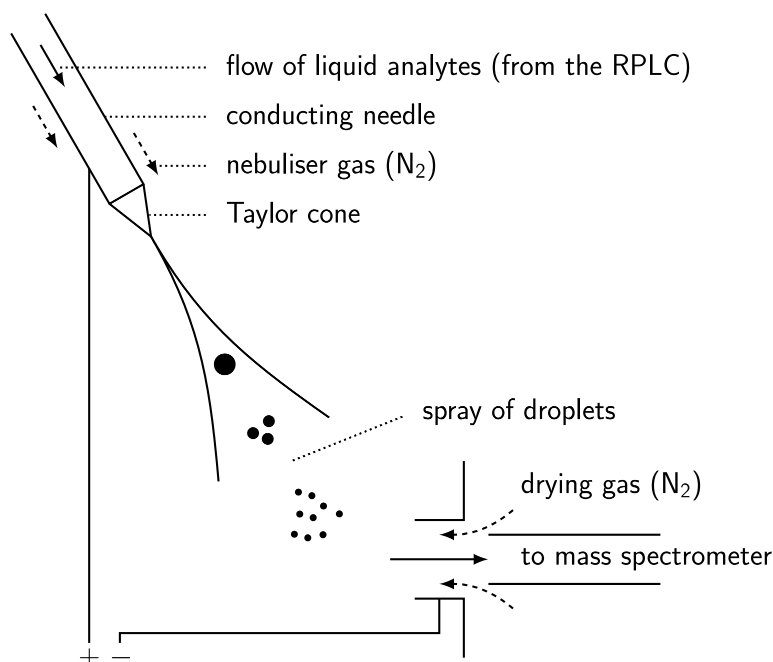


Figure 3.4: Principle of electrospray ionisation (ESI).

charged particles will be drawn to the positive wall of the needle. This accumulation of positive charges in the liquid leads to a Taylor cone (Taylor, 1964), which under the influence of the high electric field and Coulomb's law, ends in a fine spray of droplets containing positively charged analytes. A flow of (warm) N_2 nebuliser and drying gas

helps evaporate the liquid in the droplets, until they are so small that the Coulomb forces between the like charges exceed the surface tension and the droplet breaks up into smaller droplets. This process continues until the radius of the droplet is less than 10 nm at which point direct emission of an ion into the gas phase takes place (ion evaporation model) (Iribarne and Thomson, 1976; Nguyen and Fenn, 2007). The outlet of the RPLC can easily be coupled to the inlet of the ESI source (on-line). The mobile phase of the RPLC contains organic solvents which evaporate easily and the addition of acid (e.g. formic acid) lowers the pH to assist protonation of the analytes.

Unlike MALDI, where mostly singly charged ions are produced, ESI produces multiply charged ions. This results in a shorter m/z range.

Mass analysers

Once the charged analytes are transferred to the gas phase, they enter the heart of the mass spectrometer, called the mass analyser, where they are separated based on their m/z ratio. Nowadays, there is a wide range of mass analysers available with the most popular being the time-of-flight (TOF) analysers (Wiley and McLaren, 1955), quadrupole mass filters (Paul and Steinwedel, 1953), quadrupole ion traps (Paul and Steinwedel, 1960; March, 1997; Jonscher and Yates, 1997), orbitraps (Makarov, 2000; Zubarev and Makarov, 2013) and Fourier-transform ion cyclotron resonance (FTICR) mass analysers (Comisarow and Marshall, 1974). In some cases, multiple mass analysers are combined to form a hybrid mass spectrometer.

The performance of mass spectrometers is often assessed by a number of properties (de Hoffmann and Stroobant, 2007):

Mass range: The m/z range over which a mass analyser can measure ions with sufficient sensitivity to be practically useful.

Analysis speed: The rate at which the analyser measures over a particular mass range, the number of spectra per unit of time.

Transmission: The ratio of the number of ions reaching the detector and the number of ions entering the mass analyser.

Mass accuracy: The accuracy of the m/z provided by the mass analyser. It is the difference that is observed between the theoretical m/z and the measured m/z.

Resolution or resolving power: The ability to distinguish two peaks of slightly different m/z.

Linear dynamic range: The range over which the ion signal is linear with the analyte concentration.

Detailed explanations about and comparisons between mass analysers would be (i) beyond the scope of this thesis and (ii) quickly outdated, since technical advancements in the field of mass spectrometry (both software and hardware) are fast and numerous. That being

said, since a Bruker Daltonics amaZon ETD ion trap mass spectrometer (`www.Bruker.com`) was used for the analyses described in this thesis, further explanations will use the 3D ion trap mass spectrometer as example.
All things considered, there is no "ideal" mass spectrometer that does everything perfectly and one should select the most optimal mass spectrometer for the envisioned task.

Tandem mass spectrometry

In general, tandem mass spectrometry (MS/MS) is the practice of carrying out one mass analysis after another. The first mass-selective operation is meant to isolate an ion, called the parent or precursor ion, while the second mass-selective operation will determine the m/z ratios of the fragment ions formed by fragmentation of the precursor ion (March, 1997).

The core of a 3D ion trap mass spectrometer is roughly the size of a tennis ball and consists of three hyperbolic electrodes: a ring and two endcaps (Figure 3.5) (Jonscher and Yates, 1997). In this chamber –the actual ion trap– ions with different m/z ratios are accumulated and trapped using constant DC and radio frequency (RF) oscillating AC electric fields. Once trapped, ions are sequentially ejected from low m/z to high m/z by ramping up the RF voltage. The ejected ions are detected by an electron multiplier to produce a mass spectrum (Stafford, 2002).

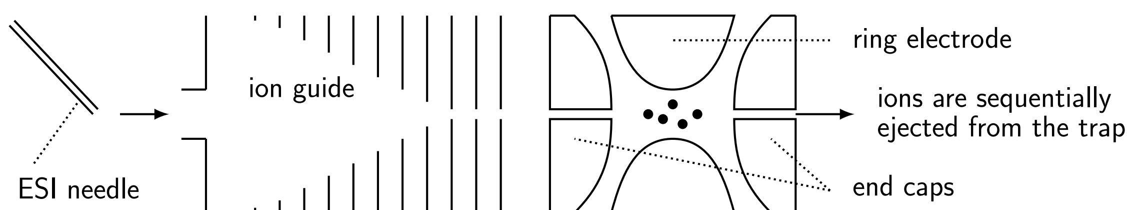


Figure 3.5: Schematic representation of an ion trap. Ions are guided from the electrospray ionisation (ESI) source to the ion trap. They enter through an opening in the end cap and are accumulated in the centre of the trap. Once accumulated/trapped, the ions are sequentially ejected from low m/z to high m/z by ramping up the radio frequency amplitude.

If an ion of interest (parent ion or precursor ion) is observed in this MS spectrum, the ion trap chamber is filled again with ions, but this time, all ions with an m/z ratio higher or lower than the parent ion are ejected, so that only the parent ions remain in the chamber. These precursor ions are then allowed to collide with the He buffer gas in the chamber. This collision-induced dissociation (CID) results in a fragmentation of the peptide backbone and, in the case of singly charged precursor ions, in the formation of b ions, when the charge is retained at the N-terminal or y ions, when the charge is retained at the C-terminal (Figure 3.6) (Roepstorff and Fohlman, 1984; Biemann, 1990; Paizs and Suhai, 2005). The presence of basic residues in the fragments greatly influences the formation of the b and y ions i.e. the fragment containing the most basic site is most likely to retain the charge. Since trypsin is the enzyme of choice in many proteomics experiments, and trypsin yields

peptides that terminate in basic arginine and lysine residues, fragmentation of these peptides tends to favour prominent y series (Tabb *et al.*, 2004; Olsen *et al.*, 2004). Doubly charged precursor ions can produce both a b ion and a y ion, but fragment ions that contain two basic residues may accept both protons during fragmentation, which, in turn, leads to a doubly charged fragment ion (Tabb *et al.*, 2004).

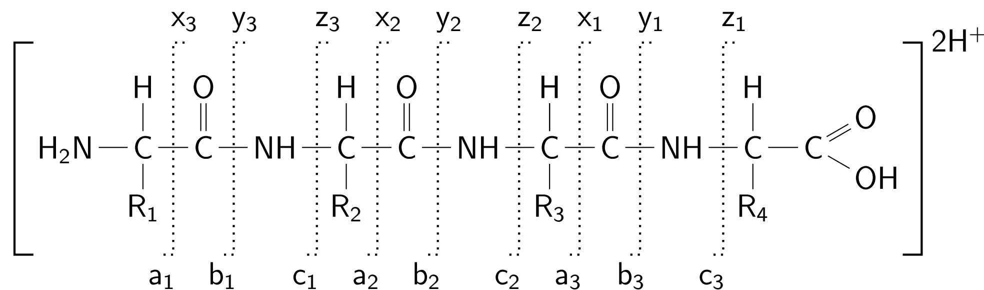


Figure 3.6: Theoretical fragmentation of a doubly charged parent ion. Charges can be mobile as illustrated by the square brackets around the ion. Collision-induced dissociation leads to fragmentation of the peptide backbone in *b* ions (charge retained at the N-terminal) and/or *y* ions (charge retained at the C-terminal). The presence of basic residues in the fragments greatly influences the formation of the *b* and *y* ions and fragments that contain two basic residues may even accept both protons during fragmentation.

All aspects of this procedure are controlled by software and many of them can be modified by the user to optimise the quality of the data e.g. one could set the mass spectrometer to select the 10 most intense peaks of each MS scan for MS/MS analysis, but to exclude them if they were already selected during the past 15 seconds. This would force the mass spectrometer to skip the most abundant precursors and to select less abundant ones instead.
At the end of the mass spectrometry experiment, the user has MS spectra (with precursors mass information) and, for a number of precursors, MS/MS spectra (with fragmentation information).

3.2.3 Data analysis

The next step in the process is to assign peptide sequences to the experimental spectra. Generally speaking, there are three ways to go about this: *de novo* analysis, spectral library searching and protein database searching.

It is possible to derive the peptide sequence directly from the fragmentation spectrum, without the assistance of a protein databases. This is called *de novo* analysis and the idea behind it is that the mass difference between e.g. fragment ions b_3 and b_2 is the mass of an amino acid residue. By overlapping all b ions and matching the mass differences to amino acids, the amino acids sequence can be determined (Figure 3.7). The same can be done with the y ions and, if done correctly, the sequence should be the same.

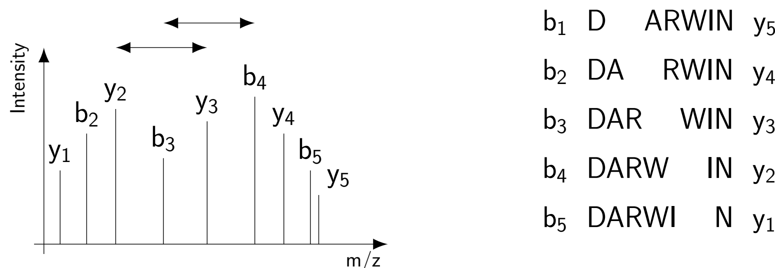


Figure 3.7: *De novo* analysis of the collision-induced fragmentation spectrum of the DARWIN peptide. The difference between each *b* or *y* ion is one amino acid residue e.g. b_3 and b_4 differ by the amino acid Tryptophan (W), which corresponds to a mass difference of 186.07931 Da. Likewise, y_2 and y_3 also differ by 186.07931 Da. Calculating the differences between the *b* and *y* peaks eventually reveals a (partial) peptide sequence.

Although *de novo* analysis sounds straightforward, the spectra have to be of excellent quality and there are a number of extra difficulties, like mass shifts due to loss of water or ammonia and incomplete *b* or *y* series. Furthermore, evergrowing datasets have made manual annotation of spectra increasingly difficult. Fortunately, specialised computer algorithms have been developed to cope with many of these issues (Ma *et al.*, 2003; Frank and Pevzner, 2005; Seidler *et al.*, 2010).

The spectral library searching approach takes experimental MS/MS spectra and compares them against a spectral library which is compiled from a large collection of experimentally observed MS/MS spectra identified in previous experiments (Yates *et al.*, 1998; Craig *et al.*, 2006b; Frewen *et al.*, 2006). Spectral library search tools like SpectraST (Lam *et al.*, 2008) use a set of user defined parameters (e.g. mass tolerance windows) to compare each experimental spectrum to a limited list of candidate spectra from the library (Figure 3.8). The best match is then retained.

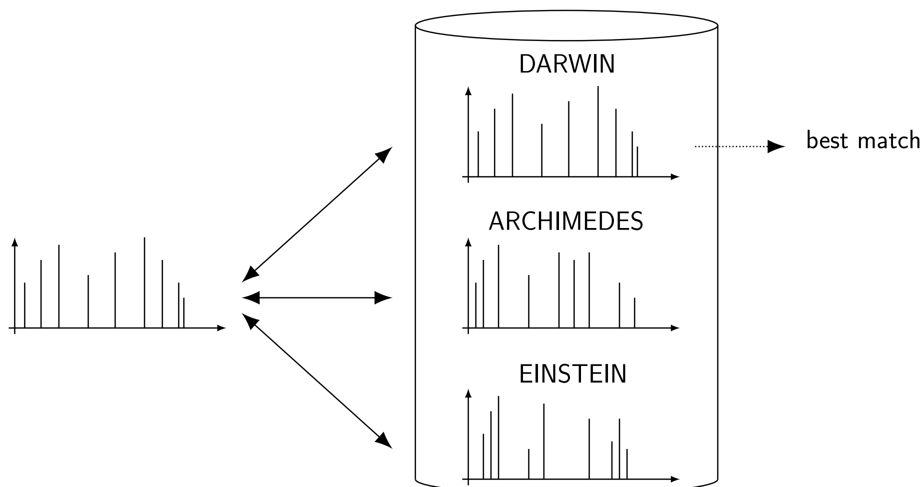


Figure 3.8: Spectral library searching approach. An experimental spectrum is compared to a subsection of identified spectra from the spectral library. Each match is scored and the best match is retained.

Although spectral library matching outperforms protein database searching in terms of speed, error rates, and sensitivity of peptide identification (Lam *et al.*, 2007), the technique is still underused. The most notable limitation is with the spectral libraries. Only previously identified peptides whose experimental spectra are uploaded to a library can be identified. Furthermore, spectra that were generated under specific conditions (modifications, mass spectrometer, ...) can not always be matched to spectra generated under other conditions. Finally, spectra from lower abundant peptides are often scarce.

In the mid-1990s, an almost revolutionary breakthrough in protein identification was made with the development of computer algorithms that converted the character-based representations of amino acid sequences in a protein database to a fragmentation pattern that could then be used to match fragment ions in a MS/MS spectrum. One of the widely used early algorithms was SEQUEST, first described by Eng *et al.* in 1994.
It should be made clear that the (almost simultaneous) development of these algorithms did not happen by chance at that particular time. On the contrary, since these tools relied heavily on the availability of protein databases and since protein sequences are encoded in DNA, it comes as no surprise that the development of these database search tools coincided with the enormous efforts to sequence the genomes of several model organisms such as *Haemophilus influenzae*, the first bacterial genome, by Fleischmann *et al.* (1995), the *Escherichia coli* K12 genome by Blattner *et al.* (1997), the first eukaryotic genome (*Saccharomyces cerevisiae*) by Goffeau *et al.* (1996) and the Human Genome Project, which was initiated in 1990 (Watson, 1990). This is a perfect illustration of the strong link between the "omics" sciences: the huge leap forward in the field of genomics enabled major progress in protein identification strategies.

The above mentioned algorithms marked the beginning of the protein database searching approach and since then, a large number of computational tools have been developed, with Mascot (Perkins *et al.*, 1999), X!Tandem (Craig and Beavis, 2004) and OMSSA (Geer *et al.*, 2004) being amongst the most popular. They all share the same basic idea (Figure 3.9): the search program takes experimental MS/MS spectra and compares them to a subset of theoretical MS/MS spectra that were created *in silico* from a protein database, using a number of user defined parameters (e.g. mass tolerances of parent and fragment ions, allowing for tryptic peptides only and fixed and variable posttranslational modifications) (Nesvizhskii, 2010, 2007).
The degree of similarity between the experimental spectrum and a theoretical *in silico* spectrum is scored by the search program and all theoretical spectra are then ranked according to their score. Most search tools consider only the top scoring peptide spectrum match (PSM) for each experimental MS/MS spectrum as the potential peptide identification.

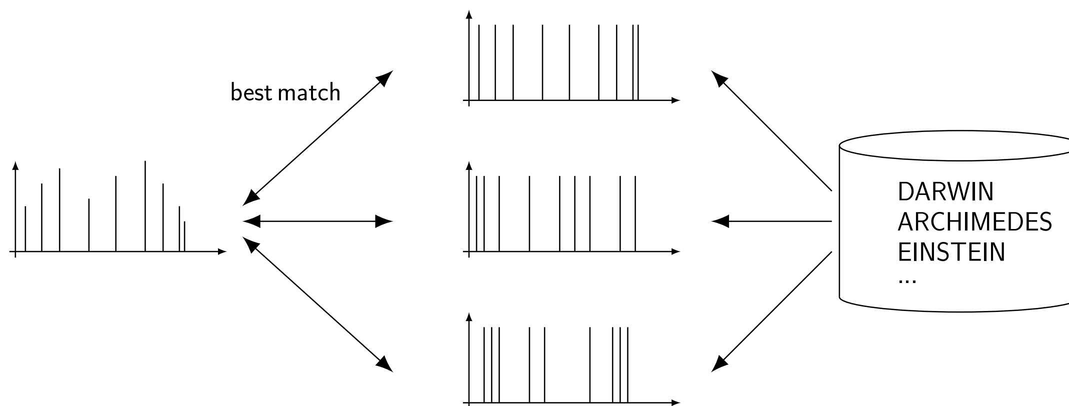


Figure 3.9: Protein database searching approach. The amino acid sequence of a peptide in the database is translated to a theoretical fragmentation spectrum. The experimental spectrum is then compared to a subset of the computer generated spectra and each peptide spectrum match is scored. Finally, the best match is retained.

It is important to realise that the protein database search tool will produce a PSM for nearly all experimental spectra, but that the majority of these PSMs are incorrect (especially when the dataset was created on low mass accuracy instruments or when peptide sequences are not in the protein database) (Choi *et al.*, 2008b; Nesvizhskii, 2010). It is therefore necessary to assess the confidence of the PSMs and to estimate error rates associated with filtering the data using various thresholds (Carr *et al.*, 2004; Choi and Nesvizhskii, 2008b). Assessing the confidence of PSMs can be roughly divided into single-spectrum and global (whole dataset) modelling approaches. The expectation value (E-value) is the most used single-spectrum statistical confidence measure and refers to the expected number of peptides with scores equal to or better than the observed best match score under the assumption that peptides are matching the acquired MS/MS spectrum by random chance (Fenyö and Beavis, 2003; Choi and Nesvizhskii, 2008a). It is achieved by modelling the distribution of the search tool score(s) constructed separately for each individual experimental spectrum from all candidate peptides in the searched protein database that were scored against it (Choi and Nesvizhskii, 2008b). However, the conversion of a search tool score into an E-value does not control the overall identification error rates, since its construction does not specifically involve steps for multiple testing correction and, although a necessity when the analysis involves simultaneous processing of multiple MS/MS spectra, classical multiple testing corrections like the Bonferroni correction were never designed for such large datasets and give overly conservative results (Choi and Nesvizhskii, 2008a; Choi *et al.*, 2008b; Nesvizhskii, 2010). Therefore, additional analysis and modelling is required to obtain statistical measures more suitable for filtering of very large PSM datasets (Choi *et al.*, 2008b; Nesvizhskii, 2010).

When dealing with large datasets, the most reported statistical measure is the false dis-

covery rate (FDR) (Benjamini and Hochberg, 1995). The FDR, which deals with the PSM dataset at a global level, can be defined as the expected number of incorrect PSMs amongst all PSMs that pass a certain search score threshold (Choi and Nesvizhskii, 2008a; Käll *et al.*, 2008).

The most popular way to estimate the global FDR is to use a target/decoy database approach (Figure 3.10). Here, the number of incorrect PSMs against the original, target protein database is estimated by searching a decoy database, which contains "fake" proteins sequences (e.g. reversed, scrambled or randomised protein sequences from the target database). This decoy database is searched separately or simultaneously as a concatenated target+decoy database. The number of decoy PSMs above the threshold score can then be used as a measure of the number of incorrect target PSMs and the desired FDR can be achieved by selecting a different threshold score. The most simple FDR estimation can thus be defined as N_{decoy}/N_{target} where N_{decoy} and N_{target} are the number of hits above the search score threshold T_s in the decoy database and the target database, respectively (Jeong *et al.*, 2012). Depending on the way the search is performed (separate or as a composite database) and the way the decoy database is constructed [reversed, scrambled or randomised (Elias and Gygi, 2010)] the use of a correction factor (Käll *et al.*, 2008), an alternative formula (Elias and Gygi, 2007; Navarro and Vázquez, 2009; Wang *et al.*, 2009a), an optimal decoy construction and search strategy (Bianco *et al.*, 2009) and a form of FDR quality control (Vaudel *et al.*, 2011b) have been suggested.

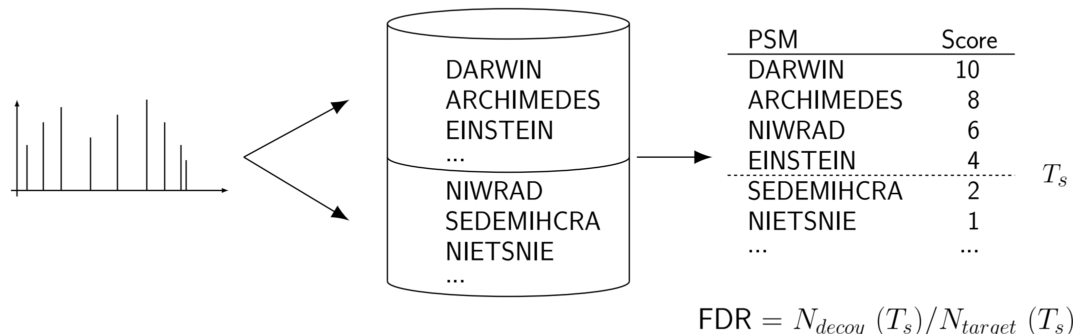


Figure 3.10: Global false discovery rate (FDR) estimation by using a concatenated target+decoy database database searching approach. Once the peptide spectrum matches are ranked by score, the FDR is estimated by taking the number of hits in the decoy database (reversed sequences) that pass the threshold score T_s divided by the number of hits in the target database (normal sequences) that pass the same threshold score. The desired FDR can be achieved by adjusting T_s.

An alternative way to estimate the global FDR is to calculate the posterior probability of correct identification for each individual PSM by means of a mixture model-based approach (Efron *et al.*, 2001) and to use those probabilities to derive global FDR estimates (Figure 3.11). This approach is implemented in the PeptideProphet (Keller *et al.*, 2002) and was later updated with three modifications (Ma *et al.*, 2012): (i) the optional use of a

decoy database to assist the expectation-maximisation (EM) algorithm (Dempster *et al.*, 1977) in establishing a more reliable distribution of incorrect PSMs (Choi and Nesvizhskii, 2008b), (ii) a non-parametric option that allows accurate fitting of the models without prior selection of the optimal distribution models to best fit the data (Choi *et al.*, 2008b) and (iii), the implementation of an adaptive method in which a new discriminant function is learned from the data, instead of using fixed weighting coefficients (Ding *et al.*, 2008). To improve the discrimination between correct and incorrect PSMs, the calculation of the

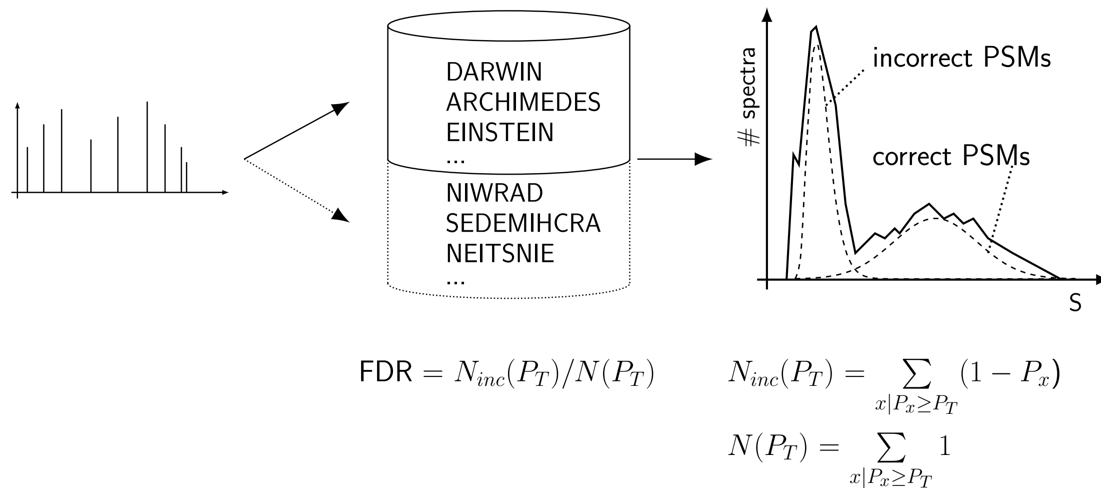


Figure 3.11: Global false discovery rate (FDR) estimation by calculating posterior probabilities P for each peptide spectrum match. A discriminant score S (calculated from the search tool scores) is plotted and two distributions (dashed lines) are fitted to the observed data. Posterior probabilities are computed using this data along with auxiliary information (e.g. mass accuracy). A global FDR can then be derived from these probabilities. The desired FDR can be achieved by moving the threshold probability P_T up or down. The use of a decoy database is optional with this approach.

posterior probabilities not only takes into account the search tools score(s) (recalculated into a discriminant score S), but also auxiliary information like the number of missed cleavages (NMC), the number of tryptic termini (NTT) and the mass accuracy (ΔM; the difference between the measured and calculated mass of the peptide in the first MS step) (Nesvizhskii, 2010).

Much research has gone into the identification of correct PSMs and the estimation of a FDR at that level, but ultimately, the main purpose of most proteomic analyses is the identification of proteins. This final step of the data analysis basically involves the assembly of the identified peptides into (a list of) proteins. As was the case at the PSM level, a FDR is also estimated at the proteins level. Here, however, a common observation is that correct PSMs tend to identify a relatively small group of (abundant) proteins, while each incorrect PSM tends to (randomly) match a new (incorrect) protein. This results in a higher FDR at the protein level, even when the PSM FDR was low. To further complicate matters, the same peptide sequence can be present in multiple proteins (e.g.

isoforms, homologous proteins, redundant entries in the database). This is often referred to as the "protein inference problem". These shared (degenerate) peptides can lead to ambiguities in identifying sample proteins (Nesvizhskii and Aebersold, 2005; Li *et al.*, 2009).

Different approaches have been described to estimate the FDR at the protein level. The target+decoy strategy (popular for FDR estimations at the PSM level), can be extended to the protein level. In its simplest form, this is done by estimating the number of incorrect protein identifications by counting the decoy proteins. Considering that a decoy protein is by definition wrong, but a target protein is considered correct if identified by at least one true positive PSM (regardless of the number of decoy PSM that match as well), the N_{decoy}/N_{target} estimation tends to be overly pessimistic as the number of decoy proteins is often higher than the number of incorrect target proteins. A tool called MAYU (Reiter *et al.*, 2009) also uses the target+decoy approach but combines it with a hypergeometric distribution to estimate more reliable FDRs, even for very large (multi-replicate) datasets. Although an improvement, MAYU still requires preprocessing of the data to solve the protein inference problem.

ProteinProphet (Nesvizhskii *et al.*, 2003) builds on the posterior probabilities that were calculated for PSMs by PeptideProphet (Keller *et al.*, 2002) and uses them to calculate protein posterior probabilities which, in turn, can be used to filter the data to a desired FDR. Although there are a number of ways to assign a probability to a protein (e.g. take the highest PSM probability or assign probability 1 if the protein is identified by at least two good peptides), the method implemented in ProteinProphet uses a cumulative score (Feng *et al.*, 2007; Li *et al.*, 2009). Equation (3.1) illustrates the calculation of the protein posterior probability based on the PSM posterior probabilities of the peptides that are associated with that protein.

Where p' is the NSP modified PSM posterior probability:

$$\text{protein posterior probability} = 1 - \prod_i (1 - p'_i) \tag{3.1}$$

In Equation (3.1), p' is a modification of the original PSM posterior probability, since using the original probabilities would overestimate the probability for a number of proteins, notably those proteins identified by only one (incorrect) peptide. To deal with this problem, ProteinProphet will re-evaluate the PSM posterior probabilities based on the number of sibling peptides (NSP). Practically, for proteins that are recognised by multiple different peptides, this will lead to an increase in the PSM posterior probabilities, while for proteins that are only recognised by a single peptide, this will result in a decrease of the peptide posterior probability. Logically, NSP correction of the PSM posterior probability will penalise all proteins that are identified by a single peptide. However, since this would

be overly negative for correct single peptide identifications [which can be 20-60% of all correctly identified proteins (Higdon and Kolker, 2007; Nesvizhskii, 2010)], ProteinProphet will also take into account the quality of the peptide i.e. if the PSM posterior probability is very high, the penalisation will be less severe. This goes somewhat against the established "two peptide rule", but recent research showed that one very good peptide is often better than two or more not-so-good ones (Gupta and Pevzner, 2009). Once the protein posterior probabilities have been calculated, the FDR is estimated much like it was at the peptide level, as shown by Equation (3.2), and as was also the case with the PeptideProphet, FDRs can be estimated without the use of decoy databases.

Where P_x is the protein posterior probability above a certain probability threshold P_T:

$$\text{protein FDR} = \frac{\sum\limits_{x|P_x \geq P_T} (1 - P_x)}{\sum\limits_{x|P_x \geq P_T} 1} \tag{3.2}$$

A final issue to address is the protein inference problem. The drastic decision to omit all shared peptides and only accept non-shared peptides would be much too conservative. ProteinProphet employs a number of algorithms that look at the proteins that have a particular degenerate peptide in common and, based on additional evidence like the number of non-degenerate peptides that match to a protein, a weight (0 to 1) is given to that peptide (Figure 3.12).

These assigned weights, combined with the NSP information and empirically determined minimum thresholds for weight and PSM posterior probability (0.5 and 0.2, respectively) are then taken into account when calculating the protein posterior probability (Nesvizhskii *et al.*, 2003).

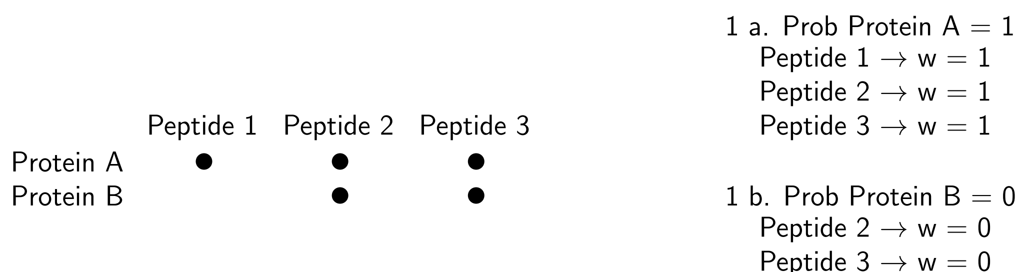


Figure 3.12: Illustration of shared peptides – the protein inference problem. Proteins A and B share peptides 2 and 3, but only Protein A contains Peptide 1. Peptides 1, 2 and 3 combine to support the presence of Protein A, as is reflected by the weight assigned to these peptides (w = 1). Similarly, the lack of conclusive evidence for Protein B is reflected in the weight (w = 0) assigned to Peptides 2 and 3.

All this combined, results in the smallest possible list of protein groups and subgroups that can account for all the observed peptides. This is the parsimony principle, also referred to as Occam's (or Ockham's) razor (Ockham, 1495; Nesvizhskii and Aebersold, 2005). This

complete list can then be shortened by setting a probability threshold that corresponds to the desired FDR.

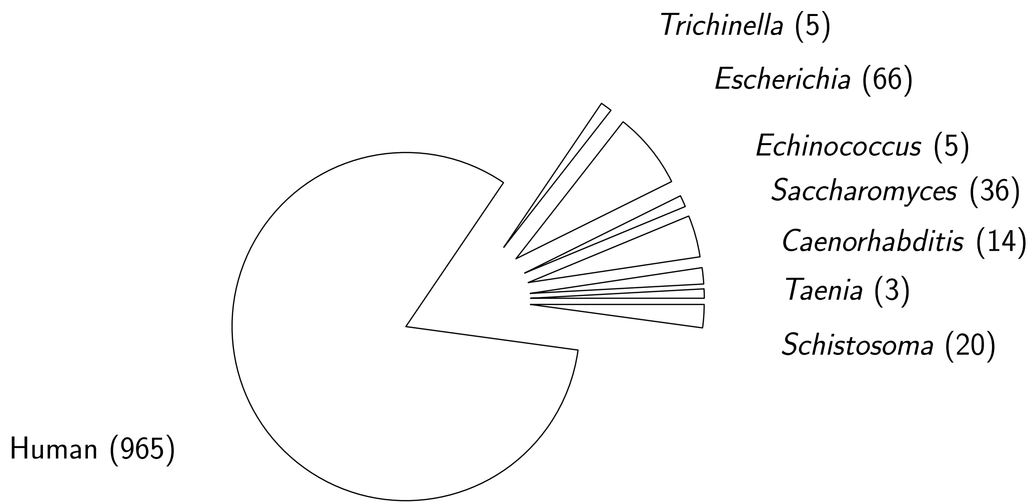
X!Tandem, PeptideProphet and ProteinProphet are combined (along with a number of other tools) in the Trans-Proteomic Pipeline (TPP) (Keller *et al.*, 2005; Deutsch *et al.*, 2010). Although the TPP is convenient and widely used, a number of research groups have developed tools and pipelines that are equally good or better at various aspects of the data analysis e.g. MS-GFDB (Kim *et al.*, 2010), MAYU (Reiter *et al.*, 2009), SearchGUI (Vaudel *et al.*, 2011a), PeptideShaker (`https://code.google.com/p/peptide-shaker/`) (Barsnes *et al.*, 2011) and a feedback framework that outputs a list of proteins and a list of corresponding peptides at the same time by iteratively updating the two lists with a feedback from the inferred proteins to the selection of correct peptides (Shi and Wu, 2012).

Since computational power is vital to the efficient processing of (very) large datasets and many desktop computers simply can no longer process data as fast as it is produced, advancements in the field of scientific workflow design and cloud computing have also made their entry into the field of proteomics. Contrary to pipelines, which allow for sequential processing of one dataset at a time, these advanced workflows, combined with cloud computing, allow for parallel processing, making the data analysis much faster and (if desired) fully automated (Oinn *et al.*, 2004; de Bruin *et al.*, 2012; Mohammed *et al.*, 2012).

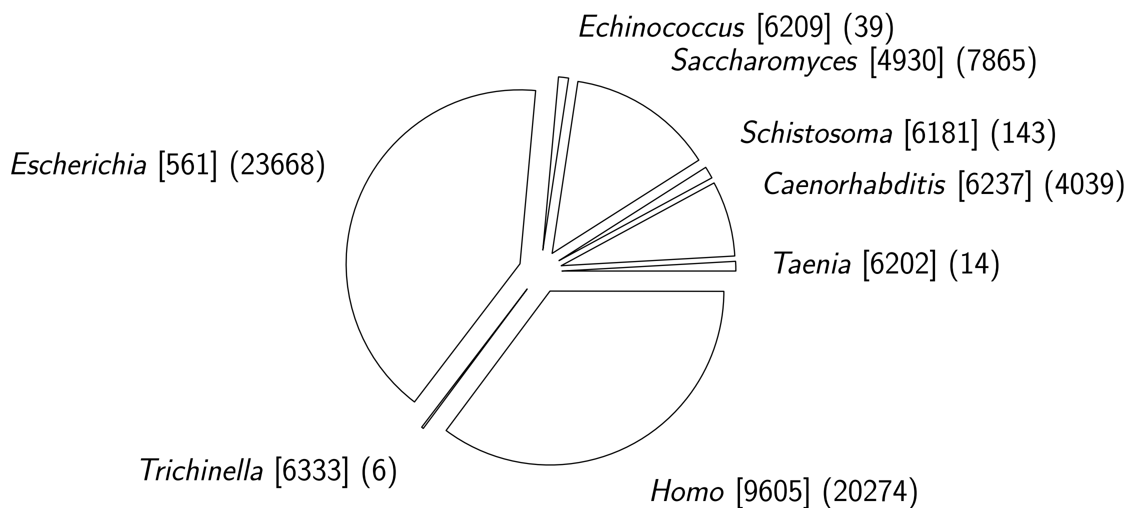
3.3 Helminth proteomics

In comparison to human and, to a lesser extent, *Escherichia* and *Saccharomyces*, (large scale) proteomic studies on helminths (and especially cestodes) have not received much attention (Figure 3.13a). A comparison of the number of reviewed proteins per genus reveals similar results, only here, *Escherichia* proteins are very well represented, mainly due to the large number of different strains (Figure 3.13b). These two figures help illustrate how research in the field of proteomics is currently dominated by a small number of organisms.

A major bottleneck for proteomic experiments involving helminths (and cestodes like *Taenia* in particular) has been the incompleteness of protein databases due to the lack of genomic information and comprehensive expressed sequence tag (EST) libraries.
In 2006, researchers at the Universidad Nacional Autonoma de Mexico started "The genome project of *Taenia solium*" (Aguilar-Díaz *et al.*, 2006) and although there was mention of an advanced draft status in 2012 (Olson *et al.*, 2012), the genome was not presented until April 2013, when the high quality reference genome of *Echinococcus multilocularis* was



(a) Illustration of the limited amount of proteomic studies done on helminths compared to human, *Escherichia* and *Saccharomyces*. Results obtained with a search on PubMed (`http://www.ncbi.nlm.nih.gov/pubmed`) on December 7, 2013 using the query '(proteomic[Title]) AND (*genus*[Title])'. The number of hits for each query is shown within parentheses.



(b) Illustration of the number of reviewed proteins per genus. Results obtained with a search on UniProtKB (`http://www.uniprot.org`) on December 17, 2013 using the query 'taxonomy:"*genus* [*taxonomy code*]"'. The number of reviewed proteins (UniProtKB/Swiss-Prot) for each query is shown within parentheses, while the taxonomy code is shown within square brackets.

Figure 3.13: Two illustrations of the underrepresentation of proteomic studies on helminths (a) and the number of reviewed helminth proteins (b) compared to some of the "big players" in the field e.g. *Homo sapiens*, *Escherichia* and *Saccharomyces*.

presented, along with the *E. granulosus*, *T. solium* and *Hymenolepis microstoma* genomes for comparison (Tsai *et al.*, 2013). Until then, Cestoda (tapeworms) were the only class for which no genome sequence had been available so far. Trematoda (flukes) and Nematoda (round worms) all had members of their respective class sequenced (Berriman *et al.*, 2009; Young *et al.*, 2012; *C. elegans* Sequencing Consortium, 1998).
Cestode proteomics is currently in its infancy and hopefully, the field will quickly see a major boost now that cestode genomics have made such promising progress.

In the case of *T. solium*, most studies have focussed on characterising one (or very few) protein(s), mostly due to their importance as vaccine candidates or diagnostic antigens in antibody detecting assays (Deckers and Dorny, 2010) e.g. diagnostic antigen GP50 (Hancock *et al.*, 2004), T24 (Hancock *et al.*, 2006) and the 8 kDa protein family (Hancock *et al.*, 2003; Ferrer *et al.*, 2007), which are the three protein groups that represent the seven diagnostic bands in the EITB (Tsang *et al.*, 1989). In recent years, advancements in technologies and data analysis have facilitated the study of larger groups of proteins e.g. in relationship to immunological significance (Salazar-Anton and Lindh, 2011; Diaz-Masmela *et al.*, 2013) or related to a specific parasite stage like activated *T. solium* oncospheres (Santivañez *et al.*, 2010) or *T. solium* metacestode excretion/secretion proteins [Chapters 6 and 7 and Victor *et al.* (2012b, 2013)]. *Echinococcus*, the other genus in the Taeniidae family, has received a bit more attention, especially *E. granulosus* (Monteiro *et al.*, 2010; Aziz *et al.*, 2011; Virginio *et al.*, 2012; Parkinson *et al.*, 2012). Furthermore, a number of papers that present an overview of *Echinococcus* related molecular cross talk and host-parasite interactions have been published (Siracusano *et al.*, 2008a,b; Brehm, 2010; Siracusano *et al.*, 2012). Since the *Taenia* and *Echinococcus* genera are so closely related, it is reasonable to assume that a (large) number of proteins will be homologous in sequence and function [e.g. González *et al.* (2007)] and that much of the molecular cross talk described for *Echinococcus* will also be applicable to *Taenia*. This concept of homology is even likely to be expandable to other studied helminths like *Schistosoma* (Mutapi, 2012; Hong *et al.*, 2011; Liu *et al.*, 2009, 2007, 2006).

Incomplete protein databases can never offer full proteome coverage, which limits the number of identifiable proteins. In the absence of a (reference) genome, this limitation can be (partially) circumvented by relying on the sequence homology between related species and genera and by using ESTs (Nagaraj *et al.*, 2007; Parkinson and Blaxter, 2009). ESTs have the advantage of already being closer to proteins than genomic DNA, but their generation was not overly efficient and fast. Fortunately, the same "next generation sequencing" technologies that have replaced Sanger sequencing (Sanger *et al.*, 1977) in whole genome analyses, have also made it possible to sequence cDNA libraries in a high throughput manner [RNA-Seq (Nagalakshmi *et al.*, 2008)]. This technology was rapidly

embraced by the proteomics field and new tools to generate customised protein databases from RNA-Seq data are being developed e.g. customProDB (Wang and Zhang, 2013), enabling an increased proteome coverage for non-sequenced organisms without the cost and difficulties of sequencing and analysing a whole genome.

A number of websites and repositories have attempted to group relevant information about the current state of helminth genome, transcriptome and proteome publications[2]:

- 959 Nematode Genomes:
  `http://www.nematodes.org/nematodegenomes/index.php/Main_Page`
- Wellcome Trust Sanger Institute – Helminth genomes:
  `http://www.sanger.ac.uk/resources/downloads/helminths/`
- HelmDB – An integrative database of curated transcriptomes of key nematodes and trematodes for functional annotation-improvement:
  `http://www.helmdb.org`
- HelmCoP – An Online Resource for Helminth Functional Genomics and Drug and Vaccine Targets Prioritisation:
  `http://www.nematode.net/helmcop.html`
- Helminth Secretome database:
  `http://www.biolinfo.org/hsd/`
- The National Centre for Biotechnology Information (NCBI):
  `http://www.ncbi.nlm.nih.gov/`

While proteomic analyses with state-of-the-art mass spectrometers and software can be a very powerful tool in biological research, the above section illustrates that their use is not so straightforward when it comes to studying one of the many organisms whose genomes have not yet been sequenced. There are significant difficulties and most of them are not easily bypassed and will likely involve a number of compromises.
It is important to realise that proteomic studies often fail to identify every single protein in a sample, even when it is a sequenced organism. This will unfortunately be even more the case when one of the main requisites, a complete protein library, is not even available.

[2] This list is not a complete overview of all helminth data repositories. All website links are functional in January 2014.

CHAPTER 4

Rationale and Objectives

4

The previous chapters have highlighted the shortcomings of many of the diagnostic tools currently used for the diagnosis of *T. solium* as well as the difficulties related to the study and characterisation of proteins that are produced by a non-sequenced organism. Specifically related to the detection of *T. solium* porcine cysticercosis, studying the proteins that are excreted and secreted by the metacestode larval stages (cysticerci) is nonetheless a vital step in the direction of a more specific assay, one that does not cross react with other *Taenia* spp. that may co-exist in the pig.
Studying protein samples like excretion/secretion proteins can be performed with the tools discussed in Chapter 3. However, the biggest bottleneck is the fact that the *Taenia* protein databases are incomplete due to a lack of genomic information. This, in turn, may have an effect on the number of identified proteins and the false discovery rate, the metric that estimates the number of incorrect identifications.

The main objective of this thesis is to analyse *T. solium* excretion/secretion proteins with the aim of identifying species-specific diagnostic antigens.
To achieve this, the main objective was divided into three sub-objectives:

1. To investigate whether false discovery rate estimations (a quality metric which estimates the number of incorrect identifications) are also applicable in the case of partially sequenced organisms, like *T. solium*, where many high-quality spectra fail to identify the correct peptides, simply because these peptides are not present in the searched protein database (Chapter 5).

2. To use protein sequences from other (closely related) helminths and expressed sequence tags to increase the proteome coverage of the incomplete protein database

and identify *T. solium* excretion/secretion proteins (Chapters 6 and 7).

3. To identify *T. solium*-specific peptides/proteins by exploring different options to compare the *T. solium* excretion/secretion proteome with proteomes of other *Taenia* spp. (Chapter 8).

CHAPTER 5

Partially Sequenced Organisms, Decoy Searches and False Discovery Rates

5

Abstract: Tandem mass spectrometry is commonly used to identify peptides, typically by comparing the experimental spectra with those predicted from a protein sequence database and scoring these matches. The most reported quality metric for a set of peptide identifications is the false discovery rate (FDR), the fraction of expected false identifications in the set. This metric has so far only been used for completely sequenced organisms or known protein mixtures. It is worthwhile to investigate whether FDR estimations are also applicable in the case of partially sequenced organisms, where many high-quality spectra fail to identify the correct peptides because the latter are not present in the searched sequence database. Using real data from human plasma and simulated partial sequence databases derived from two complete human sequence databases with different levels of redundancy, the mixture model-based approach in PeptideProphet was found to be robust for partial databases, particularly if used in combination with decoy sequences, regardless of how these decoy sequences were created (reversed, scrambled or randomised). It is therefore recommended to use this method when estimating the FDR from incompletely sequenced organisms.

Adapted from:
Victor, B., Gabriël, S., Kanobana, K., Mostovenko, E., Polman, K., Dorny, P., Deelder, A. M., and Palmblad, M. Partially Sequenced Organisms, Decoy Searches and False Discovery Rates. *Journal of Proteome Research* 11, 3 (2012), 1991–1995.

5.1 Introduction

While most published work in proteomics has been on humans and extensively sequenced model organisms, only limited sequence information is available for the remainder of the millions of extant biological species. Researchers studying one of the non- or partially sequenced organisms are faced with a number of difficulties, including the use of incomplete protein databases and robustly estimating global false discovery rates.

The problem of the incomplete databases can be addressed to some degree by adding protein sequences from other species of the same genus which are likely to be highly homologous e.g. one can study *Taenia saginata* (166 entries in NCBInr) and supplement the sequences from this species with the 401 entries for *Taenia solium* and the 119 entries for *Taenia hydatigena*. Given the known homology between proteins [like the TEG-Tsag protein and EG10 (González *et al.*, 2007)], one could even opt to add the 1,498 *Echinococcus granulosus* protein sequences. This inevitably creates a degree of redundancy in the database, but increases the chances of finding an exact match to a peptide in one of the homologous sequences searched. Similar approaches can be taken when studying many other organisms.

Global false discovery rates (FDRs), in the field of proteomics, can be defined as the expected number of incorrect peptide spectrum matches (PSMs) amongst all PSMs that pass a certain search score threshold (Choi and Nesvizhskii, 2008a; Käll *et al.*, 2008). The number of incorrect PSMs against the original, target database can be estimated by using a target-decoy strategy where a decoy database is searched separately or simultaneously as a composite target+decoy database. The number of decoy PSMs above the threshold score can be used as a measure of the number of incorrect target PSMs. The most simple FDR estimation can thus be defined as N_{decoy}/N_{target} where N_{decoy} and N_{target} are the number of hits above the search score threshold in the decoy database and the target database, respectively. Depending on the way the search is performed (separate or as a composite database) and the way the decoy database is constructed (reversed, scrambled or randomised) the use of a correction factor (Käll *et al.*, 2008), an alternative formula (Elias and Gygi, 2007; Navarro and Vázquez, 2009; Wang *et al.*, 2009a), an optimal decoy construction and search strategy (Bianco *et al.*, 2009) and a form of FDR quality control (Vaudel *et al.*, 2011b) have been suggested. A different way to estimate the FDR is a mixture model-based approach that calculates posterior probabilities for each PSM and uses those probabilities to derive global FDR estimates. This approach is implemented in the PeptideProphet (Keller *et al.*, 2002) and was later updated with the optional use of a decoy database to assist the expectation-maximisation (EM) algorithm in establishing a more reliable distribution of incorrect PSMs (Choi and Nesvizhskii, 2008b).

Nearly all published work on FDR estimations has been done using complete protein databases from fully sequenced organisms, or known well-defined protein mixtures. When working with incomplete databases (from partially sequenced organisms), FDR estimation may be more challenging. It is reasonable to assume that a significant number of high-quality spectra coming from the studied organism will not find a match in the target database, simply because the protein is not in that database. A fraction of these high-quality spectra will instead match a decoy sequence (or another peptide in the target database). It is not a trivial assumption that this larger fraction of falsely matched high quality spectra will *not* have any influence on FDR estimations, whether counting decoy hits or using a mixture model method.

In this study, the PeptideProphet FDR estimation using incomplete databases with or without the use of different decoy database designs was investigated.

5.2 Materials and methods

5.2.1 Databases

Two target databases with varying redundancy were used in this study, both downloaded from UniProt (`http://www.uniprot.org/`) on July 20th 2011 using "*Homo sapiens* reviewed:yes" as query. The first database contained 25,800 canonical Swiss-Prot protein sequences. The second database contained 41,551 canonical and isoform Swiss-Prot protein sequences. To model the effect of incomplete target databases, partial databases approximately covering 1 to 99% of both databases were constructed by scanning the full database and including each sequence with a probability ranging between 0.01 and 0.99. Reversed decoy databases were constructed using the decoyfasta script included with the Trans-Proteomic Pipeline (TPP) version 4.4 rev 1 (Keller *et al.*, 2005). Scrambled decoy databases were created with Perl scripts and the randomised decoy databases were constructed using the make_random Markov chain sequence generator (`http://www.ms-utils.org/make_random.html`) based on transitional probabilities from both target databases. Each partial target database was searched (i) as a composite database with an equally sized reversed ($n = 1$), scrambled ($n = 10$) and randomised decoy database ($n = 10$) and (ii) without the use of any decoy databases.

5.2.2 Sample preparation

As a representative complex proteomics sample with a few abundant and many less abundant proteins, a human plasma sample was prepared. Human plasma from a healthy individual was separated on a 4-12% NuPAGE Novex Bis-Tris gel (Invitrogen, Carlsbad, CA) and stained overnight with the Colloidal Blue Staining Kit (Invitrogen). The gel

lane was cut in 47 identical slices using a custom designed gel cutter (Gel Company Inc., San Francisco, CA) and all slices were transferred to a 96-well plate (Greiner Bio-One, Frickenhausen, Germany; one slice/well). Gel slices were washed three times: first with 100 µL 25 mM ammonium bicarbonate (ABC; Sigma-Aldrich Corp.), then with 100 µL 30% acetonitrile (Biosolve, Valkenswaard, The Netherlands) in ABC and finally with 100 µL 100% acetonitrile. Cystines were reduced in 75 µL 10 mM DTT in ABC for 30 minutes at 56°C and cysteines were alkylated in 75 µL 50 mM iodoacetamide (Sigma-Aldrich Corp.) in ABC for 20 minutes (in the dark) at ambient temperature. Gel slices were washed as above and proteins were digested with sequencing grade porcine trypsin (5 ng/µL 25 mM ABC; Promega, Southampton, UK; 30 µL/well) at 37°C for 6h. Supernatant from every two wells was combined and transferred to a new plate (first well was left single, resulting in 24 fractions). Digestions were quenched with the addition of 4 µL 5% trifluoroacetic acid (TFA). Gel slices were incubated for another hour at 37°C in 0.1% TFA and the second supernatant was added to the first. The plate was stored at −20°C pending further analysis.

5.2.3 Liquid chromatography – tandem mass spectrometry

All fractions were analysed by LC-MS/MS using a splitless NanoLC-Ultra 2D plus (Eksigent, Dublin, CA) for parallel ultra-high pressure liquid chromatography (UHPLC) with an additional loading pump for fast sample loading and desalting. The UHPLC system was configured with 300 µm-i.d. 5-mm PepMap C18 trap columns (Dionex, Sunnyvale, CA), 15-cm 300 µm-i.d. ChromXP C18 columns supplied by Eksigent and running 45 min. linear gradients from 4 to 33% acetonitrile in 0.05% formic acid. The UHPLC system was coupled on-line to an amaZon ETD high-capacity 3D ion trap (Bruker Daltonics, Bremen, Germany). After each MS scan, up to 10 abundant multiply charged species in m/z 300-1,300 were automatically selected for MS/MS but excluded for one minute after being selected twice. The modes of resolution were “Enhanced mode” and “UltraScan” for MS and MS/MS, respectively. The UHPLC system was controlled using the HyStar 3.4 (Bruker) with a plug-in from Eksigent and the amaZon ion trap by trapControl 7.0 (Bruker). DataAnalysis (Bruker) converted 1,000 MS/MS spectra with the highest signal intensity from each of the 24 raw files to MGF format. The resulting MGF files were merged and converted to mzXML with msconvert.

5.2.4 Analysis pipeline

The human plasma dataset was searched against the different databases with X!Tandem (2009.10.01.1) (Craig and Beavis, 2004) using the following parameters: precursor mass tolerance window was set between −2.0 and 4.0 Da, while fragment tolerance was set to 0.4 Da, modifications were set for carbamidomethylation of cysteine and oxidation of

methionine, k-scoring was enabled, maximum missed cleavages was set to two and scoring was done for b and y ions. The X!Tandem output files were translated to the pepXML file format with the tandem2xml converter. Posterior probabilities and FDRs were estimated using PeptideProphet with all parameters at their default values except the -d decoy option was used when including decoys to provide an initial discriminant score distribution for the false matches. These tools are included in version 4.4 rev 1 of the Trans-Proteomic Pipeline.

5.3 Results and discussion

5.3.1 Dataset

After searching the plasma dataset against the Swiss-Prot database containing only canonical sequences with X!Tandem, PeptideProphet estimated a total of 2,525 correct peptide assignments amongst 367 spectra from a singly charged precursor; 12,557 spectra from a doubly charged precursor; 12,618 spectra from a triply charged precursor and 1,066 spectra from a quadruply charged precursor. At FDR = 1%, PeptideProphet predicts 1,752 correct PSMs resulting in 1,703 unique peptides. This dataset represents a complex sample with a number of abundant proteins, that are also redundantly represented in the database.

5.3.2 PeptideProphet false discovery rate estimations for composite databases

Figure 5.1 shows the deviations in estimated FDRs (Δ_{FDR}) calculated relative to the (median) FDR estimation of the full database. The Δ_{FDR} is plotted for composite databases consisting of target databases (1-100% of the total size) without redundancy (Figures 5.1a to 5.1c) or with (limited) redundancy (Figures 5.1d to 5.1f) and an equally sized decoy database (reversed, scrambled or randomised) and at three different probability thresholds ($p > 0.90$, $p > 0.95$ and $p > 0.99$, respectively). An alternative presentation of the data is available in the supporting information (Figure A.1).

From the data, it is clear that mixture model-based FDR estimations for partial databases are stable, reproducible and very similar to the FDR of the full database, irrespective of the manner in which the decoy database was constructed, until the databases become so small that the mixture models can no longer be correctly fitted to the observed distributions. Appending the initial (partial) target database with protein sequences from closely related species to increase the proteome coverage and, consequently, the number of correct PSMs has no negative effect on the fitting of the mixture models or the subsequent estimation of posterior probabilities and FDRs.

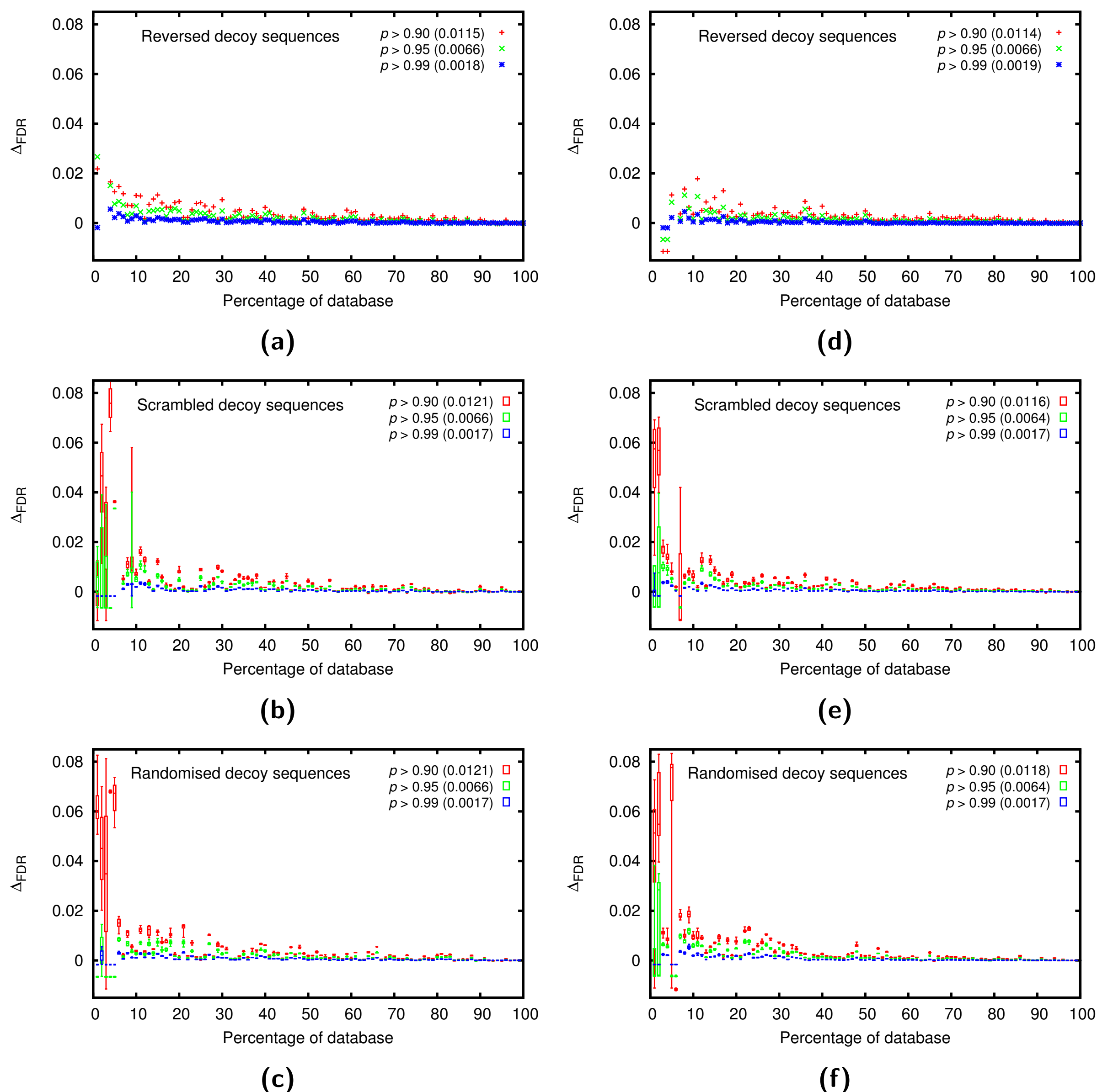


Figure 5.1: Deviations in estimated false discovery rates (Δ_{FDR}) calculated relative to the (median) FDR estimation of the full database at three different probability thresholds. The composite databases consisted of a target database with only canonical sequences (a-c) or canonical sequences and isoforms (d-f) and an equally sized reversed ($n = 1$), scrambled ($n = 10$) or randomised ($n = 10$) decoy database. The boxes in figures b, c, e and f show the median, the 25th and the 75th percentile, while the error bars show the minimum and maximum values. Values within parentheses are the FDR estimations (a and d) or the median FDR estimations (b, c, e and f) for the complete databases.

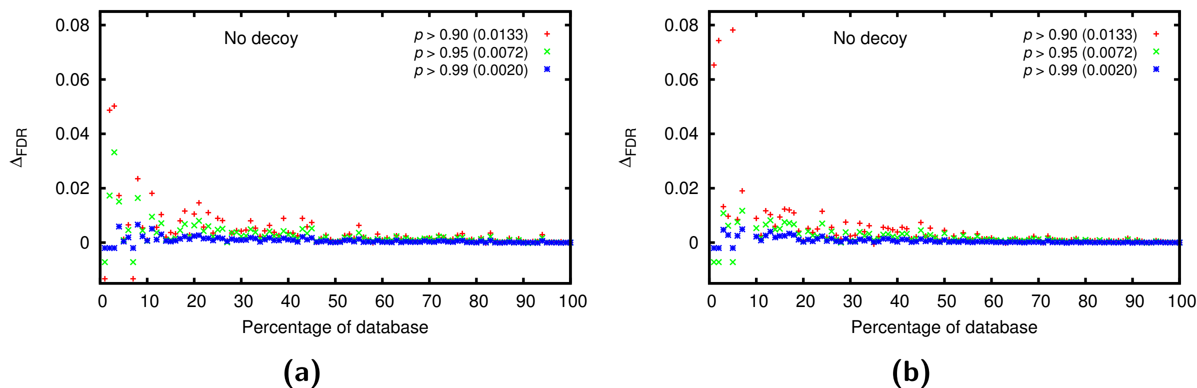


Figure 5.2: Deviations in estimated false discovery rates (Δ_{FDR}) calculated relative to the median FDR estimation of the full database at three different probability thresholds for target databases with only canonical sequences (a) or canonical sequences and isoforms (b) without any decoy database. Values within parentheses are the FDR estimations for the complete databases.

5.3.3 PeptideProphet false discovery rate estimations without decoy databases

The original implementation of PeptideProphet did not use decoy databases and their use is still optional. We therefore also investigated the use of PeptideProphet mixture modelling in the context of partially sequenced organisms without the use of decoy databases.

Figure 5.2 shows the deviations in estimated false discovery rates (Δ_{FDR}) calculated relative to the FDR estimation of the full database. The Δ_{FDR} is plotted for target databases (1-100% of the total size) without redundancy (Figure 5.2a) or with (limited) redundancy (Figure 5.2b) and at three different probability thresholds ($p > 0.90$, $p > 0.95$ and $p > 0.99$, respectively). An alternative presentation of the data is available in the supporting information (Figure A.2).

The results are largely similar to when using a decoy database, which again demonstrates the robustness of the approach. However, in cases where the dataset is challenging or when working with very limited partial target databases, it is recommended to use of a decoy database to assist the EM algorithm in establishing a more reliable fitting of the mixture models (Choi and Nesvizhskii, 2008b).

5.4 Conclusions

When describing results from a proteomics experiment, an estimate of the FDR is always recommended and often required by the scientific journal or data repository. If the database search was done with a complete sequence database (i.e. a database assumed to represent all proteins that can possibly be found in the sample) and the FDR estimated using established tools, the estimation is generally accepted without further discussion. In this short paper, it was demonstrated that mixture model-based FDR estimations using

incomplete databases from partially sequenced organisms are stable, highly reproducible and comparable to the estimations based on complete databases.
Based on this, it is recommended to use of PeptideProphet mixture model FDR estimations in combination with a decoy database strategy when studying organisms whose genome is not yet (completely) known.

CHAPTER 6

Proteomic Analysis of *Taenia solium* Metacestode Excretion/Secretion Proteins

Abstract: In order to improve the current diagnostic tools and to get a better understanding of the interaction between *Taenia solium* metacestodes and their host, there is a need for more information about the proteins that are released (excreted and secreted) by the parasite. However, the lack of a complete protein database for *T. solium* is an important limitation in mass spectrometry-based proteomics experiments. To increase the proteome coverage of the *T. solium* database, protein sequences from other (closely related) helminths were added. This composite database, along with one-dimensional polyacrylamide gel electrophoresis, reversed-phase liquid chromatography and tandem mass spectrometry were utilised to analyse the excretion/secretion proteins produced by cysticerci from infected pigs. A total of 76 proteins including 27 already described *T. solium* proteins, 17 host proteins and 32 proteins likely to be of *T. solium* origin, but identified using sequences from other helminths were identified. This is the first report of the *T. solium* metacestode excretion/secretion proteome.

Adapted from:
Victor, B., Kanobana, K., Gabriël, S., Polman, K., Deckers, N., Dorny, P., Deelder, A. M., and Palmblad, M. Proteomic Analysis of *Taenia solium* Metacestode Excretion/Secretion Proteins. *Proteomics* 12, 11 (2012), 1860–1869.

6.1 Introduction

Taenia solium metacestodes (and many other helminths/helminth stages) are able to establish and maintain an infection in the host, while being attacked by the host immune system. Helminths have therefore developed diverse mechanisms to protect themselves in this hostile environment. These mechanisms often depend on the production of excretion/secretion proteins (ESPs) (Hewitson *et al.*, 2009). Parasite defence mechanisms and ESPs have been studied and described for other helminths of public health importance such as *Echinococcus* (Siracusano *et al.*, 2008b; Monteiro *et al.*, 2010) and *Schistosoma* (Guillou *et al.*, 2007; Liu *et al.*, 2009), but the excretion/secretion proteome of *T. solium* has not been well characterised so far. A recent proteomic study investigated *T. solium* oncospheres, as this is the stage that determines the success or failure of the infection (Santivañez *et al.*, 2010). From a diagnostic point of view, ESPs are of specific interest. Furthermore, given the importance of the ESPs for the survival of the parasite, they might also be good vaccine candidates or anthelmintic drug targets.

In the case of *T. solium*, a few specific ESPs [e.g. diagnostic antigen GP50 (Hancock *et al.*, 2004)] have been extensively studied and successfully used in antibody detecting enzyme-linked immunoelectrotransfer blots (EITB) and ELISAs. The drawback is that antibody detection does not allow the differentiation between an active infection and a past infection or contact with oncospheres (García *et al.*, 2001). Other methods, such as monoclonal antibody-based antigen detection ELISAs demonstrate the presence of viable *T. solium* metacestodes by detecting circulating ESPs in the blood, cerebrospinal fluid and urine (Deckers and Dorny, 2010), but current tests for porcine cysticercosis are not able to differentiate between different *Taenia* spp. A more thorough knowledge of *T. solium* ESPs would be an important step in improving current diagnostic techniques.

At the time of publication, a complete proteomic analysis of many parasite ESPs was still hampered by incomplete protein databases due to the lack of genomic information. Although a *T. solium* genome project had been initiated in 2006 (Aguilar-Díaz *et al.*, 2006), no data were yet in the public domain. To get around this limitation, our approach relied on homology to other helminth species and supplemented the *T. solium* database with protein sequences from other helminths. Furthermore, our study took full advantage of the recent advancements in reversed-phase liquid chromatography and tandem mass spectrometry (LC-MS/MS) that have enabled high throughput analysis of large sample numbers.

This study presented the first proteomic analysis of the *T. solium* metacestode ESPs using one-dimensional polyacrylamide gel electrophoresis (1D-PAGE) and LC-MS/MS.

6.2 Materials and methods

6.2.1 Parasite material and excretion/secretion proteins

Taenia solium metacestodes were obtained from naturally infected pigs in Zambia and Peru. In each country, five pigs were selected based on positive tongue palpation. The animals were humanely euthanised and at least 200 cysts were dissected from the muscle tissues of each pig. Cysts were washed three times in sterile phosphate-buffered saline (PBS), two times in sterile PBS containing penicillin (100 units/mL), streptomycin (100 µg/mL) and amphotericin B (0.25 µg/mL) (Invitrogen, Carlsbad, CA) and incubated in culture dishes (BD, Franklin Lakes, NJ) at a concentration of 50 cysts/20 mL of culture medium (RPMI-1640 with L-glutamine and 25 mM HEPES (Invitrogen) supplemented with penicillin (100 units/mL), streptomycin (100 µg/mL) and amphotericin B (0.25 µg/mL)) for 6h at 37℃ in an atmosphere of 5% CO_2. Because it was likely to contain high amounts of host proteins, the culture medium was discarded after 6h, replaced with fresh medium and incubated for another 18h. Culture media from 200 cysts and from the same animal were pooled, centrifuged for 15 minutes at $3000 \times g$, supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche, Indianapolis, IN) and stored at −80℃ pending further processing (ESP 24h). After removal of the 24h culture medium, cysts were washed three times in culture medium and incubated for another 24h at 50 cysts/20 mL. The collection of the 48h fraction was done as described above (ESP 48h). Each fraction (per pig, country and time point) was concentrated to a final volume of 2 mL with an Amicon Stirr Cell (Millipore, Billerica, MA) using a 1 kDa ultrafiltration membrane (Millipore) and Macrosep 1K UF Spin Filters (Pall, Port Washington, NY). Concentrated fractions were aliquoted (100 µL) and stored at −20℃. Cyst viability was assessed by incubating 20 cysts/pig in 10% porcine bile in culture medium and subsequent evaluation of cyst evagination.

6.2.2 One-dimensional polyacrylamide gel electrophoresis

Aliquots were precipitated and washed in ice-cold acetone and resuspended in half the original volume with 10 mM DTT (Sigma-Aldrich Corp., St. Louis, MO) and 3% sodium dodecyl sulfate (SDS). When fully resuspended, samples were mixed with NuPAGE LDS Sample Buffer (4x) and separated on a 4-12% NuPAGE Novex Bis-Tris gel (Invitrogen). Gels were stained overnight with the Colloidal Blue Staining Kit (Invitrogen).

6.2.3 In-gel trypsin digest

Each gel lane was cut in 48 identical slices using a custom designed gel cutter (Gel Company Inc., San Francisco, CA) and all slices were transferred to a 96-well plate (Greiner Bio-One,

Frickenhausen, Germany; one slice/well). Gel slices were washed three times: first with 100 µL 25 mM ammonium bicarbonate (ABC; Sigma-Aldrich Corp.), then with 100 µL 50% acetonitrile (Biosolve, Valkenswaard, The Netherlands) in ABC and finally with 100 µL 100% acetonitrile. Cystines were reduced by adding 75 µL 10 mM DTT in ABC for 30 minutes at 56℃. Gel slices were shrunk with 100 µL of 100% acetonitrile and cysteines were alkylated with the addition of 75 µL 50 mM iodoacetamide (Sigma-Aldrich Corp.) in ABC for 20 minutes (in the dark) at ambient temperature. Gel slices were washed and shrunk again with the addition of 100 µL 25 mM ABC followed by 100 µL 100% acetonitrile. Proteins were digested with sequencing grade porcine trypsin (0.125 µg in 25 µL 25 mM ABC; Promega, Southampton, UK; 25 µL/well) at 37℃ for 6h. From each well, 20 µL of supernatant was extracted and 15 µL of Milli-Q water (Millipore) was added for another incubation of 30 minutes at 37℃. A second supernatant sample of 15 µL was extracted and the two supernatants were pooled. TFA (Sigma-Aldrich Corp.) was added to a final concentration of 1% to each well to quench the reaction. The total volume per well was reduced to around 15 µL with a vacuum centrifuge and plates were stored at −20℃ pending further analysis.

6.2.4 Liquid chromatography – tandem mass spectrometry

Peptides were analysed on an amaZon ETD ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an Apollo II ESI source and coupled on-line with a NanoLC-Ultra 2D plus LC (Eksigent, Dublin, CA, USA). A total volume of 10 µL was first desalted on a C18 PepMap300 trap column (5 µm, 300 µm × 5 mm; Dionex, Sunnyvale, CA) and then separated on a 15 cm C18 ChromXP analytical column (3 µm, 150 × 0.3 mm; Eksigent) by a 45-minute linear gradient from 4% to 35% acetonitrile, 0.05% formic acid (Sigma-Aldrich Corp.) with a flow rate of 4 µL/minute. Mass spectra were acquired in m/z 300-1,300 with the ion charge control (ICC) set to 200,000 and the maximum accumulation time set to 200 ms. The 10 most abundant precursors were selected for CID MS/MS in m/z 100-2,000 with the ICC and maximum accumulation time set to 200,000 and 50 ms respectively, but were excluded for 0.5 minutes after having been selected once. Singly charged ions were excluded for MS/MS. The LC system was controlled by HyStar 3.2 and the ion trap by trapControl 7.0.

6.2.5 Data analysis

Raw LC-MS/MS data were converted to line spectra mzXML files[1] with the Bruker compassXport tool version 3.0.3. All further data processing was done with the Trans-Proteomic Pipeline (TPP) 4.4 rev 1 (Keller *et al.*, 2005). Database searching was performed

[1] The datasets supporting the results of this article are available in the PRIDE repository at `http://www.ebi.ac.uk/pride` with accession numbers 19232 – 19267.

with X!Tandem (2009.10.01.1) (Craig and Beavis, 2004) using the following parameters: precursor mass tolerance window was set between −2.0 and 4.0 Da, while fragment tolerance was set to 0.4 Da, modifications were set for carbamidomethylation of cysteine and oxidation of methionine, k-scoring was enabled, maximum missed cleavages was set to two and scoring was done for *b* and *y* ions. The X!Tandem output files were converted to the pepXML file format with tandem2xml (no cut-off was set on X!Tandem hyperscore or E-value). X!Tandem pepXML files from different pigs but corresponding to the same time point and country were merged into PeptideProphet (Keller *et al.*, 2002) and analysed with the decoy option enabled (Choi and Nesvizhskii, 2008b). The resulting posterior probabilities for the peptide spectrum matches (PSMs) were further refined by iProphet (Shteynberg *et al.*, 2011). Finally, ProteinProphet (Nesvizhskii *et al.*, 2003) was used to compute a probability that each protein was present in the sample and to estimate a global false discovery rate (FDR). ProteinProphet output was filtered using the probability threshold that corresponded to an FDR $< 1\%$ and proteins with an individual probability of zero were discarded. Filtered protein lists were combined and proteins were organised based on homology using a rough tree (MAFFT; v6.857b) (Katoh and Toh, 2008). Further classification was based on Gene Ontology biological process and molecular function information available from the UniProt database (`http://www.uniprot.org/`). SignalP 3.0 server (Bendtsen *et al.*, 2004b) and SecretomeP 2.0 Server (Bendtsen *et al.*, 2004a) were used to predict classical (signal peptide triggered) and non-classical protein secretions, respectively. The number of spectrum IDs calculated by ProteinProphet was used to estimate the relative protein abundance.

6.2.6 Databases

The target database used in this study was constructed from three separate databases. The first database (48,245 protein sequences) contained protein sequences from all *Taenia* species as well as protein sequences from the more studied *Echinococcus* and *Schistosoma* helminths. *Trichinella* spp. protein sequences were also included because it is a common parasite in pigs. This database was extracted from the National Centre for Biotechnology Information non-redundant (NCBInr) database on June 15 2011 with a Perl script that (i) looks up the species identification number for all *Taenia*, *Echinococcus*, *Schistosoma* and *Trichinella* species in the names.dmp file (`ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz`), (ii) finds the matching GenInfo Identifiers in the gi_taxid_prot.dmp file (`ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid_prot.dmp.gz`) and (iii) extracts the corresponding protein sequences from the NCBInr database (`ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz`). The second database was a *Sus scrofa* database downloaded from the UniProt website (`http://www.uniprot.org/`) on June 15 2011 that contained 1,388 SwissProt protein sequences. The third database,

the common Repository of Adventitious Proteins database (112 protein sequences; `ftp://ftp.thegpm.org/fasta/cRAP/crap.fasta`) was included to detect common and/or accidental contaminations in the protein samples. A decoy database with 49,745 reversed sequences was created using decoyfasta and target and decoy databases were concatenated into one final database.

6.3 Results and discussion

6.3.1 *Taenia solium* metacestodes and *in vitro* excretion/secretion protein production

On carcass inspection, Zambian pigs were found to be more massively infected than Peruvian pigs. Cyst viability was 72% and 100% for Peru and Zambia, respectively. After each incubation step, plates were checked. Dead cysts were observed and removed in the Peruvian cultures only. The 24h fractions from two Zambian pigs were lost due to a failed ammonium sulphate precipitation. This method was abandoned and all remaining samples were concentrated with filters.

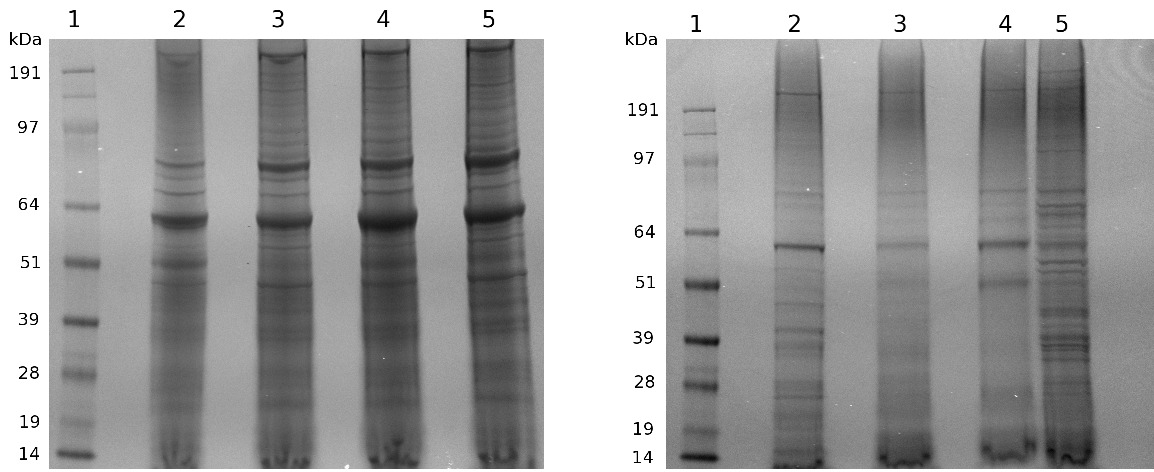
6.3.2 One-dimensional polyacrylamide gel electrophoresis

Figure 6.1a shows a protein profile of ESP 24h and ESP 48h obtained from cysts coming from two Peruvian pigs. ESP fractions from all Peruvian pigs showed almost identical profiles, irrespective of the incubation time. Figure 6.1b illustrates the variation in protein profiles which was observed between ESP 24h and ESP 48h fractions obtained from the Zambian pigs.
The discrepancies between profiles from Zambia and Peru could be attributed to differences in ESP production, as well as to proteins leaking from dead/damaged metacestodes or differences in concentration. The aim of the study was to be as complete as possible, not to draw a direct comparison between fractions or countries. Therefore, the differences are not interpreted as erroneous or problematic, but are considered beneficial to the completeness of the study.

6.3.3 Data analysis

X!Tandem pepXML files from five pigs/ESP 24h Peru, five pigs/ESP 48h Peru, five pigs/ESP 48h Zambia and three pigs/ESP 24h Zambia were merged into PeptideProphet, refined with iProphet and further analysed with ProteinProphet. The ProteinProphet output with FDR $< 1\%$ is available in Tables 3-6 in the supporting information[2]). Table 6.1

[2] `http://onlinelibrary.wiley.com/doi/10.1002/pmic.201100496/suppinfo`



(a) Lane 1: protein size marker; lanes 2 and 3: 24h and 48h ESP fractions from pig 4; lanes 4 and 5: 24h and 48h ESP fractions from pig 5.

(b) Lane 1: protein size marker; lane 2: 48h ESP fraction from pig 1; lane 3: 48h ESP fraction from pig 2; lanes 4 and 5: 24h and 48h ESP fractions from pig 5.

Figure 6.1: Two examples of one-dimensional polyacrylamide gel electrophoresis illustrating (a) the similarities between different excretion/secretion protein (ESP) fractions from cysts obtained from Peruvian pigs and (b) the differences in protein profiles observed between excretion/secretion protein fractions obtained from cysts from Zambian pigs.

summarises the probability thresholds used to filter the output, the estimated FDRs and the estimated number of correct protein groups, as reported by ProteinProphet. The mixture model in PeptideProphet failed for charge (1+) and (for ESP 24h Zambia) also for charge (3+).

Table 6.1: The ProteinProphet estimations for the four merged fractions.

Merged fraction	Probability threshold	Estimated false discovery rate (%)	Estimated number of correct protein groups
ESP 24h Peru	0.90	0.5	82
ESP 48h Peru	0.90	0.4	85
ESP 24h Zambia	0.80	0.8	30
ESP 48h Zambia	0.90	0.5	49

The combined, filtered and organised output resulted in 76 protein groups containing one or more (homologous) proteins from *T. solium* and/or other organisms (Table 7 in the supporting information[3]). Keratin contaminations were omitted from the list. The inclusions of multiple homologous proteins in a protein group is a consequence of the way the database was constructed. However, it does not imply independent evidence for multiple isoforms of the proteins, although, in the case of the 8 kDa protein family, it is highly likely that multiple proteins from this family were present in the samples. SignalP and SecretomeP analyses were done on all proteins. The former reports yes (Y) or no (N)

[3]`http://onlinelibrary.wiley.com/doi/10.1002/pmic.201100496/suppinfo`

for both the neural networks (NN) and hidden Markov models (HMM), while the latter interprets a NN-score above 0.5 as possible secretion (Y).

6.3.4 Proteins identified in *Taenia solium* excretion/secretion proteins

Table 6.2 lists the 76 protein groups that were found in the *T. solium* ESPs, organised by Gene Ontology annotations for biological process and molecular function. For simplicity, the protein groups are represented by one protein.

The objective of this study was not to compare different time points and/or geographical locations, but to provide a broader coverage and identify ubiquitous components of the ES proteome. Therefore, time points and geographical locations of the proteins were added to Table 6.2 for completeness only. The results of SignalP and SecretomeP show a number of proteins that are negative for both analyses. These proteins are mostly found in Peruvian ESPs and may be a result of leakage due to cyst damage/death, since dead cysts were observed in Peruvian cultures.

Gene Ontology group 1 includes proteins with little information on molecular function or biological processes in the UniProtKB. However, this group includes proteins that are widely used in immunodiagnostic assays. Group 2 contains proteins related to muscle (paramyosin), microtubule-based processes (dynein light chain), protein polymerisation (tubulin) and tegument/cytoskeleton (H17g protein and actin). Group 3 are the stress response/chaperone proteins and group 4 lists proteins involved in various metabolic activities e.g. glycolysis (phosphoglycerate kinase, fructose-bisphosphate aldolase, enolase, glucose phosphate isomerase, GADPH) and gluconeogenesis (phosphoenolpyruvate carboxykinase) as well as biosynthetic processes/protein synthesis (Na^{+}/K^{+}-transporting ATPase, nucleoside diphosphate kinase, elongation factor 1 alpha). Group 5 entails signal transduction, group 6 presents protein folding and group 7 contains transport proteins. Group 8 contains the enzymatic antioxidant system of Taeniidae (cytosolic Cu/Zn-superoxide dismutase, glutathione S-transferase, 2-cys peroxiredoxin) and proteins with oxidoreductase activity (ferritin, aldo-keto reductase). Group 9 holds proteolysis related proteins with serine-type (trypsin-like protein, trypsin), cysteine-type (cathepsin L-like cysteine proteinase) and threonine-type (proteasome subunit alpha) endopeptidase activity. Trypsin was introduced during the protein digestion step. Group 10 contains proteins with cysteine-type (alpha-2-HS-glycoprotein (Fragment), immunogenic protein Ts11) and serine-type endopeptidase inhibitor activity (alpha-1-antitrypsin, leukocyte elastase inhibitor, inter-alpha-trypsin inhibitor). Group 11 are binding proteins e.g. fatty acid (fatty acid binding protein), calcium (annexin, calcium binding protein), actin (actin

Table 6.2: Proteins ($n = 76$) identified in *Taenia solium* metacestode excretion/secretion proteins organised by Gene Ontology annotations for biological process and molecular function.

Gene Ontology classification Protein	Closest organism	gi/sp code	Peru[a)] 24h	Peru[a)] 48h	Zambia[a)] 24h	Zambia[a)] 48h	SigP/SecP[b)]
1) No Gene Ontology information							
Hypothetical protein	*T. solium*	21726976	-	-	-	+	N/Y
Antigen EM/EG13	*Echinococcus* spp.	multiple	+	-	-	-	N/N
8 kDa protein family [c)]	*T. solium*	multiple	-/+	-/+	+	-/+	-/-
Antigen	*T. solium*	29893185	+	+	-	-	N/Y
P27	*T. solium*	333101790	+	+	-	+	N/Y
Apolipoprotein AI binding protein	*E. multilocularis*	223640019	-	+	-	-	N/Y
Immunogenic protein	*T. solium*	multiple	+	+	-	-	Y/Y
T24	*T. solium*	37786712	+	+	+	+	-/N
Hypothetical protein	*T. solium*	21912540	+	+	-	-	N/N
Diagnostic antigen GP50	*T. solium*	multiple	+	+	+	+	-/-
Small heat-shock protein	*T. solium*	21665905	+	+	-	-	N/Y
2) Microtubule-based movement/motor activity/tegument-cytoskeleton							
Paramyosin	*T. solium*	multiple	-	-	-	+	N/-
H17g protein	*T. solium*	34368418	+	+	+	-	N/Y
Actin	multiple genera[c)]	multiple	-/+	-/+	-/+	-/+	-/-
Dynein light chain	multiple genera	multiple	-/+	+	-	-/+	N/-
Tubulin	multiple genera	multiple	-/+	-/+	-	-/+	-/-
3) Response to stress/chaperone							
Heat shock protein 70	multiple genera	multiple	+	+	-	-/+	N/N
Heat shock protein 90 alpha	*T. asiatica*	124783119	+	+	-	+	N/N
Heat shock protein 90	*T. asiatica*	124783236	+	+	-	+	N/Y
4) Metabolism/biosynthetic processes							
Phosphoglycerate kinase	multiple genera	multiple	+	-/+	-	+	N/-
Fructose-bisphosphate aldolase	multiple genera	multiple	-/+	+	-/+	-/+	N/N
Enolase	multiple genera	multiple	-/+	-/+	-/+	-/+	N/-
Cytosolic malate dehydrogenase	*T. solium*	323361126	+	+	+	+	N/N
Glucose phosphate isomerase	*Echinococcus* spp.	multiple	+	-/+	-	+	N/Y
Glycogen phosphorylase	*T. spiralis*	316972087	-	+	-	-	N/N
GAPDH[d)]	*T. solium*	149364041	+	+	-	+	N/Y
Triosephosphate isomerase	*T. solium*	38258647	+	+	+	+	N/N
Phosphoenolpyruvate carboxykinase	*T. solium*	283466482	+	+	+	+	N/N
Glucosidase	multiple genera	multiple	+	+	-	-	Y/-
Adenosylhomocysteinase	*S. scrofa*	Q710C4	-	+	-	-	N/N
Cell polarity protein	*S. mansoni*	256075333	-	+	-	-	N/N
Carbonic anhydrase 3	*S. scrofa*	Q5S1S4	+	+	-	-	N/N
Creatine kinase M-type	*S. scrofa*	Q5XLD3	+	+	-	-	N/N
Na^+/K^+-transporting ATPase subunit alpha	*T. solium*	74794482	+	+	-	+	N/Y
Nucleoside diphosphate kinase	*Schistosoma* spp.	multiple	+	+	-	-	N/N
Elongation factor 1 alpha	*T. solium*	multiple	-	+	-	-	N/N
5) Signal transduction							
GTPase	multiple genera	multiple	-/+	-/+	-	-	N/Y
ADP-ribosylation factor	multiple genera	multiple	+	-/+	-	-	N/-
Multiple pdz domain protein	*S. mansoni*	256082156	-	-	-	+	N/N
CAMP-dependent protein kinase regulatory subunit	*S. japonicum*	multiple	+	+	-	-	N/Y
14-3-3 protein	*Echinococcus* spp.	multiple	+	+	-/+	+	N/-
6) Protein folding							
Cyclophilin	multiple genera	multiple	+	+	-/+	+	N/N
Calreticulin	*T. solium*	14029538	+	-	-	-	Y/N

6

Table 6.2: Continued.

Gene Ontology classification / Protein	Closest organism	gi/sp code	Peru[a) 24h	Peru[a) 48h	Zambia[a) 24h	Zambia[a) 48h	SigP/SecP[b)
7) Transport							
Glucose transporter TGTP2	*T. solium*	1480799	+	-	-	-	Y/N
Apolipoprotein A-I	*S. scrofa*	P18648	+	-	-	-	Y/Y
Transthyretin	*S. scrofa*	P50390	+	-	-	-	Y/Y
Hemopexin	*S. scrofa*	P50828	+	+	-	-	Y/Y
Serum albumin	*S. scrofa*	P08835	+	+	+	+	Y/Y
Transferrin	*S. scrofa*	multiple	+	+	-/+	-/+	-/-
8) Cell redox homeostasis/detoxification/oxidoreductase activity							
Thioredoxin	*E. granulosus*	29337032	+	+	-	+	N/N
2-Cys peroxiredoxin	*T. solium*	multiple	-/+	-/+	+	-/+	N/Y
Cytosolic Cu/Zn-superoxide dismutase	*T. solium*	18252397	+	+	-	+	N/N
Glutathione S-transferase	*T. solium*	multiple	+	+	-/+	-/+	N/-
Ferritin	*T. saginata*	1297064	+	+	-	+	N/N
Aldo-keto reductase	*Schistosoma* spp.	multiple	+	+	-	-	N/-
9) Proteolysis/endopeptidase activity							
Trypsin-like protein	*T. solium*	311335041	+	+	+	+	Y/Y
Cathepsin L-like cysteine proteinase	*T. solium*	multiple	+	+	-	-	Y/Y
Proteasome subunit alpha 2 (T01 family)	*S. mansoni*	256083548	-	+	-	+	N/N
Haptoglobin	*S. scrofa*	Q8SPS7	+	+	+	-	Y/Y
Trypsin	*S. scrofa*	P00761	+	+	+	+	N/Y
10) Endopeptidase inhibitor activity							
Immunogenic protein Ts11	*T. solium*	7339849	+	+	-	-	Y/Y
Alpha-1-antitrypsin	*S. scrofa*	P50447	+	+	-	-	Y/Y
Leukocyte elastase inhibitor	*S. scrofa*	P80229	+	-	-	-	N/Y
Inter-alpha-trypsin inhibitor heavy chain	*S. scrofa*	P79263	+	+	-	-	Y/Y
Alpha-2-HS-glycoprotein (Fragment)	*S. scrofa*	P29700	+	+	-	-	Y/Y
11) Binding (miscellaneous)							
Cytosolic fatty acid binding protein	*T. solium*	82412213	+	+	+	+	N/N
Calcium binding protein	*Schistosoma* spp.	multiple	-	+	-	-	N/Y
Annexin	*T. solium*	multiple	+	-/+	-/+	+	N/-
Actin-binding protein	multiple genera	multiple	+	+	-	-	-/-
Tuftelin interacting protein	*S. mansoni*	256076696	+	-	-	-	N/N
Hemoglobin subunit beta	*S. scrofa*	P02067	+	+	-	-	N/N
Ig lambda chain C region	*S. scrofa*	P01846	+	+	+	+	N/Y
12) Translation							
Ubiquitin	multiple genera	multiple	+	+	-	-/+	N/Y
13) Cell division							
Cell division control protein	*S. mansoni*	multiple	-	+	-	-	N/N
14) Defence response to bacterium							
Protegrin (and related)	*S. scrofa*	multiple	-	-	+	-	-/-
15) Catalytic activity							
Transketolase	*E. multilocularis*	27526313	+	+	-	-	N/Y

[a) Proteins present in Peru/Zambia in the 24h/48h fraction are described as + while - indicates absence. -/+ indicates there were multiple entries for this protein with variable presence in the fractions.

[b) SignalP and SecretomeP (SigP/SecP) results are described as yes (Y) or no (N). SignalP is Y when both the neural network and the hidden Markov models agree. (-) is reported when not all individual proteins in that groups have the same result.

[c) For simplicity, the 8 kDa protein family is presented as one protein although it is highly likely that multiple proteins from this family were present in the samples.

[d) GAPDH: glyceraldehyde-3-phosphate dehydrogenase.

binding protein), DNA (tuftelin interacting protein), oxygen (hemoglobin) and antigen binding proteins (Ig lambda chain C region). Group 12, 13, 14 and 15 contain proteins involved in translation, cell division, defence response and catalytic activity, respectively.

6.3.5 Relative protein abundance

Protein abundance showed similar patterns in Peru ESP 24h and 48h: porcine albumin and phosphoenolpyruvate carboxykinase were most abundant followed by members of the 8 kDa protein family, immunogenic protein, trypsin-like protein, enolase, 2-cys peroxiredoxin and 14-3-3 protein. Host proteins (serotransferrin, Ig) were present in both ESPs although more abundant in the 24h fraction. Zambia ESP 24h showed high abundance for members of the 8 kDa protein family and porcine albumin, followed by immunogenic protein, trypsin-like protein, cyclophilin, porcine Ig, phosphoenolpyruvate carboxykinase and porcine serotransferrin. Zambia ESP 48h showed highest abundance for members of the 8 kDa protein family and phosphoenolpyruvate carboxykinase, followed by immunogenic protein, porcine albumin, glutathione S-transferase, enolase and 2-cys peroxiredoxin.
The abundance of albumin in the ESPs can be explained by its abundance in blood (despite multiple washings) and possibly because it was secreted by the cysts. Secretion of host albumin has been described in *Taenia crassiceps* metacestodes where it may be involved in osmoregulation (Aldridge *et al.*, 2006).

6.3.6 *Taenia solium* proteins

Of the 76 proteins listed in Table 6.2, 27 have been described in literature specifically for *T. solium*. Many of these proteins have been studied due to their importance as diagnostic antigens in antibody detecting assays (Deckers and Dorny, 2010) e.g. diagnostic antigen GP50 (Hancock *et al.*, 2004), T24 (Hancock *et al.*, 2006) and the 8 kDa protein family (Hancock *et al.*, 2003; Ferrer *et al.*, 2007), which are the three protein groups that represent the 7 diagnostic bands in the EITB (Tsang *et al.*, 1989). Members of the 8 kDa protein family are believed to be hydrophobic ligand binding proteins (Hancock *et al.*, 2003) and a similarity to the related *E. granulosus* AgB might indicate an immunoregulatory role, enabling the parasite to reduce an inflammatory response (Shepherd *et al.*, 1991) and manipulate the Th1-Th2 balance (Riganò *et al.*, 2001). Paramyosin (Vargas-Parada and Laclette, 2003), 14-3-3 protein (Santivañez *et al.*, 2010), actin (Campos *et al.*, 1990), P27, small heat-shock protein (Ferrer *et al.*, 2005b) and phosphoenolpyruvate carboxykinase have been described as immunoreactive proteins recognised by NCC-positive human sera (Salazar-Anton and Lindh, 2011). Paramyosin in particular is also reported to inhibit complement formation (Laclette *et al.*, 1992). P27 is homologous to P-29 (*E. granulosus*) and Antigen 6 (*E. multilocularis*). Several studies have shown similarities between sequences and epitopes of P-29, Antigen S, Antigen 6 and even Antigen 5, which possesses similar

protease activity as the *T. solium* trypsin-like protein (Lorenzo *et al.*, 2003; Siles-Lucas *et al.*, 1998; González *et al.*, 2000a; Ben Nouir *et al.*, 2009). The enzymatic antioxidant system of Taeniidae involves Cu/Zn-superoxide dismutase (Castellanos-González *et al.*, 2002), glutathione S-transferase (Vibanco-Pérez *et al.*, 2002; Plancarte *et al.*, 2004; Torres-Rivera and Landa, 2008; Nguyen *et al.*, 2010) and 2-cys peroxiredoxin (Vaca-Paniagua *et al.*, 2009). These proteins have been considered as targets for drugs and vaccines (Vaca-Paniagua *et al.*, 2008). Cysteine proteases help the parasite to evade the host immune system in multiple ways: they possess the ability to lyse host immunoglobulins (White *et al.*, 1997a) and *in vitro* tests have proven their capability to deplete CD4+ lymphocytes by inducing apoptosis (Molinari *et al.*, 2000; Tato *et al.*, 2004). Furthermore, cathepsin L-like cysteine peptidase has been used for differential diagnosis of infections with *Taenia* species with a loop-mediated isothermal amplification assay (Nkouawa *et al.*, 2009). Annexin B2 is known to be associated with inflammatory reaction (Gao *et al.*, 2007). Other studies have shown anticoagulant activity (Lu *et al.*, 2010). Furthermore, secreted annexin B1 [antigen cC1 (Hongli *et al.*, 2002)] was shown to induce apoptosis in eosinophils, a novel strategy to prevent the host's immune attack (Yan *et al.*, 2008), and was suggested as a candidate for the development of a vaccine against *T. solium* cysticercosis (Guo *et al.*, 2004). Calreticulin controls intracellular Ca^{2+} homeostasis, but can also assist in secretion, protein synthesis and control of protein folding (Mendlovic *et al.*, 2004). Cytosolic malate dehydrogenase is involved in the citric acid metabolism and has recently been studied because it might potentially serve as a target for drugs (Nava *et al.*, 2011). Glucose transporter TGTP2 is believed to mediate sugar uptake and is predominantly localised on the surface of the cyst (Rodríguez-Contreras *et al.*, 1998). It is however not recognised by the host immune system (Rodríguez-Contreras *et al.*, 2002).

6.3.7 Proteins matching sequences from other species

The interpretation of the 49 proteins that are not identified by *T. solium* requires careful consideration to assess if they are likely to be of helminth or host origin.

In most cases, spectra matching proteins from other helminths ($n = 32$) can be assumed to derive from *T. solium* proteins not present in the *T. solium* database. A number of proteomic studies have been carried out on other helminth parasites like *Schistosoma* spp. (Guillou *et al.*, 2007; Pérez-Sánchez *et al.*, 2008; Liu *et al.*, 2009) and *E. granulosus* metacestodes (Monteiro *et al.*, 2010). Many of those proteins are identified in this study as well and it is reasonable to assume that the proposed function is similar. Although a few of these proteins (e.g. enolase, tubulin and phosphoglycerate kinase) are identified also by *S. scrofa*, the proteins are unlikely to be of host origin due to their nature or function and presence in other helminths.

A total of 17 proteins were identified only by *S. scrofa* and because most of them are commonly found in blood (e.g. Ig, hemoglobin, endopeptidase inhibitors and transport proteins), they are almost certainly of host origin. They may be considered as contamination, although importantly, some host proteins (e.g. albumin) are also known to be taken up and secreted by cysts (Aldridge *et al.*, 2006).

6.4 Conclusions

This study presented the first report of the *T. solium* metacestode excretion/secretion proteome. Since the genome of *T. solium* was not yet available at the time of publication, and only 401 protein entries were found in the NCBInr database, our approach relied on homology to other helminths to increase the proteome coverage. The 76 reported proteins included 27 already described *T. solium* proteins, 17 host proteins and 32 proteins likely to be of *T. solium* origin, but identified using sequences from other helminths, effectively demonstrating the value of this approach. However, until a reference genome for *T. solium* becomes available, it is likely that several proteins in the ESPs remained unidentified. The fact that many of the proteins found in this study are involved in parasite survival strategies and have been described in other helminths as well indicates that the excretion/secretion proteome is not all that different between species or even related genera. Therefore, it is reasonable to assume that the proposed protein functions are similar. Albeit the close homology to another protein may limit the specificity of diagnostic tests based on that protein, it also has the potential to widen the usable range of a vaccine or drug.

CHAPTER 7

Use of Expressed Sequence Tags for the Identification of *Taenia solium* Metacestode Excretion/Secretion Proteins

Abstract: In the previous study, a custom protein database containing protein sequences from related helminths was used to identify *Taenia solium* metacestode excretion/secretion proteins. An alternative or complementary approach would be to use expressed sequence tags combined with BLAST and protein mapping to supercontigs of *Echinococcus granulosus*, a closely related cestode. In this study, this approach is evaluated and the results are compared to those obtained in the previous study. A total of 297 proteins organised in 106 protein groups were identified. Of the 106 protein groups, 58 groups were newly identified, while 48 groups confirmed previous findings. Blast2GO analysis revealed that the majority of the proteins were involved in catalytic activities and binding. Comparable findings were obtained in recent studies on other helminth genera like *Echinococcus*, *Schistosoma* and *Clonorchis*, indicating similarities between helminth excretion/secretion proteomes.

Adapted from:

Victor, B., Dorny, P., Kanobana, K., Polman, K., Lindh, J., Deelder, A. M., Palmblad, M., and Gabriël, S. Use of Expressed Sequence Tags as an Alternative Approach for the Identification of *Taenia solium* Metacestode Excretion/Secretion Proteins. *BMC Research Notes* 6, 1 (2013), 224.

7.1 Introduction

Proteomic experiments involving liquid chromatography and tandem mass spectrometry (LC-MS/MS) typically attempt to match the generated experimental spectra to *in silico* spectra from a (target) protein database. Ideally, this database contains every protein likely to be in the sample, but obtaining such an all-including protein database proves difficult when there is little to no genomic information available, as was the case for *T. solium* until recently (Tsai *et al.*, 2013). In the previous study, we bypassed this limitation by using a custom database with known proteins from related helminths (*Taenia*, *Echinococcus*, *Schistosoma* and *Trichinella*) as a target database in the LC-MS/MS experiments [Chapter 6 and Victor *et al.* (2012b)]. In that study, translated expressed sequence tags (ESTs) were deliberately not used, because the aim there was to investigate to usefulness of a target database made up of protein sequences originating mostly from (closely) related helminths.

The usefulness of ESTs for the identification of helminth proteins has already been described for e.g. *Haemonchus contortus* (Yatsuda *et al.*, 2003; Millares *et al.*, 2012) and *Echinococcus granulosus* (Monteiro *et al.*, 2010). In the case of *T. solium*, ESTs from different parasite stages have been made available by different research groups, both published (Almeida *et al.*, 2009; Lundström *et al.*, 2010) and unpublished (Huang J. *et al.*, Analysis of *Taenia solium* and *Taenia saginata* adult gene expression profile, 2009 and Aguilar-Diaz H. *et al.*, *Taenia solium* larva/adult ESTs, 2007).

In this study, *T. solium* ESTs combined with the Basic Local Alignment Search Tool (BLAST) and protein mapping to supercontigs of *E. granulosus* (a member of the Taeniidae family) were used to investigate whether this would increase the number of *T. solium* metacestode excretion/secretion protein identifications compared to the previous study.

7.2 Materials and methods

7.2.1 Generation of the dataset

The *in vitro* production of the *T. solium* metacestode excretion/secretion proteins from Peru and Zambia at 24h and 48h and the generation of line spectra mzXML files[1] have been described in Chapter 6 and Victor *et al.* (2012b).

[1] The datasets supporting the results of this article are available in the PRIDE repository at `http://www.ebi.ac.uk/pride` with accession numbers 19232 – 19267.

7.2.2 Database design and data analysis

To construct the target database, 30,700 expressed sequence tags were downloaded from the National Centre for Biotechnology Information (NCBI) website in April 2012 and a six frame translation was performed using transeq (Rice *et al.*, 2000). A *Sus scrofa* database with 1,388 Swiss-Prot sequences (`http://www.uniprot.org/`) and the common Repository of Adventitious Proteins database (112 protein sequences; `ftp://ftp.thegpm.org/fasta/cRAP/crap.fasta`) were also included to assist detection of host proteins and accidental contaminations, respectively. A decoy database with 185,700 reversed sequences was created using decoyfasta. These databases were fused into one final database. Database searching with X!Tandem (2010.10.01.1) (Craig and Beavis, 2004) and subsequent analyses with PeptideProphet (Keller *et al.*, 2002; Choi and Nesvizhskii, 2008b), iProphet (Shteynberg *et al.*, 2011) and ProteinProphet (Nesvizhskii *et al.*, 2003) were also performed as previously described [Chapter 6 and Victor *et al.* (2012b)]. All above mentioned tools, except transeq, are included with the Trans-Proteomic Pipeline v4.5 RAPTURE rev 2 (Keller *et al.*, 2005). The identified translated ESTs were further filtered to a false discovery rate of $< 1\%$ and ESTs with an individual probability of zero were discarded. The remaining ESTs were blasted against the NCBI nonredundant database (E-value $<$ 1e-10) and for each recognised EST, the best matching protein was retained. The resulting proteins were then screened by mapping the proteins to the *E. granulosus* supercontigs using TBLASTN (`http://www.sanger.ac.uk/cgi-bin/blast/submitblast/Echinococcus`). Identifications with a Score $>$ 200 were considered valid. Identifications with a lower score were manually evaluated and proteins originating from *T. solium* were retained. This step also helped to filter out host contaminations. Finally, proteins were grouped based on homology. All proteins that could not be grouped and were identified by only one EST were also discarded. SignalP 4.1 server (Petersen *et al.*, 2011) and SecretomeP 2.0 Server (Bendtsen *et al.*, 2004a) were used to predict classical (signal peptide triggered) and non-classical protein secretions, respectively. Finally, Blast2GO was used for Gene Ontology (GO) annotations (biological process, molecular function and cellular component) and the construction of level 2 pie charts (Conesa *et al.*, 2005). In order to gain more specific information, the largest categories were analysed to levels 3 and 4.

7.3 Results and discussion

7.3.1 Identified proteins and Gene Ontology annotation

In this study, 297 proteins (from 1,787 translated ESTs) were identified and organised in 106 protein groups based on homology (Additional file 1[2]). For simplicity, each protein

[2] `http://www.biomedcentral.com/content/supplementary/1756-0500-6-224-s1.pdf`

group is represented by one protein. The groups were further organised by Gene Ontology annotation information on biological process and molecular function. A total of 48 protein groups are labelled with an asterisk, indicating that they were also identified in the previous study (Additional file 2[3]) [Chapter 6 and Victor *et al.* (2012b)]. For brevity, Table 7.1 shows only the 58 newly identified protein groups along with the SignalP and SecretomeP analysis result. The former reports yes (Y) or no (N) for both the neural networks (NN) and hidden Markov models (HMM), while the latter interprets a NN-score above 0.5 as possible secretion (Y). For a number of proteins/protein groups, no Gene Ontology information was available. Nonetheless, many of them, like the 8 kDa protein family (Hancock *et al.*, 2003), have been extensively studied and used in diagnostic assays.

Most of the identified protein groups could be categorised in miscellaneous binding activities (e.g. actin binding, calcium binding and metal ion binding), various metabolic processes, gluconeogenesis (triosephosphate isomerase, enolase, phosphoenolpyruvate carboxykinase and phosphoglucose isomerase), glycolysis (glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase and fructose-bisphosphate aldolase) and proteins with (endo) peptidase activity, including cysteine-type (calpain, UDP-glucose 4-epimerase and cathepsin), threonine-type (proteasome subunits) and serine-type endopeptidase activity (trypsin-like protein). Endopeptidase inhibitors with both serine-type (kunitz protein 8 and leukocyte elastase inhibitor) and cysteine-type endopeptidase inhibitor activity (immunogenic protein Ts11) and components of the enzymatic antioxidant system of Taeniidae (Cu/Zn superoxide dismutase, glutathione S-transferase and peroxiredoxin) were also identified (Vaca-Paniagua *et al.*, 2009).

Gene Ontology level 2 pie charts were created for biological process (Figure 7.1a), molecular function (Figure 7.1b) and cellular component (Figure 7.1c). To avoid overly busy charts, the sequence filter was set to 10. The two largest categories of the biological process chart were cellular and metabolic processes. Others included biological regulation, response to stimulus, multicellular organismal processes and cellular component organisation or biogenesis. Further investigation of the general cellular and metabolic processes revealed primary and cellular metabolic processes at level 3 and protein, cellular macromolecule and cellular nitrogen compound metabolic processes at level 4 (Additional file 3, tab 1[4]). Molecular function was clearly divided between binding and catalytic activity. GO level 3 showed protein binding and hydrolase activity while level 4 entailed mostly nucleotide binding, hydrolase activity (acting on acid anhydrides), cation binding, peptidase activity, cytoskeletal and identical protein binding (Additional file 3, tab 2[4]). The level 2 pie chart for the cellular component indicated cell and organelle as the largest categories. Further

[3]`http://www.biomedcentral.com/content/supplementary/1756-0500-6-224-s2.pdf`
[4]`http://www.biomedcentral.com/content/supplementary/1756-0500-6-224-s3.xls`

Table 7.1: Protein groups ($n = 58$) newly identified in *Taenia solium* metacestode excretion/secretion proteins, organised by Gene Ontology annotation information on biological processes and molecular functions. For simplicity, all protein groups are represented by one protein.

Gene Ontology classification Protein group	Closest organism	GI code	Proteins[a)]	ESTs[b)]	SigP/SecP[c)]
1) No Gene Ontology classification					
Major egg antigen	*Clonorchis sinensis*	358336515	1	2	N/Y
ES1 protein homolog	Multiple[d)]	-	3	5	N/Y
Phosphoglyceride transfer protein	*Taenia asiatica*	124782980	1	7	N/Y
Alpha-2-macroglobulin-like protein 1	*Clonorchis sinensis*	358333571	1	5	Y/Y
Aldose 1-epimerase	*Clonorchis sinensis*	358334888	1	2	N/Y
SJCHGC02626 protein	*Schistosoma japonicum*	-	3	7	N/Y
Hypothetical protein	*Schistosoma mansoni*	256079415	1	4	Y/N
TSP1	*Echinococcus multilocularis*	209967595	1	4	N/Y
Putative major vault protein	*Echinococcus granulosus*	62178032	1	2	N/Y
2) Binding (miscellaneous)					
Filamin	Multiple	-	3	7	-/-
Methionyl-tRNA synthetase cytoplasmic	*Clonorchis sinensis*	358255967	1	5	N/N
SJCHGC09631 protein	*Schistosoma* spp.	-	2	2	N/-
four and a half LIM domains protein 3	*Clonorchis sinensis*	358341124	1	7	N/Y
Alpha-actinin isoform B	*Taenia asiatica*	124783372	1	3	N/Y
Calumenin	*Taenia asiatica*	124784033	1	2	N/Y
Calcium-binding protein	*Schistosoma mansoni*	256071353	1	2	N/Y
Lysyl oxidase-like	*Schistosoma mansoni*	256072781	1	2	N/Y
Porphobilinogen synthase	Multiple	-	3	3	N/-
Phosphoglucomutase-1	*Clonorchis sinensis*	358337844	1	2	N/N
Fibrillar collagen	Multiple	-	9	16	-/-
3) Glycolysis/Metabolic processes (miscellaneous)					
Adenylosuccinate synthetase	*Schistosoma mansoni*	387912858	1	4	N/Y
Adenylate kinase	Multiple	-	2	4	N/Y
UDP-glucose pyrophosphorylase 2	*Schistosoma* spp.	-	2	2	N/-
Hypothetical protein SINV_09109	*Solenopsis invicta*	322793762	1	2	N/Y
Aspartate aminotransferase	Multiple	-	3	5	N/-
Lactate dehydrogenase A	*Taenia solium*	318054471	1	6	N/Y
SJCHGC05968 protein	Multiple	-	2	2	N/Y
Methylthioadenosine phosphorylase	Multiple	-	2	2	N/-
Ornithine aminotransferase	Multiple	-	3	3	N/-
Endoglycoceramidase	Multiple	-	3	3	-/Y
Aminoacylase	Multiple	-	2	2	N/N
Glucose-6-phosphate 1-dehydrogenase-like	*Sus scrofa*	350595984	1	2	N/N
Phosphoglycerate mutase	Multiple	-	3	7	N/N
4) (Endo)peptidase activity					
Calpain	Multiple	-	5	5	N/-
UDP-glucose 4-epimerase	Multiple	-	2	2	N/-
Dipeptidyl-peptidase	Multiple	-	2	3	N/-
Glutamate carboxypeptidase 2	*Clonorchis sinensis*	358331956	1	3	N/Y
5) Endopeptidase inhibitor activity					
Kunitz protein 8	Multiple	-	2	3	-/Y

Table 7.1: Continued.

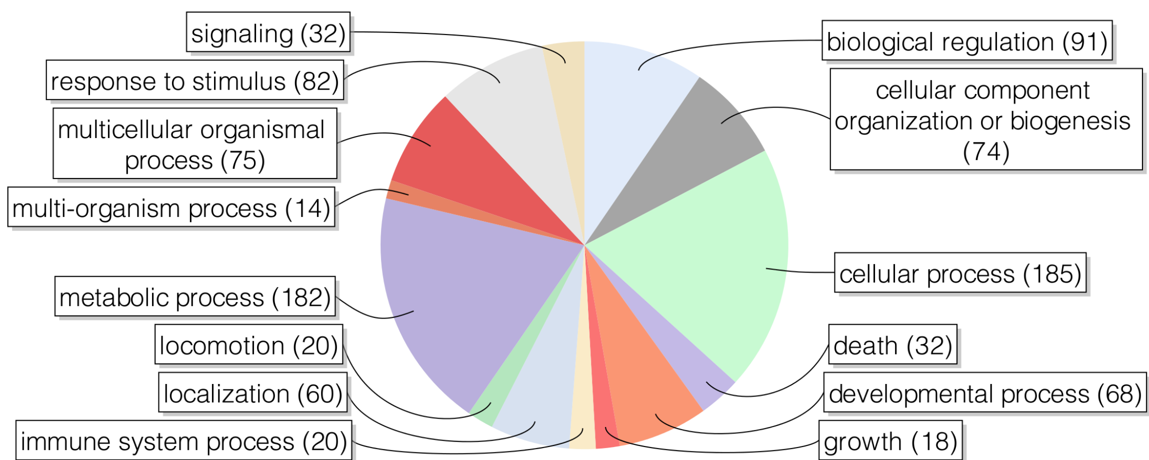
Gene Ontology classification Protein group	Closest organism	GI code	Proteins[a)]	ESTs[b)]	SigP/SecP[c)]
6) Cell redox homeostasis/Oxidation-reduction related					
Carbonyl reductase	*Schistosoma* spp.	-	3	3	N/N
Methionine sulfoxide reductase	Multiple[d)]	-	2	4	N/Y
procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3	Multiple	-	3	6	-/-
7) Transport					
Charged multivesicular body protein	Multiple	-	3	4	N/N
SJCHGC06082 protein	Multiple	-	2	8	N/-
Glycolipid transfer protein-like protein	*Taenia asiatica*	124782916	1	2	N/N
Gamma-soluble NSF attachment protein	Multiple	-	4	10	N/-
Sodium/glucose cotransporter	Multiple	-	2	7	N/N
8) Motor activity/Cytoskeleton and Microtubule related					
Tubulin polymerization-promoting protein	Multiple	-	2	2	N/-
Myophilin	Multiple	-	2	19	N/-
9) Miscellaneous Gene Ontology classification					
Translation initiation factor 5A	Multiple	-	2	4	N/N
Ubiquitin-conjugating enzyme	Multiple	-	4	10	N/Y
Protein-l-isoaspartate o-methyltransferase	*Schistosoma mansoni*	256081696	1	2	N/Y
Protein DJ-1-like	Multiple	-	2	4	N/Y
6-phosphogluconolactonase	Multiple	-	2	4	N/Y
SJCHGC02435 protein	*Schistosoma japonicum*	56756018	1	5	Y/Y
Family T2 unassigned peptidase	*Schistosoma mansoni*	256088374	1	4	Y/Y
3'(2'), 5'-bisphosphate nucleotidase	Multiple	-	2	2	-/Y
RAB GDP dissociation inhibitor alpha	Multiple	-	2	3	N/N
Laminin	Multiple	-	2	2	-/N

[a)] The number of proteins in each protein group.
[b)] The number of expressed sequence tags that were matched to proteins in this protein group.
[c)] SignalP and SecretomeP (SigP/SecP) results are described as yes (Y) or no (N). SignalP is Y when both the neural network and the hidden Markov models agree. (-) is reported when not all individual proteins in that groups have the same result.
[d)] 'Multiple' indicates that different (helminth) genera have identified proteins in that protein group.

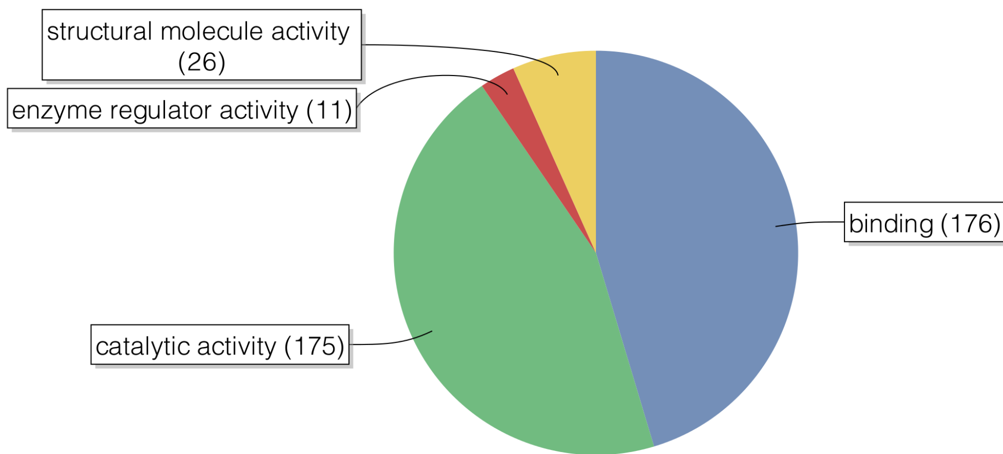
analyses showed mostly cell part and membrane-bound organelle, and intracellular (part) GO terms at levels 3 and 4, respectively (Additional file 3, tab 3[4]).

Human keratin and porcine trypsin were identified in all samples. As keratin is a common contamination and trypsin was deliberately added during the LC-MS/MS experiments, both were omitted from the final results.

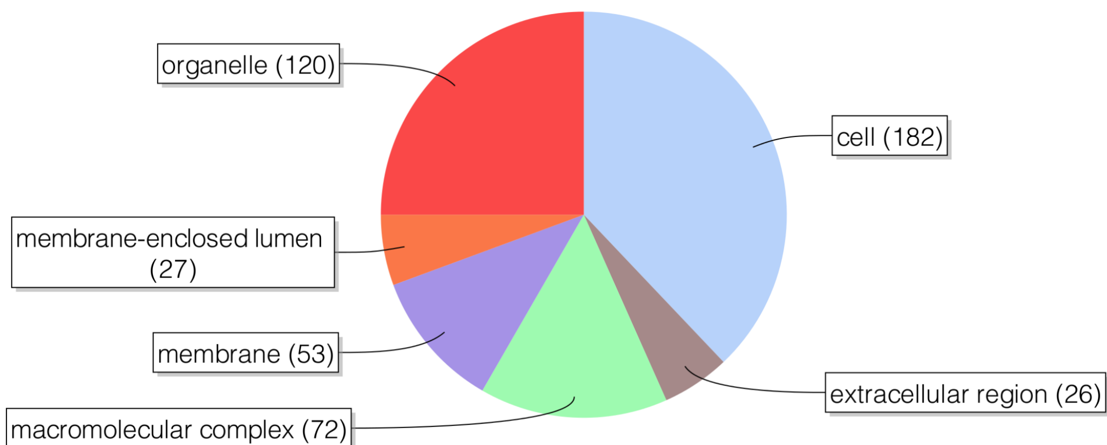
The presence of intracellular/non-secreted proteins in the ESPs is interesting and has been observed in other ESP studies before (Mulvenna *et al.*, 2010; Virginio *et al.*, 2012). Although it is highly likely that the majority of those proteins are indeed excreted or secreted by the parasite, the possibility that they are the result of leakage due to cyst damage or death should not be excluded.



(a) Gene Ontology level 2 pie chart – biological processes.



(b) Gene Ontology level 2 pie chart – molecular functions.



(c) Gene Ontology level 2 pie chart – cellular components.

Figure 7.1: Gene Ontology level 2 pie charts displaying the biological processes (a), the molecular functions (b) and the cellular components (c) of the 297 proteins that were identified in the *Taenia solium* metacestode excretion/secretion proteins. Values within parentheses are the number of sequences associated with each Gene Ontology term. All charts were created using Blast2GO with the sequence filter set to 10.

In general, the findings reported in this study are comparable to recent studies on other helminth genera like *Echinococcus* (Virginio *et al.*, 2012), *Schistosoma* (Liu *et al.*, 2009) and *Clonorchis* (Zheng *et al.*, 2011), indicating that excretion/secretion proteomes are not very different between helminth genera/species.

7.3.2 Comparison between the two studies

When comparing the level 2 GO terms identified in both studies (Table 7.2), all GO terms from the previous study were identified here as well. Additionally, we identified six new GO terms with the EST analyses: rhythmic process (GO:0048511), antioxidant activity (GO:0016209), molecular transducer activity (GO:0060089), protein binding transcription factor activity (GO:0000988), receptor activity (GO:0004872) and synapse (GO:0045202). Although a direct comparison between numbers should be avoided (due to proteins having multiple GOs and the presence of homologous proteins in the proteins groups, especially in the previous study where it is a logical result of the target database construction), the general levels of abundance (= proteins in each GO term) are largely comparable between the two studies e.g. in both studies, cellular process, metabolic process and biological stimulation are the largest groups for 'biological process' while binding and catalytic activity are the largest groups for 'molecular function' and cell and organelle are the largest groups for 'cellular component'.

The six new GO terms were identified by a very small number of proteins and may be a result of proteins being linked to multiple GO terms. This is supported by the fact that the proteins linked to these GO terms are homologous to other proteins identified in both studies, so none of these GO terms was identified by a 'new' protein group.

7.4 Conclusions

In this study, a library of translated ESTs combined with BLAST and mapping strategies was used not only to confirm previously identified *T. solium* metacestode excretion/secretion proteins, but to identify several new proteins as well, thereby effectively increasing the overall number of protein identifications.

The larger and more complete the EST database, the better proteomic coverage likely obtained. No ESTs from other Taeniidae were used in this study, since the available *T. solium* ESTs were already a merge of EST submissions by different groups and were therefore likely to offer decent proteome coverage. However, in cases where only a small EST library is available with low coverage, one could also include protein sequences and/or ESTs from related organisms in a combined database. This may be particularly advantageous in proteomic studies on less studied, unsequenced, organisms. It should be noted that research on non-sequenced organisms mostly relies on homology to already

Table 7.2: Gene Ontology level 2 annotations identified in this study alongside the ones identified in the previous study. Although a direct comparison between numbers should be avoided (due to proteins having multiple GOs and the presence of homologous proteins in the proteins groups, especially in the previous study where it is a logical result of the target database construction), the general levels of abundance are largely comparable between the two studies. Additionally, this study revealed six new GO annotations.

Gene Ontology information		Current (EST) study	Previous study[a)]
Biological process			
cellular process	GO:0009987	185	162
metabolic process	GO:0008152	182	150
biological regulation	GO:0065007	91	153
response to stimulus	GO:0050896	82	147
multicellular organismal process	GO:0032501	75	107
cellular component organisation or biogenesis	GO:0071840	74	92
developmental process	GO:0032502	68	85
localisation	GO:0051179	60	88
signaling	GO:0023052	32	55
death	GO:0016265	32	65
immune system process	GO:0002376	20	59
locomotion	GO:0040011	20	29
growth	GO:0040007	18	28
multi-organism process	GO:0051704	14	65
reproduction	GO:0000003	10	36
biological adhesion	GO:0022610	9	13
viral reproduction	GO:0016032	8	12
cell proliferation	GO:0008283	5	24
cell killing	GO:0001906	2	21
rhythmic process	GO:0048511	2	-
Molecular function			
binding	GO:0005488	176	168
catalytic activity	GO:0003824	175	129
structural molecule activity	GO:0005198	26	20
enzyme regulator activity	GO:0030234	11	21
electron carrier activity	GO:0009055	7	13
antioxidant activity	GO:0016209	7	-
transporter activity	GO:0005215	6	14
molecular transducer activity	GO:0060089	4	-
protein binding transcription factor activity	GO:0000988	3	-
nucleic acid binding transcription factor activity	GO:0001071	2	17
receptor activity	GO:0004872	1	-
Cellular component			
cell	GO:0005623	182	168
organelle	GO:0043226	120	155
macromolecular complex	GO:0032991	72	79
membrane	GO:0016020	53	76
membrane-enclosed lumen	GO:0031974	27	63
extracellular region	GO:0005576	26	71
extracellular matrix	GO:0031012	10	11
synapse	GO:0045202	5	-
cell junction	GO:0030054	2	12

[a)] Chapter 6 and Victor *et al.* (2012b)

existing proteins from other (preferably closely related) organisms. Therefore, there is no possibility of finding unique proteins, unless (i) *de novo* sequencing is performed on the good quality unmatched experimental spectra or (ii) ESTs that were identified by spectra but remained unmatched during BLAST are further investigated.

The results of SignalP and SecretomeP show a number of proteins that are negative for both analyses. This indicates that some identified proteins were neither secreted through the classical pathways (SignalP), nor through the non-classical pathway (SecretomeP). The possibility of leaked proteins in the ESPs should therefore not be overlooked, especially since dead cysts were observed in the ESP cultures in Peru (but not in Zambia). However, the presence of proteins that are not predicted to be secreted by SignalP or SecretomeP analysis has been reported in a number of other studies as well. Liu *et al.* (2009) estimated that 48% of the excretory/secretory proteome of adult *Schistosoma japonicum* flukes were released through an unknown mechanism. Similar observations were made by e.g. Knudsen *et al.* (2005) for *Schistosoma mansoni*, by Mulvenna *et al.* (2009) for *Ancylostoma caninum* and by Craig *et al.* (2006a) for *Teladorsagia circumcincta*.

Finally, it is important to realise that, although the mapping to the *E. granulosus* supercontigs helped to remove *S. scrofa* host proteins (e.g. albumin, protegrin and hemopexin), some may still be present. Heat shock protein 70, for example, is identified both in *S. scrofa* and *E. granulosus*.

CHAPTER 8

The Search for Differential Proteins

Abstract: Circulating *Taenia solium* antigens are currently detected with two ELISAs that were originally developed to detect *T. saginata* cysticerci in cattle, implying a cross-reaction with *T. solium* antigens. Unfortunately, *T. asiatica* and *T. hydatigena*, two *Taenia* species that can co-occur in pigs also cross-react, thereby seriously impairing the usefulness of antigen detection in pigs. Diagnosis of porcine cysticercosis would benefit greatly from a *T. solium*-specific assay. To achieve this, a comparison between the different *Taenia* proteomes could result in the identification of unique *T. solium* peptides/proteins. Three comparison methods are explored, of which a novel spectral comparison pipeline showed the most potential. The preliminary results of this comparison hint at members of the 8 kDa protein family and phosphoenolpyruvate carboxykinase.

8

Victor, B.[1], Yilmaz, Ş.[2], Borloo, J.[3], Geldhof, P.[3], Deelder, A. M.[4], Palmblad, M.[4], Gabriël, S.[1], Dorny, P.[1,3], Martens, L.[2,5]

[1] Veterinary Helminthology Unit, Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[2] Department of Medical Protein Research, Flanders Interuniversity Institute for Biotechnology, Ghent, Belgium
[3] Laboratory of Parasitology, Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium
[4] Centre for Proteomics and Metabolomics, Leiden University Medical Centre, Leiden, The Netherlands
[5] Faculty of Medicine and Health Sciences, Department of Biochemistry, Ghent University, Ghent, Belgium

Manuscript in preparation.

8.1 Introduction

Circulating antigens that are excreted and secreted by viable *Taenia* spp. cysticerci are currently detected by two different monoclonal antibody-based antigen detecting ELISAs: the HP10 ELISA (Harrison *et al.*, 1989) and the B158/B60 ELISA (Brandt *et al.*, 1992; Van Kerckhoven *et al.*, 1998; Dorny *et al.*, 2000). Both assays were originally developed to detect excretion/secretion proteins from *T. saginata* cysticerci in cattle. From this it follows that their use in humans and pigs is based on cross reactions with *T. solium* antigens. Although convenient for the detection of human (neuro)cysticercosis, the use of these assays is compromised in pigs, more specifically, in regions where *T. solium*, *T. hydatigena* and/or *T. asiatica* are sympatric.

Diagnosis of porcine cysticercosis would benefit from a more specific assay. To achieve this, one needs to not only study the excretion/secretion proteins (ESPs) from *T. solium*, but compare these proteins with the ESPs from the other *Taenia* parasites as well. The main objective of such a comparison is to identify unique proteins and target them in a modified or new antigen detection assay.

There are at least three possible ways to compare the ESPs.
A first method is to analyse the ESPs of the other *Taenia* species in the same way the ESPs for *T. solium* were analysed [Chapter 6 and Victor *et al.* (2012b)] and to compare the protein output lists. For the analysis of the *T. solium* ESPs, the protein database was supplemented with protein sequences from closely related helminths (including other *Taenia* species) and expressed sequence tags. This increased the proteome coverage significantly and allowed a more detailed proteomic study of *T. solium*. However, for comparative studies, especially between closely related *Taenia* species, this approach quickly fails due to the necessary emphasis on the small differences in protein sequence that separate these species. This option is not considered to be suitable for the task at hand and is therefore immediately discarded.

A second option is to separate the ESPs with two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and to identify spots that are present in one sample, but not in another. These differential spots can then be further analysed with liquid chromatography and tandem mass spectrometry (LC-MS/MS). Although the identification of the proteins after LC-MS/MS will likely be difficult because of the above mentioned reasons, it is much more targeted then the first option i.e. at least now the analysis is focussed on a much smaller subgroup of proteins that appear to be unique on gel. However, 2D-PAGE is not without its shortcomings (see also Section 3.2.1) e.g. the tendency to identify only the most abundant proteins. Obviously, this immediately puts a number of possible candidate

proteins out of the equation.

A third option is to analyse the ESPs with LC-MS/MS, but then to employ a new methodology that bypasses the identification stage and instead considers all spectra from all analyses, ensuring that none of the acquired information is lost at the all-important comparison stage. In other words, the experimental spectra will be compared directly with each other in order to identify which spectra are unique for the organism of interest. The comparison phase uses Spearman's rank correlation coefficients (ρ) and in addition, the analysis methodology also identifies contaminating host proteins and filters spectra by their quality. More specifically, this last step will flag all spectra unlikely to be of peptide origin (e.g. spectra from polymers) as well as overall poor spectra (e.g. very few peaks). This is achieved by training the SpecQual software tool with known identified and unidentified spectra recorded on the same instrument, but from different samples (Flikka *et al.*, 2006). The final output of this methodology is a complete set of across-organism matched spectra with contaminant and quality status annotations. From this set, unique and relevant spectra for any of the organisms can be readily deduced, although logically, in this case, unique *T. solium* spectra are of particular importance. Ultimately, identifications should be obtained for (at least some of) the differential spectra and this still requires a protein database to search against. This is where the recently published *T. solium* genome becomes relevant (Tsai *et al.*, 2013). From this genome, and the preceding expressed sequence tag (EST) libraries, it is possible to derive a protein database that should ideally cover the entire proteome. This database can then be used to identify the peptides and proteins that correspond to the unique spectra that were retained after the analysis with the spectral comparison tools.

Unfortunately, no *T. asiatica* material was available, so only *T. solium* and *T. hydatigena* ESPs were compared.

8.2 Materials and methods

8.2.1 Parasite material and excretion/secretion proteins

T. solium metacestode collection from a naturally infected pig and 24h *in vitro* ESP production were done in Zambia as previously described [Chapter 6 and Victor *et al.* (2012b)]. *T. hydatigena* metacestodes of various sizes ($n = 14$) were collected from the omentum and liver of naturally infected Zambian goats and carefully separated from the host tissue. ESPs were prepared as described for *T. solium*, except that the number of cysts per culture dish was modified according to the size of the cysts. *T. solium* metacestode viability was assessed by incubating 20 cysts in 10% porcine bile in culture medium and

subsequent evaluation of scolex evagination. No 'bile evagination test' was performed for *T. hydatigena*, but the metacestode movement was monitored. Protein concentration of the ESPs was determined with a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific Inc., Rockford, IL). Species confirmation of the *Taenia* metacestodes was done with a polymerase chain reaction assay followed by enzymatic digestion (PCR-RFLP) as described by Devleesschauwer *et al.* (2013).

8.2.2 Two-dimensional polyacrylamide gel electrophoresis

The 2D-PAGE protocol was modified from Borloo *et al.* (2013). Briefly, 200 µg of *T. solium* and *T. hydatigena* ESPs were precipitated by adding four volumes of ice-cold acetone. This was vortexed briefly and incubated at −20℃ for one hour. Protein pellet was recovered by centrifugation at 5,000×g for five minutes. The supernatant was discarded and the pellet was resolubilised in 8 M urea, 2 M thiourea, 2% w/v CHAPS, 20 mM dithiothreitol (DTT) and 0.2% v/v carrier ampholytes (pH 3-10; GE Healthcare Bio-Sciences AB; Uppsala, Sweden). Dissolved proteins were centrifuged again at 8,000×g for five minutes to remove any remaining insoluble proteins. The protein solution was applied to 7 cm, pH 3-10, ImmobilineTM DryStrip IEF strips (GE Healthcare), covered with mineral oil and left overnight for rehydration. An Ettan IPGphor3 instrument (GE Healthcare) was used for the isoelectric focussing (IEF) with an initial three hour focussing period at 300 V, followed by a five hour linear gradient from 300 to 3500 V and a final 18 h at 3500 V, yielding a total voltage load of approximately 73 kVh. Proteins in the IEF strips were reduced and alkylated prior to the second dimension separation, i.e. strips were incubated for 15 minutes at room temperature in a 50 mM Tris-HCl, pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v sodium dodecyl sulfate (SDS) and 2% w/v DTT solution followed by another 15 minutes in an similar solution containing 2.5% w/v iodoacetamide instead of the DTT. Second dimension separation was performed according to Laemmli (1970) i.e. a 12% bis-acrylamide gel (pH 8.8) resolving gel and a stacking gel at pH 6.8 were used. The treated IEF strips were imbedded in the stacking gel prior to the gel run. Finally, staining was done with SimplyBlue SafeStain (Invitrogen, Carlsbad, CA) according to the manufacturers' instructions.

8.2.3 One-dimensional polyacrylamide gel electrophoresis and in-gel trypsin digest

T. solium and *T. hydatigena* ESP aliquots (40 µg) were precipitated in four volumes of ice-cold acetone and pellets were resuspended in NuPAGE Sample Reducing Agent (10x) and NuPAGE LDS Sample Buffer (4x). Protein separation was done on a 4-12% NuPAGE Novex Bis-Tris gel. Gels were stained overnight with the Colloidal Blue Staining Kit (all reagents from Invitrogen). Each gel lane was cut in 48 identical slices using a

custom designed gel cutter (Gel Company Inc., San Francisco, CA) and all slices were transferred to a 96-well plate (Greiner Bio-One, Frickenhausen, Germany; one slice/well). Gel slices were washed three times: first with 100 µL 25 mM ammonium bicarbonate (ABC; Sigma-Aldrich Corp.), then with 100 µL 30% acetonitrile (Biosolve, Valkenswaard, The Netherlands) in ABC and finally with 100 µL 100% acetonitrile. Cystines were reduced by adding 75 µL 10 mM DTT in ABC for 30 minutes at 56°C. Gel slices were shrunk with 100 µL of 100% acetonitrile and cysteines were alkylated with the addition of 75 µL 55 mM iodoacetamide (Sigma-Aldrich Corp.) in ABC for 20 minutes (in the dark) at ambient temperature. Gel slices were washed and shrunk again with the addition of 100 µL 25 mM ABC followed by 100 µL 100% acetonitrile. Proteins were digested overnight at 37°C with sequencing grade porcine trypsin (5 µg/mL ABC; Promega, Southampton, UK; 30 µL/well). From each well, 20 µL of supernatant was extracted and mixed with 2 µL 5% TFA (Sigma-Aldrich Corp.), while 20 µL 0.1% TFA was added to the gel pieces for another hour of incubation. A second supernatant sample of 20 µL was extracted and the two supernatants were pooled. The total volume per well was reduced to 20 µL with a vacuum centrifuge and plates were stored at −30°C pending further analysis.

8.2.4 Liquid chromatography – tandem mass spectrometry

Peptides were analysed on an amaZon ETD ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an Apollo II ESI source and coupled on-line with a NanoLC-Ultra 2D plus LC (Eksigent, Dublin, CA, USA). A total volume of 10 µL was first desalted on a C18 PepMap300 trap column (5 µm, 300 µm × 5 mm; Dionex, Sunnyvale, CA) and then separated on a 15 cm C18 ChromXP analytical column (3 µm, 150 × 0.3 mm; Eksigent) by a 45-minute linear gradient from 4% to 35% acetonitrile, 0.05% formic acid (Sigma-Aldrich Corp.) with a flow rate of 4 µL/minute. Mass spectra were acquired in m/z 300-1,300 with the ion charge control (ICC) set to 200,000 and the maximum accumulation time set to 200 ms. The 10 most abundant precursors were selected for CID MS/MS in m/z 100-2,000 with the ICC and maximum accumulation time set to 200,000 and 50 ms respectively, but were excluded for 0.5 minutes after having been selected once. Singly charged ions were excluded for MS/MS. The LC system was controlled by HyStar 3.2 and the ion trap by trapControl 7.0.

8.2.5 Data analysis – spectral comparison

An overview of the spectral comparison workflow is shown in Figure 8.1. Raw LC-MS/MS data were converted to line spectra mzXML files with the Bruker compassXport tool (version 3.0.4) and then to MGF files with MSConvert (ProteoWizard; version 3.0.4388) (Kessner *et al.*, 2008; Chambers *et al.*, 2012) with the assumption that all spectra are either doubly or triply charged. *T. solium* spectra were searched against a

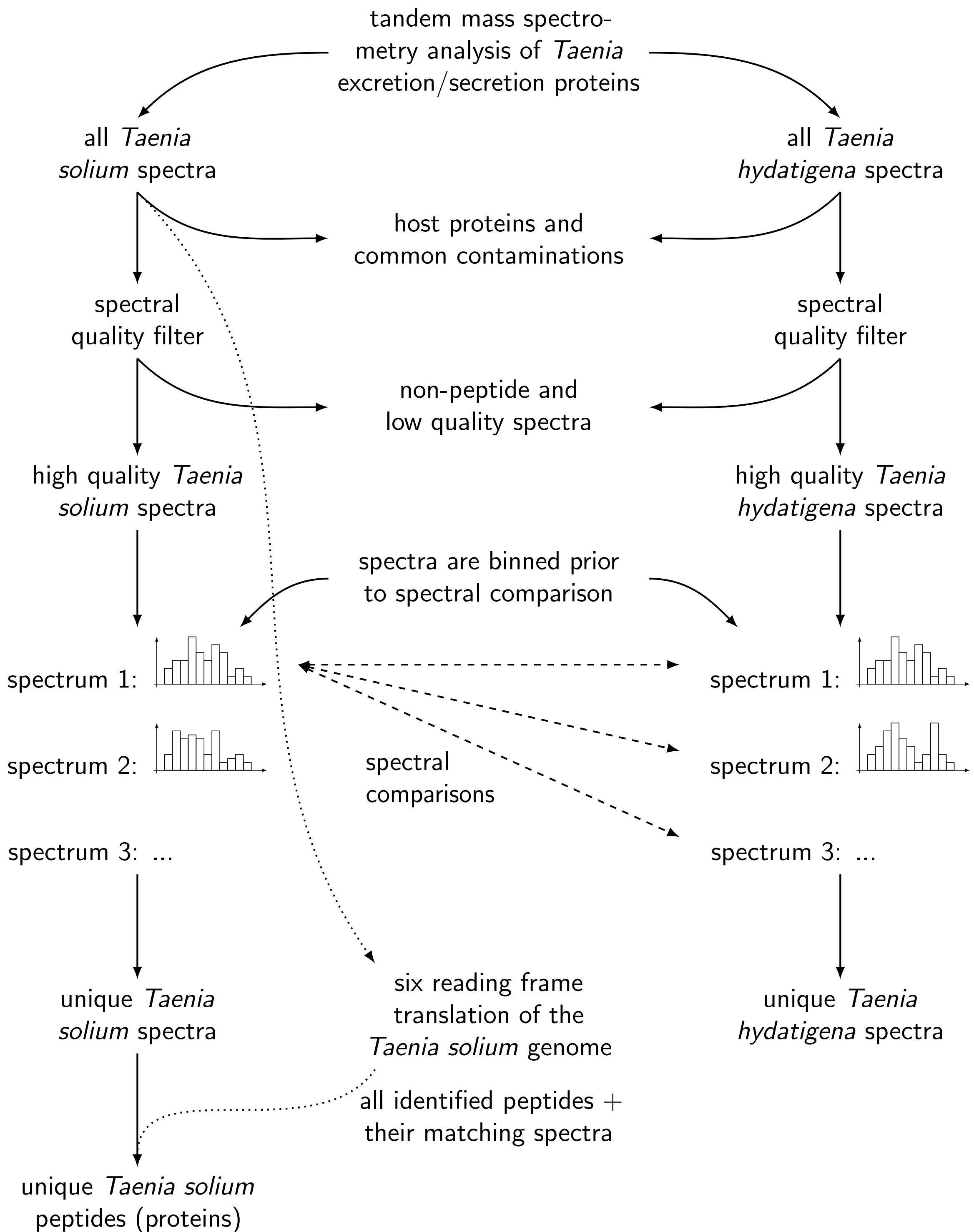


Figure 8.1: Overview of the spectral comparison methodology. First, spectra matching host proteins and common contaminations are removed. The remaining spectra are filtered to remove the non-peptide spectra and spectra of lower quality. Then, the high quality spectra are transformed to binned spectra, which are then compared between datasets with a Spearman's rank correlation coefficient threshold set to 0.2. The result is a list of unique spectra for each *Taenia* species. Finally, the spectra that remained after the host and common contaminations were removed, are searched against a six reading frame translation of the genome. The spectra that match to *T. solium* peptides are then compared to the unique spectra to allow the identification of peptides that are only matched by unique spectra.

Sus scrofa protein database (25,883 sequences; `http://www.ensembl.org/index.html`) with SearchGUI (Vaudel *et al.*, 2011a) and PeptideShaker (`https://code.google.com/p/peptide-shaker/`) (Barsnes *et al.*, 2011) with the following settings: precursor and fragment tolerances were set to 0.9 Da and 0.5 Da, respectively. Accepted number of missed cleavages was set to one. Fixed modification was carbamidomethylation of cysteine, while oxidation of methionine and pyroglutamate for N-terminal glutamine were set as variable modifications. All assigned spectra at a false discovery rate (FDR) of 1% were discarded. Similarly, *T. hydatigena* data were matched to a *Capra hircus* protein database (1,841 sequences; `http://www.uniprot.org`) and all assigned spectra at FDR 1% were discarded. Next, the remaining datasets were searched against the common Repository of Adventitious Proteins database (112 protein sequences; `ftp://ftp.thegpm.org/fasta/cRAP/crap.fasta`) to detect common and/or accidental contaminations in the protein samples. Again, assigned spectra at FDR 1% were discarded. The spectral quality filtering tool SpecQual (Flikka *et al.*, 2006) was trained with spectra from *Escherichia coli* standards (run on the same instrument) and all experimental spectra were subsequently filtered with the "cost" parameter c set to 10. Noise filtering was then applied to all remaining spectra, with the cut-off intensity determined for each spectrum individually. The similarity comparison process was initiated by transforming each individual high quality spectrum in each dataset to a binned spectrum, with bin sizes equal to two times the fragment tolerance. Each bin intensity was then adjusted to reflect the sum of the peak intensities in that bin. These transformed spectra were subsequently compared to a subset of (transformed) spectra in the other dataset, with subsets determined by allowing a retention time window of 2.5 minutes around the retention time of the spectra. The actual spectral comparison was performed with Spearman's rank correlation coefficient and all spectra with a coefficient > 0.2 were discarded. Below or equal to that value, they were considered unique. A six reading frame translation of the *T. solium* genome (`ftp://bioinformatica.biomedicas.unam.mx/`) was performed with no start codon restrictions i.e. proteins could start with any amino acid, not only methionine (AUG) (Barsnes *et al.*, 2011). After translation, DBToolKit (Martens *et al.*, 2005) was used to discard any entries shorter than five amino acids. Finally, the spectra that remained after removal of the host proteins and the common and/or accidental contaminations were searched against this six reading frame translation with search parameters as described above. The resulting set of peptide spectrum matches were subsequently compared to the list of unique spectra to identify peptides that were matched only by unique spectra.

8.3 Results and discussion

8.3.1 Parasite material and excretion/secretion proteins

Metacestode viability/scolex evagination was determined at 100% for *T. solium*. *T. hydatigena* metacestodes moved actively and were evaginating and invaginating continuously. No dead cysts were observed during the 24h culture period. Protein concentrations were determined at 4.565 mg/mL and 1.478 mg/mL for the *T. solium* and *T. hydatigena* ESPs, respectively. PCR-RFLP analysis confirmed genus and species of the collected metacestodes (Figure 8.2).

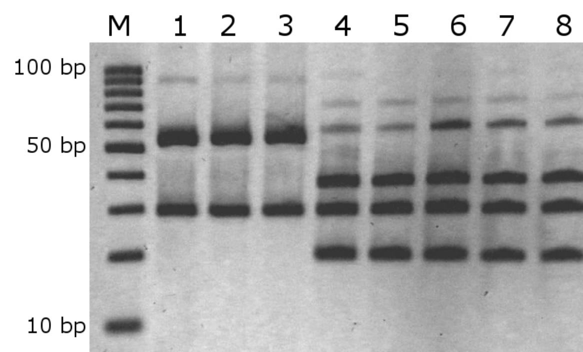
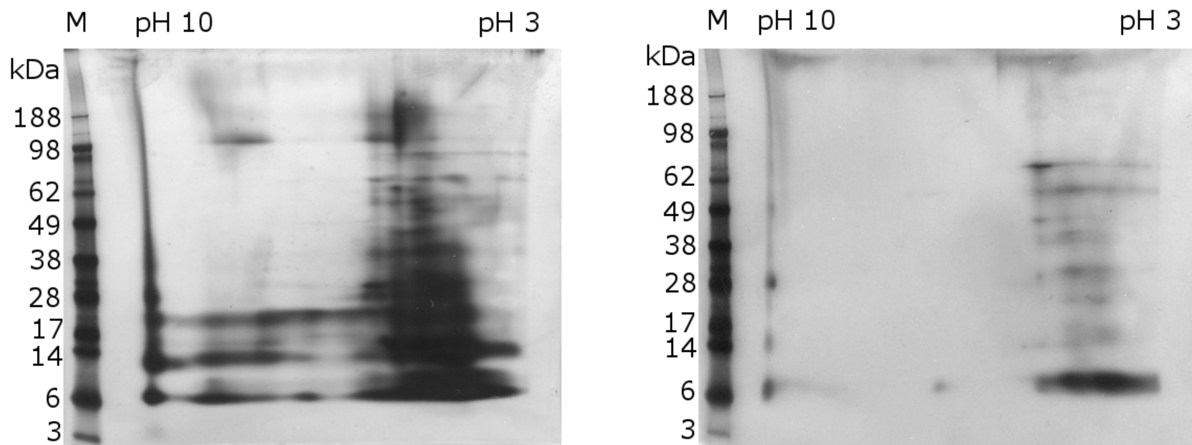


Figure 8.2: Restriction fragment length polymorphism profiles of *Taenia* metacestodes; M: DNA size marker; Lanes 1-3: *Taenia solium*; Lanes 4-8: *Taenia hydatigena*.

8.3.2 Two-dimensional polyacrylamide gel electrophoresis

Despite multiple attempts, the basic protocol did not yield a suitable result. A large number of adjustments were made to the sample preparation, the protein concentration, the second dimension separation system and the staining method, but none of these modifications resulted in gels that enabled us to repeatedly spot the same differences. For informational purposes only, ESP profiles of *T. solium* and *T. hydatigena* are shown in Figure 8.3a and Figure 8.3b, respectively. Isoelectric focussing was done on 7 cm, pH 3-10, ImmobilineTM DryStrip IEF strips (GE Healthcare), while the second dimension separation was performed with NuPAGE Novex 4-12% Bis-Tris ZOOM protein gels (Invitrogen, Carlsbad, CA). Silver staining was done according to Schaegger (2006).

Comparison of both proteomes with 2D-PAGE and subsequent LC-MS/MS analysis of the visually unique proteins was the second of the three suggested options. The advantage of this option over the first one, is that by first (visually) identifying the proteins of interest, the approach becomes much more targeted. Unfortunately, the many shortcomings of 2D-PAGE make for a realistic chance of missing potentially interesting proteins. Nonetheless, the option was explored, but ultimately abandoned due to issues with reproducibility and



(a) 2D-PAGE profile of *Taenia solium* excretion/secretion proteins.

(b) 2D-PAGE profile of *Taenia hydatigena* excretion/secretion proteins.

Figure 8.3: Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) profiles of *Taenia solium* (a) and *Taenia hydatigena* (b) excretion/secretion proteins. Lane M is a protein size marker. The pH values mark the direction in which the isoelectric focussing strip was inserted into the gel. The profiles were not reproducible and therefore, did not allow for a reliable selection of differential proteins.

(horizontal) streaking of the gel. The problem of 2D-PAGE reproducibility was one of the reasons 2D DIGE became so popular, but unfortunately, 2D DIGE was not available at the time. The streaking issue can have multiple reasons, ranging from incomplete focussing in the first dimension (too much or too little voltage) and protein overloading (protein aggregation), to sample preparation issues (contaminating salts, charged detergents, nucleic acids, lipids,...) and failure to completely solubilise all the proteins in the sample. Pinpointing the exact cause (or causes) of the streaking requires time, evaluation of various sample clean-up kits and methods and large sample volumes. However, due to the limited amounts of sample available for further experiments and the other remaining issues specific to 2D-PAGE, this was not considered a priority.

8.3.3 One-dimensional polyacrylamide gel electrophoresis

Figure 8.4 shows the protein profiles of the *T. solium* and *T. hydatigena* ESPs. Each gel lane was cut in 48 identical slices and all slices were subjected to the in-gel tryptic digest protocol.

8.3.4 Data analysis – spectral comparison

Figure 8.5 provides a visual overview of the numbers of spectra that were discarded at each step of the process. More specifically, database searching against host proteins and common contaminations reduced the number of spectra in the *T. solium* dataset from 453,532 to

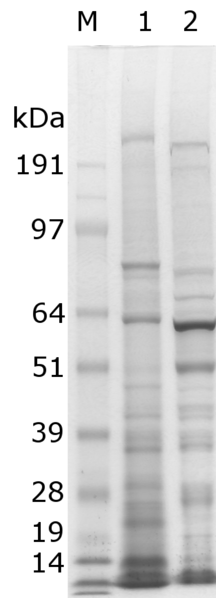


Figure 8.4: One-dimensional polyacrylamide gel electrophoresis profiles of *Taenia* metacestode excretion/secretion proteins; M: Protein size marker; Lane 1: *Taenia solium*; Lane 2: *Taenia hydatigena*.

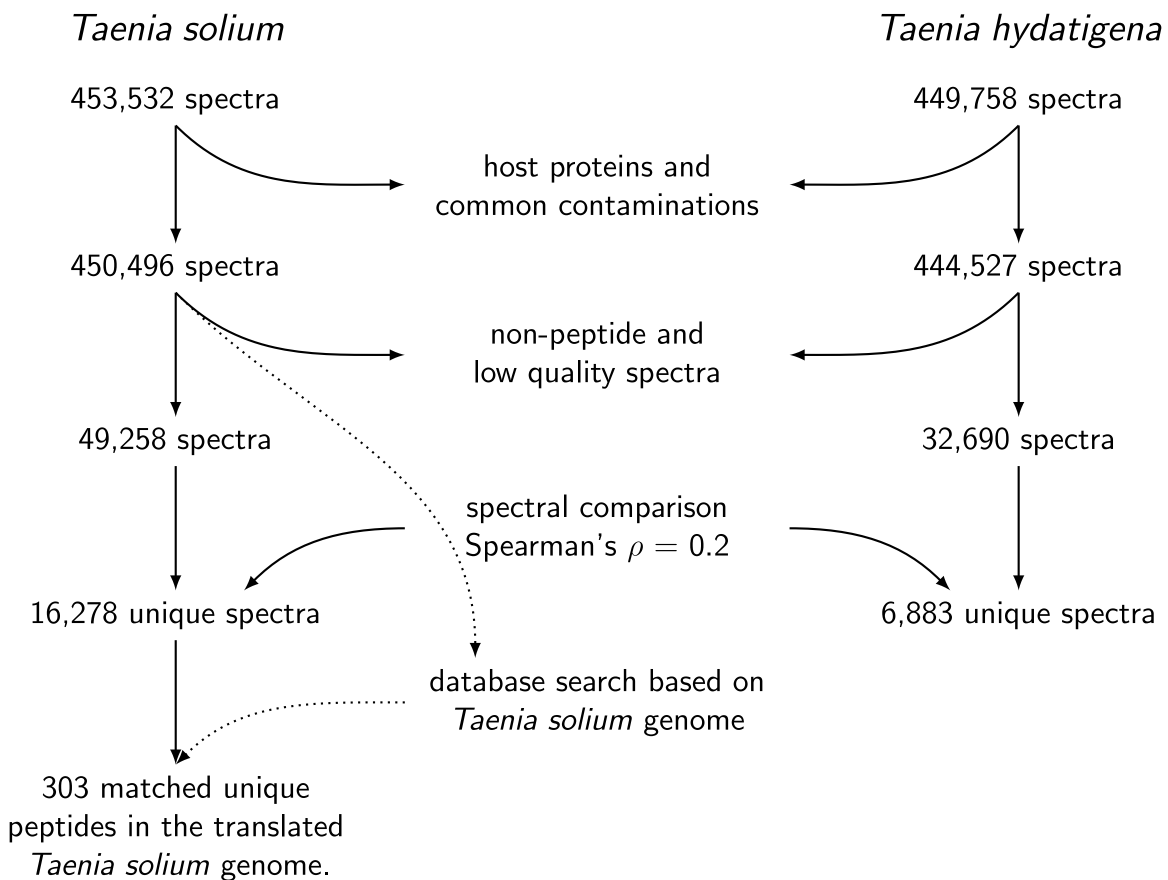


Figure 8.5: Filtering of the *Taenia solium* and *Taenia hydatigena* experimental spectra with the spectral comparison methodology.

450,496. Similarly, the *T. hydatigena* dataset was reduced from 449,758 to 444,527 spectra. The removed spectra mainly matched to host albumin, human keratin and porcine trypsin. The spectral quality filter (at c = 10) further reduced the number of spectra to 49,258 and 32,690 for *T. solium* and *T. hydatigena*, respectively. Ultimately, spectral matching with a Spearman's rank correlation coefficient threshold set to 0.2 retained 16,278 and 6,883 unique spectra for *T. solium* and *T. hydatigena*, respectively. Database searching with the 450,496 *T. solium* spectra against the translated genome resulted in 9,071 peptide spectrum matches and 1,545 peptides (FDR = 1%). The 16,278 unique spectra matched 624 of these peptides of which a subset of 303 peptides were identified ONLY by unique spectra. The remaining 321 peptides were identified by both unique spectra and at least one non-unique spectrum. Logically, these were not considered as unique peptides. Although the analysis of these peptides is still ongoing, preliminary results were further explored by first setting a spectral count > 10 threshold to reduce the number of peptides to 11. Then, the sequences were analysed with BLASTP (version 2.2.29+, optimised for queries shorter than 30 residues) against (i) the National Centre for Biotechnology Information non-redundant (NCBInr) database and (ii) the protein database derived from the *T. solium* genome (TsM1_13.12.11.prot.faa; `http://www.taeniasolium.unam.mx/taenia/`) on May 4, 2014 (Table 8.1).

Since low BLASTP E-values (e.g. E-value < 1e-10) signify a better match, the most promising results in this list are peptides 1 and 2. In the NCBInr database, peptide 1 matched to "Ts18 variant 5", an 86 amino acids long protein sequence that was published by Kim *et al.* in 2011. Remarkably, in 2003, Hancock *et al.* already reported an "8 kDa diagnostic antigen Ts18 variant 5, partial" protein (GI:19879938), which is highly similar to the "Ts18 variant 5" identified in Table 8.1 (BLASTP E-value 1e-57 with a 99% identity score). In the genome derived database, peptide 1 matched an "8 kDa glycoprotein" sequence which is almost identical to a "Ts18 variant 2" (GI:347546161) also reported by Kim *et al.* (2011) (BLASTP E-value 6e-56 with a 99% identity score). The Ts18 variant proteins are part of the Ts18 clade of the 8 kDa protein family (Hancock *et al.*, 2003) of which several members were found to be abundantly present in *T. solium* metacestode excretion/secretion proteins [Chapters 6 and 7 and Victor *et al.* (2012b, 2013)]. However, a family of genes encoding 8 kDa antigens was identified in both *Echinococcus* and *Taenia* species (Jia *et al.*, 2011), so whether (monoclonal) antibodies are able to differentiate between the closely related 8 kDa proteins remains to be elucidated. Furthermore, since two different genotypes of *T. solium* have already been identified (Ito *et al.*, 2003), it is not unlikely that this might result in differences in the proteome as well. That being said, the "Ts18 variant 5" proteins reported by both Hancock *et al.* and Kim *et al.* came from cysts that were sourced in China. The next result (peptide 2) matches to phosphoenolpyruvate carboxykinase in both databases.

Table 8.1: Peptides ($n = 11$) identified using the spectral comparison method. The sequences were analysed with BLASTP (optimised for queries shorter than 30 residues) against the NCBInr database and the genome derived TsM1_13.12.11.prot.faa protein database (`http://www.taeniasolium.unam.mx/taenia/`) on May 4, 2014. Where applicable, the top result is mentioned as well as the corresponding BLASTP E-value for each database.

Peptide sequence	Spectral count	Top protein match in BLASTP analysis [Organism] Identification	E-value Database
1) EIEYIHNWFFHDDPIGK	11	Ts18 variant 5 [*Taenia solium*] GI:347546167	9e-10 NCBInr
		8 kDa glycoprotein [*Taenia solium*] TsM_000105300	8e-13 genome
2) AINPEAGFFGVAPGTNHK	11	Phosphoenolpyruvate carboxykinase [*Hymenolepis microstoma*] GI:555931973	3e-09 NCBInr
		Phosphoenolpyruvate carboxykinase [*Taenia solium*] TsM_000763700	2e-12 genome
3) GEQGPPGPQGPAGALGPAGS-IGDAGER	14	Collagen alpha-2(I) chain [*Echinococcus granulosus*] GI:576700155	7e-04 NCBInr
		Collagen alpha 2(I) chain [*Taenia solium*] TsM_000807400	2e-08 genome
4) NNYLYETTYFTK	16	Fasciclin-1 [*Taenia solium*] GI:576902497	0.002 NCBInr
		Gynecophoral canal protein [*Taenia solium*] TsM_000655200	5e-07 genome
5) VFLLHSSQGQGSR	13	Collagen type XI alpha 2 [*Echinococcus granulosus*] GI:556518320	5.5 NCBInr
		Collagen type XI alpha 2 [*Taenia solium*] TsM_000632700	3e-07 genome
6) SIIYAAIACGPR	17	Putative vesicle-associated membrane protein [*Echinococcus granulosus*] GI:576700607	0.23 NCBInr
		Vesicle associated membrane protein [*Taenia solium*] TsM_000458800	1e-06 genome
7) SWGDIVSHADK	10	No match in NCBInr	
		N acetylated alpha linked acidic dipeptidase 2 [*Taenia solium*] TsM_000691600	7e-06 genome
8) YIAEERPCGDK	11	Trypsin-like protein [*Taenia solium*] GI:311335041	0.035 NCBInr
		Mastin [*Taenia solium*] TsM_000238300	1e-05 genome
9) VMCVGDDIAWMR	19	Hypothetical protein LOTGIDRAFT_178121 [*Lottia gigantea*] GI:556109577	0.039 NCBInr
		Phosphoenolpyruvate carboxykinase [*Taenia solium*] TsM_000763700	0.007 genome
10) WAVCQVGILLAR	19	No match in NCBInr	
		RhoGAP [*Taenia solium*] TsM_000922100	0.72 genome
11) VVDGLAMDEWGK	10	PglZ domain containing protein [*Sporomusa ovata*] GI:544738427	1.8 NCBInr
		Cytosolic malate dehydrogenase [*Taenia solium*] TsM_000048200	0.025 genome

Although the top NCBInr result in the BLASTP analysis originated from *Hymenolepis microstoma*, the (partial) protein from *T. solium* was also in the list (GI:283466482; E-value 7e-09). The *H. microstoma* protein was published by Tsai *et al.* (2013) (the same paper that also presented the *T. solium* draft genome), while the *T. solium* protein was mentioned by Knapp *et al.* (2011), who used the gene that codes for this protein (amongst others) to study the phylogenetic relationships within the Taeniidae family (*Echinococcus* and *Taenia*) and to trace back the evolutionary histories of these tapeworms. A phosphoenolpyruvate carboxykinase protein was also recognised by serum samples from neurocysticercosis patients (Salazar-Anton and Lindh, 2011), which highlights the immunogenicity and potential diagnostic use of the protein. However, the study did not examine cross-reactivity with other helminth infections. Peptide 9 also matched phosphoenolpyruvate carboxykinase in the genome derived database albeit with a higher E-value and, although the hypothetical protein LOTGIDRAFT_178121 was the top results in the NCBInr database, all other results in the list were also phosphoenolpyruvate carboxykinase. However, the E-value were equally high, making the identified match much less confident. Finally, and as was also mentioned for the 8 kDa proteins above, whether (monoclonal) antibodies will be able to differentiate between the closely related carboxykinases remains to be determined.

Spectral comparison and matching of unique spectra to a six reading frame translation of the *T. solium* genome was the third of the three suggested options. Although this novel methodology is still under active development, the initial results show promise and the method offers a number of advantages over the above mentioned comparison methods. The first one is that this methodology employs all available information (i.e. spectra), no matter how low in abundance they are, before any identification of proteins is attempted. The obvious advantage there is that protein identification through database searching mainly focusses on the abundant proteins (i.e. peptides/proteins identified by many spectra) and peptides identified by very few spectra are often classified as incorrect. A second advantage involves 2D-PAGE limitations. The spectral comparison methodology is perfectly suitable for use in combination with the latest advancements in shotgun proteomics and multidimensional liquid chromatography. These gel-free strategies effectively sidestep the limitations observed with 2D-PAGE (or even 1D-PAGE for that matter). A third advantage is the fact that, in theory, one could make do with only the protein database of the organism of interest i.e. the methodology does not depend on protein identifications for the other organism in the comparison. However, there is definitely value in knowing what the unique spectra of the other organisms match to as well, especially since it could be very interesting to find out which of the identified spectra unique for *T. solium* can not be matched to the *T. hydatigena* proteins database.

As stated above, the method is still under development. Going back to Figure 8.1, it is easy to see that there are a number of steps in the spectral comparison process that offer the possibility of setting thresholds (e.g. a different FDR for host proteins and contaminations, another C parameter in SpecQual and different Spearman's thresholds). Finding the optimal "sweet spot" will require further evaluation and even then, it is possible that re-evaluation will be necessary for every dataset, especially when the data is created on another mass spectrometer. Furthermore, it should be made clear that unique proteins should also be evaluated with *T. asiatica* data. It is possible that unique proteins identified here turn out to be shared by *T. asiatica*. Therefore, further development and evaluation must include *T. asiatica* metacestode ESPs as well.

8.4 Conclusions

In this study, three methods were considered that could all potentially identify proteins that are unique for *T. solium*.
The first method involved LC-MS/MS analysis of the ESPs followed by protein identification through database searching and comparison of the identified proteins. Since there is no complete protein database for *T. hydatigena* (and no curated database for *T. solium* for that matter), protein identification is already compromised and supplementing the protein database with sequences from other *Taenia* would obviously achieve the opposite of what was intended in the experimental set-up i.e. a (large) number of proteins would be shared between both databases, inhibiting any identification of differential proteins. Therefore, this option was immediately discarded. The second option involved 2D-PAGE comparison of both proteomes, followed by LC-MS/MS analysis of the visually unique proteins. Although a workhorse in many proteomics labs, 2D-PAGE remains labour intensive, expensive (both in the financial sense as well as in the use of valuable proteins samples), subject to many issues, and overall, unsuitable for the intended purpose. In the end, the third option was the only one remaining and involved spectral comparison and matching of unique spectra to a six reading frame translation of the *T. solium* genome. The fact that no tools are currently available to facilitate this type of analysis only serves to highlight the novelty of this approach. There is huge potential in this method, not only for this study on *Taenia*. Nonetheless, additional work will be required before this approach can be presented in a ready-to-use format.

CHAPTER 9

General Discussion

9.1 Introduction

The main objective of this thesis was to analyse *Taenia solium* excretion/secretion proteins (ESPs) with the aim of identifying species-specific diagnostic antigens by means of a high throughput proteomics approach involving liquid chromatography and tandem mass spectrometry analysis. Although the generic workflow of a proteomics experiment is rather straightforward, it quickly became apparent that studying an organism that was neither *Homo sapiens* nor one of the common model organisms like *Escherichia coli* or yeast, was going to be quite the opposite. The performed study in Chapter 5 illustrated that the bioinformatics tools that assist with the data analysis (e.g. global false discovery rate estimations) are commonly developed and evaluated using "ideal world" conditions. However, as is often the case, real world conditions are not entirely ideal, making re-evaluation of such tools worthwhile. Furthermore, in Chapters 6 and 7, it was shown that by supplementing an incomplete protein database with proteins sequences from (closely) related helminths or by using expressed sequence tags, it was possible to significantly increase the number of protein identifications for *T. solium*. These two approaches can definitely be combined and should be usable when studying other organisms for which limited information is available. Finally, in Chapter 8, the need for a species-specific antigen detection assay for porcine cysticercosis was addressed by evaluating a number of possible options to identify one or more (novel) diagnostic target(s). Here, however, the database strategies that were used for the previous analyses proved unsuitable when the intended use was a comparative analysis. This fact, combined with the limitations and shortcomings of 2D-PAGE required the development and evaluation of novel tools, which, although still under active development, already show very promising results.

In this final chapter, the encountered shortcomings of current bioinformatics are further discussed, the value of the database strategies in unsequenced organisms and the resulting identifications are discussed in a broader context and the potential road towards a species-specific assay is further explored.

9.2 Proteomics bioinformatics

The false discovery rate (FDR) is a mandatory metric in proteomics literature and is used to determine how many identifications are included in the published list of proteins. Given the importance of this metric, it made sense to establish if and how it would be influenced by incomplete databases. The research described in Chapter 5 and Victor *et al.* (2012a) deals with this particular issue and after evaluation of the empirical evidence, the mixture model-based FDR estimation came out as being a stable, robust and reproducible method, regardless of the way the decoy databases were constructed. However, while preparing that paper, it was remarkable to see how most of the research done on FDR estimations was performed by groups who work mainly on bioinformatics or mass spectrometry related to the well studied model organisms. Very little is written about the behaviour of the current bioinformatic tools in relation to less than ideal conditions.
Arguably, the FDR is of lesser importance when simply trying to identify one or even a few proteins, but it does become more relevant when studying the whole proteome. It should be stressed that the bioinformatics behind LC-MS/MS data analysis are not suitable to be used as most other tools in the lab, i.e. as a simple "means to an end" like PCR or ELISA. The currently available tools still require a great deal of critical user intervention and at least a working understanding of how the algorithms work. Unfortunately, none of the tools generate helpful errors in that regard, making it very tempting to blindly rely on them [the black box concept (Cauer, 1941)].

The fact that (i) LC-MS/MS is indeed more and more becoming a common component in biological studies and (ii) universally accepted and widely available computational tools for validation of (published) data are absent, has prompted the publication of strict guidelines that list the minimal informational requirements for publishing large datasets with protein identifications. Originally, these guidelines were limited to the *Molecular & Cellular Proteomics* journal[1] (Carr *et al.*, 2004), but they are slowly being updated and adopted by other proteomics journals like *Proteomics*[2] and the *Journal of Proteome Research*[3] as well (Bradshaw *et al.*, 2006; Barnouin, 2011). However, an increasing amount of proteomics studies are being submitted to non-proteomics journals, where even basic

[1] http://www.mcponline.org
[2] http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1615-9861
[3] http://pubs.acs.org/journal/jprobs

information is often missing and the peer review process is less likely to pick up on this. It would therefore be very desirable to extend these guidelines to the publication of all proteomics data, regardless of the journal. Going even further, strongly encouraging (or even requiring!) that authors submit all data to a public repository for other researchers to use/review would undoubtedly be beneficial to scientific progress in any field.

9.3 *Taenia solium* protein analysis

Having established the best way to estimate the FDR, the next issue to tackle was how to make an incomplete database as complete as possible. Two methods were explored and described in Chapters 6 and 7 and Victor *et al.* (2012b, 2013). The main conclusion of this work is that proteins from (closely) related organisms and expressed sequence tags (ESTs) can be used to significantly increase the proteome coverage of the protein database. This was evidenced in the first study by the identification of 32 proteins in the ESPs of *T. solium* by matching proteins from other helminths on top of the 27 that were identified by matching to *T. solium* proteins. In the second study, an additional 58 proteins were identified when a library of translated ESTs was combined with BLAST and mapping strategies. A further observation after the EST analysis was that, although the absolute number of protein identification had been increased, none of the new proteins identified a new Gene Ontology (GO) term. The six additional GO terms were identified by proteins homologous to other proteins identified in both studies. A logical conclusion here is that both approaches are largely complementary and therefore, when faced with an incomplete protein database and having only one organism to study (i.e. no comparisons will be made) it is advised to add both ESTs and related protein sequences.
Since the sequence of *T. solium* has recently been published (Tsai *et al.*, 2013), a better approach would now be to derive proteins from the genome by performing a six reading frame translation of the genome or by running gene prediction software like AUGUSTUS (Stanke and Waack, 2003; Stanke and Morgenstern, 2005) and to use these predictions as a protein database. However, both approaches used in this thesis still hold value for the study of other non-sequenced organisms.

Since parasite modulation of the immune system is most likely to be effected through the release of soluble mediators which interact with host immune cells and molecules (Hewitson *et al.*, 2009), it is logical to find an overlap between these mediators and excretion/secretion proteins. For example, cystatins (cysteine protease inhibitors) that inhibit host cysteine proteases from processing and presenting parasite antigens were identified in the *T. solium* ESPs (Dainichi *et al.*, 2001). *Taenia* metacestodes also produce their own cysteine proteinases, which have been identified as a predominant enzyme involved in host immunoglobulin degradation, most likely to serve as a source of nutrients (White

et al., 1997a). Also, since reactive oxygen species production by phagocytes is a primary pathway of immune attack against parasites, it was not surprising to find evidence of the enzymatic antioxidant system of Taeniidae e.g. Cu/Zn-superoxide dismutase, glutathione S-transferase and 2-cys peroxiredoxin (Castellanos-González *et al.*, 2002; Vibanco-Pérez *et al.*, 2002; Plancarte *et al.*, 2004; Torres-Rivera and Landa, 2008; Vaca-Paniagua *et al.*, 2009; Nguyen *et al.*, 2010). A complete explanation of the function and role of all identified proteins would be very difficult, since it is possible that not all proteins have a purpose in immune regulation as some of them may be a result of cyst damage and would normally not be present in the excretion/secretion proteome of a viable metacestode. Furthermore, many proteins have multiple functions! However, it can be agreed upon that many of the proteins secreted by viable cysts serve to direct the host immune response towards a T-helper 2 (Th2) non-inflammatory response in a certain way (Hewitson *et al.*, 2009; Moreau and Chauvin, 2010; Peón *et al.*, 2013; Singh *et al.*, 2013a).

The presence of host proteins like immunoglobulin and albumin in the ESPs may sound slightly odd at first, but simply dismissing them as mere sample contamination would be unwise. Experimental work on *Taenia crassiceps* showed that cysticerci secrete varying amounts of albumin while maintaining a constant albumin concentration in the cyst fluid. This form of albumin cycling may be an evolved mechanism providing cysticerci with a means of osmoregulation (Aldridge *et al.*, 2006). More recently, Navarrete-Perea *et al.* (2014) also confirmed the presence of serum albumin (as well as hemoglobin, haptoglobin and immunoglobulin G) in the vesicular fluid of *T. solium* cysticerci that were dissected from the central nervous system and the skeletal muscle of naturally infected pigs.

Another important consideration is the existence of two distinct *T. solium* genotypes. Based on differences in mitochondrial DNA sequences, Ito *et al.* (2003) have shown that isolates from Asia from one cluster, while isolates from Latin America and Africa formed another. That these two genotypes also represent (slightly) different proteomes is probable, but since all ESP samples used in this thesis came from Peru and Zambia, there was no way of verifying that.

Although it is certain that the work presented in this thesis has identified proteins that had not yet been reported in *T. solium* ESPs, it is important to remain critical, especially towards those proteins that are not predicted to be secreted by SignalP and SecretomeP analysis. Those proteins could be a result of cyst damage and subsequent leakage in the culture medium, although a number of publications have mentioned the presence of such proteins in trematodes and nematodes as well [e.g. *Schistosoma mansoni* (Knudsen *et al.*, 2005), *Schistosoma japonicum* (Liu *et al.*, 2009), *Ancylostoma caninum* (Mulvenna *et al.*, 2009) and *Teladorsagia circumcincta* (Craig *et al.*, 2006a)].

In principle, a more "conservative" list of identified proteins could be constructed by filtering out the non-secretory proteins, but given the fact that (i) protein sequences from NCBInr may be incomplete (truncated) and could therefore lack the signal peptide and (ii) many other studies have also reported "non-secretory" proteins in their parasite ESPs, no such filtering was performed for our analyses. However, for informational purposes, SigP/SecP analyses were added to the tables.
Furthermore, repeating the entire ESP production is not always an option given the sometimes less than ideal local working circumstances.

Since the main objective of this thesis was to analyse *Taenia solium* excretion/secretion proteins with the aim of identifying species-specific diagnostic antigens, it is worthwhile to reflect upon a recent paper in which Diaz-Masmela *et al.* (2013) presented seven antigens that were exclusively recognised by pooled sera from cysticercotic pigs, namely, tropomyosin 2, alpha-1 tubulin, beta-tubulin 2, annexin B1, small heat-shock protein, 14-3-3 protein, and cAMP-dependent protein kinase. A number of these proteins were also identified in the studies described in this thesis, but annexin B1 and cAMP-dependent protein kinase are of particular interest to the authors as they possess highly specific regions that make them suitable candidates for *T. solium*-specific diagnosis.
Although the approach taken by Diaz-Masmela *et al.* is certainly interesting and shows great potential, it would have been helpful if the authors had included more information on the number of samples that was used in each pool. It is therefore possible that the sample size was not large enough to draw conclusions. Also, it would be interesting to see how each individual serum sample reacted to the antigens, i.e. if one of the animals would react to an antigen not recognised by the others, one could wonder whether the seven suggested proteins would get recognised as well had the sample size been larger. Logically, this would include running many more 2D-PAGE gels and given the known issue with 2D-PAGE reproducibility, the decision to pool the samples is justifiable. Nonetheless, knowing that the immunological response between pigs, as measured by IgG antibodies, is never identical and that the protein profiles of the cyst fluid are also variable (mainly due to differences in the glycosylation) (Esquivel-Velázquez *et al.*, 2011a), the logical conclusion would be that further validation is definitely required. Indeed, using more samples to further evaluate the proteins was suggested by the authors in their conclusions, but the inclusion of sera from pigs positive for *Taenia asiatica* would undoubtedly be beneficial as well.

9.4 Towards a *Taenia solium*-specific antigen detection assay

As is also mentioned in Chapter 2 and Chapter 8, diagnosis of porcine cysticercosis would benefit greatly from a species-specific test, not only because that would make epidemiological studies and assessment of control measures a great deal more reliable since it would solve the cross reaction of the current assays with *T. hydatigena* and *T. asiatica*, but also because it could be a major step towards the development of a "pen-side" test that can be used by farmers, butchers and "middlemen" to identify infected animals and subsequently remove them from the food chain. Of course, such a test would ideally be 100% specific and 100% sensitive, very fast, very simple, just as functional in the lab as in the field, cheap, and if possible, not dependant on a cold chain.

Realistically however, such a test will likely never exist. Any assay is always a compromise, but depending on the intended purpose, certain characteristics can be stressed more than others e.g. a very simple and quick lateral flow immunochromatographic assay that is 85% sensitive can be preferred over a laborious and expensive assay that is 99% sensitive when it is being used to identify and remove infected animals from the food chain at the slaughter slabs in the field. Alternatively, infected animals can be identified a few months prior to slaughter and treated on the spot with a single dose of 30 mg/kg oxfendazole. By the time they are slaughtered, the 17 day withdrawal period (due to the presence of drug residues) will have passed (Moreno *et al.*, 2012), the cysts will be dead (except for some in the brain!) (Gonzalez *et al.*, 1996, 1997, 1998) and the meat will have become aesthetically suitable for human consumption (Gonzalez *et al.*, 1998; Sikasunge *et al.*, 2008). Furthermore, there is a protective effect for up to three months after treatment (Gonzalez *et al.*, 2001), so the chance of reinfection after treatment is very limited, especially if combined with vaccination at the time of treatment.
Such a "pen-side" test need not even be limited to porcine cysticercosis. Having a "bed-side" test would also be very useful in a clinical setting, especially where screenings or follow-ups of people are being conducted in relatively simple field hospitals!

9

The above scenario is highly desirable, but there is still a long way to go.
Candidate proteins need to be identified first and to this end, the development of the spectral comparison (highlighted in Chapter 8) will need to be continued and should also incorporate *T. asiatica* ESPs as well as a better way to match the unique spectra to (unique) proteins. To this end, it would be interesting to look into deep transcriptional sequencing (RNA-Seq) techniques to construct libraries for *T. solium*, *T. hydatigena* and *T. asiatica* that cover gene expression across the genome, and then translate these into custom protein databases for identification with tools like customProDB, developed by

Wang and Zhang (2013). RNA-Seq analysis is cheaper than whole genome sequencing and eliminates many of the difficulties encountered when predicting proteins from DNA, since the sequenced mRNAs are already much closer to the translated protein sequence. Furthermore, the efficacy of such an approach has recently been proven by Menschaert *et al.* (2013).

Once specific proteins have been identified, they must be either isolated from the ESP samples by e.g. a number of HPLC separation steps, produced as recombinant proteins or synthetically constructed based on the amino acid sequence of the protein in the database. The best option is of course to use the native proteins, since they have the proper glycosylations as well as the linear and conformational epitopes. However, separating the ESPs until only the proteins of interest remain will likely proved (very) difficult. Moreover, it requires a constant source of ESPs which in turn depends on obtaining viable cysticerci. Alternatively, the production of recombinant proteins is an option worth considering. For example, the GP50 protein, one of the seven diagnostic bands in the EITB assay (Tsang *et al.*, 1989), has been produced as a recombinant protein using the baculovirus expression system in an antigenically active form i.e. correct formation of disulfide bonds (Hancock *et al.*, 2004). Reduction of these disulfide bonds resulted in a loss of antibody reactivity with infection serum samples, indicating the importance of the conformational epitopes. Furthermore, glycosylations also play a role in antigen - antibody recognition (Obregón-Henao *et al.*, 2001), but unfortunately, glycosylation in insect cells is different from glycosylation in mammalian cells and is likely to be different from the glycosylation that occurs in *T. solium* (Bueno *et al.*, 2005). This may result in a reduced sensitivity compared to native proteins. Synthetic peptide production is a further option with proven potential (Scheel *et al.*, 2005), but synthetic peptides generally perform worse than recombinant proteins, most likely because part of the antibody response is directed towards conformational epitopes (Ferrer *et al.*, 2012; Rodriguez *et al.*, 2012).

Since monoclonal antibodies have long been the tool of choice for the development of antigen detecting assays they can be produced for this purpose as well. Mice can be immunised with purified proteins or protein fractions. Since IgG antibodies are less prone to non-specific reactions, are generally more stable and are more easily linked to biotin or horseradish peroxidase than IgM antibodies, immunised mice should be selected based on maximal IgG and minimal IgM titers. An alternative option is the production of nanobodies [the antigen recognising part of a camelid heavy chain antibody (Hamers-Casterman *et al.*, 1993; Muyldermans and Lauwereys, 1999)]. Previous experimental work by Deckers *et al.* (2009) already proved their potential, but unfortunately, no functional assay could be developed. Since the ultimate goal is to make a species-specific assay for *T. solium*, it is recommended to evaluate any possible cross reactions in more detail. To this end, an extra validation step can be included in which the *Taenia* ESPs will be separated

by 1D/2D-PAGE, blotted on nitrocellulose paper and probed with (i) the monoclonal antibodies/nanobodies and (ii) serum samples from pigs with various helminth infections (e.g. *Trichinella* spp., *Ascaris suum*, *Echinococcus* spp. etc.). The purpose of these blots is to identify those proteins that are commonly recognised by various sera and to evaluate if there are any unexpected cross reactions with the proteins originally identified as unique as well as to evaluate any cross reaction between the monoclonal antibodies/nanobodies and other *Taenia* ESPs. The above described steps are schematically presented in Figure 9.1.

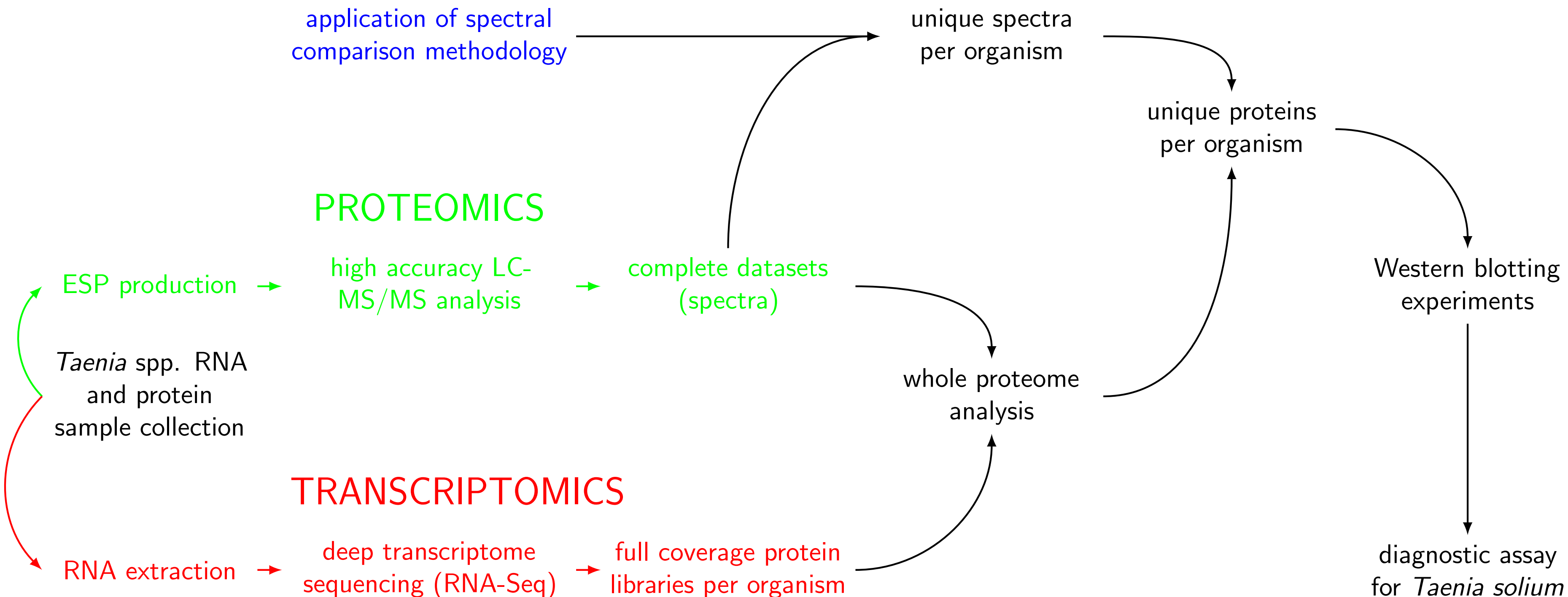


Figure 9.1: Overview of the envisaged road towards a species-specific diagnostic assay for *Taenia solium*. Excretion/secretion proteins (ESPs) from *Taenia solium*, *Taenia hydatigena* and *Taenia asiatica* are analysed with tandem mass spectrometry (MS/MS) to obtain complete spectral datasets. RNA-Seq analyses of all three *Taenia* transcripts and subsequent translation will result in three full coverage protein databases. The MS/MS datasets will be run through the spectral comparison algorithms, and in parallel they will be searched against the translated RNA-Seq databases. It should thus be possible to extract an overview of the three complete proteomes and identify which spectra are unique to each organism. By linking the spectra that are unique to *T. solium* with their identifications from the RNA-Seq database search, it should be possible to determine which proteins (and transcripts) are unique to *T. solium*. These specific proteins will then be used as targets for the production of monoclonal antibodies and/or nanobodies and be further evaluated with Western blot assays. This will ultimately lead to a species-specific diagnostic assay for *Taenia solium*.

9.5 Conclusions and recommendations

The studies in this thesis have (i) demonstrated that mixture model-based FDR estimations can be used when studying organisms whose genome is not yet (completely) known, (ii) presented the first proteomic analysis of the *T. solium* metacestode ESPs using one-dimensional polyacrylamide gel electrophoresis and LC-MS/MS and (iii) illustrated the potential of the novel spectral comparison methodology for the identification of unique *T. solium* proteins, which may, in turn, lead to a *T. solium*-specific diagnostic assay.

A logical recommendation is therefore to continue with the development and evaluation of the spectral comparison pipeline with the aim of developing a new assay that targets unique proteins.

CHAPTER 10

Bibliography

Abbas, A. K., Murphy, K. M. Sher, A., 1996. Functional diversity of helper T lymphocytes. Nature 383 (6603), 787–793.

Abdallah, C., Dumas-Gaudot, E., Renaut, J. Sergeant, K., 2012. Gel-based and gel-free quantitative proteomics approaches at a glance. Int. J. Plant Genomics 2012, 494572.

Aguilar-Díaz, H., Bobes, R. J., Carrero, J. C., Camacho-Carranza, R. *et al.*, 2006. The genome project of *Taenia solium*. Parasitol. Int. 55 Suppl, S127–S130.

Alban, A., David, S. O., Bjorkesten, L., Andersson, C. *et al.*, 2003. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics 3 (1), 36–44.

Aldridge, Jr, J. R., Jennette, M. A. Kuhn, R. E., 2006. Uptake and secretion of host proteins by *Taenia crassiceps* metacestodes. J. Parasitol. 92 (5), 1101–1102.

Allan, J. C., Avila, G., Garcia Noval, J., Flisser, A. Craig, P. S., 1990. Immunodiagnosis of taeniasis by coproantigen detection. Parasitology 101 Pt 3, 473–477.

Allan, J. C., Craig, P. S., Garcia Noval, J., Mencos, F. *et al.*, 1992. Coproantigen detection for immunodiagnosis of echinococcosis and taeniasis in dogs and humans. Parasitology 104 (Pt 2), 347–356.

Allan, J. C., Velasquez-Tohom, M., Fletes, C., Torres-Alvarez, R. *et al.*, 1997. Mass chemotherapy for intestinal *Taenia solium* infection: effect on prevalence in humans and pigs. Trans. R. Soc. Trop. Med. Hyg. 91 (5), 595–598.

Allan, J. C., Velasquez-Tohom, M., Torres-Alvarez, R., Yurrita, P. Garcia-Noval, J., 1996. Field trial of the coproantigen-based diagnosis of *Taenia solium* taeniasis by enzyme-linked immunosorbent assay. Am. J. Trop. Med. Hyg. 54 (4), 352–356.

Almeida, C. R., Stoco, P. H., Wagner, G., Sincero, T. C. M. *et al.*, 2009. Transcriptome analysis of *Taenia solium* cysticerci using Open Reading Frame ESTs (ORESTES). Parasites Vectors 2 (1), 35.

Anderson, L. Seilhamer, J., 1997. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 18 (3–4), 533–537.

Andrews, P., Thomas, H., Pohlke, R. Seubert, J., 1983. Praziquantel. Med. Res. Rev. 3 (2), 147–200.

Assana, E., Kyngdon, C. T., Gauci, C. G., Geerts, S. *et al.*, 2010. Elimination of *Taenia solium* transmission to pigs in a field trial of the TSOL18 vaccine in Cameroon. Int. J. Parasitol. 40 (5), 515–519.

Assana, E., Lightowlers, M. W., Zoli, A. P. Geerts, S., 2013. *Taenia solium* taeniosis/cysticercosis in Africa: risk factors, epidemiology and prospects for control using vaccination. Vet. Parasitol. 195 (1–2), 14–23.

Avila, G., Benitez, M., Aguilar-Vega, L. Flisser, A., 2003. Kinetics of *Taenia solium* antibodies and antigens in experimental taeniosis. Parasitol. Res. 89 (4), 284–289.

Aziz, A., Zhang, W., Li, J., Loukas, A. *et al.*, 2011. Proteomic characterisation of *Echinococcus granulosus* hydatid cyst fluid from sheep, cattle and humans. J. Proteomics 74 (9), 1560–1572.

Baig, S., Damian, R. T., Morales-Montor, J., Olecki, P. *et al.*, 2005. Characterization of excretory/secretory endopeptidase and metallo-aminopeptidases from *Taenia crassiceps* metacestodes. J. Parasitol. 91 (5), 983–987.

Barcelos, I. S. C., Moura, L. P., Costa, V. P., Ferreira, M. S. Costa-Cruz, J. M., 2007. *Taenia solium* metacestode immunodominant peptides recognized by IgG antibodies in cerebrospinal fluid and serum paired samples from patients with active and inactive neurocysticercosis. Mem. Inst. Oswaldo Cruz 102 (6), 713–717.

Barnouin, K., 2011. Guidelines for experimental design and data analysis of proteomic mass spectrometry-based experiments. Amino Acids 40 (2), 259–260.

Barsnes, H., Vaudel, M., Colaert, N., Helsens, K. *et al.*, 2011. compomics-utilities: an open-source Java library for computational proteomics. BMC Bioinformatics 12, 70.

Bastos, A. L., Marchiori, E., Gasparetto, E. L., Andrade, B. H. *et al.*, 2007. Pulmonary and cardiac cysticercosis: helical CT findings. Br. J. Radiol. 80 (951), e58–e60.

Beckett, P., 2012. The basics of 2d dige. Methods Mol. Biol. 854, 9–19.

Ben Nouir, N., Gianinazzi, C., Gorcii, M., Müller, N. *et al.*, 2009. Isolation and molecular characterization of recombinant *Echinococcus granulosus* P29 protein (recP29) and its assessment for the post-surgical serological follow-up of human cystic echinococcosis in young patients. Trans. R. Soc. Trop. Med. Hyg. 103 (4), 355–364.

Bendtsen, J. D., Jensen, L. J., Blom, N., Von Heijne, G. Brunak, S., 2004a. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng., Des. Sel. 17 (4), 349–356.

Bendtsen, J. D., Nielsen, H., von Heijne, G. Brunak, S., 2004b. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340 (4), 783–795.

Benjamini, Y. Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57 (1), 289–300.

Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A. *et al.*, 2009. The genome of the blood fluke *Schistosoma mansoni*. Nature 460 (7253), 352–358.

Bianco, L., Mead, J. A. Bessant, C., 2009. Comparison of novel decoy database designs for optimizing protein identification searches using ABRF sPRG2006 standard MS/MS data sets. J. Proteome Res. 8 (4), 1782–1791.

Bickerstaff, E. R., Cloake, P. C. P., Hughes, B. Smith, W. T., 1952. The racemose form of cerebral cysticercosis. Brain 75 (1), 1–18.

Biemann, K., 1990. Appendix 5. nomenclature for peptide fragment ions (positive ions). Methods Enzymol. 193, 886–887.

Biondi, G. F., Mucciolo, R. G., Nunes, C. M. Richtzenhain, L. J., 1996. Immunodiagnosis of swine cysticercosis by indirect ELISA employing a heterologous antigen from *Taenia crassiceps* metacestode. Vet. Parasitol. 64 (4), 261–266.

Blattner, F. R., Plunkett, 3rd, G., Bloch, C. A., Perna, N. T. *et al.*, 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277 (5331), 1453–1462.

Bobes, R. J., Hernández, M., Márquez, C., Fragoso, G. *et al.*, 2006. Subarachnoidal and intraventricular human neurocysticercosis: application of an antigen detection assay for the diagnosis and follow-up. Trop. Med. Int. Health 11 (6), 943–950.

Borloo, J., De Graef, J., Peelaers, I., Nguyen, D. L. *et al.*, 2013. In-depth proteomic and glycomic analysis of the adult-stage *Cooperia oncophora* excretome/secretome. J. Proteome Res. 12 (9), 3900–3911.

Botero, D., Uribe, C. S., Sanchez, J. L., Alzate, T. *et al.*, 1993. Short course albendazole treatment for neurocysticercosis in Columbia. Trans. R. Soc. Trop. Med. Hyg. 87 (5), 576–577.

Bradshaw, R. A., Burlingame, A. L., Carr, S. Aebersold, R., 2006. Reporting protein identification data: the next generation of guidelines. Mol. Cell. Proteomics 5 (5), 787–788.

Brandt, J., Geerts, S., De Deken, R., Kumar, V. *et al.*, 1992. A monoclonal antibody-based ELISA for the detection of circulating excretory-secretory antigens in *Taenia saginata* cysticercosis. Int. J. Parasitol. 22, 471–477.

Brehm, K., 2010. The role of evolutionarily conserved signalling systems in *Echinococcus multilocularis* development and host-parasite interaction. Med. Microbiol. Immunol. 199 (3), 247–259.

Bueno, E. C., Scheel, C. M., Vaz, A. J., Machado, L. R. *et al.*, 2005. Application of synthetic 8-kD and recombinant GP50 antigens in the diagnosis of neurocysticercosis by enzyme-linked immunosorbent assay. Am. J. Trop. Med. Hyg. 72 (3), 278–283.

Bursey, C. C., McKenzie, J. A. Burt, M. D., 1980. Polyacrylamide gel electrophoresis in the differentiation of *Taenia* (cestoda) by total protein. Int. J. Parasitol. 10 (3), 167–174.

Bustos, J. A., Rodriguez, S., Jimenez, J. A., Moyano, L. M. *et al.*, 2012. Detection of *Taenia solium* taeniasis coproantigen is an early indicator of treatment failure for taeniasis. Clin. Vaccine Immunol. 19 (4), 570–573.

Campbell, G., García, H. H., Nakao, M., Ito, A. Craig, P. S., 2006. Genetic variation in *Taenia solium*. Parasitol. Int. 55 Suppl, S121–S126.

Campos, A., Bernard, P., Fauconnier, A., Landa, A. *et al.*, 1990. Cloning and sequencing of two actin genes from *Taenia solium* (Cestoda). Mol. Biochem. Parasitol. 40 (1), 87–93.

Carabin, H., Budke, C. M., Cowan, L. D., Willingham, A. L. Torgerson, P. R., 2005. Methods for assessing the burden of parasitic zoonoses: echinococcosis and cysticercosis. Trends Parasitol. 21 (7), 327–333.

Carpio, A., Kelvin, E. A., Bagiella, E., Leslie, D. *et al.*, 2008. Effects of albendazole treatment on neurocysticercosis: a randomised controlled trial. J. Neurol. Neurosurg. Psychiatry 79 (9), 1050–1055.

Carr, S., Aebersold, R., Baldwin, M., Burlingame, A. *et al.*, 2004. The need for guidelines in publication of peptide and protein identification data - Working group on publication guidelines for peptide and protein identification data. Mol. Cell. Proteomics 3 (6), 531–533.

Carrette, O., Burkhard, P. R., Sanchez, J.-C. Hochstrasser, D. F., 2006. State-of-the-art two-dimensional gel electrophoresis: a key tool of proteomics research. Nat. Protoc. 1 (2), 812–823.

Castellanos-González, A., Jiménez, L. Landa, A., 2002. Cloning, production and characterisation of a recombinant Cu/Zn superoxide dismutase from *Taenia solium.* Int. J. Parasitol. 32 (9), 1175–1182.

Cauer, W., 1941. Theorie der linearen Wechselstromschaltungen. Akademische Verlagsgesellschaft Becker & Erler, cop., Leipzig.

C. elegans Sequencing Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282 (5396), 2012–2018.

Chambers, M. C., Maclean, B., Burke, R., Amodei, D. *et al.*, 2012. A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30 (10), 918–920.

Chang, G. Y. Keane, J. R., 2001. Visual loss in cysticercosis: analysis of 23 patients. Neurology 57 (3), 545–548.

Choi, H., Fermin, D. Nesvizhskii, A. I., 2008a. Significance analysis of spectral count data in label-free shotgun proteomics. Mol. Cell. Proteomics 7 (12), 2373–2385.

Choi, H., Ghosh, D. Nesvizhskii, A. I., 2008b. Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling. J. Proteome Res. 7 (1), 286–292.

Choi, H. Nesvizhskii, A. I., 2008a. False discovery rates and related statistical concepts in mass spectrometry-based proteomics. J. Proteome Res. 7 (1), 47–50.

Choi, H. Nesvizhskii, A. I., 2008b. Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. J. Proteome Res. 7 (1), 254–265.

Chung, J.-Y., Bahk, Y. Y., Huh, S., Kang, S.-Y. *et al.*, 1999. A recombinant 10-kDa protein of *Taenia solium* metacestodes specific to active neurocysticercosis. J. Infect. Dis. 180 (4), 1307–1315.

Chung, J.-Y., Yun, D.-H., Eom, K. S., Kang, S.-Y. *et al.*, 2002. *Taenia solium*: identification of specific antibody binding regions of metacestode 10-kDa protein. Exp. Parasitol. 100 (2), 87–94.

Colli, B. O., Valença, M. M., Carlotti, Jr, C. G., Machado, H. R. Assirati, Jr, J. A., 2002. Spinal cord cysticercosis: neurosurgical aspects. Neurosurg. Focus 12 (6), e9.

Comisarow, M. B. Marshall, A. G., 1974. Fourier transform ion cyclotron resonance spectroscopy. Chem. Phys. Lett. 25 (2), 282–283.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J. *et al.*, 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (18), 3674–3676.

Conlan, J. V., Vongxay, K., Khamlome, B., Dorny, P. *et al.*, 2012. A cross-sectional study of *Taenia solium* in a multiple taeniid-endemic region reveals competition may be protective. Am. J. Trop. Med. Hyg. 87 (2), 281–291.

Correa, D., Sandoval, M. A., Harrison, L. J. S., Parkhouse, R. M. E. *et al.*, 1989. Human neurocysticercosis: comparison of enzyme immunoassay capture techniques based on monoclonal and polyclonal antibodies for the detection of parasite products in cerebrospinal fluid. Trans. R. Soc. Trop. Med. Hyg. 83 (6), 814–816.

Coyle, C. M. Tanowitz, H. B., 2009. Diagnosis and treatment of neurocysticercosis. Interdiscip. Perspect. Infect. Dis. 2009, 180742.

Craig, H., Wastling, J. M. Knox, D. P., 2006a. A preliminary proteomic survey of the *in vitro* excretory/secretory products of fourth-stage larval and adult *Teladorsagia circumcincta*. Parasitology 132 (Pt 4), 535–543.

Craig, R. Beavis, R. C., 2004. TANDEM: Matching proteins with tandem mass spectra. Bioinformatics 20 (9), 1466–1467.

Craig, R., Cortens, J. C., Fenyo, D. Beavis, R. C., 2006b. Using annotated peptide mass spectrum libraries for protein identification. J. Proteome Res. 5 (8), 1843–1849.

Dainichi, T., Maekawa, Y., Ishii, K., Zhang, T. *et al.*, 2001. Nippocystatin, a cysteine protease inhibitor from *Nippostrongylus brasiliensis*, inhibits antigen processing and modulates antigen-specific immune response. Infect. Immun. 69 (12), 7380–7386.

de Aluja, A. S., Herrera, G. S., Hernández, M., Plancarte, A. *et al.*, 2011. Limits of the therapeutic properties of synthetic S3Pvac anti-cysticercosis vaccine. Vet. Parasitol. 177 (1–2), 90–96.

de Aluja, A. S., Villalobos, A. N., Plancarte, A., Rodarte, L. F. *et al.*, 1996. Experimental *Taenia solium* cysticercosis in pigs: characteristics of the infection and antibody response. Vet. Parasitol. 61 (1–2), 49–59.

de Aluja, A. S., Villalobos, N. M., Nava, G., Toledo, A. *et al.*, 2005. Therapeutic capacity of the synthetic peptide-based vaccine against *Taenia solium* cysticercosis in pigs. Vaccine 23 (31), 4062–4069.

de Bruin, J. S., Deelder, A. M. Palmblad, M., 2012. Scientific workflow management in proteomics. Mol. Cell. Proteomics 11 (7), M111.010595.

de Hoffmann, E. Stroobant, V., 2007. Mass Spectrometry: Principles and Applications, Third Edition. John Wiley & Sons.

Deckers, N. Dorny, P., 2010. Immunodiagnosis of *Taenia solium* taeniosis/cysticercosis. Trends Parasitol. 26 (3), 137–144.

Deckers, N., Saerens, D., Kanobana, K., Conrath, K. *et al.*, 2009. Nanobodies, a promising tool for species-specific diagnosis of *Taenia solium* cysticercosis. Int. J. Parasitol. 39 (5), 625–633.

Del Brutto, O. H., 2012. Diagnostic criteria for neurocysticercosis, revisited. Pathog. Glob. Health 106 (5), 299–304.

Del Brutto, O. H., Nash, T. E. García, H. H., 2012. Cysticerci-related single parenchymal brain enhancing lesions in non-endemic countries. J. Neurol. Sci. 319 (1–2), 32–36.

Del Brutto, O. H., Rajshekhar, V., White, Jr, A. C., Tsang, V. C. *et al.*, 2001. Proposed diagnostic criteria for neurocysticercosis. Neurology 57 (2), 177–183.

Del Brutto, O. H., Roos, K. L., Coffey, C. S. García, H. H., 2006. Meta-analysis: Cysticidal drugs for neurocysticercosis: albendazole and praziquantel. Ann. Intern. Med. 145 (1), 43–51.

Del Brutto, O. H., Santibañez, R., Noboa, C. A., Aguirre, R. *et al.*, 1992a. Epilepsy due to neurocysticercosis: analysis of 203 patients. Neurology 42 (2), 389–392.

Del Brutto, O. H., Sotelo, J., Aguirre, R., Díaz-Calderón, E. Alarcón, T. A., 1992b. Albendazole therapy for giant subarachnoid cysticerci. Arch. Neurol. 49 (5), 535–538.

Dempster, A. P., Laird, N. M. Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1), 1–38.

Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T. *et al.*, 2010. A guided tour of the Trans-Proteomic Pipeline. Proteomics 10 (6), 1150–1159.

Devleesschauwer, B., Ale, A., Torgerson, P., Praet, N. *et al.*, 2014. The burden of parasitic zoonoses in Nepal: a systematic review. PLoS Neglected Trop. Dis. 8 (1), e2634.

Devleesschauwer, B., Aryal, A., Tharmalingam, J., Joshi, D. D. *et al.*, 2013. Complexities in using sentinel pigs to study *Taenia solium* transmission dynamics under field conditions. Vet. Parasitol. 193 (1–3), 172–178.

Diaz-Masmela, Y., Fragoso, G., Ambrosio, J. R., Mendoza-Hernández, G. *et al.*, 2013. Immunodiagnosis of porcine cysticercosis: Identification of candidate antigens through immunoproteomics. Vet. J. 198 (3), 656–660.

Ding, Y., Choi, H. Nesvizhskii, A. I., 2008. Adaptive discriminant function analysis and reranking of MS/MS database search results for improved peptide identification in shotgun proteomics. J. Proteome Res. 7 (11), 4878–4889.

Ditzel, J. Schwartz, M., 1967. Worm cure without tears. The effect of niclosamide on taeniasis saginata in man. Acta Med. Scand. 182 (5), 663–664.

Dorny, P., Brandt, J. Geerts, S., 2004a. Immunodiagnostic approaches for detecting *Taenia solium*. Trends Parasitol. 20 (6), 259–260.

Dorny, P., Brandt, J., Zoli, A. Geerts, S., 2003. Immunodiagnostic tools for human and porcine cysticercosis. Acta Trop. 87 (1), 79–86.

Dorny, P., Phiri, I. K., Vercruysse, J., Gabriël, S. *et al.*, 2004b. A Bayesian approach for estimating values for prevalence and diagnostic test characteristics of porcine cysticercosis. Int. J. Parasitol. 34 (5), 569–576.

Dorny, P., Vercammen, F., Brandt, J., Vansteenkiste, W. *et al.*, 2000. Sero-epidemiological study of *Taenia saginata* cysticercosis in Belgian cattle. Vet. Parasitol. 88 (1–2), 43–49.

D'Souza, P. E. Hafeez, M., 1999. Detection of *Taenia solium* in pigs by ELISA with an excretory-secretory antigen. Vet. Res. Commun. 23, 293–298.

Efron, B., Tibshirani, R., Storey, J. D. Tusher, V., 2001. Empirical bayes analysis of a microarray experiment. J. Am. Stat. Assoc. 96 (456), 1151–1160.

Elias, J. E. Gygi, S. P., 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4 (3), 207–214.

Elias, J. E. Gygi, S. P., 2010. Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol. Biol. 604, 55–71.

Eng, J. K., McCormack, A. L. Yates, 3rd, J. R., 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5 (11), 976–989.

Engvall, E. Perlmann, P., 1971. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. Immunochemistry 8 (9), 871–874.

Eom, K. S., 2006. What is Asian *Taenia*? Parasitol. Int. 55 Suppl, S137–S141.

Eom, K. S., Jeon, H.-K. Rim, H.-J., 2009. Geographical distribution of *Taenia asiatica* and related species. Korean J. Parasitol. 47 Suppl, S115–S124.

Eom, K. S. Rim, H. J., 1993. Morphologic descriptions of *Taenia asiatica* sp. Korean J. Parasitol. 31 (1), 1–6.

Escobar, A., 1983. The pathology of neurocysticercosis. In: Cysticercosis of the Central Nervous System. Charles C. Thomas, Springfield, pp. 27–54.

Esquivel-Velázquez, M., Larralde, C., Morales, J. Ostoa-Saloma, P., 2011a. Protein and antigen diversity in the vesicular fluid of *Taenia solium* cysticerci dissected from naturally infected pigs. Int. J. Biol. Sci. 7 (9), 1287–1297.

Esquivel-Velázquez, M., Ostoa-Saloma, P., Morales-Montor, J., Hernández-Bello, R. Larralde, C., 2011b. Immunodiagnosis of neurocysticercosis: ways to focus on the challenge. J. Biomed. Biotechnol. 2011, 516042.

Feng, J., Naiman, D. Q. Cooper, B., 2007. Probability model for assessing proteins assembled from peptide sequences inferred from tandem mass spectrometry data. Anal. Chem. 79 (10), 3901–3911.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. Whitehouse, C. M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. Science 246 (4926), 64–71.

Fenyö, D. Beavis, R. C., 2003. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal. Chem. 75 (4), 768–774.

Ferrer, E., Bonay, P., Foster-Cuevas, M., González, L. *et al.*, 2007. Molecular cloning and characterisation of Ts8B1, Ts8B2 and Ts8B3, three new members of the *Taenia solium* metacestode 8 kDa diagnostic antigen family. Mol. Biochem. Parasitol. 152 (1), 90–100.

Ferrer, E., Cortéz, M. M., Cabrera, Z., Rojas, G. *et al.*, 2005a. Oncospheral peptide-based ELISAs as potential seroepidemiological tools for *Taenia solium* cysticercosis/neurocysticercosis in Venezuela. Trans. R. Soc. Trop. Med. Hyg. 99 (8), 568–576.

Ferrer, E., González, L. M., Foster-Cuevas, M., Cortéz, M. M. *et al.*, 2005b. *Taenia solium*: characterization of a small heat shock protein (Tsol-sHSP35.6) and its possible relevance to the diagnosis and pathogenesis of neurocysticercosis. Exp. Parasitol. 110 (1), 1–11.

Ferrer, E., Sánchez, J., Milano, A., Alvarez, S. *et al.*, 2012. Diagnostic epitope variability within *Taenia solium* 8 kDa antigen family: implications for cysticercosis immunodetection. Exp. Parasitol. 130 (1), 78–85.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A. *et al.*, 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269 (5223), 496–512.

Flikka, K., Martens, L., Vandekerckhove, J., Gevaert, K. Eidhammer, I., 2006. Improving the reliability and throughput of mass spectrometry-based proteomics by spectrum quality filtering. Proteomics 6 (7), 2086–2094.

Flisser, A., 1994. Taeniasis and cysticercosis due to *Taenia solium*. Prog. Clin. Parasitol. 4, 77–116.

Flisser, A., 2013. State of the art of *Taenia solium* as compared to *Taenia asiatica*. Korean J. Parasitol. 51 (1), 43–49.

Flisser, A., Gonzalez, D., Shkurovich, M., Madrazo, I. *et al.*, 1990. Praziquantel treatment of porcine brain and muscle *Taenia solium* cysticercosis. 1. radiological, physiological and histopathological studies. Parasitol. Res. 76 (3), 263–269.

Flisser, A., Madrazo, I., Plancarte, A., Schantz, P. *et al.*, 1993. Neurological symptoms in occult neurocysticercosis after single taeniacidal dose of praziquantel. Lancet 342 (8873), 748.

Flisser, A., Pérez-Montfort, R. Larralde, C., 1979. The immunology of human and animal cysticercosis: a review. Bull. W. H. O. 57 (5), 839–856.

Flisser, A., Sarti, E., Lightowlers, M. W. Schantz, P. M., 2003. Neurocysticercosis: regional status, epidemiology, impact and control measures in the Americas. Acta Trop. 87 (1), 43–51.

Flisser, A., Viniegra, A. E., Aguilar-Vega, L., Garza-Rodriguez, A. *et al.*, 2004. Portrait of human tapeworms. J. Parasitol. 90 (4), 914–916.

Frank, A. Pevzner, P., 2005. PepNovo: *de novo* peptide sequencing via probabilistic network modeling. Anal. Chem. 77 (4), 964–973.

Frewen, B. E., Merrihew, G. E., Wu, C. C., Noble, W. S. MacCoss, M. J., 2006. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. Anal. Chem. 78 (16), 5678–5684.

Gabriël, S., Blocher, J., Dorny, P., Abatih, E. N. *et al.*, 2012. Added value of antigen ELISA in the diagnosis of neurocysticercosis in resource poor settings. PLoS Neglected Trop. Dis. 6 (10), e1851.

Gao, Y.-J., Yan, H.-L., Ding, F.-X., Lu, Y.-M. Sun, S.-H., 2007. Annexin B1 at the host-parasite interface of the *Taenia solium* cysticercus: secreted and associated with inflammatory reaction. Acta Trop. 101 (3), 192–199.

García, H. H., Del Brutto, O. H. Cysticercosis Working Group in Peru, 2005. Neurocysticercosis: updated concepts about an old disease. Lancet Neurol. 4 (10), 653–661.

García, H. H., Dorny, P., Castillo, Y., Pretell, E. J. *et al.*, 2010. Circulating antigen levels follow post-treatment evolution of subarachnoid neurocysticercosis. J. Neuroparasitol. 1, 1–3.

García, H. H., Evans, C. A. W., Nash, T. E., Takayanagui, O. M. *et al.*, 2002. Current consensus guidelines for treatment of neurocysticercosis. Clin. Micr 15 (4), 747–756.

García, H. H., Gilman, R. H., Catacora, M., Verastegui, M. *et al.*, 1997. Serologic evolution of neurocysticercosis patients after antiparasitic therapy. cysticercosis working group in peru. J. Infect. Dis. 175 (2), 486–489.

García, H. H., Gilman, R. H., Gonzalez, A. E., Pacheco, R. *et al.*, 1999. Human and porcine *Taenia solium* infection in a village in the highlands of Cusco, Peru. The Cysticercosis Working Group in Peru. Acta Trop. 73 (1), 31–36.

García, H. H., Gilman, R. H., Gonzalez, A. E., Verastegui, M. *et al.*, 2003a. Hyperendemic human and porcine *Taenia solium* infection in Perú. Am. J. Trop. Med. Hyg. 68 (3), 268–275.

García, H. H., Gonzalez, A. E., Del Brutto, O. H., Tsang, V. C. *et al.*, 2007. Strategies for the elimination of taeniasis/cysticercosis. J. Neurol. Sci. 262 (1–2), 153–157.

García, H. H., Gonzalez, A. E., Evans, C. A. W., Gilman, R. H. Cysticercosis Working Group in Peru, 2003b. *Taenia solium* cysticercosis. Lancet 362 (9383), 547–556.

García, H. H., Gonzalez, A. E., Gilman, R. H., Moulton, L. H. *et al.*, 2006. Combined human and porcine mass chemotherapy for the control of *T. solium*. Am. J. Trop. Med. Hyg. 74 (5), 850–855.

García, H. H., Gonzalez, A. E., Gilman, R. H., Palacios, L. G. *et al.*, 2001. Short report: transient antibody response in *Taenia solium* infection in field conditions – a major contributor to high seroprevalence. Am. J. Trop. Med. Hyg. 65 (1), 31–32.

García, H. H., Herrera, G., Gilman, R. H., Tsang, V. C. *et al.*, 1994. Discrepancies between cerebral computed tomography and western blot in the diagnosis of neurocysticercosis. Am. J. Trop. Med. Hyg. 50 (2), 152–157.

García, H. H., Parkhouse, R. M. E., Gilman, R. H., Montenegro, T. *et al.*, 2000. Serum antigen detection in the diagnosis, treatment, and follow-up of neurocysticercosis patients. Trans. R. Soc. Trop. Med. Hyg. 94 (6), 673–676.

García, H. H., Rodriguez, S., Gilman, R. H., Gonzalez, A. E. *et al.*, 2012. Neurocysticercosis: is serology useful in the absence of brain imaging? Trop. Med. Int. Health 17 (8), 1014–1018.

Gasser, R. B. Chilton, N. B., 1995. Characterisation of taeniid cestode species by PCR-RFLP of ITS2 ribosomal DNA. Acta Trop. 59 (1), 31–40.

Gasser, R. B., Zhu, X. Woods, W., 1999. Genotyping *Taenia* tapeworms by single-strand conformation polymorphism of mitochondrial DNA. Electrophoresis 20 (14), 2834–2837.

Gavidia, C. M., Verastegui, M. R., García, H. H., Lopez-Urbina, T. *et al.*, 2013. Relationship between serum antibodies and *Taenia solium* larvae burden in pigs raised in field conditions. PLoS Neglected Trop. Dis. 7 (5), e2192.

Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L. *et al.*, 2004. Open mass spectrometry search algorithm. J. Proteome Res. 3 (5), 958–964.

Geysen, D., Kanobana, K., Victor, B., Rodriguez-Hidalgo, R. *et al.*, 2007. Validation of meat inspection results for *Taenia saginata* cysticercosis by PCR-restriction fragment length polymorphism. J. Food Prot. 70 (1), 236–240.

Gilman, R. H., Gonzalez, A. E., Llanos-Zavalaga, F., Tsang, V. C. W. *et al.*, 2012. Prevention and control of *Taenia solium* taeniasis/cysticercosis in Peru. Pathog. Global Health 106 (5), 312–318.

Gilmore, J. M. Washburn, M. P., 2010. Advances in shotgun proteomics and the analysis of membrane proteomes. J. Proteomics 73 (11), 2078–2091.

Giri, S. Parija, S. C., 2012. A review on diagnostic and preventive aspects of cystic echinococcosis and human cysticercosis. Trop. Parasitol. 2 (2), 99–108.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W. *et al.*, 1996. Life with 6000 genes. Science 274 (5287), 546, 563–567.

Gomes, A. B., Soares, K. A., Bueno, E. C., Espíndola, N. M. *et al.*, 2007. Comparative evaluation of different immunoassays for the detection of *Taenia solium* cysticercosis in swine with low parasite burden. Mem. Inst. Oswaldo Cruz 102 (6), 725–731.

Gonzalez, A. E., 1997. Evaluation of a control programme for *Taenia solium* cysticercosis targetting human and porcine health. Ph.D. thesis, University of Reading, UK.

Gonzalez, A. E., Cama, V., Gilman, R. H., Tsang, V. C. *et al.*, 1990. Prevalence and comparison of serologic assays, necropsy, and tongue examination for the diagnosis of porcine cysticercosis in Peru. Am. J. Trop. Med. Hyg. 43 (2), 194–149.

Gonzalez, A. E., Falcon, N., Gavidia, C., García, H. H. *et al.*, 1997. Treatment of porcine cysticercosis with oxfendazole: a dose-response trial. Vet. Rec. 141 (16), 420–422.

Gonzalez, A. E., Falcon, N., Gavidia, C. M., García, H. H. *et al.*, 1998. Time-response curve of oxfendazole in the treatment of swine cysticercosis. Am. J. Trop. Med. Hyg. 59 (5), 832–836.

Gonzalez, A. E., García, H. H., Gilman, R. H., Gavidia, C. M. *et al.*, 1996. Effective, single-dose treatment or porcine cysticercosis with oxfendazole. Am. J. Trop. Med. Hyg. 54 (4), 391–394.

Gonzalez, A. E., García, H. H., Gilman, R. H., Lopez, M. T. *et al.*, 1995. Treatment of porcine cysticercosis with albendazole. Am. J. Trop. Med. Hyg. 53 (5), 571–574.

Gonzalez, A. E., García, H. H., Gilman, R. H. Tsang, V. C., 2003. Control of *Taenia solium.* Acta Trop. 87 (1), 103–109.

Gonzalez, A. E., Gauci, C. G., Barber, D., Gilman, R. H. *et al.*, 2005a. Vaccination of pigs to control human neurocysticercosis. Am. J. Trop. Med. Hyg. 72 (6), 837–839.

Gonzalez, A. E., Gavidia, C. M., Falcon, N., Bernal, T. *et al.*, 2001. Protection of pigs with cysticercosis from further infections after treatment with oxfendazole. Am. J. Trop. Med. Hyg. 65 (1), 15–18.

Gonzalez, A. E., López-Urbina, T., Tsang, B. Y., Gavidia, C. M. *et al.*, 2005b. Short report: secondary transmission in porcine cysticercosis: description and their potential implications for control sustainability. Am. J. Trop. Med. Hyg. 73 (3), 501–503.

Gonzalez, A. E., Verastegui, M., Noh, J. C., Gavidia, C. *et al.*, 1999. Persistence of passively transferred antibodies in porcine *Taenia solium* cysticercosis. Cysticercosis Working Group in Peru. Vet. Parasitol. 86 (2), 113–138.

González, G., Spinelli, P., Lorenzo, C., Hellman, U. *et al.*, 2000a. Molecular characterization of P-29, a metacestode-specific component of *Echinococcus granulosus* which is immunologically related to, but distinct from, antigen 5. Mol. Biochem. Parasitol. 105 (2), 177–184.

González, L., Montero, E., Morakote, N., Puente, S. *et al.*, 2004. Differential diagnosis of *Taenia saginata* and *Taenia saginata asiatica* taeniasis through PCR. Diagn. Microbiol. Infect. Dis. 49 (3), 183–188.

González, L. M., Ferrer, E., Spickett, A., Michael, L. M. *et al.*, 2007. The *Taenia saginata* homologue of the major surface antigen of *Echinococcus* spp. is immunogenic and 97% identical to its *Taenia solium* homologue. Parasitol. Res. 101 (6), 1541–1549.

González, L. M., Montero, E., Harrison, L. J., Parkhouse, R. M. Gárate, T., 2000b. Differential diagnosis of *Taenia saginata* and *Taenia solium* infection by PCR. J. Clin. Microbiol. 38 (2), 737–744.

González, L. M., Montero, E., Sciutto, E., Harrison, L. J. S. *et al.*, 2002. Differential diagnosis of *Taenia saginata* and *Taenia solium* infections: from DNA probes to polymerase chain reaction. Trans. R. Soc. Trop. Med. Hyg. 96 Suppl 1, S243–S250.

Görg, A., Weiss, W. Dunn, M. J., 2004. Current two-dimensional electrophoresis technology for proteomics. Proteomics 4 (12), 3665–3685.

Greene, R. M., Hancock, K., Wilkins, P. P. Tsang, V. C., 2000. *Taenia solium*: molecular cloning and serologic evaluation of 14- and 18-kDa related, diagnostic antigens. J. Parasitol. 86 (5), 1001–1007.

Guezala, M.-C., Rodrìguez, S., Zamora, H., García, H. H. *et al.*, 2009. Development of a species-specific coproantigen ELISA for human *Taenia solium* taeniasis. Am. J. Trop. Med. Hyg. 81 (3), 433–437.

Guillou, F., Roger, E., Moné, Y., Rognon, A. *et al.*, 2007. Excretory-secretory proteome of larval *Schistosoma mansoni* and *Echinostoma caproni*, two parasites of *Biomphalaria glabrata*. Mol. Biochem. Parasitol. 155 (1), 45–56.

Guo, Y.-J., Sun, S.-H., Zhang, Y., Chen, Z.-H. *et al.*, 2004. Protection of pigs against *Taenia solium* cysticercosis using recombinant antigen or in combination with DNA vaccine. Vaccine 22 (29–30), 3841–3847.

Gupta, N. Pevzner, P. A., 2009. False discovery rates of protein identifications: a strike against the two-peptide rule. J. Proteome Res. 8 (9), 4173–4181.

Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y. Aebersold, R., 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9390–9395.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, 1999a. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17 (10), 994–999.

Gygi, S. P., Rochon, Y., Franza, B. R. Aebersold, R., 1999b. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19 (3), 1720–1730.

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G. *et al.*, 1993. Naturally occurring antibodies devoid of light chains. Nature 363 (6428), 446–448.

Hancock, K., Khan, A., Williams, F. B., Yushak, M. L. *et al.*, 2003. Characterization of the 8-kilodalton antigens of *Taenia solium* metacestodes and evaluation of their use in an enzyme-linked immunosorbent assay for serodiagnosis. J. Clin. Microbiol. 41 (6), 2577–2586.

Hancock, K., Pattabhi, S., Greene, R. M., Yushak, M. L. *et al.*, 2004. Characterization and cloning of GP50, a *Taenia solium* antigen diagnostic for cysticercosis. Mol. Biochem. Parasitol. 133 (1), 115–124.

Hancock, K., Pattabhi, S., Whitfield, F. W., Yushak, M. L. *et al.*, 2006. Characterization and cloning of T24, a *Taenia solium* antigen diagnostic for cysticercosis. Mol. Biochem. Parasitol. 147 (1), 109–117.

Handali, S., Gonzalez, A. E., Hancock, K., García, H. H. *et al.*, 2004. Porcine antibody responses to *Taenia solium* antigens rGp50 and sTs18var1. Am. J. Trop. Med. Hyg. 71 (3), 322–326.

Harnett, W., 2014. Secretory products of helminth parasites as immunomodulators. Mol. Biochem. Parasitol.

Harrison, L. J., Delgado, J. Parkhouse, R. M., 1990. Differential diagnosis of *Taenia saginata* and *Taenia solium* with DNA probes. Parasitology 100 Pt 3, 459–461.

Harrison, L. J. S., Joshua, G. W., Wright, S. H. Parkhouse, R. M. E., 1989. Specific detection of circulating surface/secreted glycoproteins of viable cysticerci in *Taenia saginata* cysticercosis. Parasite Immunol. 11 (4), 351–370.

Hewitson, J. P., Grainger, J. R. Maizels, R. M., 2009. Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. Mol. Biochem. Parasitol. 167 (1), 1–11.

Higdon, R. Kolker, E., 2007. A predictive model for identifying proteins by a single peptide match. Bioinformatics 23 (3), 277–280.

Hoberg, E. P., 2006. Phylogeny of *Taenia*: Species definitions and origins of human parasites. Parasitol. Int. 55 Suppl, S23–S30.

Hoberg, E. P., Jones, A., Rausch, R. L., Eom, K. S. Gardner, S. L., 2000. A phylogenetic hypothesis for species of the genus *Taenia* (eucestoda : Taeniidae). J. Parasitol. 86 (1), 89–98.

Hong, Y., Peng, J., Jiang, W., Fu, Z. *et al.*, 2011. Proteomic analysis of *Schistosoma japonicum* schistosomulum proteins that are differentially expressed among hosts differing in their susceptibility to the infection. Mol. Cell. Proteomics 10 (8), M110.006098.

Hongli, Y., Shuhan, S., Ruiwen, C. Yingjun, G., 2002. Cloning and functional identification of a novel annexin subfamily in *Cysticercus cellulosae*. Mol. Biochem. Parasitol. 119 (1), 1–5.

Huerta, M., de Aluja, A. S., Fragoso, G., Toledo, A. *et al.*, 2001. Synthetic peptide vaccine against *Taenia solium* pig cysticercosis: successful vaccination in a controlled field trial in rural Mexico. Vaccine 20 (1–2), 262–266.

Iribarne, J. V. Thomson, B. A., 1976. On the evaporation of small ions from charged droplets. J. Chem. Phys. 64 (6), 2287–2294.

Ito, A., Nakao, M., Ito, Y., Yuzawa, I. *et al.*, 1999. Neurocysticercosis case with a single cyst in the brain showing dramatic drop in specific antibody titers within 1 year after curative surgical resection. Parasitol. Int. 48 (1), 95–99.

Ito, A., Okamoto, M., Li, T., Wandra, T. *et al.*, 2011. The first workshop towards the control of cestode zoonoses in Asia and Africa. Parasites Vectors 4, 114.

Ito, A., Plancarte, A., Ma, L., Kong, Y. *et al.*, 1998. Novel antigens for neurocysticercosis: simple method for preparation and evaluation for serodiagnosis. Am. J. Trop. Med. Hyg. 59, 291–294.

Ito, A., Wandra, T., Yamasaki, H., Nakao, M. *et al.*, 2004. Cysticercosis/taeniasis in Asia and the Pacific. Vector Borne Zoonotic Dis. 4 (2), 95–107.

Ito, A., Yamasaki, H., Nakao, M., Sako, Y. *et al.*, 2003. Multiple genotypes of *Taenia solium*–ramifications for diagnosis, treatment and control. Acta Trop. 87 (1), 95–101.

Jayashi, C. M., Gonzalez, A. E., Castillo Neyra, R., Rodríguez, S. *et al.*, 2013. Validity of the Enzyme-linked Immunoelectrotransfer Blot (EITB) for naturally acquired porcine cysticercosis. Vet. Parasitol. 199 (1–2), 42–49.

Jayashi, C. M., Kyngdon, C. T., Gauci, C. G., Gonzalez, A. E. Lightowlers, M. W., 2012. Successful immunization of naturally reared pigs against porcine cysticercosis with a recombinant oncosphere antigen vaccine. Vet. Parasitol. 188 (3–4), 261–267.

Jeong, K., Kim, S. Bandeira, N., 2012. False discovery rates in spectral identification. BMC Bioinformatics 13 Suppl 16, S2.

Jeri, C., Gilman, R. H., Lescano, A. G., Mayta, H. *et al.*, 2004. Species identification after treatment for human taeniasis. Lancet 363 (9413), 949–950.

Jia, W.-Z., Yan, H.-B., Lou, Z.-Z., Ni, X.-W. *et al.*, 2011. Genetic variation of the 8-kDa glycoprotein family from *Echinococcus granulosus*, *Taenia multiceps* and *Taenia hydatigena*. Chin. Med. J. (Engl. Ed.) 124 (18), 2849–2856.

Jonscher, K. R. Yates, 3rd, J. R., 1997. The quadrupole ion trap mass spectrometer–a small solution to a big challenge. Anal. Biochem. 244 (1), 1–15.

Käll, L., Storey, J. D., MacCoss, M. J. Noble, W. S., 2008. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J. Proteome Res. 7 (1), 29–34.

Kar, K. Chambers, R., 2008. Handbook on community-led total sanitation. Institute of Development Studies and Plan International (UK).

Karas, M., Bachmann, D., Bahr, U. Hillenkamp, F., 1987. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. Int. J. Mass Spectrom. Ion Processes 78, 53–68.

Katoh, K. Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Briefings Bioinf. 9 (4), 286–298.

Katz, N., Rocha, R. S. Chaves, A., 1979. Preliminary trials with praziquantel in human infections due to *Schistosoma mansoni*. Bull. W. H. O. 57 (5), 781–785.

Keilbach, N. M., de Aluja, A. S. Sarti-Gutierrez, E., 1989. A programme to control taeniasis-cysticercosis (*T. solium*): experiences in a Mexican village. Acta Leiden. 57 (2), 181–189.

Keller, A., Eng, J., Zhang, N., Li, X.-J. Aebersold, R., 2005. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol. Syst. Biol. 1, 1–8.

Keller, A., Nesvizhskii, A. I., Kolker, E. Aebersold, R., 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74 (20), 5383–5392.

Kessner, D., Chambers, M., Burke, R., Agus, D. Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24 (21), 2534–2536.

Kim, S., Mischerikow, N., Bandeira, N., Navarro, J. D. *et al.*, 2010. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. Mol. Cell. Proteomics 9 (12), 2840–2852.

Kim, S.-H., Bae, Y.-A., Yang, Y., Hong, S.-T. Kong, Y., 2011. Paralogous proteins comprising the 150 kDa hydrophobic-ligand-binding-protein complex of the *Taenia solium* metacestode have evolved non-overlapped binding affinities toward fatty acid analogs. Int. J. Parasitol. 41 (11), 1207–1215.

Kim, T. S., Jung, Y., Na, B. K., Kim, K. S. Chung, P. R., 2000. Molecular cloning and expression of Cu/Zn-containing superoxide dismutase from *Fasciola hepatica*. Infect. Immun. 68 (7), 3941–3948.

Klose, J., 1975. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26 (3), 231–243.

Knapp, J., Nakao, M., Yanagida, T., Okamoto, M. *et al.*, 2011. Phylogenetic relationships within *Echinococcus* and *Taenia* tapeworms (Cestoda: Taeniidae): an inference from nuclear protein-coding genes. Mol. Phylogenet. Evol. 61 (3), 628–638.

Knochenmuss, R., 2006. Ion formation mechanisms in UV-MALDI. Analyst 131 (9), 966–986.

Knudsen, G. M., Medzihradszky, K. F., Lim, K.-C., Hansell, E. McKerrow, J. H., 2005. Proteomic analysis of *Schistosoma mansoni* cercarial secretions. Mol. Cell. Proteomics 4 (12), 1862–1875.

Ko, R. C. Ng, T. F., 1998. Evaluation of excretory/secretory products of larval *Taenia solium* as diagnostic antigens for porcine and human cysticercosis. J. Helminthol. 72 (2), 147–154.

Krecek, R. C., Michael, L. M., Schantz, P. M., Ntanjana, L. *et al.*, 2008. Prevalence of *Taenia solium* cysticercosis in swine from a community-based study in 21 villages of the Eastern Cape Province, South Africa. Vet. Parasitol. 154 (1–2), 38–47.

Krecek, R. C., Michael, L. M., Schantz, P. M., Ntanjana, L. *et al.*, 2011. Corrigendum to "Prevalence of *Taenia solium* cysticercosis in swine from a community-based study in 21 villages of the Eastern Cape Province, South Africa" [Vet. Parasitol. 154 (2008) 38–47]. Vet. Parasitol. 183 (1–2), 198–200.

Kuchler, K., Rubartelli, A. Holland, B., 1997. Unusual Secretory Pathways: From Bacteria to Man. Springer Berlin Heidelberg.

Kumar Garg, R., Kumar Singh, M. Misra, S., 2000. Single-enhancing CT lesions in Indian patients with seizures: a review. Epilepsy Res. 38 (2–3), 91–104.

Laclette, J. P., Shoemaker, C. B., Richter, D., Arcos, L. *et al.*, 1992. Paramyosin inhibits complement C1. J. Immunol. 148 (1), 124–128.

Laemmli, U. K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227 (5259), 680–685.

Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K. *et al.*, 2007. Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics 7 (5), 655–667.

Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K. *et al.*, 2008. Building consensus spectral libraries for peptide identification in proteomics. Nat. Methods 5 (10), 873–875.

Laurent, J. M., Vogel, C., Kwon, T., Craig, S. A. *et al.*, 2010. Protein abundances are more conserved than mRNA abundances across diverse taxa. Proteomics 10 (23), 4209–4212.

Le Riche, P. D. Sewell, M. M., 1977. Differentiation of *Taenia saginata* and *Taenia solium* by enzyme electrophoresis. Trans. R. Soc. Trop. Med. Hyg. 71 (4), 327–328.

Le Riche, P. D. Sewell, M. M., 1978. Differentiation of taeniid cestodes by enzyme electrophoresis. Int. J. Parasitol. 8 (6), 479–483.

Le Roch, K. G., Johnson, J. R., Florens, L., Zhou, Y. *et al.*, 2004. Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. Genome Res. 14 (11), 2308–2318.

Lee, E.-G., Kim, S.-H., Bae, Y.-A., Chung, J.-Y. *et al.*, 2007. A hydrophobic ligand-binding protein of the *Taenia solium* metacestode mediates uptake of the host lipid: implication for the maintenance of parasitic cellular homeostasis. Proteomics 7 (21), 4016–4030.

Lee, Y.-M., Handali, S., Hancock, K., Pattabhi, S. *et al.*, 2011. Serologic diagnosis of human *Taenia solium* cysticercosis by using recombinant and synthetic antigens in QuickELISA™. Am. J. Trop. Med. Hyg. 84 (4), 587–593.

Levine, M. Z., Calderón, J. C., Wilkins, P. P., Lane, W. S. *et al.*, 2004. Characterization, cloning, and expression of two diagnostic antigens for *Taenia solium* tapeworm infection. J. Parasitol. 90 (3), 631–638.

Levine, M. Z., Lewis, M. M., Rodriquez, S., Jimenez, J. A. *et al.*, 2007. Development of an enzyme-linked immunoelectrotransfer blot (EITB) assay using two baculovirus expressed recombinant antigens for diagnosis of *Taenia solium* taeniasis. J. Parasitol. 93 (2), 409–417.

Li, Y. F., Arnold, R. J., Li, Y., Radivojac, P. *et al.*, 2009. A bayesian approach to protein inference problem in shotgun proteomics. J. Comput. Biol. 16 (8), 1183–1193.

Lidwell, W., Holden, K. Butler, J., 2010. Universal Principles of Design. Rockport Publishers, Minneapolis.

Lightowlers, M. W., 2003. Vaccines for prevention of cysticercosis. Acta Trop. 87 (1), 129–135.

Lightowlers, M. W., 2006. Cestode vaccines: origins, current status and future prospects. Parasitology 133 Suppl, 27–42.

Lightowlers, M. W., 2010. Eradication of *Taenia solium* cysticercosis: a role for vaccination of pigs. Int. J. Parasitol. 40 (10), 1183–1192.

Lightowlers, M. W., 2013. Control of *Taenia solium* taeniasis/cysticercosis: past practices and new possibilities. Parasitology 140 (13), 1566–1577.

Lilley, K. S., Razzaq, A. Dupree, P., 2002. Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. Curr. Opin. Chem. Biol. 6 (1), 46–50.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. *et al.*, 1999. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17 (7), 676–682.

Linnæus, C., 1758. Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis, 10th Edition. Vol. 1. Laurentii Salvii (Holmiae).

Liu, F., Cui, S.-J., Hu, W., Feng, Z. *et al.*, 2009. Excretory/secretory proteome of the adult developmental stage of human blood fluke, *Schistosoma japonicum*. Mol. Cell. Proteomics 8 (6), 1236–1251.

Liu, F., Hu, W., Cui, S.-J., Chi, M. *et al.*, 2007. Insight into the host-parasite interplay by proteomic study of host proteins copurified with the human parasite, *Schistosoma japonicum*. Proteomics 7 (3), 450–462.

Liu, F., Lu, J., Hu, W., Wang, S.-Y. *et al.*, 2006. New perspectives on host-parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. PLoS Pathog. 2 (4), e29.

Llompart Pou, J. A., Gené, A., Ayestarán, J. I. Saus, C., Jul 2005. Neurocysticercosis presenting as sudden death. Acta Neurochir. 147 (7), 785–786.

Lobato, R. D., Lamas, E., Portillo, J. M., Roger, R. *et al.*, 1981. Hydrocephalus in cerebral cysticercosis. Pathogenic and therapeutic considerations. J. Neurosurg. 55 (5), 786–793.

Lorenzo, C., Salinas, G., Brugnini, A., Wernstedt, C. *et al.*, 2003. *Echinococcus granulosus* antigen 5 is closely related to proteases of the trypsin family. Biochem. J. 369 (Pt 1), 191–188.

Lu, Y.-M., Wang, N., Wang, J.-J., Wang, K.-H. Sun, S.-H., 2010. Expression, purification, and characterization of a novel Ca(2+)- and phospholipid-binding protein annexin B2. Mol. Biol. Rep. 37 (3), 1591–1596.

Lundgren, D. H., Hwang, S.-I., Wu, L. Han, D. K., 2010. Role of spectral counting in quantitative proteomics. Expert Rev. Proteomics 7 (1), 39–53.

Lundström, J., Salazar-Anton, F., Sherwood, E., Andersson, B. Lindh, J., 2010. Analyses of an Expressed Sequence Tag library from *Taenia solium*, cysticerca. PLoS Neglected Trop. Dis. 4 (12), e919.

Ma, B., Zhang, K., Hendrie, C., Liang, C. *et al.*, 2003. PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry. Rapid Commun. Mass Spectrom. 17 (20), 2337–2742.

Ma, K., Vitek, O. Nesvizhskii, A. I., 2012. A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. BMC Bioinformatics 13 Suppl 16, S1.

MacGillivray, A. J. Rickwood, D., 1974. The heterogeneity of mouse-chromatin nonhistone proteins as evidenced by two-dimensional polyacrylamide-gel electrophoresis and ion-exchange chromatography. Eur. J. Biochem. 41 (1), 181–190.

Maizels, R. M., Balic, A., Gomez-Escobar, N., Nair, M. *et al.*, 2004. Helminth parasites–masters of regulation. Immunol. Rev. 201, 89–116.

Maizels, R. M., Pearce, E. J., Artis, D., Yazdanbakhsh, M. Wynn, T. A., 2009. Regulation of pathogenesis and immunity in helminth infections. J. Exp. Med. 206 (10), 2059–2066.

Makarov, A., 2000. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal. Chem. 72 (6), 1156–1162.

Malla, N., Kaur, R., Ganguly, N. K., Sawhney, I. M. S. Mahajan, R. C., 2005. Utility of specific IgG4 response in saliva and serum samples for the diagnosis and follow up of human neurocysticercosis. Nepal Med. Coll. J. 7 (1), 1–9.

Mann, M., 2006. Functional and quantitative proteomics using SILAC. Nat. Rev. Mol. Cell Biol. 7 (12), 952–958.

Manoutcharian, K., Rosas, G., Hernandez, M., Fragoso, G. *et al.*, 1996. Cysticercosis: identification and cloning of protective recombinant antigens. J. Parasitol. 82 (2), 250–254.

March, R. E., 1997. An introduction to quadrupole ion trap mass spectrometry. J. Mass Spectrom. 32 (4), 351–369.

Martens, L., Vandekerckhove, J. Gevaert, K., 2005. DBToolkit: processing protein databases for peptide-centric proteomics. Bioinformatics 21 (17), 3584–3585.

Martins-de Souza, D., 2014. Proteomics, metabolomics, and protein interactomics in the characterization of the molecular features of major depressive disorder. Dialogues Clin. Neurosci. 16 (1), 63–73.

Mayta, H., Gilman, R. H., Prendergast, E., Castillo, J. P. *et al.*, 2008. Nested PCR for specific diagnosis of *Taenia solium* taeniasis. J. Clin. Microbiol. 46 (1), 286–289.

Mayta, H., Talley, A., Gilman, R. H., Jimenez, J. *et al.*, 2000. Differentiating *Taenia solium* and *Taenia saginata* infections by simple hematoxylin-eosin staining and pcr-restriction enzyme analysis. J. Clin. Microbiol. 38 (1), 133–137.

McSorley, H. J., Hewitson, J. P. Maizels, R. M., 2013. Immunomodulation by helminth parasites: defining mechanisms and mediators. Int. J. Parasitol. 43 (3–4), 301–310.

Mendlovic, F., Ostoa-Saloma, P., Solís, C. F., Martínez-Ocaña, J. *et al.*, 2004. Cloning, characterization, and functional expression of *Taenia solium* calreticulin. J. Parasitol. 90 (4), 891–893.

Menschaert, G., Van Criekinge, W., Notelaers, T., Koch, A. *et al.*, 2013. Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Mol. Cell. Proteomics 12 (7), 1780–1790.

Millares, P., Lacourse, E. J., Perally, S., Ward, D. A. *et al.*, 2012. Proteomic profiling and protein identification by MALDI-TOF mass spectrometry in unsequenced parasitic nematodes. PLoS One 7 (3), e33590.

Minden, J. S., Dowd, S. R., Meyer, H. E. Stühler, K., 2009. Difference gel electrophoresis. Electrophoresis 30 Suppl 1, S156–S161.

Mkupasi, E. M., Ngowi, H. A., Sikasunge, C. S., Leifsson, P. S. Johansen, M. V., 2013a. Efficacy of ivermectin and oxfendazole against *Taenia solium* cysticercosis and other parasitoses in naturally infected pigs. Acta Trop. 128 (1), 48–53.

Mkupasi, E. M., Sikasunge, C. S., Ngowi, H. A. Johansen, M. V., 2013b. Efficacy and safety of anthelmintics tested against *Taenia solium* cysticercosis in pigs. PLoS Neglected Trop. Dis. 7 (7), e2200.

Mohammed, Y., Mostovenko, E., Henneman, A. A., Marissen, R. J. *et al.*, 2012. Cloud parallel processing of tandem mass spectrometry based proteomics data. J. Proteome Res. 11 (10), 5101–5108.

Molinari, J. L., Mejia, H., White, Jr, A. C., Garrido, E. *et al.*, 2000. *Taenia solium*: a cysteine protease secreted by metacestodes depletes human CD4 lymphocytes *in vitro*. Exp. Parasitol. 94 (3), 133–142.

Molinari, J. L., Meza, R., Suárez, B., Palacios, S. *et al.*, 1983. *Taenia solium*: immunity in hogs to the *Cysticercus*. Exp. Parasitol. 55 (3), 340–357.

Monteiro, K. M., de Carvalho, M. O., Zaha, A. Ferreira, H. B., 2010. Proteomic analysis of the *Echinococcus granulosus* metacestode during infection of its intermediate host. Proteomics 10 (10), 1985–1999.

Montero, E., González, L., Puente, S. Gárate, T., 2003. Diagnosis of human taeniasis by multiplex-PCR. Med. Clin. (Barcelona) 120 (1), 37.

Moreau, E. Chauvin, A., 2010. Immunity against helminths: interactions with the host and the intercurrent infections. J. Biomed. Biotechnol. 2010, 428593.

Moreno, L., Lopez-Urbina, M. T., Farias, C., Domingue, G. *et al.*, 2012. A high oxfendazole dose to control porcine cysticercosis: pharmacokinetics and tissue residue profiles. Food Chem. Toxicol. 50 (10), 3819–3825.

Morgan, D. W., 1982. Toxicity study of oxfendazole in pregnant sows. Vet. Rec. 111 (8), 161–163.

Moses, A., 1911. Dos metodos biolojicos de diagnostico nas cisticercozes. Mem. Inst. Oswaldo Cruz 3, 322–326.

Motoyama, A. Yates, 3rd, J. R., 2008. Multidimensional LC separations in shotgun proteomics. Anal. Chem. 80 (19), 7187–7193.

Mulvenna, J., Hamilton, B., Nagaraj, S. H., Smyth, D. *et al.*, 2009. Proteomics analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, *Ancylostoma caninum*. Mol. Cell. Proteomics 8 (1), 109–121.

Mulvenna, J., Sripa, B., Brindley, P. J., Gorman, J. *et al.*, 2010. The secreted and surface proteomes of the adult stage of the carcinogenic human liver fluke *Opisthorchis viverrini*. Proteomics 10 (5), 1063–1078.

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M. *et al.*, 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380 (9859), 2197–2223.

Murrell, K. D. (Ed.), 2005. WHO/FAO/OIE Guidelines for the surveillance, prevention and control of taeniosis/cysticercosis. Office International des Epizooties (OIE), Paris.

Mutapi, F., 2012. Helminth parasite proteomics: from experimental models to human infections. Parasitology, 1–10.

Muyldermans, S. Lauwereys, M., 1999. Unique single-domain antigen binding fragments derived from naturally occurring camel heavy-chain antibodies. J. Mol. Recognit. 12 (2), 131–140.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C. *et al.*, 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320 (5881), 1344–1349.

Nagaraj, S. H., Gasser, R. B. Ranganathan, S., 2007. A hitchhiker's guide to expressed sequence tag (EST) analysis. Briefings Bioinf. 8 (1), 6–21.

Nagaraj, S. H., Gasser, R. B. Ranganathan, S., 2008. Needles in the EST haystack: large-scale identification and analysis of excretory-secretory (ES) proteins in parasitic nematodes using expressed sequence tags (ESTs). PLoS Negl. Trop. Dis. 2 (9), e301.

Nakao, M., Okamoto, M., Sako, Y., Yamasaki, H. *et al.*, 2002. A phylogenetic hypothesis for the distribution of two genotypes of the pig tapeworm *Taenia solium* worldwide. Parasitology 124 (Pt 6), 657–662.

Nakao, M., Sako, Y. Ito, A., 2003. The mitochondrial genome of the tapeworm *Taenia solium*: a finding of the abbreviated stop codon U. J. Parasitol. 89 (3), 633–635.

Nash, T., 2003. Human case management and treatment of cysticercosis. Acta Trop. 87 (1), 61–69.

Nash, T. E. García, H. H., 2011. Diagnosis and treatment of neurocysticercosis. Nat. Rev. Neurol. 7 (10), 584–594.

Nava, G., Laclette, J. P., Bobes, R., Carrero, J. C. *et al.*, 2011. Cloning, sequencing and functional expression of cytosolic malate dehydrogenase from *Taenia solium*: Purification and characterization of the recombinant enzyme. Exp. Parasitol. 128 (3), 217–224.

Navarrete-Perea, J., Moguel, B., Mendoza-Hernández, G., Fragoso, G. *et al.*, 2014. Identification and quantification of host proteins in the vesicular fluid of porcine *Taenia solium* cysticerci. Exp. Parasitol.

Navarro, P. Vázquez, J., 2009. A refined method to calculate false discovery rates for peptide identification using decoy databases. J. Proteome Res. 8 (4), 1792–1796.

Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M. *et al.*, 2011. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics 11 (4), 535–553.

Nesvizhskii, A. I., 2007. Protein identification by tandem mass spectrometry and sequence database searching. Methods Mol. Biol. 367, 87–119.

Nesvizhskii, A. I., 2010. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. J. Proteomics 73 (11), 2092–2123.

Nesvizhskii, A. I. Aebersold, R., 2005. Interpretation of shotgun proteomic data: the protein inference problem. Mol. Cell. Proteomics 4 (10), 1419–1440.

Nesvizhskii, A. I., Keller, A., Kolker, E. Aebersold, R., 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75 (17), 4646–4658.

Ngowi, H. A., Carabin, H., Kassuku, A. A., Mlozi, M. R. S. *et al.*, 2008. A health-education intervention trial to reduce porcine cysticercosis in Mbulu District, Tanzania. Prev. Vet. Med. 85 (1–2), 52–67.

Nguekam, A., Zoli, A., Vondou, L., Pouedet, S. M. *et al.*, 2003. Kinetics of circulating antigens in pigs experimentally infected with *Taenia solium* eggs. Vet. Parasitol. 111 (4), 323–332.

Nguyen, H. A., Bae, Y.-A., Lee, E.-G., Kim, S.-H. *et al.*, 2010. A novel sigma-like glutathione transferase of *Taenia solium* metacestode. Int. J. Parasitol. 40 (9), 1097–1106.

Nguyen, S. Fenn, J. B., 2007. Gas-phase ions of solute species from charged droplets of solutions. Proc. Natl. Acad. Sci. U. S. A. 104 (4), 1111–1117.

Nkouawa, A., Sako, Y., Nakao, M., Nakaya, K. Ito, A., 2009. Loop-mediated isothermal amplification method for differentiation and rapid detection of *Taenia* species. J. Clin. Microbiol. 47 (1), 168–174.

Nunes, C. M., Biondi, G. F., Heinemann, M. B. Richtzenhain, L. J., 2000. Comparative evaluation of an indirect ELISA test for diagnosis of swine cysticercosis employing antigen from *Taenia solium* and *Taenia crassiceps* metacestodes. Vet. Parasitol. 93 (2), 135–140.

Obregón-Henao, A., Gil, D. L., Gómez, D. I., Sanzón, F. *et al.*, 2001. The role of N-linked carbohydrates in the antigenicity of *Taenia solium* metacestode glycoproteins of 12, 16 and 18 kD. Mol. Biochem. Parasitol. 114 (2), 209–215.

Ockham, W., 1495. Quaestiones et decisiones in quattuor libros Sententiarum Petri Lombardi. Centilogium theologicum. Johannes Trechsel, Lugdunum.

O'Farrell, P. H., 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250 (10), 4007–4021.

Oinn, T., Addis, M., Ferris, J., Marvin, D. *et al.*, 2004. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20 (17), 3045–3054.

Olsen, J. V., Ong, S.-E. Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol. Cell. Proteomics 3 (6), 608–614.

Olson, P. D., Zarowiecki, M., Kiss, F. Brehm, K., 2012. Cestode genomics - progress and prospects for advancing basic and applied aspects of flatworm biology. Parasite Immunol. 34 (2–3), 130–150.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B. *et al.*, 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 1 (5), 376–386.

Paizs, B. Suhai, S., 2005. Fragmentation pathways of protonated peptides. Mass Spectrom. Rev. 24 (4), 508–548.

Pardini, A. X., Peralta, R. H., Vaz, A. J., Machado, L. D. R. Peralta, J. M., 2002. Use of *Taenia crassiceps* cysticercus antigen preparations for detection of antibodies in cerebrospinal fluid samples from patients with neurocysticercosis (*Taenia solium*). Clin. Diagn. Lab. Immunol. 9 (1), 190–193.

Parkinson, J. Blaxter, M., 2009. Expressed sequence tags: an overview. Methods Mol. Biol. 533, 1–12.

Parkinson, J., Wasmuth, J. D., Salinas, G., Bizarro, C. V. *et al.*, 2012. A transcriptomic analysis of *Echinococcus granulosus* larval stages: Implications for parasite biology and host adaptation. PLoS Neglected Trop. Dis. 6 (11), e1897.

Paul, W. Steinwedel, H., 1953. Ein neues Massenspektrometer ohne Magnetfeld. Z. Naturforsch. 8 (7), 448–450.

Paul, W. Steinwedel, H., 1960. Us patent number 2939952.

Pawlowski, Z. S., Allan, J. C. Sarti, E., 2005. Control of *Taenia solium* taeniasis/cysticercosis: from research towards implementation. Int. J. Parasitol. 35, 1221–1232.

Pearson, R. D. Hewlett, E. L., 1985. Niclosamide therapy for tapeworm infections. Ann. Intern. Med. 102 (4), 550–551.

Peón, A. N., Espinoza-Jiménez, A. Terrazas, L. I., 2013. Immunoregulation by *Taenia crassiceps* and its antigens. Biomed Res. Int. 2013, 498583.

Pereira, F., Niu, X. deMello, A. J., 2013. A nano LC-MALDI mass spectrometry droplet interface for the analysis of complex protein samples. PLoS One 8 (5), e63087.

Pérez-Sánchez, R., Valero, M. L., Ramajo-Hernández, A., Siles-Lucas, M. *et al.*, 2008. A proteomic approach to the identification of tegumental proteins of male and female *Schistosoma bovis* worms. Mol. Biochem. Parasitol. 161 (2), 112–123.

Perkins, D. N., Pappin, D. J., Creasy, D. M. Cottrell, J. S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20 (18), 3551–3567.

Petersen, T. N., Brunak, S., von Heijne, G. Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8 (10), 785–786.

Peterson, A., Hohmann, L., Huang, L., Kim, B. *et al.*, 2009. Analysis of RP-HPLC loading conditions for maximizing peptide identifications in shotgun proteomics. J. Proteome Res. 8 (8), 4161–4168.

Phiri, I. K., Ngowi, H., Afonso, S., Matenga, E. *et al.*, 2003. The emergence of *Taenia solium* cysticercosis in Eastern and Southern Africa as a serious agricultural problem and public health risk. Acta Trop. 87 (1), 13–23.

Pinto, P. S. A., Vaz, A. J., Germano, P. M. L. Nakamura, P. M., 2000. Performance of the ELISA test for swine cysticercosis using antigens of *Taenia solium* and *Taenia crassiceps* cysticerci. Vet. Parasitol. 88 (1–2), 127–130.

Pittella, J. E., 1997. Neurocysticercosis. Brain Pathol. 7 (1), 681–693.

Plancarte, A., Rendon, J. L. Landa, A., 2004. Purification, characterization and kinetic properties of the *Taenia solium* glutathione S-transferase isoform 26.5 kDa. Parasitol. Res. 93 (2), 137–144.

Pondja, A., Neves, L., Mlangwa, J., Afonso, S. *et al.*, 2010. Prevalence and risk factors of porcine cysticercosis in Angónia District, Mozambique. PLoS Neglected Trop. Dis. 4 (2), e594.

Prabhakaran, V., Rajshekhar, V., Murrell, K. D. Oommen, A., 2004. *Taenia solium* metacestode glycoproteins as diagnostic antigens for solitary cysticercus granuloma in Indian patients. Trans. R. Soc. Trop. Med. Hyg. 98 (8), 478–484.

Prabhakaran, V., Rajshekhar, V., Murrell, K. D. Oommen, A., 2007. Conformation-sensitive immunoassays improve the serodiagnosis of solitary cysticercus granuloma in Indian patients. Trans. R. Soc. Trop. Med. Hyg. 101 (6), 570–577.

Praet, N., Kanobana, K., Kabwe, C., Maketa, V. *et al.*, 2010a. *Taenia solium* cysticercosis in the Democratic Republic of Congo: How does pork trade affect the transmission of the parasite? PLoS Neglected Trop. Dis. 4 (9), e817.

Praet, N., Rodriguez-Hidalgo, R., Speybroeck, N., Ahounou, S. *et al.*, 2010b. Infection with versus exposure to *Taenia solium*: What do serological test results tell us? Am. J. Trop. Med. Hyg. 83 (2), 413–415.

Praet, N., Speybroeck, N., Manzanedo, R., Berkvens, D. *et al.*, 2009. The disease burden of *Taenia solium* cysticercosis in Cameroon. PLoS Neglected Trop. Dis. 3 (3), e406.

Praet, N., Verweij, J. J., Mwape, K. E., Phiri, I. K. *et al.*, 2013. Bayesian modelling to estimate the test characteristics of coprology, coproantigen ELISA and a novel real-time PCR for the diagnosis of taeniasis. Trop. Med. Int. Health 18 (5), 608–614.

Prasad, K. N., Chawla, S., Prasad, A., Tripathi, M. *et al.*, 2006. Clinical signs for identification of neurocysticercosis in swine naturally infected with *Taenia solium*. Parasitol. Int. 55 (2), 151–154.

Proaño-Narvaez, J. V., Meza-Lucas, A., Mata-Ruiz, O., García-Jerónimo, R. C. Correa, D., 2002. Laboratory diagnosis of human neurocysticercosis: double-blind comparison of enzyme-linked immunosorbent assay and electroimmunotransfer blot assay. J. Clin. Microbiol. 40 (6), 2115–2118.

Rabilloud, T., 2002. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2 (1), 3–10.

Rabilloud, T. Lelong, C., 2011. Two-dimensional gel electrophoresis in proteomics: a tutorial. J. Proteomics 74 (10), 1829–1841.

Rajshekhar, V., 1991. Etiology and management of single small CT lesions in patients with seizures: understanding a controversy. Acta Neurol. Scand. 84 (6), 465–470.

Rajshekhar, V., 2004. Purging the worm: management of *Taenia solium* taeniasis. Lancet 363 (9413), 912.

Rajshekhar, V. Chandy, M. J., 1997. Validation of diagnostic criteria for solitary cerebral cysticercus granuloma in patients presenting with seizures. Acta Neurol. Scand. 96 (2), 76–81.

Rajshekhar, V., Joshi, D. D., Doanh, N. Q., van De, N. Xiaonong, Z., 2003. *Taenia solium* taeniosis/cysticercosis in Asia: epidemiology, impact and issues. Acta Trop. 87 (1), 53–60.

Ramahefarisoa, R. M., Rakotondrazaka, M., Jambou, R. Carod, J., 2010. Comparison of ELISA and PCR assays for the diagnosis of porcine cysticercosis. Vet. Parasitol. 173 (3–4), 336–339.

Reiter, L., Claassen, M., Schrimpf, S. P., Jovanovic, M. *et al.*, 2009. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. Mol. Cell. Proteomics 8 (11), 2405–2417.

Restrepo, B. I., Llaguno, P., Sandoval, M. A., Enciso, J. A. Teale, J. M., 1998. Analysis of immune lesions in neurocysticercosis patients: central nervous system response to helminth appears Th1-like instead of Th2. J. Neuroimmunol. 89 (1–2), 64–72.

Rice, P., Longden, I. Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16 (6), 276–277.

Riganò, R., Profumo, E., Bruschi, F., Carulli, G. *et al.*, 2001. Modulation of human immune response by *Echinococcus granulosus* antigen B and its possible role in evading host defenses. Infect. Immun. 69 (1), 288–296.

Rodriguez, S., Dorny, P., Tsang, V. C. W., Pretell, E. J. *et al.*, 2009. Detection of *Taenia solium* antigens and anti-*T. solium* antibodies in paired serum and cerebrospinal fluid samples from patients with intraparenchymal or extraparenchymal neurocysticercosis. J. Infect. Dis. 199 (9), 1345–1352.

Rodriguez, S., Wilkins, P. Dorny, P., 2012. Immunological and molecular diagnosis of cysticercosis. Pathog. Global Health 106 (5), 286–298.

Rodríguez-Contreras, D., de, l. T. P., Velasco, J., Shoemaker, C. B. Laclette, J. P., 2002. The *Taenia solium* glucose transporters TGTP1 and TGTP2 are not immunologically recognized by cysticercotic humans and swine. Parasitol. Res. 88 (3), 280–282.

Rodríguez-Contreras, D., Skelly, P. J., Landa, A., Shoemaker, C. B. Laclette, J. P., 1998. Molecular and functional characterization and tissue localization of 2 glucose transporter homologues (TGTP1 and TGTP2) from the tapeworm *Taenia solium*. Parasitology 117, 579–588.

Rodriguez-Hidalgo, R., Benitez-Ortiz, W., Praet, N., Saa, L. R. *et al.*, 2006. Taeniasis-cysticercosis in Southern Ecuador: assessment of infection status using multiple laboratory diagnostic tools. Mem. Inst. Oswaldo Cruz 101 (7), 779–782.

Rodriguez-Hidalgo, R., Geysen, D., Benitez-Ortiz, W., Geerts, S. Brandt, J., 2002. Comparison of conventional techniques to differentiate between *Taenia solium* and *Taenia saginata* and an improved polymerase chain reaction-restriction fragment length polymorphism assay using a mitochondrial 12S rDNA fragment. J. Parasitol. 88 (5), 1007–1011.

Roepstorff, P. Fohlman, J., 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. 11 (11), 601.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B. *et al.*, 2004. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3 (12), 1154–1169.

Sakai, H., Sone, M., Castro, D. M., Nonaka, N. *et al.*, 1998. Seroprevalence of *Taenia solium* cysticercosis in pigs in a rural community of Honduras. Vet. Parasitol. 78 (3), 233–238.

Salazar-Anton, F. Lindh, J., 2011. *Taenia solium*: A two-dimensional Western blotting method combined with the use of an EST-library for the identification of immunogenic proteins recognized by sera from neurocysticercosis patients. Exp. Parasitol. 128 (4), 371–376.

Salgado, P., Rojas, R. Sotelo, J., 1997. Cysticercosis. Clinical classification based on imaging studies. Arch. Intern. Med. 157 (17), 1991–1997.

Sanger, F., Nicklen, S. Coulson, A. R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74 (12), 5463–5467.

Santivañez, S. J., Hernández-González, A., Chile, N., Oleaga, A. *et al.*, 2010. Proteomic study of activated *Taenia solium* oncospheres. Mol. Biochem. Parasitol. 171 (1), 32–39.

Sarti, E., Flisser, A., Schantz, P. M., Gleizer, M. *et al.*, 1997. Development and evaluation of a health education intervention against *Taenia solium* in a rural community in Mexico. Am. J. Trop. Med. Hyg. 56 (2), 127–132.

Sarti, E. Rajshekhar, V., 2003. Measures for the prevention and control of *Taenia solium* taeniosis and cysticercosis. Acta Trop. 87 (1), 137–143.

Sato, M. O., Yamasaki, H., Sako, Y., Nakao, M. *et al.*, 2003. Evaluation of tongue inspection and serology for diagnosis of *Taenia solium* cysticercosis in swine: usefulness of ELISA using purified glycoproteins and recombinant antigen. Vet. Parasitol. 111 (4), 309–322.

Schaegger, H., 2006. Tricine-SDS-PAGE. Nat. Protoc. 1 (1), 16–22.

Schantz, P. M., 2006. Progress in diagnosis, treatment and elimination of echinococcosis and cysticercosis. Parasitol. Int. 55 Suppl, S7–S13.

Schantz, P. M., Cruz, M., Sarti, E. Pawlowski, Z., 1993. Potential eradicability of taeniasis and cysticercosis. Bull. Pan. Am. Health. Organ. 27 (4), 397–403.

Schantz, P. M. Sarti-Gutierrez, E., 1989. Diagnostic methods and epidemiologic surveillance of *Taenia solium* infection. Acta Leiden. 57 (2), 153–163.

Scheel, C. M., Khan, A., Hancock, K., García, H. H. *et al.*, 2005. Serodiagnosis of neurocysticercosis using synthetic 8-kD proteins: comparison of assay formats. Am. J. Trop. Med. Hyg. 73 (4), 771–776.

Schrimpf, S. P., Weiss, M., Reiter, L., Ahrens, C. H. *et al.*, 2009. Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. PLoS Biol. 7 (3), e48.

Sciutto, E., Chavarria, A., Fragoso, G., Fleury, A. Larralde, C., 2007a. The immune response in *Taenia solium* cysticercosis: protection and injury. Parasite Immunol. 29 (12), 621–636.

Sciutto, E., Fragoso, G., Hernández, M., Rosas, G. *et al.*, 2013. Development of the S3Pvac vaccine against porcine *Taenia solium* cysticercosis: a historical review. J. Parasitol. 99 (4), 686–692.

Sciutto, E., Hernández, M., García, G., de Aluja, A. S. *et al.*, 1998a. Diagnosis of porcine cysticercosis: a comparative study of serological tests for detection of circulating antibody and viable parasites. Vet. Parasitol. 78 (3), 185–194.

Sciutto, E., Martínez, J. J., Huerta, M., Avila, R. *et al.*, 2003. Familial clustering of *Taenia solium* cysticercosis in the rural pigs of Mexico: hints of genetic determinants in innate and acquired resistance to infection. Vet. Parasitol. 116 (3), 223–229.

Sciutto, E., Martínez, J. J., Villalobos, N. M., Hernandez, M. *et al.*, 1998b. Limitations of current diagnostic procedures for the diagnosis of *Taenia solium* cysticercosis in rural pigs. Vet. Parasitol. 79 (4), 299–313.

Sciutto, E., Morales, J., Martínez, J. J., Toledo, A. *et al.*, 2007b. Further evaluation of the synthetic peptide vaccine S3Pvac against *Taenia solium* cysticercosis in pigs in an endemic town of Mexico. Parasitology 134 (Pt 1), 129–133.

Sciutto, E., Rosas, G., Hernandez, M., Morales, J. *et al.*, 2007c. Improvement of the synthetic tri-peptide vaccine (S3Pvac) against porcine *Taenia solium* cysticercosis in search of a more effective, inexpensive and manageable vaccine. Vaccine 25 (8), 1368–1378.

Seidler, J., Zinn, N., Boehm, M. E. Lehmann, W. D., 2010. *De novo* sequencing of peptides by MS/MS. Proteomics 10 (4), 634–649.

Shepherd, J. C., Aitken, A. McManus, D. P., 1991. A protein secreted *in vivo* by *Echinococcus granulosus* inhibits elastase activity and neutrophil chemotaxis. Mol. Biochem. Parasitol. 44 (1), 81–90.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. Mann, M., 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 1 (6), 2856–2860.

Shevchenko, A., Wilm, M., Vorm, O. Mann, M., 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. 68 (5), 850–858.

Shi, J. Wu, F.-X., 2012. A feedback framework for protein inference with peptides identified from tandem mass spectra. Proteome Sci. 10 (1), 68.

Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K. *et al.*, 2011. iProphet: Multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol. Cell. Proteomics 10 (12), M111.007690.

Sikasunge, C. S., Johansen, M. V., Willingham, 3rd, A. L., Leifsson, P. S. Phiri, I. K., 2008. *Taenia solium* porcine cysticercosis: viability of cysticerci and persistency of antibodies and cysticercal antigens after treatment with oxfendazole. Vet. Parasitol. 158 (1–2), 57–66.

Siles-Lucas, M., Gottstein, B. Felleisen, R. S., 1998. Identification of a differentially expressed *Echinococcus multilocularis* protein Em6 potentially related to antigen 5 of *Echinococcus granulosus*. Parasite Immunol. 20 (10), 473–481.

Siles-Lucas, M., Merli, M. Gottstein, B., 2008. 14-3-3 proteins in *Echinococcus*: their role and potential as protective antigens. Exp. Parasitol. 119 (4), 516–523.

Singh, A. K., Prasad, K. N., Prasad, A., Tripathi, M. *et al.*, 2013a. Immune responses to viable and degenerative metacestodes of *Taenia solium* in naturally infected swine. Int. J. Parasitol. 43 (14), 1101–1107.

Singh, A. K., Singh, S. K., Prasad, K. N., Singh, A. *et al.*, 2013b. Evaluation of ELISA, neck muscle, tongue and eyelid examinations for the diagnosis of swine cysticercosis in a highly endemic area of north India. Exp. Parasitol. 134 (3), 313–317.

Siracusano, A., Delunardo, F., Teggi, A. Ortona, E., 2012. Host-parasite relationship in cystic echinococcosis: an evolving story. Clin. Dev. Immunol. 2012, 639362.

Siracusano, A., Margutti, P., Delunardo, F., Profumo, E. *et al.*, 2008a. Molecular cross-talk in host-parasite relationships: the intriguing immunomodulatory role of *Echinococcus* antigen B in cystic echinococcosis. Int. J. Parasitol. 38 (12), 1371–1376.

Siracusano, A., Riganò, R., Ortona, E., Profumo, E. *et al.*, 2008b. Immunomodulatory mechanisms during *Echinococcus granulosus* infection. Exp. Parasitol. 119 (4), 483–489.

Smyth, J. D. McManus, D. P., 2007. The Physiology and Biochemistry of Cestodes. Cambridge University Press.

Sotelo, J., del Brutto, O. H., Penagos, P., Escobedo, F. *et al.*, 1990. Comparison of therapeutic regimen of anticysticercal drugs for parenchymal brain cysticercosis. J. Neurol. 237 (2), 69–72.

Sotelo, J., Rosas, N. Palencia, G., 1986. Freezing of infested pork muscle kills cysticerci. JAMA 256 (7), 893–894.

Stafford, Jr, G., 2002. Ion trap mass spectrometry: a personal perspective. J. Am. Soc. Mass Spectrom. 13 (6), 589–596.

Stanke, M. Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33 (Web Server issue), W465–W467.

Stanke, M. Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 Suppl 2, ii215–ii225.

Suss, R. A., Maravilla, K. R. Thompson, J., 1986. MR imaging of intracranial cysticercosis: comparison with CT and anatomopathologic features. AJNR Am. J. Neuroradiol. 7 (2), 235–242.

Svensson, H., 1961. Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. I. The differential equation of solute concentrations at a steady state and its solution for simple cases. Acta Chem. Scand. 15 (2), 325–341.

Tabb, D. L., Huang, Y., Wysocki, V. H. Yates, 3rd, J. R., 2004. Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides. Anal. Chem. 76 (5), 1243–1248.

Takayanagui, O. M. Odashima, N. S., 2006. Clinical aspects of neurocysticercosis. Parasitol. Int. 55 Suppl, S111–S115.

Tato, P., Fernández, A. M., Solano, S., Borgonio, V. *et al.*, 2004. A cysteine protease from *Taenia solium* metacestodes induce apoptosis in human CD4+ T-cells. Parasitol. Res. 92 (3), 197–204.

Taylor, G., 1964. Disintegration of water drops in an electric field. Proc. R. Soc. London, Ser. A 280 (1382), 383–397.

Telléz-Girón, E., Ramos, M. C. Montante, M., 1981. Effect of flubendazole on *Cysticercus cellulosae* in pigs. Am. J. Trop. Med. Hyg. 30 (1), 135–138.

Timms, J. F. Cramer, R., 2008. Difference gel electrophoresis. Proteomics 8 (23–24), 4886–4897.

Toledo, A., Fragoso, G., Rosas, G., Hernández, M. *et al.*, 2001. Two epitopes shared by *Taenia crassiceps* and *Taenia solium* confer protection against murine *T. crassiceps* cysticercosis along with a prominent T1 response. Infect. Immun. 69 (3), 1766–1773.

Toledo, A., Larralde, C., Fragoso, G., Gevorkian, G. *et al.*, 1999. Towards a *Taenia solium* cysticercosis vaccine: an epitope shared by *Taenia crassiceps* and *Taenia solium* protects mice against experimental cysticercosis. Infect. Immun. 67 (5), 2522–2530.

Torgerson, P. R., de Silva, N. R., Fèvre, E. M., Kasuga, F. *et al.*, 2014. The global burden of foodborne parasitic diseases: an update. Trends Parasitol. 30 (1), 20–26.

Torgerson, P. R. Macpherson, C. N. L., 2011. The socioeconomic burden of parasitic zoonoses: global trends. Vet. Parasitol. 182 (1), 79–95.

Torres, A., Plancarte, A., Villalobos, A. N., de Aluja, A. S. *et al.*, 1992. Praziquantel treatment of porcine brain and muscle *Taenia solium* cysticercosis. 3. effect of 1-day treatment. Parasitol. Res. 78 (2), 161–164.

Torres-Corzo, J., Rodriguez-della Vecchia, R. Rangel-Castilla, L., 2006. Bruns syndrome caused by intraventricular neurocysticercosis treated using flexible endoscopy. J. Neurosurg. 104 (5), 746–748.

Torres-Rivera, A. Landa, A., 2008. Cooperative kinetics of the recombinant glutathione transferase of *Taenia solium* and characterization of the enzyme. Arch. Biochem. Biophys. 477 (2), 372–378.

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garciarrubio, A. *et al.*, 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496 (7443), 57–63.

Tsang, V. C., Brand, J. A. Boyer, A. E., 1989. An enzyme-linked immunoelectrotransfer blot assay and glycoprotein antigens for diagnosing human cysticercosis (*Taenia solium*). J. Infect. Dis. 159 (1), 50–59.

Tsang, V. C., Pilcher, J. A., Zhou, W., Boyer, A. E. *et al.*, 1991. Efficacy of the immunoblot assay for cysticercosis in pigs and modulated expression of distinct IgM/IgG activities to *Taenia solium* antigens in experimental infections. Vet. Immunol. Immunopathol. 29 (1–2), 69–78.

Unlü, M., Morgan, M. E. Minden, J. S., 1997. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18 (11), 2071–2077.

Vaca-Paniagua, F., Parra-Unda, R. Landa, A., 2009. Characterization of one typical 2-Cys peroxiredoxin gene of *Taenia solium* and *Taenia crassiceps*. Parasitol. Res. 105 (3), 781–787.

Vaca-Paniagua, F., Torres-Rivera, A., Parra-Unda, R. Landa, A., 2008. *Taenia solium*: antioxidant metabolism enzymes as targets for cestocidal drugs and vaccines. Curr. Top. Med. Chem. 8 (5), 393–399.

Van Kerckhoven, I., Vansteenkiste, W., Claes, M., Geerts, S. Brandt, J., 1998. Improved detection of circulating antigen in cattle infected with *Taenia saginata* metacestodes. Vet. Parasitol. 76 (4), 269–274.

Vargas-Parada, L. Laclette, J. P., 2003. Gene structure of *Taenia solium* paramyosin. Parasitol. Res. 89 (5), 375–378.

Varshavsky, A., 1996. The N-end rule: functions, mysteries, uses. Proc. Natl. Acad. Sci. U. S. A. 93 (22), 12142–12149.

Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A. Martens, L., 2011a. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. Proteomics 11 (5), 996–999.

Vaudel, M., Burkhart, J. M., Sickmann, A., Martens, L. Zahedi, R. P., 2011b. Peptide identification quality control. Proteomics 11 (10), 2105–2114.

Vaz, A. J., Nakamura, P. M., Camargo, M. E., Camargo, E. D. Ferreira, A. W., 1996. Dot-ELISA for the detection of anti-*Cysticercus cellulosae* antibodies in cerebrospinal fluid using a new solid phase (resin-treated polyester fabric) and *Cysticercus longicollis* antigens. Rev. Inst. Med. Trop. Sao Paulo 38 (6), 391–396.

Venkataraman, S. Vijayan, G. P., 1983. Uncommon manifestation of human cysticercosis with muscular pseudohypertrophy. Trop. Geogr. Med. 35 (1), 75–77.

Verster, A., Du Plessis, T. A. Van Den Heever, L. W., 1976. The effect of gamma radiation on the cysticerci of *Taenia solium*. Onderstepoort J. Vet. Res. 43 (1), 23–26.

Vesterberg, O. Svensson, H., 1966. Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. IV. Further studies on the resolving power in connection with separation of myoglobins. Acta Chem. Scand. 20 (3), 820–834.

Vibanco-Pérez, N., Jiménez, L., Mendoza-Hernández, G. Landa, A., 2002. Characterization of a recombinant mu-class glutathione S-transferase from *Taenia solium*. Parasitol. Res. 88 (5), 398–404.

Victor, B., Dorny, P., Kanobana, K., Polman, K. *et al.*, 2013. Use of expressed sequence tags as an alternative approach for the identification of *Taenia solium* metacestode excretion/secretion proteins. BMC Res. Notes 6 (1), 224.

Victor, B., Gabriël, S., Kanobana, K., Mostovenko, E. *et al.*, 2012a. Partially sequenced organisms, decoy searches and false discovery rates. J. Proteome Res. 11 (3), 1991–1995.

Victor, B., Kanobana, K., Gabriël, S., Polman, K. *et al.*, 2012b. Proteomic analysis of *Taenia solium* metacestode excretion-secretion proteins. Proteomics 12 (11), 1860–1869.

Virginio, V. G., Monteiro, K. M., Drumond, F., de Carvalho, M. O. *et al.*, 2012. Excretory/secretory products from *in vitro*-cultured *Echinococcus granulosus* protoscoleces. Mol. Biochem. Parasitol. 183 (1), 15–22.

Viswanathan, S., Unlü, M. Minden, J. S., 2006. Two-dimensional difference gel electrophoresis. Nat. Protoc. 1 (3), 1351–1358.

Vogel, C. Marcotte, E. M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. 13 (4), 227–232.

Wang, G., Wu, W. W., Zhang, Z., Masilamani, S. Shen, R.-F., 2009a. Decoy methods for assessing false positives and false discovery rates in shotgun proteomics. Anal. Chem. 81 (1), 146–159.

Wang, X. Zhang, B., 2013. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. Bioinformatics 29 (24), 3235–3237.

Wang, Z., Gerstein, M. Snyder, M., 2009b. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (1), 57–63.

Watson, J. D., 1990. The human genome project: past, present, and future. Science 248 (4951), 44–49.

Weinberg, M., 1909. Recherche des anticorps specifiques dans la distomatose et la cysticercose. C. R. Soc. Biol. Paris 66, 219–221.

White, Jr, A. C., 2000. Neurocysticercosis: updates on epidemiology, pathogenesis, diagnosis, and management. Annu. Rev. Med. 51, 187–206.

White, Jr, A. C., Baig, S. Chappell, C. L., 1997a. Characterization of a cysteine proteinase from *Taenia crassiceps* cysts. Mol. Biochem. Parasitol. 85 (2), 243–253.

White, Jr, A. C., Robinson, P. Kuhn, R., 1997b. *Taenia solium* cysticercosis: host-parasite interactions and the immune response. Chem. Immunol. 66, 209–230.

White, Jr, A. C., Tato, P. Molinari, J. L., 1992. Host-parasite interactions in *Taenia solium* cysticercosis. Infect. Agents Dis. 1 (4), 185–193.

WHO, 1995. WHO model prescribing information: Drugs used in parasitic diseases - second edition. Tech. rep., World Health Organisation, Geneva.

WHO, 2006. The control of neglected zoonotic diseases: a route to poverty alleviation: report of a joint WHO/DFID-AHP meeting, 20 and 21 september 2005, WHO Headquarters, Geneva, with the participation of FAO and OIE. Tech. rep., World Health Organisation, Geneva.

WHO, 2007. Global plan to combat neglected tropical diseases (2008–2015). Tech. rep., World Health Organisation, Geneva.

WHO, 2013. Sixty-sixth World Health Assembly (WHA66.12). Tech. rep., World Health Organisation, Geneva.

Wiley, W. C. McLaren, I. H., 1955. Time-of-flight mass spectrometer with improved resolution. Rev. Sci. Instrum. 26 (12), 1150–1157.

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D. *et al.*, 1995. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol. Genet. Eng. Rev. 13, 19–50.

Wilkins, P. P., Allan, J. C., Verastegui, M., Acosta, M. *et al.*, 1999. Development of a serologic assay to detect *Taenia solium* taeniasis. Am. J. Trop. Med. Hyg. 60 (2), 199–204.

Wilson, M., Bryan, R. T., Fried, J. A., Ware, D. A. *et al.*, 1991. Clinical evaluation of the cysticercosis enzyme-linked immunoelectrotransfer blot in patients with neurocysticercosis. J. Infect. Dis. 164 (5), 1007–1009.

Wolters, D. A., Washburn, M. P. Yates, 3rd, J. R., 2001. An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73 (23), 5683–5690.

Wu, C. C. MacCoss, M. J., 2002. Shotgun proteomics: tools for the analysis of complex biological systems. Curr. Opin. Mol. Ther. 4 (3), 242–250.

Yamasaki, H., Allan, J. C., Sato, M. O., Nakao, M. *et al.*, 2004. DNA differential diagnosis of taeniasis and cysticercosis by multiplex PCR. J. Clin. Microbiol. 42 (2), 548–553.

Yamashita, M. Fenn, J. B., 1984. Electrospray ion-source - another variation on the free-jet theme. J. Phys. Chem. 88 (20), 4451–4459.

Yan, H.-L., Xue, G., Mei, Q., Ding, F.-X. *et al.*, 2008. Calcium-dependent proapoptotic effect of *Taenia solium* metacestodes annexin B1 on human eosinophils: a novel strategy to prevent host immune response. Int. J. Biochem. Cell. Biol. 40 (10), 2151–2163.

Yates, 3rd, J. R., Morgan, S. F., Gatlin, C. L., Griffin, P. R. Eng, J. K., 1998. Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. Anal. Chem. 70 (17), 3557–3565.

Yatsuda, A. P., Krijgsveld, J., Cornelissen, A. W. C. A., Heck, A. J. R. de Vries, E., 2003. Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. J. Biol. Chem. 278 (19), 16941–16951.

Yoshino, K., 1934. On the subjective symptoms caused by the parasitism of *Taenia solium* and its development in man. Taiwan Igakkai Zasshi 33, 183–194.

Young, N. D., Jex, A. R., Li, B., Liu, S. *et al.*, 2012. Whole-genome sequence of *Schistosoma haematobium*. Nat. Genet. 44 (2), 221–225.

Zea-Vera, A., Cordova, E. G., Rodriguez, S., Gonzales, I. *et al.*, 2013. Parasite antigen in serum predicts the presence of viable brain parasites in patients with apparently calcified cysticercosis only. Clin. Infect. Dis. 57 (7), e154–e159.

Zhen, Y., Xu, N., Richardson, B., Becklin, R. *et al.*, 2004. Development of an LC-MALDI method for the analysis of protein complexes. J. Am. Soc. Mass Spectrom. 15 (6), 803–822.

Zheng, M., Hu, K., Liu, W., Hu, X. *et al.*, 2011. Proteomic analysis of excretory secretory products from *Clonorchis sinensis* adult worms: molecular characterization and serological reactivity of a excretory-secretory antigen-fructose-1,6-bisphosphatase. Parasitol. Res. 109 (3), 737–744.

Zoli, A., Shey-Njila, O., Assana, E., Nguekam, J. P. *et al.*, 2003. Regional status, epidemiology and impact of *Taenia solium* cysticercosis in Western and Central Africa. Acta Trop. 87 (1), 35–42.

Zubarev, R. A. Makarov, A., 2013. Orbitrap mass spectrometry. Anal. Chem. 85 (11), 5288–5296.

CHAPTER 11

Summary

Chapter 1 introduces the two main topics of this thesis. The first one is the *Taenia solium* taeniasis/cysticercosis complex, a neglected zoonotic disease affecting both human health and economy in large parts of Africa, Asia and Latin America. The second topic is proteomics, or more specifically, the use of liquid chromatography and tandem mass spectrometry to study proteins.

Chapter 2 further focusses on the *T. solium* literature by exploring the morphology, the life cycle and the different infections related to the life cycle stages: taeniasis and porcine and human (neuro)cysticercosis. Then, the clinical aspects of human neurocysticercosis are discussed as well as the impact of cysticercosis on human health and economy in large parts of many developing countries. The main part is this chapter is, however, dedicated to diagnosis. The use of diagnostic tools varies depending on the life cycle stage of the parasite and the intended use. However, most of the available tools have certain shortcomings (labour intensive, expensive, lack of sensitivity and/or specificity,...). Specific attention goes to immunological assays for the detection of (neuro)cysticercosis. These tools can be divided into antigen detection (used for the detection of living parasites) and antibody detection (detection of host reaction to exposure to the parasite). Two antigen detection assays are currently being used, but both assays were originally designed to detect antigens produced by *Taenia saginata* parasites in cattle. The fact that they are used in the detection of *T. solium* (neuro)cysticercosis is due to a cross reaction between *Taenia* species. However, the same cross reaction compromises the usability of the assay in pigs where not only *T. solium* can be present, but other *Taenia* species like *T. hydatigena* and *T. asiatica* as well.

Chapter 3 provides a more thorough introduction into the field of proteomics, which focusses on the study of proteins and is located between other "omics" fields like genomics (study of DNA), transcriptomics (study of the transcriptome or mRNA) and metabolomics (study of metabolites). The literature study on proteomics is based on a generic workflow, similar to the one that was followed for the analyses described in this thesis. Basically, samples are collected and separated using several gel-based methods. Once separated, the proteins (of choice) are digested with an enzyme i.e. they are cut into smaller peptides, usually with trypsin. Since this process again adds complexity to the samples, the peptides are further separated, this time using liquid chromatography. Since mass spectrometry is the measurement of the mass-to-charge ratio (*m/z*) of gas-phase ions, peptides need to be ionised and transferred to the gas phase before they can be analysed. Sample ionisation is explained using the examples of matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) and several types of mass spectrometers are briefly mentioned. For the studies in this thesis, a Bruker Daltonics amaZon ETD ion trap mass spectrometer was used to perform tandem mass spectrometry (MS/MS). This means that a first scan (MS) is meant to isolate an ion, called the precursor ion, while the second scan (MS/MS) will determine the *m/z* ratios of the fragment ions formed by fragmentation of the precursor ion. The result of these actions is MS spectra (with precursors mass information) and, for a number of precursors, MS/MS spectra (with fragmentation information). The rest of the chapter deals with the methods to link these spectra to peptides and ultimately to proteins and elaborates on the metrics that are used to estimate probability of correctness of each spectrum match and the number of incorrect identification in the entire dataset, also known as the false discovery rate (FDR). As a closing note, the state of parasite proteomics is discussed.

The main objective of this thesis discussed in **Chapter 4** and involves the analysis of *T. solium* excretion/secretion proteins with the aim of identifying species-specific diagnostic antigens. This objective is broken up into three sub-objectives.

The first sub-objective is outlined in **Chapter 5** and deals with the effect of incomplete protein databases on the FDR estimations. Global FDRs can be defined as the expected number of incorrect peptide spectrum matches (PSMs) amongst all PSMs that pass a certain search score threshold, but the problem is that nearly all published work on FDR estimations has been done using complete protein databases from fully sequenced organisms, or known well-defined protein mixtures. When working with incomplete databases (from partially sequenced organisms), FDR estimation may become more challenging i.e it is reasonable to assume that a significant number of high-quality spectra coming from the studied organism will not find a match in the target database, simply because the protein is not in that database. A fraction of these spectra will instead match a decoy sequence (or

another peptide in the target database). To study the effect of incomplete target databases, partial databases approximately covering 1 to 99% of the full database were combined with equally sized reversed, scrambled or randomised decoy databases. In this study, it was demonstrated that mixture model-based FDR estimations using incomplete databases from partially sequenced organisms are stable, highly reproducible and comparable to the estimations based on complete databases, irrespective of the manner in which the decoy database was constructed, until the databases become so small that the mixture models can no longer be correctly fitted to the observed distributions.

The second sub-objective involves the use of protein sequences from other (closely related) helminths and expressed sequence tags to increase the proteome coverage of the incomplete protein database and identify *T. solium* excretion/secretion proteins (ESPs). These ESPs are often involved in diverse mechanisms to evade the host immune system and, from a diagnostic point of view, ESPs are also of interest as they can often be detected in the blood. A more thorough knowledge of *T. solium* ESPs would be an important step in improving current diagnostic techniques. At the time of publication, a complete proteomic analysis of many parasite ESPs was still hampered by incomplete protein databases due to the lack of genomic information. To get around this limitation, our approach relied on homology to other helminth species and supplemented the *T. solium* database with protein sequences from other helminths. This study, described in **Chapter 6**, resulted in the identification of 76 proteins including 27 already described *T. solium* proteins, 17 host proteins and 32 proteins likely to be of *T. solium* origin, but identified using sequences from other helminths, effectively demonstrating the value of this approach. In a second study, described in **Chapter 7**, translated *T. solium* expressed sequence tags (ESTs) combined with the Basic Local Alignment Search Tool (BLAST) and protein mapping to supercontigs of *Echinococcus granulosus* (a member of the Taeniidae family) were used to investigate whether this would increase the number of *T. solium* metacestode ESP identifications compared to the previous study. Here, 106 protein groups were identified and the groups were further organised by Gene Ontology (GO) annotation information on biological process and molecular function. A total of 48 protein groups were also identified in the previous study while 58 protein groups were new identifications. Despite this difference in absolute protein identifications, all GO terms from the previous study were identified here as well and only six new GO terms were identified with the EST analyses. However, they were identified by a very small number of proteins and may be a result of proteins being linked to multiple GO terms. This is supported by the fact that the proteins linked to these GO terms are homologous to other proteins identified in both studies, so none of these GO terms was identified by a "new" protein group.

The third sub-objective is the comparison of the *T. solium* excretion/secretion proteome

with other *Taenia* proteomes in order to identify unique proteins (**Chapter 8**). These unique proteins can then be targeted in a *T. solium*-specific assay. Three methods of proteome comparison are explored. A first option is to analyse the ESPs of the other *Taenia* species (i.e *Taenia asiatica* and *Taenia hydatigena*) in the same way the ESPs for *T. solium* were analysed (Chapter 6) and to compare the protein output lists. However, the then used approach relied on homology between related species and it therefore fails here, when comparing *Taenia* species. A second option is to separate the ESPs with two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and to identify spots that are present in one sample, but not in another. However, 2D-PAGE is not without its shortcomings. Although this method was explored, it was later abandoned due to too many problems. A third option is to analyse the ESPs with LC-MS/MS, but then to employ a new methodology that bypasses the identification stage and instead considers all spectra from all analyses, ensuring that none of the acquired information is lost at the all-important comparison stage. In other words, the experimental spectra are compared directly with each other in order to identify which spectra are unique for the organism of interest. These unique spectra can then be matched to a protein database derived from the recently published *T. solium* genome to identify unique peptides and proteins. Although still under active development, this option already shows great potential. However, further refinement is required.

The general discussion in **Chapter 9** is the final chapter in this thesis and here, the performed research is reviewed in a broader context. Improvements to the way data is published in non-proteomics journals are suggested and the obtained ESP identifications are reviewed in light of recently made progress by other research groups. Finally, an overview of the workflow that could lead to a species-specific assay is presented and discussed.

CHAPTER 12

Samenvatting

Hoofdstuk 1 introduceert de twee hoofdonderwerpen van dit proefschrift. Het *Taenia solium* taeniase/cysticercose complex is het eerste onderwerp: een verwaarloosde zoönose met gevolgen voor zowel de volksgezondheid als de economie in grote delen van Afrika, Azië en Latijns-Amerika. Het tweede onderwerp gaat over proteomics, meer bepaald het gebruik van vloeistofchromatografie en tandem massaspectrometrie voor de studie van eiwitten.

Hoofdstuk 2 richt zich op de *T. solium* literatuur en onderzoekt de morfologie, de levenscyclus en de verschillende infecties die gerelateerd zijn aan de stadia van de levenscyclus: taeniasis en porciene en humane (neuro)cysticercose. Vervolgens worden de klinische aspecten van humane neurocysticercose besproken, evenals de impact van cysticercose op de volksgezondheid en de economie in grote delen van veel ontwikkelingslanden. Het grootste deel van dit hoofdstuk is gewijd aan de diagnose. Het gebruik van diagnostische technieken hangt af van de fase van de levenscyclus van de parasiet en het beoogde doel. De meeste technieken hebben echter tekortkomingen (arbeidsintensief, prijs, gebrek aan gevoeligheid en/of specificiteit, ...). Specifieke aandacht gaat naar immunologische testen voor de detectie van (neuro)cysticercose. Deze testen kunnen worden onderverdeeld in antigeen detectie technieken (die de levende parasiet aantonen) en antilichaam detectie technieken (die de reactie van de gastheer op de aanwezigheid van de parasiet aantonen). Twee verschillende antigeen detectie technieken worden momenteel gebruikt, maar beide technieken werden oorspronkelijk ontworpen om antigenen van *Taenia saginata* bij runderen te detecteren. Een kruisreactie tussen *Taenia* soorten verklaart waarom de technieken ook worden gebruikt voor de detectie van *T. solium* (neuro)cysticercose. Dezelfde kruisreactie bemoeilijkt echter het gebruik van de test bij varkens waar niet alleen

T. solium aanwezig kan zijn, maar ook andere *Taenia* soorten zoals *T. hydatigena* en *T. asiatica*.

Hoofdstuk 3 bevat een grondige inleiding over proteomics. Deze richt zich op de studie van eiwitten en situeert zich tussen andere "omics" gebieden zoals genomics (DNA-onderzoek), transcriptomics (studie van de DNA transcriptie of mRNA) en metabolomics (studie van afbraakproducten). De literatuurstudie betreffende proteomics is gebaseerd op een generieke workflow, vergelijkbaar met degene die werd gevolgd voor de in dit proefschrift beschreven analyses. Eens de eiwitstalen verzameld zijn, worden ze gescheiden met behulp van verschillende op gelscheiding gebaseerde methoden. Eenmaal gescheiden, worden de eiwitten "geknipt" met een enzym (meestal trypsine) in kleinere peptiden. Vermits dit proces opnieuw complexiteit van de stalen teweegbrengt, worden de peptiden verder gescheiden, ditmaal met vloeistofchromatografie. Aangezien massaspectrometrie het meten is van de massa-tot-lading verhouding (m/z) van ionen in de gasfase, moeten peptiden worden geïoniseerd en overgedragen naar de gasfase alvorens te kunnen worden geanalyseerd. Ionisatie wordt toegelicht aan de hand van matrix geassisteerde laser desorptie/ionisatie (MALDI) en electrospray ionisatie (ESI). Tevens worden verschillende soorten massaspectrometers kort vermeld. Voor de studies in dit proefschrift, werd een Bruker Daltonics Amazon ETD ion trap massaspectrometer gebruikt om tandem massaspectrometrie (MS/MS) uit te voeren. Dit betekent dat een eerste scan (MS) een (precursor) ion isoleert, terwijl de tweede scan (MS/MS) de m/z verhouding van de fragment ionen gevormd door fragmentatie van de precursor ionen bepaalt. Het resultaat van deze acties is MS spectra (met precursor informatie) en, voor een aantal precursors, MS/MS spectra (met fragmentatie informatie). De rest van het hoofdstuk behelst het koppelen van deze spectra aan peptiden en eiwitten en de statistische technieken die worden gebruikt om de probabiliteit te bepalen of een dergelijke spectrum-peptide link correct is. Tevens wordt een schatting gemaakt van het aantal mogelijk niet correcte identificaties. Om af te sluiten worden de moeilijkheden van eiwitanalyses bij parasieten besproken.

Het hoofdobjectief van dit proefschrift (**hoofdstuk 4**) omvat een analyse van *T. solium* excretie/secretie eiwitten met als doel het identificeren van *T. solium*-specifieke diagnostische antigenen. Dit objectief is onderverdeeld in drie sub-objectieven.

Een eerste sub-objectief is omschreven in **hoofdstuk 5** en gaat over het effect van onvolledige eiwit databases op de "false discovery rate" (FDR). De FDR kan worden gedefinieerd als het verwachte aantal onjuiste peptide identificaties binnen alle peptide identificaties die een bepaalde score-drempel overschrijden. Het probleem hierbij is dat bijna al het gepubliceerde werk steunt op FDR schattingen gebaseerd op complete eiwit databases van volledig gesequeneerde organismen of bekende welomschreven eiwitstalen.

Bij het werken met onvolledige databases (van gedeeltelijk gesequeneerde organismen), wordt een FDR schatting minder eenvoudig en kan men veronderstellen dat een aanzienlijk aantal van de spectra afkomstig van het bestudeerde organisme, een peptide in de database niet vinden, simpelweg omdat het eiwit niet in die database aanwezig is. Een fractie van deze spectra zal in plaats daarvan overeenkomen met een decoy peptide (of een ander peptide in de database). Om het effect van onvolledige databases te bestuderen, werden gedeeltelijke databases van 1 tot 99% van de volledige database gecombineerd met even grote omgekeerde, herschikte of gerandomiseerde decoy databases. In deze studie wordt aangetoond dat "mixture model-based" FDR schattingen met onvolledige databases stabiel, reproduceerbaar en vergelijkbaar zijn met de schattingen van volledige databases, ongeacht de wijze waarop de decoy database is opgebouwd.

Het tweede sub-objectief gaat over het gebruik van eiwit sequenties van andere (nauw verwante) wormen en "expressed sequence tags" (ESTs) om de proteoom dekking van de onvolledige eiwitdatabase te verhogen en *T. solium* excretie/secretie eiwitten (ESPs) te identificeren. Deze ESPs zijn vaak betrokken bij diverse mechanismen om het immuunsysteem van de gastheer te omzeilen en vanuit diagnostisch oogpunt zijn ESPs ook interessant omdat ze vaak kunnen worden gedetecteerd in het bloed. Een grondigere kennis van *T. solium* ESPs zou een belangrijke stap zijn in de verbetering van de huidige diagnostische technieken. Op het moment van publicatie, werd een volledige proteoom analyse van de vele parasieten nog steeds gehinderd door onvolledige eiwit databases. Om deze beperking te omzeilen, is onze aanpak gebaseerd op homologie met andere wormsoorten en werd de *T. solium* databank aangevuld met eiwit sequenties van andere wormen. Deze studie, beschreven in **hoofdstuk 6**, resulteerde in de identificatie van 76 eiwitten waarvan 27 reeds gekende *T. solium* eiwitten, 17 gastheereiwitten en 32 eiwitten die waarschijnlijk van oorsprong *T. solium* zijn, maar geïdentificeerd werden met behulp van sequenties van andere wormen. In een tweede studie, beschreven in **hoofdstuk 7**, werden *T. solium* ESTs in combinatie met de Basic Local Alignment Search Tool (BLAST) en eiwit vergelijkingen met supercontigs van *Echinococcus granulosus* (een lid van de Taeniidae familie) gebruikt om te onderzoeken of dit het aantal *T. solium* ESP identificaties zou vergroten in vergelijking met de vorige studies. Hierbij werden 106 eiwit groepen geïdentificeerd die vervolgens verder onderverdeeld werden op basis van Gene Ontology (GO) informatie over biologisch processen en moleculaire functies. Ook werden in totaal 48 eiwit groepen uit de vorige studie bevestigd, terwijl er 58 nieuwe eiwit groepen geïdentificeerd werden. Ondanks dit verschil in absolute eiwitidentificaties, werden alle GO termen uit de vorige studie hier ook geïdentificeerd en werden slechts zes nieuwe GO termen geïdentificeerd met de EST analyses. Ze werden echter geïdentificeerd door een klein aantal eiwitten die tevens homoloog zijn aan andere eiwitten die in beide studies gevonden werden. Geen van deze GO termen werd dus geïdentificeerd door een nieuwe eiwit groep.

Het derde sub-objectief is de vergelijking van het *T. solium* excretie/secretie proteoom met andere *Taenia* proteomen met als doel unieke eiwitten te identificeren (**hoofdstuk 8**). Deze unieke eiwitten kunnen vervolgens gebruikt worden in een *T. solium*-specifieke assay. Drie methoden van proteoom vergelijking werden verkend. Een eerste mogelijkheid is de ESPs van de andere *Taenia* soorten (*T. asiatica* en *T. hydatigena*) op dezelfde wijze te analyseren als de ESPs voor *T. solium* (hoofdstuk 6) en dan de eiwit identificaties te vergelijken. De toen gebruikte aanpak vertrouwde echter op homologie tussen verwante soorten. Bij het vergelijken van *Taenia* soorten schiet deze dan ook duidelijk tekort. Een tweede optie is om de ESPs te scheiden met behulp van twee-dimensionale polyacrylamide gelelektroforese (2D-PAGE) en zodoende eiwitten te identificeren die een afwijkend profiel geven tussen de twee stalen. 2D-PAGE heeft echter ook tekortkomingen. Deze methode werd wel onderzocht, maar werd later opgegeven door de vele problemen. Een derde optie is om de ESPs te analyseren met LC-MS/MS, maar dan wel met behulp van een nieuwe methode die het identificatie stadium omzeilt en in plaats daarvan alle spectra van alle analyses in overweging neemt en er zo voor zorgt dat er geen informatie verloren gaat. Met andere woorden, de experimentele spectra worden direct met elkaar vergeleken om te bepalen welke spectra uniek zijn voor het te bestuderen organisme. Deze unieke spectra kunnen dan worden gelinkt aan een eiwit databank afgeleid van het onlangs gepubliceerde *T. solium* genoom, om zo unieke peptiden en eiwitten te identificeren. Hoewel deze methode nog steeds actief wordt ontwikkeld, is er nu reeds groot potentieel merkbaar. Verdere verfijning is echter nodig.

Het laatste deel van dit proefschrift is de algemene discussie (**hoofdstuk 9**). Hierin wordt het uitgevoerd onderzoek beoordeeld in een bredere context. Verbeteringen in de manier waarop gegevens worden gepubliceerd in niet-proteomics tijdschriften worden voorgesteld en de verkregen ESP identificaties worden beoordeeld in het licht van de recent gemaakte vooruitgang door andere onderzoeksgroepen. Tot slot wordt een overzicht gegeven van de workflow die kan leiden tot een *T. solium*-specifieke test.

A

Appendix

A.1 Supporting information for Chapter 5

Figure A.1 shows the false discovery rate (FDR) estimations for composite databases consisting of target databases (1-100% of the total size) without redundancy (Figures A.1a to A.1c) or with (limited) redundancy (Figures A.1d to A.1f) and an equally sized decoy database (reversed, scrambled or randomised). The three curves represent three different probability thresholds ($p > 0.90$, $p > 0.95$ and $p > 0.99$, respectively). A $p > 0.95$ minimum probability cut-off results in an FDR around 1% which is currently commonly reported and accepted in proteomics.

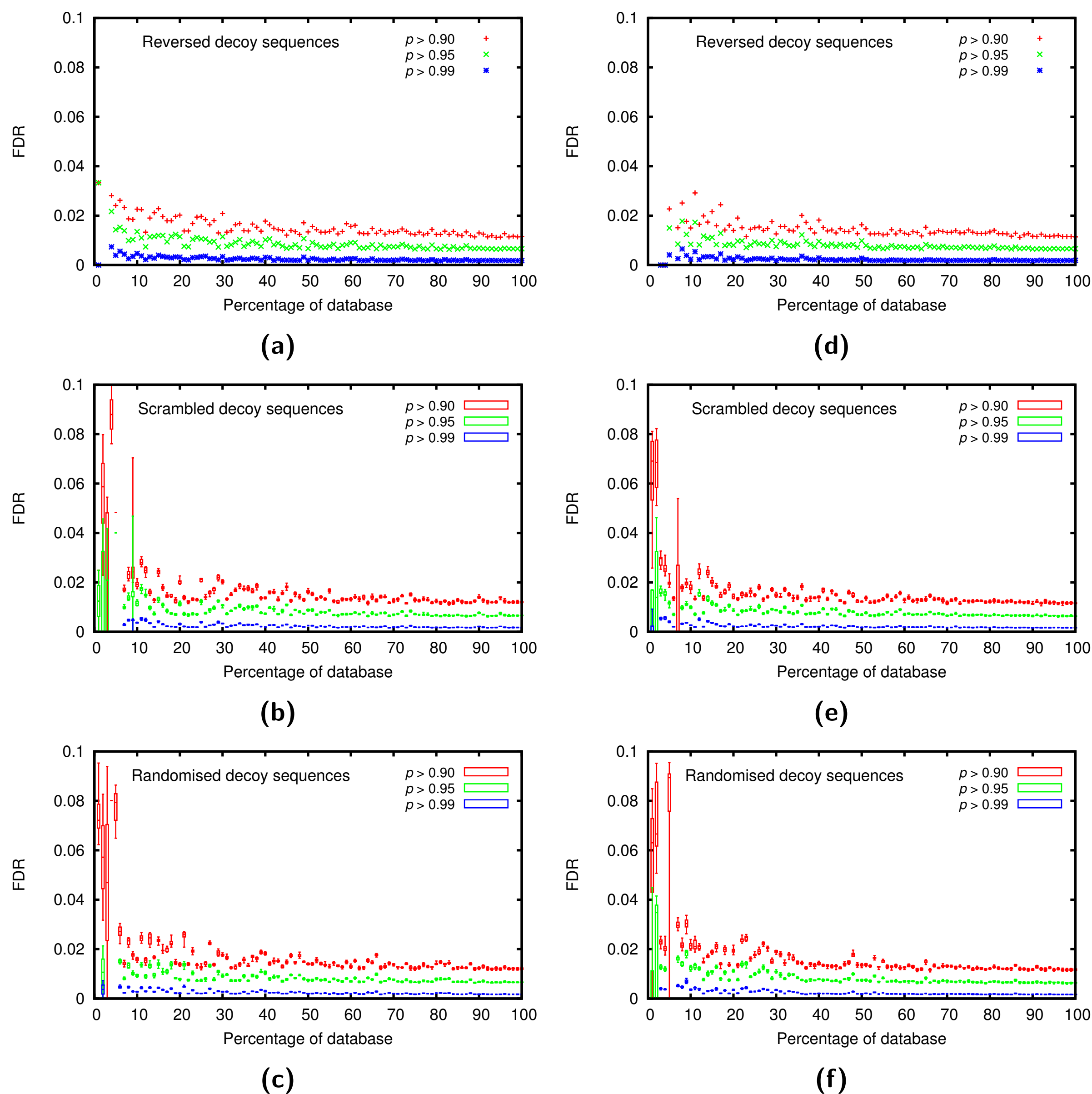


Figure A.1: False discovery rate behaviour at three different probability thresholds for composite databases consisting of a target database with only canonical sequences (a-c) or canonical sequences and isoforms (d-f) and an equally sized reversed ($n = 1$), scrambled ($n = 10$) or randomised ($n = 10$) decoy database. The boxes in figures b, c, e and f show the median, the 25th and the 75th percentile, while the error bars show the minimum and maximum values.

Figure A.2 shows the FDR estimations for target databases (1-100% of the total size) without redundancy (Figure A.2a) or with (limited) redundancy (Figure A.2b) without any decoy database at 3 different probability thresholds ($p > 0.90$, $p > 0.95$ and $p > 0.99$, respectively).

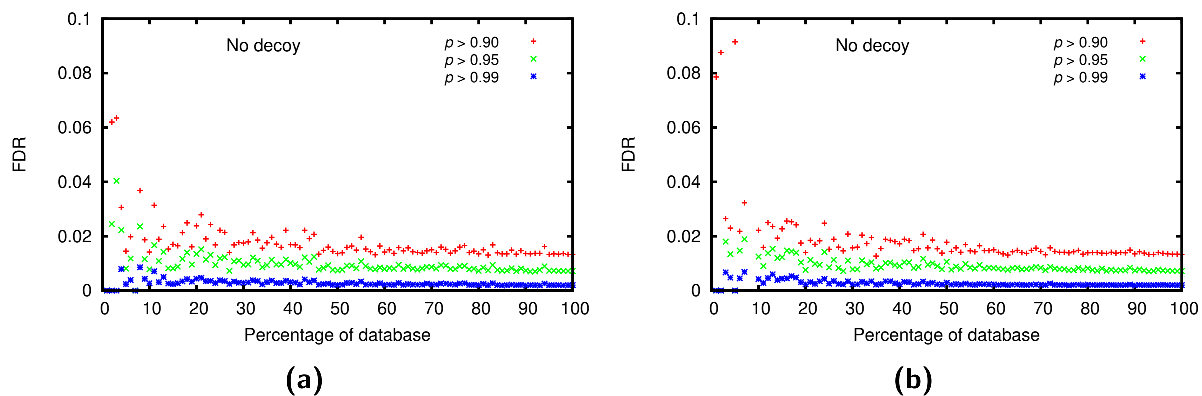


Figure A.2: False discovery rate behaviour at three different probability thresholds for target databases with only canonical sequences (a) or canonical sequences and isoforms (b) without any decoy database.