Quantitative proteomic analyses of isolate variation and virulence in *Giardia duodenalis*

by

Samantha Jane Emery

MACQUARIE University

This thesis is presented for the award of the degree of

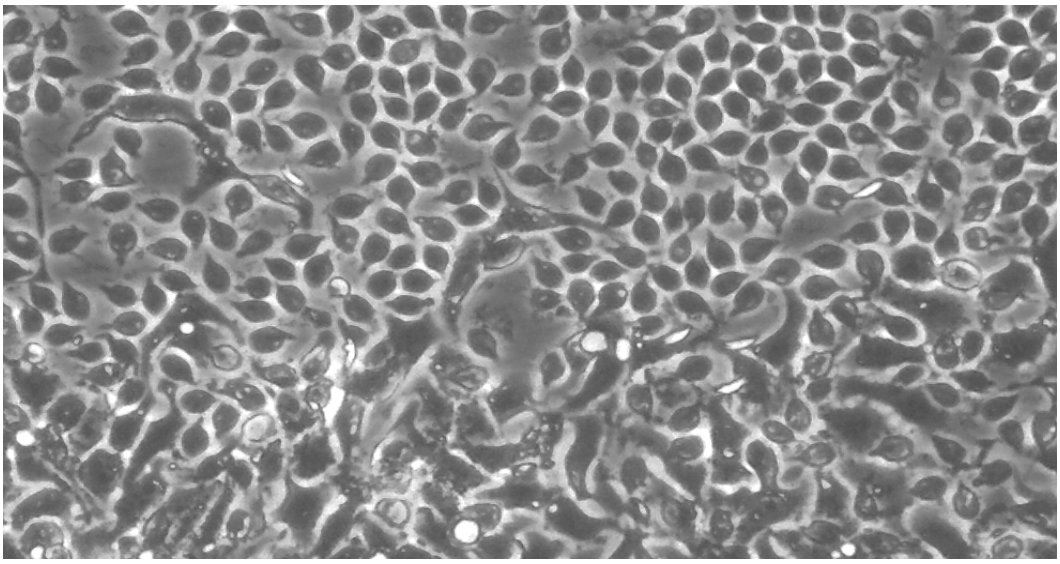
Doctor of Philosophy

Department of Chemistry and Biomolecular Sciences

Faculty of Science and Engineering

Macquarie University, Sydney, New South Wales, Australia

December 2015



G. duodenalis interacting with HT-29 intestinal epithelial cells (2015)

"All the particles aforesaid lay in a clear transparent medium, wherein I have sometimes also seen animalcules a-moving very prettily; some of 'em a bit bigger, others a bit less, than a blood globule, but all of one and the same make. Their bodies were somewhat longer than broad, and their belly was flatlike, furnisht with sundry little paws, where with they made such a stir in the clear medium and among the globules that you might e'en fancy you saw a pissabed [dandelion] running up against a wall; and albeit they made a quick motion with their paws, yet for all that they made but slow progress."

- Antony van Leeuwenhoek, description of *Giardia* in 1681

Table of Contents

Thesis Declaration

I certify that this thesis titled '**Quantitative proteomic analyses of isolate variation and virulence in *Giardia duodenalis***' is my original research. Some parts of this research were achieved in collaboration with other researchers; any form of assistance received from others has been duly acknowledged and their contribution recognised. All sources of information and cited material have been referenced in this thesis. No part of this thesis has been submitted to any other institution as a component of a degree or award. I also certify that this thesis contains no information that has been formerly written or published by anyone else except where due reference is cited in the text.

Samantha Jane Emery

December 2015

Acknowledgements

It is a truth universally acknowledged that a single PhD student in possession of a good thesis must be in want of a night's sleep. However, before then, it is important to recognise the efforts of multiple individuals behind the scenes who have been collaborating, mentoring, assisting, calming, urging, and telling me to 'lab less, write more'. Without the professional assistance, the personal camaraderie which I have received over these recent years, I doubt I would be submitting a thesis of this calibre today. Though there are many people I am grateful to, there are some to which I am particularly indebted and wish to specifically acknowledge.

Firstly, thanks goes to my principle supervisor, Professor Paul Haynes. Thank you for giving me the guidance and yet the freedom to be the architect and the engineer of my PhD project while being available to support and mentor me along the way. Thanks for the jokes, wine recommendations and being encouraging of whatever crazy colour clothes/lipstick/hair I rocked up to the lab with each day. I hope you continue to have fun teaching 'Functional Proteomics' with protein from my cute little trophozoites for many more years. Thanks also to my secondary supervisor Professor Ian Paulsen, for comments and instruction when needed and being available for when I required help.

A great deal of gratitude is owed my off-the-books, unofficial supervisor Dr Ernest Lacey, for his ongoing support and instruction as well as marvellous brand of philosophy and witticisms on academia, science, and research. Thank you for also opening up MST and its resources throughout my PhD, and letting me be a 'fly-in, fly-out' researcher. More thanks goes out to the whole staff of MST, in particular Daniel, Andrew and Alastair who have had to put up with me the longest, and therefore have been victim to the majority of my demands.

Thanks to my labmates in the Haynes lab at Macquarie University. Thank you to Iniga, who only ever left me alone to get married, and for being the buddy stuck in the same boat (whether sinking or sailing) for the last few years. Thanks to the post-docs for all their wisdom and help, Steve for teaching me how to rig up the mass spec as a newbie and Mehdi for sharing new and exciting techniques of labelled proteomics that kept me distracted and keen even until thesis submission. Thanks also to the old members Karlie and Sri, and good luck to the newer members Yunqi, Vineet and David 'Self-Sufficiency' Handler. Thanks to past and present PhD students throughout the department for the friendship and support,

in particular to Shabnam for sharing and office and therefore all my highs and lows, not to mention lunches. Thanks to my friends outside of university, especially Judy, for helping remind me there is a life outside work.

Most of all, thanks to my family. Thanks to my mum, Catherine, for her love, support and solidarity. Thanks for all the little things, and thanks even more for the big things to keep me afloat. To my Dad, David, thanks for being there personally and professionally, listening to me when I talked parasite and proteomics, and always being proud in the crowd during my presentations at our shared conferences. Thanks to my brothers, Tim and Nathan, for going first through your PhD's so I could learn pitfalls in advance, and for your special brand of brotherly 'affection'. Thanks to my sister-in-law, Darci, for providing an extra pillar in the family support platform. Last, but not least, thank you to my partner Tom: thanks for your enthusiasm, support and patience, thanks for lifting me up and telling me I'm amazing when times were tough and providing all the love you had.

Thesis Abstract

Giardia duodenalis is a parasitic protozoan with a global human infection burden of 250 million, and is therefore the largest parasitic cause of diarrheal disease worldwide. Though some cases are asymptomatic, giardiasis can be acute and chronic, with post-infection sequellae including irritable bowel syndrome, chronic fatigue, obesity and type II diabetes. Importantly, *Giardia* is problematic in children under the age of five, causing ill-thrift and failure-to-thrive. In addition, diarrheal diseases including *Giardia* constitute the second-leading cause of mortality for this age category. *Giardia* has a direct life cycle, where infective, tetranucleated cysts are transmitted via the faeco-oral route, and then excyst in the duodenum into virulent, flagellated trophozoites. The prevalence of the parasite is also due to its wide host range, with zoonotic transfer from wild, livestock and domestic animal species to humans. Efforts continue to define the mechanisms of virulence and pathophysiology, as more research is needed to elucidate the relationship between host and parasite factors.

Advances in genetic epidemiology have defined clear assemblages that segregate phylogenetically according to host range, and multiple assemblage and subassemblage genome sequences are now available. These genome sequences have provided the databases necessary for bottom-up, or shotgun, proteomics, and as such have expanded possibilities for quantitative analyses in this parasite. This thesis aimed to provide a thorough quantitative proteomic foundation to enhance the *Giardia* research field both biologically and technically. To achieve this, the thesis consists of four experimental investigations into aspects of parasite variation and virulence, all of which have generated quantitative proteomic data.

Firstly, two different protein sample preparation and fractionation methods were compared for label-free quantitative proteomics. These were applied to two *G. duodenalis* assemblage A1 isolates with different phenotypes, in order to investigate possible sources of isolate

variation. The optimised protocol generated from this initial investigation was applied in later studies, which are also contained within this thesis. In addition, phenotypes associated with pathogenicity correlated with up-regulation of known virulence factors in *Giardia*.

Following this initial investigation, quantitative data was generated using the same label-free approach for eight assemblage A isolates, which constituted the first comprehensive proteomic baseline for this taxonomic group. Isolates of diverse host, geographic and subassemblage origins were analysed using mass spectrometry to characterise their common proteomes and isolate-specific variations. In addition, both the A1 and A2 subassemblage genome databases were evaluated for peptide to spectrum matching, which demonstrated the importance of subassemblage databases to improve identifications from the *Giardia* variable genome.

The third study investigated isolate variation in the biological context of the process of differentiation in *G. duodenalis*. Label-free quantitative proteomics was used to analyse the proteomes of cysts and trophozoites from two genome-alternate subassemblage A1 isolates. This is the first post-genomic analysis of the life cycle beyond the genome isolate, WB. A range of isolate-independent, universal encystation markers were identified, as well as several indications of isolate-specific life-cycle adaptations which may impact reinfection success in subsequent generations.

Finally, the last experiment in this thesis investigated disease induction using *in vitro* host-parasite interaction models between intestinal epithelial cell (IEC) lines and trophozoites. We used isobaric Tandem Mass Tags (TMT) to sensitively quantitate changes in trophozoites which were either allowed to attach to host-cell monolayers, or were exposed to host-cell secretions alone. This is the first use of TMT label technologies for quantitative proteomics in *Giardia*. This has demonstrated that distinct protein cascades are induced by both levels of host-signals, and also that induction of virulence factors is not dependent on parasite attachment to host cells.

Through these experiments, this thesis demonstrates that a range of quantitative proteomic approaches are suitable for *G. duodenalis,* all of which are capable of providing important insight into key aspects of parasite biology. These studies provide an important proteomic complement for genomic and transcriptomic data currently available in the literature, which is necessary for undertaking a systems biology approach to understanding *Giardia*.

List of Manuscripts

Publications listed in chronological order.

I. Neilson, K. A., George, I. S., Emery, S. J., Muralidharan, S*., et al.*, Analysis of rice proteins using SDS-PAGE shotgun proteomics. *Methods in molecular biology* 2014, *1072*, 289-302.

II. Emery, S. J., van Sluyter, S., Haynes, P. A., Proteomic analysis in *Giardia duodenalis* yields insights into strain virulence and antigenic variation. *Proteomics* 2014, *14*, 2523-2534. ***(Chapter 2)***

III. Emery, S. J., Lacey, E., Haynes, P. A., Quantitative proteomic analysis of *Giardia duodenalis* assemblage A: A baseline for host, assemblage, and isolate variation. *Proteomics* 2015, *15*, 2281-2285. ***(Chapter 3)***

IV. Emery, S. J., Lacey, E., Haynes, P. A., Data from a proteomic baseline study of Assemblage A in Giardia duodenalis. *Data in brief* 2015, *5*, 23-27. ***(Chapter 3)***

V. Emery, S. J., Pascovici, D., Lacey, E., Haynes, P. A., The generation gap: Proteome changes and strain variation during encystation in *Giardia duodenalis*. *Molecular and biochemical parasitology,* 2015, *201*, 47-56. ***(Chapter 4)***

VI. Emery, S. J., Mirzaei, M., Vuong, D., Pascovici, D., Chick, J. M., Lacey, E., Haynes, P. A., Induction of virulence factors in *Giardia duodenalis* independent of host attachment. *Scientific reports*, 2016, *6*, 20765 ***(Chapter 5)***

VII. Emery, S.J., Lacey, E., Haynes, P.A., Quantitative proteomics in *Giardia duodenalis* – current achievements and future challenges, *Molecular and Biochemical Parasitology*, 2016 (Accepted 18/07/16) ***(Chapter 1)***

Conference Presentations

1. Emery, S.J., van Sluyter, S. & Haynes, P.A. (2012) Shotgun proteomics in *Giardia duodenalis* for strain variation and virulence, *17th Lorne Proteomics Symposium,* Lorne, Australia, 2-5 February 2012 (Poster)

2. Emery, S.J., van Sluyter, S. & Haynes, P.A. (2012) Proteomic analyses in Giardia duodenalis: insights for virulence and variation, 2012 ASP Annual Conference, Launceston, Australia, 2-5 July 2012 (Oral Contributed Paper)

3. Emery, S.J., Lacey, E. & Haynes, P.A. (2013) The Generation Gap: a longitudinal study of proteome changes in *Giardia duodenalis* cysts and trophozoites, *18th Lorne Proteomics Symposium,* Lorne, Australia, 7-10 February 2013 (Poster)

4. Emery, S.J., Lacey, E. & Haynes, P.A. (2013) Quantitative proteomics in *Giardia duodenalis*: investigating host origin, assemblage and isolate variation, *2013 World Association for Veterinary Parasitology,* Perth, Australia, 25-29 August 2013 (Oral Contributed Paper) ** WAAVP/ASP award for most meritous Student Oral Presentation, Perth (2014)*

5. Emery, S.J., Lacey, E. & Haynes, P.A. (2014) Label-free quantitative proteomics in *Giardia duodenalis*, *5th International Giardia and Cryptosporidium Conference,* 27-30 May, Uppsala, Sweden 2014 (Oral Contributed Paper)

6. Emery, S.J., Vuong, D., Gill, J., & Lacey, E. (2014) GiaTox: a simple chromagenic method of screening drug susceptibility in *Giardia duodenalis*, *5th International Giardia and Cryptosporidium Conference,* 27-30 May, Uppsala, Sweden 2014 (Oral Contributed Paper)

7. Emery, S.J., Lacey, E. & Haynes, P.A. (2014) The Generation Gap: a longitudinal study of proteome changes in *Giardia duodenalis* cysts and trophozoites, 2014 50th Anniversary ASP Annual Conference, Canberra, Australia, 30th June – 3rd July 2014 (Oral Contributed Paper)

8. Emery, S.J., Mirzaei M., Vuong, D., Chick, J.M., Lacey, E., & Haynes, P.A. (2014) Quantitative proteomic analysis of *Giardia duodenalis* during host-cell interactions, 2014 2nd Proteomics and Beyond Symposium, Sydney, Australia 2014 (Poster)

9. Emery, S.J., Mirzaei M., Vuong, D., Lacey, E., & Haynes, P.A. (2014) Quantitative proteomics in *Giardia*, 2014 2nd Proteomics and Beyond Symposium, Sydney, Australia 2014 (Oral Contributed Paper)

10. Emery, S.J., Mirzaei M., Vuong, D., Chick, J.M., Lacey, E., & Haynes, P.A. (2014) Discovery of viruence factors in *Giardia duodenalis* host-cell interactions, *20th Lorne Proteomics Symposium,* Lorne, Australia, 5-8 February 2015 (Poster)

11. Emery, S.J., Lacey A.E., Vuong, D., Crombie, A., Lacey, E. & Piggott, A.M. (2015) Towards the discovery of novel sites of anthelmintic action, 2015 NZSP & ASP Annual Conference, Auckland, New Zealand 2015 (2min oral and poster presentation)

12. Emery, S.J., Mirzaei M., Vuong, D., Pascovici, D., Chick, J.M., Lacey, E., & Haynes, P.A. (2014) Differential stimulation of *Giardia duodenalis* trophozoites between host soluble signals and host cell attachment during in vitro interactions, 2015 NZSP & ASP Annual Conference, Auckland, New Zealand 2015 (Symposium Presentation)

Abbreviations

1-DE	One Dimensional Gel Electrophoresis
2-DE	Two Dimensional Gel Electrophoresis
ADI	Arginine Deiminase
ACN	Acetonitrile
ANK	Ankyrin
BCA	Bicinchoninic Acid
CI	Co-incubation
CW	Cyst Wall
CWP	Cyst Wall Protein
DDA	Data-Dependent Acquisition
DTT	Dithiolthreitol
DIA	Data-Independent Acquisition
EDTA	Ethylenediaminetetraacetic Acid
ERIF	Environmentally Resistant Infective Form
ER	Endoplasmic Reticulum
ES	Enrichment Score
ESI	Electrospray Ionisation
ESV	Encystation Specific Vesicle
EV	Extracellular Vesicle
FASP	Filter Aided Separation of Proteins
FDR	False Discovery Rate
FeS	Iron Sulphur
GalNAc	β(1-3)-N-acetyl-d-galactosamine
GFP	Green Fluorescent Protein
GO	Gene Ontology
GPF	Gas Phase Fractionation

GPM	Global Proteome Machine
HAT	Histone Acetylase
HCD	High Energy Collisional dissociation
HCMP	High Cysteine Membrane Protein
HDAC	Histone Deacetylase
HPLC	High Performance Liquid Chromatography
HSF	Host Soluble Factors
HSP	Heat Shock Protein
ICAT	Isotope Coded Affinity Tags
IEC	Intestinal Epithelial Cell
IL	Interleukin
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
LC-MS/MS	Liquid Chromatography Tandem Mass Spectrometry
LTQ	Linear Trap Quadrapole
MALDI	Matrix Assisted Laser Desorption/Ionisation
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
m/z	Mass to Charge ratio
NanoLC-MS/MS	Nanoflow Liquid Chromatography Tandem Mass Spectrometry
NEK	NIMA related Kinase
NO	Nitric Oxide
NSAF	Normalised Spectral Abundance Factor
nSpC	Normalised Spectral Counting
OCT	Ornithine Carbamoyl Transferase
ORF	Open Reading Frame
OST	Oligosaccharyltransferase
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline

P.I	Post Induction
PDI	Protein Disulfide Isomerase
PDX	Peroxiredoxin
PFOR	Pyruvate flavodoxin oxidoreductase
PLP	Pyridoxamine phosphate
PRIDE	PRoteomics IDEntifications
PTM	Post Translation Modification
PV	Peripheral Vesicle
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
SAGE	Serial Analysis of Gene Expression
SDS	Sodium Dodecyl Suphase
SCX	Strong Cation Exchange
SILAC	Stable Isotope Labels with Amino Acid in Culture
SNP	Single Nucleotide Polymorphism
SpC	Spectral Count
SRM	Selected Reaction Monitoring
SWATH	Sequential Window Acquisition of all Theoretical mass spectra
TCP-1	Chaperonin T Complex 1
TFE	Trifluoroethanol
TMT	Tandem-Mass-Tag
UBCE	Ubiquitin Carrier Enzyme
UniProt	The Universal Protein Resource
UPL-1	Uridine phosphorylase 1
VSP	Variant-specific Surface Protein
WGA	Wheat Germ Agglutinin

CHAPTER 1

A comprehensive literature review of the proteomic studies currently available in Giardia duodenalis. *This review details the biological outcomes of proteomic analyses in* Giardia*, as well as evaluates the quantitative technologies utilised for these studies.*

1. Quantitative proteomics in *Giardia duodenalis* –achievements and challenges.

Each experimental chapter of this thesis contains an individual introduction detailing the contents and relevant literature associated with the chapter. In ***Chapter 1*** a broader view of proteomics in *Giardia* is presented, with a particular focus on post-genomic studies and quantitative technologies.

Chapter 1 has been accepted as a review manuscript for submission to *Molecular and Biochemical Parasitology*. The full reference list for ***Chapter 1*** can be found in the bibliography at the end of this thesis.

In terms of contributions, I was responsible for writing the manuscript, with assistance in editing by Professor Paul Haynes and Dr Ernest Lacey.

1.1 Abstract

Giardia duodenalis is a protozoan parasite of vertebrates and a major contributor to the global burden of diarrheal diseases and gastroenteritis. The publication of multiple genome sequences in the *G. duodenalis* species complex has provided important insights into parasite biology, and hence made post-genomic technologies, including proteomics, significantly more accessible. The aims of proteomics are to identify and quantify proteins present in a cell, and assign functions to them within the context of dynamic biological systems. In *Giardia*, proteomics in the post-genomic era has transitioned from reliance on gel-based systems to utilisation of a diverse array of techniques based on LC-MS/MS technologies. Together, these have generated crucial foundations for subcellular proteomes, elucidated intra- and inter-assemblage isolate variation, and identified pathways and markers in differentiation, host-parasite interactions and drug resistance. However, proteomics in *Giardia* is an emerging field, with considerable shortcomings evident from the published research. These include a bias towards assemblage A, a lack of emphasis on quantitative analytical techniques, and limited information on post-translational protein modifications. Additionally, there are multiple areas of research for which proteomic data is not available to add value to published transcriptomic data. The challenge of amalgamating data in the systems biology paradigm necessitates the further generation of large, high-quality and quantitative datasets to accurately model parasite biology. Furthermore, the next generation of quantitative proteomics will continue to provide data of enormous relevance to the understanding of *Giardia* and giardiasis, but only if the experiments involved are well-designed and targeted. This review surveys the current proteomic research available for *Giardia* and evaluates their technical and quantitative approaches, while contextualising their biological insights for parasite pathology, isolate variation and eukaryotic evolution. Finally, we propose areas of priority for future proteomic research to explore fundamental questions in *Giardia*, including the analysis of

post-translational modifications, and the design of MS-based assays for validation of differentially expressed proteins in these large datasets.

1.2 Introduction

Giardia duodenalis is an anaerobic gastrointestinal protozoan, and is the most common cause of parasitic diarrheal disease in humans. *G. duodenalis* transmission occurs via faecal-oral direct contact, and posseses a simple, direct lifecycle of environmentally resistant cysts and flagellated trophozoites which colonise the small intestine after ingestion [1]. It is estimated that *G. duodenalis* afflicts 250 million people worldwide at any one time, with high prevalence in a variety of wild, domestic and livestock animals [2]. Such estimates likely underestimate rates of *G. duodenalis* infection, with up to half of infections presenting as asymptomatic. Though clinical giardiasis is often self-limiting, infections can be chronic with long-term clinical sequellae, including failure-to-thrive in children, induction of irritable bowel syndrome (IBS), arthritis and chronic fatigue [3-5]. Though our understanding of how *G. duodenalis* causes disease has improved, it remains problematic to define pathophysiology between parasite, commensal bacteria and host [6, 7]. There are also significant gaps in defining the scale of variation in *G. duodenalis*. These include differences in virulence [8], infectivity [9], and host range [2], and additional difficulties in relating these back to variations observed at the genome level.

Research into *G. duodenalis* has utilised an array of post-genomic technologies as they have improved in availability, precision and resolution. These tools explore parasite biology and disease pathogenesis on a global scale, using quantitative techniques to identify and characterise pathways, biomarkers and virulence factors. This has been facilitated by the publication of multiple genome sequences for *G. duodenalis*, which have provided the scaffold to underpin large datasets from genetic, transcriptomic and proteomic experiments. The ultimate approach involves systems biology; compiling large, high quality datasets from DNA, RNA and proteins to enable systems modelling for understanding *Giardia* biology globally [10, 11]. While quantitative proteomics is indispensable for systems biology, it also remains unparalleled in identifying, exploring and placing functional relevance in biological scenarios.

A previous review by Steuart [12] focused on the emerging shift between pre- and post-genomic proteomics in *Giardia*, and also captured the concomitant transition in the wider proteomic field from gel-based systems, to LC-MS/MS technologies [13, 14]. Though gel-based technologies will always play an important role in proteomics, they are now more commonly employed on a case-specific basis, such as absence of a sequenced genome. There are several common criticisms of two dimensional gel electrophoresis (2-DE) specifically that they are more error-prone and less reproducible in quantitative scenarios [15], such that liquid chromatography tandem mass spectrometry (LC-MS/MS) proteomic methods are now more favoured. LC-MS/MS allows multidimensional HPLC to be coupled to a mass spectrometer for separation of complex samples prior to MS analysis [16], which is now a fundamental technique in bottom-up, or shotgun, proteomics. Shotgun quantitative proteomics encapsulates a diverse collection of methodologies for defining differences between biological states, samples and systems. Quantitative methods include both label-free quantitative proteomics [15] as well as labelled strategies such as stable isotope labels with amino acids in culture (SILAC) [17], isotope coded affinity tags (ICAT) [18], isobaric tags for relative and absolute quantitation (ITRAQ) [19], and tandem mass tags (TMT) [20], which multiplexes to 10 channels of labelling for quantitative comparison of multiple samples [21]. Significant inroads have also been made towards identification and analysis of post-translational modifications (PTMs) [22, 23], which is particularly relevant but under-represented in published research in both *Giardia* and other enteric protozoa [10]. A more detailed explanation of each of these proteomic technologies is available in an earlier review by Wastling [10], which also describes their application in studies within the broader field of parasitology.

From these recent advances in proteomic technologies, as well as emphasis on the new paradigms surrounding research in a systems biology setting, we believe a review of *G. duodenalis* proteomics is timely. The recent availability of multiple genome sequences for *G. duodenalis* has removed the primary obstacle to LC-MS/MS quantitative proteomics.

Hence, new studies utilising these technologies have emerged in the literature since the publication of previous reviews in *Giardia* proteomics [12], and proteomics technologies more broadly in parasitology [10]. However, considerable gaps remain in both the technologies accessed, as well as the biology investigated, which will be evaluated throughout this review. We also feel it necessary to acknowledge the significant advances in the literature regarding *Giardia* transcriptomics, particularly to highlight complementary proteomic data where available, or notably absent. A well-designed quantitative proteomic experiment presents the opportunity to explore and evaluate *Giardia* biology, including epidemiology, differentiation and pathogenesis. In light of this biological resolution, this review will evaluate recent achievements in the field, and prioritise the next generation of proteomic data necessary for *G. duodenalis*.

1.3 Genomes and Annotation

As LC-MS/MS quantitative proteomics requires an appropriate database for peptide to spectrum matching, it is important to understand the difference between available genome sequences for *G. duodenalis*. Molecular epidemiology in *G. duodenalis* have reclassified it as a species complex containing eight genetically defined assemblages [24]. Each of these assemblages (A-H) are most commonly defined by multilocus sequencing, and segregate phylogenetically according to host species and host specificity [2, 25]. Assemblages A and B are both infective for humans, with assemblage A possessing a wide host range including domestic and wild animals [2]. More recently, assemblages may be more specifically classified into subassemblages, which are also defined genetically and correlate with more precise host ranges, for example A1 and A2 subassemblages [2].

There are currently 5 genome sequences published for *G. duodenalis*, all of which are available and consistently updated through the central database of Giardiadb.org [26]. Genome sequences are available for both the A1 subassemblage from the isolate WB [27]

and, more recently, the A2 subassemblage from isolate DH [28]. Importantly, the A1 genome from isolate WB has also been physically and optically mapped [29, 30]. Two genome sequences are available for assemblage B, firstly the draft sequence originally produced for the GS isolate [31], and then the recent re-published GS genome sequence with greater coverage and depth [28]. There is also a genome sequence available for the artiodactyl-specific assemblage E from isolate P15 [32]. A summary of the genomes and their protein-coding features can be found in Table 1, and a detailed analysis of genome characteristics is available in Adam *et al*. [28]. In addition to these, two new isolates of assemblage A2 have been very recently sequenced and comparatively analysed [33], and previously the WB and Portland A1 isolates have been sequenced [34], although both these studies are yet to result in curated databases for use for –omics technologies.

Several comparative analyses for the *G. duodenalis* genomes outline the variation in support of their classifications as separate subspecies [28, 32, 33]. This is especially relevant in the context of their treatment as distinct databases for proteomics. Between the four available genome sequences (A1, A2, B and E) there are a total of 4 097 common ORFs that represent core orthologs of shared functions for essential roles for *G. duodenalis* biology [28]. However, *G. duodenalis* also possesses several multigene families that comprise what is termed the 'variable genome', consisting of the variant-specific surface proteins (VSPs), high cysteine membrane proteins (HCMPs), ankyrin-repeat proteins (formerly 21.1 proteins) and the NIMA related kinases (NEK) [32]. Many of the genes from the variable genome families, especially the VSPs, show diversification between the assemblages and subassemblages, as well as isolates in some assemblages [28, 33]. Overall, the two subassemblage A genomes sequences are more similar in ORF comparison, synteny and phylogeny to assemblage E than assemblage B genome sequences [28]. This supports the hypothesis that assemblages A and B represent two species, despite both being infective for humans [31].

Table 1: Comparative summary of protein coding and annotation information of the five sequenced genomes available as databases for *G. duodenalis*. The presence of mass spectrometry evidence for each of the assemblage or subassemblage genomes since publication is listed, as well as whether any RNA transcript data has also been produced.

Genome	**Assemblage**	**Protein Coding ORFs**	**Protein Coding ORFs**		**Proteomic Analysis***		**Transcript Data**
			Assigned Function	**No Function**	**MS Evidence**	**Quantitative**	
WB	A1 [27]	5901	2905 *(49.2%)*	2996 *(50.8%)*	Y	Y	Y
DH	A2 [28]	6724	2900 *(43.1%)*	3824 *(56.9%)*	Y	Y	Y
GS	B [31]	4470	2842 *(63.6%)*	1648 *(36.4%)*	N		Y
GS	B [28]	7477	3946 *(52.8%)*	3531 *(47.2%)*	N		Y
P15	E [32]	5008	2752 *(55.0%)*	2253 *(45.0%)*	N		Y

* This takes into account post-genomic proteomic analyses of *G. duodenalis* specifically utilising the genomes as databases.

Although the WB genome was originally used as representative for the entirety of assemblage A, multilocus sequencing showed divergence between A1 and A2 subassemblages [2], and multiple genomes from A1 and A2 isolates indicate subassemblages also may represent cryptic species [33]. Epidemiological evidence also highlighted that human infections were largely caused by the A2 subassemblage, while the A1 subassemblage possessed wide host range and transmission between humans and animals [2]. Increasing evidence of intra-assemblage genome variation is also available, from the currently two A1 [34] and three A2 isolates [33] that have been sequenced. From these analyses, the A1 subassemblage appears to be more conserved than the A2 sub-assemblage, with 7.5 single nucleotide polymorphisms (SNPs) per 100 000 between A1 genomes [34], compared to 350 SNPs per 100 000 between the two recently sequences A2 genomes [33]. For the A2 isolates, the most significant genes of variation were cysteine rich membrane proteins of the VSP and HCMP families.

Currently, there is mass spectrometry data available mainly from studies of the proteome for Assemblage A, predominantly the A1 subassemblage. This is largely attributable to the assemblage A genome as the first published [27], but also due to the historical precedence and prevalence of the WB isolate (ATCC 30957, 1979) and its clonally-derived WB C6 (ATCC 50803, 1983). However, there is a lack of mass spectrometry data available for the other assemblages B and E. There are early protein studies covering different assemblages prior to reclassification the *G. duodenalis* complex [35], and 2-DE analyses which pre-date the genome sequence for assemblage B [36], therefore no post-genomic, quantitative proteomics data are available for either assemblage B or E. A single quantitative proteomic study has been performed on Assemblage D trophozoites and cysts, but there is no assemblage-specific genome sequence available as a database [37]. By comparison, all published genome sequences have transcriptomic data available, at the very least for support of ORF assignment during genome annotation [38]. This highlights a significant

gap in the proteomic data for assemblages other than assemblage A, with no mass spectrometry based evidence available for ORF verification in the other *G. duodenalis* genome sequences. Importantly, with the demonstration of considerable inter-assemblage genetic divergence [28], there is a necessity to analyse proteomes of all assemblages to ensure understanding of *G. duodenalis* is not biologically biased for assemblage A.

Another significant issue, for post-genomic *Giardia* proteomics as well as other disciplines in systems biology, remains the functional annotation of genomes. As indicated in Table 1, there are significant numbers of protein coding ORFs without function, or 'hypothetical proteins', in all available genomes sequences for *G. duodenalis*. This includes specifically half of all proteins for WB C6, which is has been utilised the most in published *Giardia* proteomic studies. Proteomics analyses of *G. duodenalis* have already confirmed that many of these are indeed functional proteins expressed in the parasite, but no annotations are available to further analyse their biological role, nor additional functional information such as localisation, or protein-protein interaction data. For those proteins for which functional data is available, these are also spread across multiple annotation databases. Many proteins, such as members of the Protein 21.1/Ankyrin repeat and VSPs families, lack GO annotations but contain Interpro protein domain or fold information. Therefore functional analysis tools which include multiple annotation databases are preferred over GO analyses alone [39, 40]. There is also promise in using structural homology tools for inferring protein function including software such as I-TASSER [41, 42], RaptorX [43], Phyre2 [44] and HHPred [45], which offer an alternative route to traditional methods of sequence homology tools such as BLAST. The functional annotation of *Giardia* genomes, as in most genomes, is an ongoing effort; more comprehensive annotations will significantly improve the biological insights of proteomic analysis in this parasite.

1.4 Quantitative Proteomics

G. duodenalis is suitable for analyses using quantitative proteomics to study its biology, pathophysiology and unique evolutionary origins for a variety of reasons, including that it possesses a range of laboratory and in vitro systems for its biological processes, which can be explored using proteomics technologies. *Giardia* can be axenised into culture as trophozoites reproducing by binary fission, and thus its differentiation processes can be modelled *in vitro* [46]. It has also been demonstrated that *Giardia* can be co-incubated with human cell lines as *in vitro* models of host-cell attachment and disease induction [47, 48]. Once in culture, *G. duodenalis* isolates can be exposed to sub-lethal concentrations of therapeutic drugs to generate resistant isolates for metronidazole [49], quinacrine [50] and benzimidazoles [51] for the study of drug discovery, mode of action and resistance. Aside from the ability to model important biological process *in vitro*, there is also a considerable amount of information regarding *Giardia* biology, including metabolism and pathology studies, as well analyses of structure and cellular organisation [52]. Finally, the evolutionary and phylogenetic origins of *Giardia* mean it can be utilised as a model species for processes and structures within other eukaryotes [53]. By merit of its biology as well as versatility *in vitro*, *G. duodenalis* is a logical candidate for quantitative proteomic analyses to address biological or clinical questions.

An important resource for *Giardia* proteomics is its central database, Giardiadb.org [26]. Giardiadb.org is an integrated tool which combines DNA, RNA and protein level information, and is an essential step for systems biology in this parasite [10]. Proteomics datasets can be submitted to Giardiadb.org, which links back to functional databases such as UniProt and NCBI, as well as annotation tools including Gene Ontology (GO) and Interpro. Further, genes are available to search with integrated information including genomic position, functional annotation, and protein information, as well as experimental, quantitative evidence of expression. This centralised, combined and comprehensive

database allows wider dissemination of proteomics data in context to similar –omics data generated in *Giardia* research.

The following sections will review key areas of *G. duodenalis* biology, in particular post-genomic contributions which utilise shotgun proteomics. A summary of currently published proteomic studies in *G. duodenalis* is presented in Table 2. It is important to note that some of the following sections may feature limited, or even no, available proteomics data. These significant gaps will be examined and evaluated to highlight those crucial areas where the next generation of proteomic data must be generated. Complementary data, in particularly quantitative transcriptomics, will be provided where applicable

Table 2: Summary of post-genomic, proteomic studies in *G. duodenalis*; for a summary of earlier pre-genomic studies refer to Steuart, [12]. Isolate and assemblage and source of protein are indicated, along with proteomic sample method used and whether results utilised any method for quantitation of proteins identified. The last column indicates the database used to search results.

Reference	Isolate and Assemblage	Protein Source	Category	Method	Quantitation	Databases
[81] Ringvist *et al*, 2008	WB C6 (A1)	Media Fraction	Host-Parasite Interactions	2DE Gel	Protein IDs*	www.mbl.edu/Giardia
[36] Steuart *et al*, 2008	BAH 2c2, 26c11, 40c9 (A) BAH 34c8, 12c14, 15c1 (B)	Trophozoites	Strain Variation	2DE Gel	Protein IDs*	NCBInr Eukaryotic
[158] Ratner *et al*, 2008	WB (A1) MR4	Trophozoites Cysts	Differentiation PTM (Glycosylation)	WGA Affinity LC-MS/MS	Protein IDs Peak Area	*Giardia*DB N/A (WB C6)
[106] Kim *et al*, 2009	WB (A1)	Trophozoites Cysts	Differentiation	2DE Gel	Spot Density	*Giardia*DB N/A (WB C6)
[138] Alvarado & Wasserman, 2010	WB C6 (A1)	Trophozoites Cysts	Differentiation PTM (Phosphorylation)	2DE Gel	Protein IDs	NCBInr Eukaryotic
[60] Hagen *et al*, 2011	WB C6 (A1)	Ventral Disc	Cytoskeleton/Organelle	LC-MS/MS	Protein IDs	GenBank (*G. duodenalis)*
[19] Jedelsky *et al*, 2011	WB (A1)	Mitosomes	Cytoskeleton/Organelle	LC-MS/MS	ITRAQ	*Giardia*DB-1.3 (WB C6)
[61] Lauwaet *et al*, 2011	WB C6 (A1)	Basal Bodies	Cytoskeleton/Organelle	LC-MS/MS	Protein IDs	*Giardia*DB-1.4 (WB C6)
[59] Jerlstrom-Hultqvist *et al*, 2012	WB C6 (A1)	Proteasome	Cytoskeleton/Organelle	Protein-tagging LC-MS/MS	Protein IDs	NCBInr Eukaryotic
[37] Lingdan *et al*, 2012	Changchun (D)	Soluble Fraction Trophozoite	Differentiation	LC-MS/MS	ITRAQ	*Giardia*DB/Uniprot
[69] Lourenco *et al*, 2012	WB (A1)	Ventral Disc	Cytoskeleton/Organelle	1DE & 2DE Gel	Protein IDs	NCBInr Eukaryotic
[140] Lalle *et al*, 2012	WB C6	14-3-3 Protein-Binding	Differentiation PTM (Phosphorylation)	Protein-tagging LC-MS/MS	EmPAI	*Giardia*DB-1.2 (WB C6)
[124]Paz-Maldonado *et al*, 2013	WB C6 (A1) WB C6, Abz Resistant (A1)	Trophozoites	Drug Resistance	1DE & 2DE Gel	Spot Density	NCBInr Eukaryotic
[166] Nino *et al*, 2013	WB/9B10	Trophozoites	PTM (Ubiquitination)	Protein-tagging LC-MS/MS	Protein IDs	*Giardia*DB-2.5 (WB C6)

Table 2 (cont.): Summary of post-genomic, proteomic studies in *G. duodenalis*; for a summary of earlier pre-genomic studies refer to Steuart, [12]. Isolate and assemblage and source of protein are indicated, along with proteomic sample method used and whether results utilised any method for quantitation of proteins identified. The last column indicates the database used to search results.

[39] Faso *et al*, 2013	WB C6 (A1)	Trophozoites	Differentiation	LC-MS/MS	Spectral Counting	*Giardia*DB-1.2 (WB C6)
[58] Wampfler *et al,* 2013	WB C6 (A1)	PV & ESV	Cytoskeleton/Organelle	LC-MS/MS	Protein IDs	*Giardia*DB-2.2 (WB C6)
[75] Paredez *et al*, 2014	WB C6 (A1)	Actin-Binding	Cytoskeleton/Organelle	Protein-tagging LC-MS/MS	Protein IDs	*Giardia*DB N/A (WB C6)
[101] Emery *et al*, 2014	BRIS/95/HEPU/2041 (A1) BRIS/82/HEPU/106 (A1)	Trophozoites	Strain Variation	LC-MS/MS	Spectral Counting	*Giardia*DB-2.5 (WB C6)
[100] Emery *et al*, 2015	BRIS/83/HEPU 106 (A1) BRIS87/HEPU/713 (A1) OAS1 (A1) Bac2 (A1) BRIS/95/HEPU/2041 (A1) BRIS/89/HEPU/1065 (A1) WB (A1) BRIS/89/HEPU/1003 (A2)	Trophozoites	Strain Variation	LC-MS/MS	Spectral Counting	*Giardia*DB-4.0 (WB C6) *Giardia*DB-4.0 (DH)
[40] Emery *et al*, 2015	BRIS/95/HEPU/2041 (A1) BRIS/82/HEPU/106 (A1)	Trophozoites Cysts	Differentiation Strain Variation	LC-MS/MS	Spectral Counting	*Giardia*DB-4.0 (WB C6)
[57] Martincova *et al*, 2015	WB (A1)	Mitosomes	Cytoskeleton/Organelle	LC-MS/MS	Protein IDs	NCBInr
[114] Emery *et al*, 2015	BRIS/95/HEPU/2041 (A1)	Trophozoites	Host-Parasite Interactions	LC-MS/MS	TMT	*Giardia*DB-5.0 (WB C6)

* Quantitation was performed using operator-based qualitative assessment, not software or statistical methods

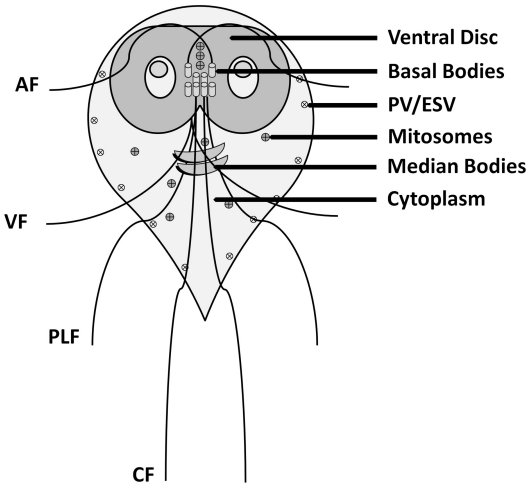
1.4.1 Subcellular Proteomics

Subcellular proteomics allows greater resolution of a targeted cell fraction and analyses of protein function in relation to cell localisation [54, 55]. Subcellular proteomics involves a combination of biochemical techniques for fractionation of protein or organelle complexes, coupled to high resolution mass spectrometry technologies [55]. Since the release of the *Giardia* genome sequence, there have been eight proteomic analyses of cytoskeletal or organelle fractions (Figure 1). These subcellular proteomic analyses have consistently provided protein identifications which greatly expanded the known organelle proteome at the time, often doubling or tripling the number of proteins assigned. These higher-resolution, organelle proteomes have improved mapping of signalling pathways, protein trafficking, sub-localisation and models for cytoskeletal assembly/disassembly during the life cycle. Furthermore, they are an important comparative resource for understanding divergence in organelles and metabolism throughout eukaryotic evolution [56].

Significantly, many of the identifications from proteomics data in several of the studies were excluded, or failed to be identified by, sequence-based predictive bioinformatics [19, 57, 58]. Besides utilising a range of fractionation techniques, these studies also cover multiple strategies for achieving high-stringency datasets, in particular eliminating contamination by non-target proteins or accessory structures. These include *in silico* filtering [58], iTRAQ isobaric ratios [19], homology searching [19, 59], and large-scale immunolocalisation of candidate proteins for confirmation [57, 58, 60, 61].

1.4.2 Cytoskeletal Proteomics

The cytoskeleton of *Giardia* is a predominately microtubule structure, comprised of the ventral disc, median bodies, basal bodies and flagella. The ventral or adhesive disc is a *Giardia*-specific structure responsible for attachment of trophozoites to the intestinal lining, to avoid peristalsis and clearance from the gastrointestinal tract [1, 56]. Previous



Subcellular Fraction	Enrichment Method	# Proteins	Reference
Ventral Disc, Lateral Crest	Detergent Extraction	18	[60] Hagen *et al*, (2011)
Ventral Disc	Detergent Extraction	11	[69] Lourenco *et al* (2012)
Basal Bodies (PFR, Ventral Disc, Median Bodies)	Density Centrifugation	75	[61] Lauwaet *et al* (2011)
Mitosomes	Density Centrifugation ITRAQ Ratio	20	[19] Jedelsky *et al* (2011)
Mitosomes	Pull-Down Assay	13	[57] Martincova *et al* (2015)
Encystation Specific Vesicle (ESV) Peripheral Vesicle (PV)	Flow-Cytometry *In silico* enrichment	ESV – 72 PV – 82	[58] Wampfler *et al* (2014)
Proteasome	Pull-Down Assay	28	[56] Jerlstrom-Hultqvist (2012)
Actin-Binding	Pull-Down Assay	57	[72] Paredez *et al* (2014)

Figure 1: Summary of subcellular proteomics in *G. duodenalis*. Upper panel shows the cell and organelle structure of trophozoites. Flagella pairs are labelled according to anterior (AF), posterior-lateral (PLF), caudal (CF) and ventral (VF). Table below indicates references for proteomic analysis and which cytoskeleton or organelle enriched fraction was analysed. The number of proteins listed in the last column considers those identifications indicated by authors as high-stringent and organelle-specific following confirmatory tests, *in silico* enrichment and contaminant exclusion.

studies of the ventral disc have identified proteins including annexins [62], giardins [63-65], striated fiber-assemblin-like protein (SALP-1) [66], median body protein (MBP) [67], and several kinases, including NEK kinases [68]. Two proteomic analyses of the ventral disc have been performed to date [60, 69]. Moreover, of the 58 putative ventral disc or lateral crest proteins identified by Hagen *et al.* [60], 18 novel disc-associated proteins were confirmed by GFP-tagging and localisation, doubling the known ventral disc proteome, and a total of 33 ventral disc and lateral crest associated proteins were compiled. Furthermore, Hagen *et al.* [60] also provided complementary insight into disc assembly/disassembly

during cell division, encystation and excystation, which had previously been analysed at the transcriptomic level [66].

Basal bodies also contribute to the cytoskeletal structure in *Giardia*, consisting of conserved nine microtubule triplet structures forming a helicoidal structure with a diameter of 250nm [61]. In *Giardia*, there are eight basal bodies corresponding to the four pairs of flagella, though basal bodies also localise to the spindle during mitosis, playing a role during replication [70, 71]. Lauwaet *et al.* [72] utilised subcellular proteomics to analyse the composition of the basal bodies, which previously had only a dozen proteins assigned including tubulins, centrins, and several signalling proteins including an aurora kinase which regulated mitosis [71, 73, 74]. The study combined a bioinformatics approach, to predict homologues in *G. duodenalis* assemblage A, with a proteomic analysis of basal body enriched fractions from interphase trophozoites. This approach identified 75 proteins, of which 65 were novel, and 13 were successfully immunolocalised [61]. Importantly, both novel homologues excluded by bioinformatics analysis were identified by proteomics and localised, while conserved homologues identified by bioinformatics failed to localise to the basal bodies. This emphasises the capability of proteomics to identify unique, novel homologues that may be excluded by sequence-based, predictive bioinformatics.

The other cytoskeletal proteomic analysis currently available in *Giardia* analysed actin-binding proteins in trophozoites, where similarly a lack of canonical actin-binding proteins in *Giardia* limited the use of predictive analyses [75]. Combining affinity purification with LC-MS/MS, a total of 57 actin-binding proteins were identified, covering a range of functions in the cytoskeleton, particularly for the flagella, protein folding and trafficking, as well as nuclear and signalling proteins [75]. Of these, 23 proteins were *Giardia*-specific with unknown functional information, suggesting a novel actin-interacting protein subset [75]. It was also demonstrated that actin is part of the *Giardial* nucleoskeleton, with the

identification of other nuclear proteins suggesting the presence of actin-based chromatin remodelling in *Giardia*.

1.4.3 Organelle Proteomics

There have been several proteomic analyses reported for organelles in *Giardia,* including the mitosomes, secretory vesicles and proteasome. In *Giardia*, mitosomes are highly reduced and adapted forms of mitochondria, having lost their genomes and been largely reduced to iron sulphur (FeS) cluster synthesis and protein import and folding [76-78]. Jedelsky, *et al.* [19] enriched mitosomes via subcellular fractionation, labelled the peptides isobarically and analysed them via LC-MS/MS, after which they used the iTRAQ quantification tags to filter contaminants from the dataset. A total of 139 putative mitosomal proteins were identified, of which 20 out of 44 were localised to the mitosomes. None of these 20 proteins had been identified previously using bioinformatic predictions [19]. Proteins identified were linked to FeS cluster assembly, molecular chaperones and membrane protein import machinery, indicating a greatly reduced proteome, which the authors predicted to be approximately 50-100 proteins [19]. A further 13 proteins were recently identified in the mitosomes by Martincova *et al.* [57] using *in vivo* enzymatic tagging. Five proteins were specific to the outer mitosomal membrane, including the identification of a novel highly diverged TIM homologue. Furthermore, Martincova *et al.* [57] identified several *Giardia*-specific proteins indicative of additional, novel metabolic processes in mitosomes beyond FeS cluster biogenesis. These studies reinforce the utility of proteomics as an exploratory tool for analysing organelles, especially where evolutionary, lineage-specific alterations occur that limit bioinformatics predictions for highly diverged homologues.

The biases in homology-based bioinformatics meant proteomics was also preferred for analysing the secretory vesicles in *Giardia*. Wampfler *et al.* [58] analysed the peripheral

vesicles (PV) and encystation specific vesicles (ESV) in order to identify novel factors that might elucidate their functional range in context of the ecological niche occupied by *Giardia*. PV and ESV were enriched by flow cytometry and analysed by LC-MS/MS, and a total of 72 ESV- and 82 PV-specific proteins were identified, many of which were novel. Interestingly, localisation of ESV proteins, including localisation and co-localisations to the endoplasmic reticulum (ER), supported the hypothesis that ESV and the ER in *Giardia* are closely physically and functionally linked [58].

The proteasome in *Giardia* is responsible for ubiquitin-mediated protein degradation and sorting of cyst-wall material during encystation [79]. The proteasome of Giardia trophozoites was tagged and purified using affinity chromatography, and a total of 30 proteins were identified, 28 of which possesses homologous components in the *Saccharomyces cerevisiae* proteasome [59]. This was the first analysis of subunit composition of the proteasome, and encompassed the identification of proteasome-associated proteins including dynamin, the ER chaperone BiP, and an ubiquitin-specific protease [59].

1.4.4 Secreted Proteins and Extracellular Vesicles

Despite considerable progress in characterising organelles and cytoskeletal fractions, a gap still remains in available data for quantitative proteomics of both secreted proteins and extracellular vesicles in *Giardia.* This is particularly relevant considering increasing evidence of host immunomodulation by several known secreted *Giardia* proteins [80]. Currently, several metabolic enzymes have been identified in the media fraction during co-incubation with host cell lines [81], for which there are strong implications for their role in repression of reactive nitrogen species (RNS) to avoid of clearance from the gastrointestinal tract [82, 83]. Moreover, the secreted cysteine protease cathepsin B is a known virulence factor which degrades interleukin-8 and reduces pro-inflammatory cell chemotaxis [84].

Cysteine proteases are secreted virulence factors in a range of parasites, and multiple uncharacterised cysteine proteases are detected in *Giardia* media through activity assays [85, 86], although the majority of these lack defined roles in pathogenesis. Overall, further work is required to specifically identify in the secreted fraction any of the 23 other cathepsins present in the *G. duodenalis* genome [26], as well the possibility of identifying other secreted protease classes.

There is also considerable merit in the identification of and subsequent analysis of extracellular vesicles (EV) in *G. duodenalis*. EV can be microvesicles, exosomes or apoptotic bodies, and are well-characterised in higher eukaryote and mammals, but only recently identified in a range of protozoan parasites including *Plasmodium* [87], *Trichomonas* [88], *Leishmania* [89-92] and *Trypanosomes* [93-95]. These extracellular vesicles contribute to a range of cell processes, many of which underpin processes during pathogenesis and key mechanisms of virulence, including immunomodulation, pathophysiology and cell adhesion [87, 96]. Currently, intracellular ESV and PV are known to play important roles within the *Giardia* trophozoite, particularly during differentiation, where empty or 'ghost' vesicles are hypothesised to be released into the media [58, 97]. The isolation of EV in *Giardia*, the identification of the types of EV, and their analysis both at the proteomic and transcriptomic level, are an untapped research area which may provide important insights into host-parasite and parasite-parasite interactions. Considering the key roles in virulence and pathogenesis performed by secreted proteins currently known, further research in this area is likely to provide significant insight in the biology of disease and pathophysiology of giardiasis.

1.5 Isolate Variation

The biological relevance and pathophysiological effects of genetic variation between assemblages, aside from host specificity and zoonotic trends, remain elusive and poorly understood. Only a few studies correlate inter-assemblage variation to clinical syndromes

[32, 98]. Differences in cyst infectivity have been demonstrated in Mongolian gerbils [9], as well as significant differences in virulence, pathophysiology and infection intensity between isolates within the A1 subassemblage in mouse models [8, 99]. In addition, inter-assemblage variation exists between genomes [32], as well as inter-subassemblage variation between A1 and A2 genomes [28], and, recently, considerable variation between A2 isolates was demonstrated [33].

There are several quantitative proteomics studies which analyse isolate variation, specifically within assemblage A. These studies predominately compared isolates from the A1 subassemblage, but also analysed trophozoites from the A2 subassemblage for the first time [100]. They also supported the recent reclassification of A1 and A2 genomes, as searching isolates against their correct subassemblage genome increased the number of high confidence peptide matches and protein identifications reported [100]. In addition, quantitative proteomics also revealed VSP variant numbers and variant subpopulation distribution differed between A1 isolates [100, 101]. Furthermore, the zoonotic, virulent, avian isolate BRIS/95/HEPU/2041 [102, 103] displayed increased abundance of key virulence factors, including cathepsin B, and greater VSP diversity [104]. Up-regulation of virulence factors is consistent with phenotypic data from mouse models, where BRIS/95/HEPU/2041, when compared to BRIS/82/HEPU/106, showed a 3-fold higher parasite load and caused a 20% weight reduction in in neonatal mice [105], as well as higher rates of villous atrophy, goblet cell hyperplasia and cell vacuolation, [8].

Quantitative proteomic data of differentiation in *G. duodenalis* is available for three A1 isolates including WB C6 [39], BRIS/95/HEPU/2041 and BRIS/82/HEPU/106 [40]. Universal encystation-regulated markers and pathways were identified between these three isolates, relating to consistent changes in cell structure and metabolism. However, a distinct pattern of VSP variant loss between trophozoites to cysts was documented in the human-derived BRIS/82/HEPU/106, and previously observed in WB C6, while

BRIS/95/HEPU/2041 retained near-complete VSP variant diversity [39, 40]. If VSP variants possess differing host specificity, it is hypothesised this retention of variants may contribute to the wide host range of the highly zoonotic BRIS/95/HEPU/2041, which is avian derived and infective for neonatal and adult mice [99], lambs and kittens [103]. Comparative proteomics also demonstrated a significant variation in up-regulation of ankyrin-repeat/protein 21.1 proteins in cysts, with 27 of higher abundance in BRIS/95/HEPU/2041 cysts compared to only one in cysts from BRIS/82/HEPU/106 [40].

As previously outlined in section 1.3, there is a lack of post-genomic mass spectrometry data for *G. duodenalis* outside of assemblage A, specifically subassemblage A1. Even within assemblage A1, the majority of studies use isolate WB C6 (ATCC 50803) as it is well-characterised and the source material for the available A1 genome sequence [27]. While there is merit to the idea of a concerted effort to identify reproducible markers at the proteomic level by utilising WB C6 exclusively, it precludes the existence of isolate variation at the proteome level which may impact on important phenotypes. The available proteomic analyses of the A subassemblage highlight that isolate variation exists outside core proteomes. It is therefore necessary to document whether these divergences also occur in other assemblages, and examine their impact on phenotype and pathology.

1.6 Differentiation

The faeco-oral life cycle of *G. duodenalis* consists of two life-cycle stages of trophozoites and environmentally resilient cysts. Differentiation involves the processes of excystation of cysts after ingestion by a host, from which excyzoites develop into trophozoites that colonise the small intestine, followed by encystation of trophozoites to cysts in the jejunum, which are shed in the faeces. The differentiation process is the biological process best characterised in *G. duodenalis* in terms of the systems biology approach, with multiple transcriptomic and proteomic data sets available in the literature. Currently, there are three

proteomic analyses of encystation reported, including comparative 2DE of cysts and trophozoites from WB [106], label-free quantitative proteomics of WB C6 trophozoites throughout the first 14 hours of encystation [39] and label-free quantitative proteomics of two genome-alternate A1 isolate trophozoites and cysts [40]. These proteomic data are complemented by transcript-level data from serial analysis of gene expression (SAGE) data throughout the *G. duodenalis* life cycle of encystation and excystation [107], and microarray data from cysts [108, 109]. Together, these studies have identified a set of reproducible encystation markers at the proteomic level [39, 40]; the consistent biological processes and markers reported between studies and isolates have been summarised and presented in Table 3.

The process of differentiation in *Giardia* involves extensive cytoskeletal remodelling, changes in genome ploidy and restructuring of parasite metabolism. Encystation results in tear-shaped, binucleate and flagellated trophozoites becoming round, tetranucleated cysts. These cysts possess a characteristic cyst wall (CW) comprised of three cyst wall proteins (CWP), CWP1-3, complexed with β(1-3)-N-acetyl-d-galactosamine (GalNAc) polymer [72, 110]. These CW components are synthesised, concentrated and transported in the ESV, and appear in the first 8-10 hours of induction of differentiation in a highly regulated trafficking process [111]. Proteins associated with the CW structure, including GalNAc polymer synthesis, are highly reproducible markers throughout proteomic analyses of encystation [39, 40, 106], with candidates also identified in transcriptomic studies. Similarly, multiple proteins have also been identified by label-free quantitative proteomics as being up-regulated in protein trafficking involving secretory vesicles and fusion with the plasma membrane or ER [39, 40]. It also appears from proteomic data that, possibly as a result of increased pressure on protein trafficking for the CW, VSP diversity is affected, especially during early encystation [39]. This may also affect the differences in VSP subpopulation composition in cysts between isolates observed at the end of differentiation [40]

Table 3: Functional clusters associated with up-regulation during differentiation from trophozoites to cysts. *Giardia* ORF's are considered an encystation marker if they have been consistently identified in at least two studies from either transcriptome [107-109] or proteome [39, 40, 106] analyses of encystation in *G. duodenalis.* All markers have been detected in isolate WB/WB C6 and at least one other *G. duodenalis* isolate, indicating they are not isolate-specific.

Functional Cluster	**Identified Encystation Markers**	**Transcript**	**Protein**	**Differential expression & functional role during encystation:**
GalNac Biosynthesis	4-alpha-glucanotransferase, amylo-alpha-1,6-glucosidase (GL50803_10885)		Y	β(1-3)-N-acetyl-d-galactosamine (GalNAc) polymer is complexed with three CWP proteins (CWP1-3). Approximately ~60 of the CW are comprised from GalNAc polymer along with the CWP. Production of the CW is a hallmark pathway for encystation, and therefore very reproducible across encystation studies.
	UDP-glucose 4-epimerase (GL50803_7982)	Y	Y	
	Glucosamine-6-phosphate deaminase (GL50803_8245)	Y	Y	
	UDP-N-acetylglucosamine pyrophosphorylase (GL50803_16217)		Y	
Cyst Wall	Cyst Wall Protein 1 (GL50803_5638)	Y	Y	
	Cyst Wall Protein 2 (GL50803_5435)	Y	Y	
Protein Trafficking	Clatherin heavy-chain (GL50803_102108)		Y	Encystation involves major protein export for the trafficking of proteins for generating the thick, extracellular matrix of the cell wall. Multiple proteins of the eukaryotic secretory system are expressed, including vescicular membrane components and proteins responsible for membrane fusion.
	Coatomer beta subunit (GL50803_88082)	Y	Y	
	Sec61 alpha (GL50803_5744)	Y	Y	
	Alpha-SNAP (putative) (GL50803_17224)	Y	Y	
	Dynamin (GL50803_14373)		Y	
Heat Shock Proteins	Cytosolic HSP70 (GL50803_88765)		Y	HSP have been reproducibly identified in all proteomic analyses of encystation, but not in any transcript analyses.
	Heat shock protein HSP 90-alpha (GL50803_98054)		Y	
Cytoskeleton	Alpha tubulin (GL50803_103676)		Y	Extensive cytoskeletal disassembly and remodelling occurs during encystation.
	Beta tubulin (GL50803_136021)		Y	
Giardins	Gamma giardin (GL50803_17230)	Y	Y	Giardins are a multigene family associated with the cytoskeleton and related structures. Multiple members are differentially expressed at the transcript and protein during the life cycle.
	Beta giardin (GL50803_4812)	Y	Y	
	Alpha-3 giardin (GL50803_11683)	Y	Y	
Kinases	NEK 2 (GL50803_5375)		Y	Multiple NEK kinases are up-regulated during differentiation, with NEK2 specifically associated with cell cycle. NEK kinases constitute about 70% of kinases in *Giardia* though the majority lack catalytic residues. Multiple NEK kinases are detected as differentially expressed across studies, though few are reproducibly identified.
	Nek kinase (GL50803_101534)		Y	
	Nek kinase (GL50803_11311)	Y	Y	
	Nek kinase (GL50803_95593)		Y	

During the encystation process the cytoskeleton is drastically remodelled, with multiple structures disassembled including the flagella and the median disc [66]. This is reflected in the reproducible observation of differential expression of tubulin proteins at the protein level, as well as giardin proteins. Giardins are a multi-gene family of annexin-related proteins associated with the cytoskeleton, particularly the membrane and flagella, and are involved in important processes including host-attachment [62, 69]. Both transcriptomic and proteomic experiments indicate there are fluctuations in expression in multiple giardins during differentiation, most likely as a result of cytoskeletal remodelling of associated structures. In concordance, multiple members of the basal bodies and adhesive disc are also differentially expressed, especially early in encystation [39]. Furthermore, both of the published studies on label-free quantitative proteomics of encystation report differential expression in ankyrin repeat proteins (protein 21.1), which may vary between isolates [39, 40], and also contain a subset of members with coiled-coil domains that associate with flagella and axonemes [112].

Another hallmark feature of encystation is the distinct down-regulation of the glycolytic pathway. This down-regulation of glycolysis during encystation has also been observed during encystation in *Entamoeba histolytica*, and also correlates with the up-regulation of cyst wall production in particular in correlation with up-regulated chitin synthesis [113, 114]. Microarray analysis in *Giardia* indicates that in terms of gene transcription, cysts are largely metabolically dormant [109]. In support of this, label-free quantitative proteomics studies have shown that the majority of the enzymes in the glycolysis pathway are down-regulated. Kinase up-regulation is also observed in proteomic analyses, particularly for the NEK kinases. NEK kinases dominate kinase-related genes in *Giardia*, though 70% lack the catalytic residues for phosphorylation, and may in fact be better classed as pseudokinases [68]. Both NEK 1 (GL50803_92498) and NEK 2 (GL50803_5375) are specifically involved in the cell cycle, and NEK 2 has been reported to be up-regulated in encystation and cysts [39, 40]. It is interesting that subsets of NEK kinases are consistently up-regulated

in proteomic analyses, as well as transcriptomic studies, which indicates this gene family may be involved in encystation. However, there is a lack of reproducibility in the specific NEK kinase ORFs identified as differentially expressed, and this variation makes it difficult to confirm specific functions.

While proteomics and transcriptomics have captured some of the biological processes occurring at the molecular level during encystation, there are also several encystation markers for which further functional data are needed. In particular, a furin precursor putative serine protease (GL50803_4653). This *Giardial* serpin 1 homologue is consistently overexpressed at transcript and protein levels in late encystation, although whether it is a pseudo-enzyme or regulates proteolytic events in encystation remain to be explored. In addition, another reductase (GL50803_88581) associated with lipid metabolism has been consistently reported in multiple transcriptomic studies [107, 108], as well as between multiple isolates in label-free proteomics [40]. In addition, both HSP70 and HSP90 are reproducibly identified in all proteomic analyses of encystation as differentially expressed, which is not the case in transcriptomic equivalents. There are also several other consistently identified proteins considered as putative encystation markers for which no functional data is available.

There is a considerable baseline data at both the proteomic and transcript level characterising the encystation process. There are several pathways that are consistently reported, though several gene families, including NEK kinases, ankyrin repeat and giardins, require further analysis. In addition, there is also evidence that isolates may emerge from the encystation process differently [40], and that *in vitro* encystation protocols may impact differential expression results [108]. A significant gap exists in the absence of proteomic data for the excystation stage of the life cycle in *Giardia*, though transcript data is available from SAGE analyses [107].

1.7 Host-Parasite Interactions

There are significant gaps in our understanding of early stages of pathogenesis in *Giardia*, including the molecular mechanisms and signals behind disease induction [6, 80]. In particular, identifying biochemical factors produced by trophozoites, and relating these to downstream symptomology requires further investigation [80]. *In vitro* host-parasite models utilising intestinal epithelial cells (IECs) and *Giardia* trophozoites are effective in simulating disease processes. The application of post-genomic technologies to analyse these models on a global scale has allowed exploration of disease induction in the parasite, as well as building our understanding of host immune responses to infection. Most importantly, it has been demonstrated that exposing *Giardia* to host signals, such as co-incubation with IECs, induces gene expression that is otherwise constitutive or basal during axenic culture [81, 115]. So far, the majority of these studies have analysed changes in gene expression at the transcript level, with significantly more data required at the protein level. There have been expression studies of transcriptional changes in *Giardia* trophozoites co-incubated with IECs Caco-2 and HCT-8 [115], as well as HT-29 cells [116]. There are also complementary studies of transcripts from IECs exposed to *Giardia* trophozoites [117, 118]. In proteomics, there have been analyses of secreted proteomes of *Giardia* trophozoites co-incubated with IECs [78], as well as a recent proteomic analysis comparing *Giardia* trophozoites co-incubated with HT-29 IECs with trophozoites incubated with HT-29 secreted products [119].

Transcriptomics of IECs co-incubated with trophozoites demonstrate induction of host immune defences, but also immunomodulation in suppression of host innate immune mechanisms. Co-incubation of differentiated Caco-2 cells with trophozoites prompted considerable up-regulation of transcripts for chemokine genes [117], which were associated with recruitment of host immune cells, rather than being pro-inflammatory. In addition, *Giardia* trophozoites suppress expression of genes involved in nitric oxide (NO) production [118]. As production of NO reduces *Giardia* viability and growth, and negatively impacts

on differentiation [82], *Giardia* has been demonstrated to mitigate NO production by consuming arginine to deprive host cells of substrates [82], and suppressing expression of host genes for NO production [118]. There is no complementary proteomics data yet published for IECs interacting with trophozoites. Considering the intensity and immediacy of fold changes observed at the transcript level in host cells, especially in cytokines [115], such data would confirm the magnitude of transcript up-regulation at the level of protein production.

G. duodenalis trophozoites co-incubated with IECs displayed increased levels of secreted proteins in the media fraction, including enolase and two enzymes involved in arginine metabolism: arginine deiminase (ADI) and ornithine carbamoyl transferase (OCT) [81]. Given that competitive arginine consumption by *Giardia* blocks NO production, these secreted enzymes have obvious implications for pathogenesis and virulence [82]. Previous transcriptomic analyses of trophozoites interacted with IEC monolayers have demonstrated co-incubation induces expression of oxygen defence genes for cellular redox homeostasis [115, 116]. Reactive oxygen species (ROS) are a known mechanism for parasite clearance by the host [120], so the up-regulation of antioxidant systems in *Giardia* is consistent with anticipation of ROS by host cells. Coinciding with previous transcript data, quantitative proteomics of trophozoites co-incubated with HT-29 cells up-regulated a diverse range of oxidoreductases [119]. Interestingly, up-regulation of oxygen defence genes were exclusive to trophozoites co-incubated with IEC monolayers rather than through exposure to host secretions, suggesting specific interactions between host-attached parasites are responsible for induction [119]. This study also reported that trophozoites exposed to host secretory products switched from an attaching to a motile population phenotype, while proteomics of these trophozoites revealed host secretions are sufficient for induction of virulence factors in the absence of trophozoite-host attachment. Host secretory products from HT-29 cells up-regulated multiple secreted and membrane associated proteins, including tenascins, cystatin, cathepsin B precursor and numerous VSP variants, the last

two of which are confirmed virulence factors in the parasite [84, 121]. Both tenascins and the cathepsin B cysteine proteases have been previously identified at the transcript level [115], including specifically with HT-29 cells [116]. The induction of virulence factors in the presence of host secretions supports the hypothesis of two independent levels of *Giardia*-host interactions – cell-to-cell interactions mediated by host attachment, and secretory-based interactions [119]. The presence of a soluble messenger based interaction also coincides with the results of Roxstrom-Lindquist, *et al.* [117], where upregulation of the cytokine CCL20 was not reliant on trophozoite attachment, but rather occurred in the presence of a secreted parasite factor.

The current consolidation of datasets from transcript and proteomic data has highlighted several important processes during early pathogenesis. Those molecular mechanisms and pathways, common between transcript and protein expression, indicate multiple level of crosstalk between host and parasite in the context of virulence biology in the parasite, and immunology of the host. However, more proteomics studies are required, especially in the analysis of the host response to *Giardia* invasion, and the analysis of host-parasite interactions outside Assemblage A. The latter of these, in particular, will explore the possibility of induction of assemblage- and isolate-specific disease factors. Furthermore, with strong indications of soluble signals and secreted protein interactions occurring between *Giardia* and the host [81, 119], there is a clear necessity to further explore and characterise the identity and nature of soluble factors.

1.8 Drug Resistance

Treatment of giardiasis is predominately through nitroheterocyclics (e.g. metronidazole, nitazoxanide and furazolidone), though benzimidazoles (albendazole and mebendazole) are also used. However, it is estimated the effectiveness of the frontline drug metronidazole is in the range of 73–100%, while for albendazole the effectiveness is reported to be in the

range of 77–97% [122], meaning that treatment failure is routinely encountered in both classes. Drug resistance to both compounds has been demonstrated in clinical isolates using mouse models [123], and drug resistant isolates have also been raised in the laboratory [49, 50]. Between the two drug classes, nitroheterocyclic resistance remains the best characterised [120], despite the fact that post-genomic transcript and proteomic analyses of drug-resistant isolates are limited or unavailable. Currently, the only reported proteomic analysis of drug resistance concerns albendazole resistance in the A1 genome isolate WB C6 and a derived resistant strain [124], utilising a 2DE-gel based approach. However, there is currently no quantitative proteomics study available in the literature concerning genome-wide protein expression of drug resistant isolates and strains.

The mode of action of nitroheterocyclics is believed to be through activation by parasite oxidoreductases, which induces oxidative damage to DNA and protein [125]. Importantly, metronidazole resistance seems to occur through regulation of the oxidoreductases and associated genes involved in the activation and detoxification of the drug [49, 126-130]. This has been recently reviewed by Ansell *et al* [120], who suggest that multiple mechanisms of metronidazole resistance can occur in *G. duodenalis*, including the possibility of epigenetic mechanisms that direct differential expression of proteins implicated in drug resistance. Drug- resistant isolates also show varying phenotypes related to fitness as well as infectivity and host adherence in mouse models [131], reinforcing the possibility of multiple modes of resistance. Quantitative proteomics, in conjunction with genomic, transcriptomic and metabolomic technologies, is ideal to profile protein expression globally, and explore interactions implicated in the antioxidant and epigenetic networks.

1.9 Protein Modifications

There are approximately 300 post-translation modifications (PTMs) currently known, which play pivotal roles in diversifying the proteome for a range of regulatory functions.

Unfortunately, there are limited studies available that analyse the status, function and dynamics of PTMs in *Giardia*, and these only utilise a fraction of the MS technologies available. The study of PTMs remains a challenge at both the technical and functional level, especially given the differing stability of PTMs and the dynamics nature of such modifications. PTMs occur in specific molecular and cellular contexts, and hence multiple spatio-temporal PTM variants of proteins exist which are particular to the specific biotic and abiotic circumstances. Conventional techniques of analysis by radio-isotope labelling, western blotting, and protein or peptide array, suffer from limitations in sensitivity and efficiency [132, 133]. Consequently, new MS strategies for the characterisation of the PTMs of specific proteins at the whole proteome level are gaining favour [23, 132]. Table 4 shows a range of amino acid residues, their known PTMs, and references which indicate their occurrence in *Giardia*. Several specific PTMs and their roles in Giardia biology and its evolutionary lineage will be discussed in detail in the following sections.

Table 4: A summary of common PTMs of proteins and their amino acid substrates, with references that demonstrate the occurrence of these modifications on *Giardia* proteins.

PTM	**Amino Acid Substrates**	**Evidence in *Giardia***
Phosphorylation	Asp, Ser, Thr, Tyr, His	[68, 138, 139]
Acetylation	Lys (Ser, Thr)	[148-151]
Glycosylation	Ser, Thr (*O*-linked), Asn (*N*-linked)	[157, 158, 161]
S-Palmitoylation	Cys	[157, 164, 181]
N-Myristoylation	Gly	[137, 182]
Ubiquitination	Lys	[165, 166, 183, 184]
Methylation	Lys	
Prenylation	Cys	[185, 186]
Disulfide Bonds	Cys	[161, 163, 187]

1.9.1 Phosphorylation

The core kinome of *Giardia* is highly reduced, with the *Giardia*-specific NEK kinases comprising 198 of the total 278 kinases, many of which lack enzymatic residues and are likely to be catalytically inert [68]. Despite this, many NEK kinases are expressed at the protein level [100], and are also up-regulated during differentiation [39, 40]. To date, few

NEK kinases have been localised, and few have defined functions, although NEK1 and NEK2 are known to be associated with regulation of the life cycle [68]. There are multiple studies that suggest a range of kinases are essential during the differentiation process between cysts and trophozoites in *Giardia* [71, 134-137]. Localisation of these kinases to cytoskeletal structures also suggests that these enzymes may be specific to discrete cytoskeletal structure in *Giardia* [68].

Immunolocalisation and western blotting of trophozoite lysates clearly indicates that serine, threonine and tyrosine phosphorylation all occur in *Giardia* [68]. However, analyses of phosphorylation substrates at the proteome level remains limited. Investigations of proteins during encystation using 2DE and phosphorylation-specific staining identified both HSP70 and the 14-3-3 proteins as phosphorylated during the differentiation process [138]. The 14-3-3 homologue in *Giardia* is of particular interest, as it is a dimeric protein conserved in eukaryotes, which specifically binds to serine/threonine phosphorylated sites. In *Giardia*, 14-3-3 is both phosphorylated and polyglycylated, with a fluctuating cell localisation dependent on life cycle stage [139]. The protein interaction network of the *Giardia* 14-3-3 protein has been analysed using LC-MS/MS [140], which identified a diverse range of 314 putative interaction partners in both trophozoites and encysting trophozoites. It is clear that phosphorylation plays an important regulatory role during encystation, however, more investigations are needed to fully understand the biological context. Recently, the 14-3-3 homologue was demonstrated as up-regulated during host-parasite interactions [119], which suggests this protein may regulate phosphorylation events occurring during pathogenesis and disease induction.

The methodology of Lalle *et al* [140] also demonstrates the particular strengths of using affinity chromatography coupled to LC-MS/MS to characterise protein-protein interactions and their networks [132]. While the targeted analysis of phosphorylation protein networks with 14-3-3 has been analysed in *Giardia*, global analysis of the phosphoproteome are

required to identify the range and diversity of substrates. Proteomic methods that facilitate this shotgun approach involve immobilized metal ion affinity chromatography (IMAC) [141] or titanium dioxide enrichment [142, 143], or the combined titanium-IMAC (Ti^{4+}-IMAC)-based enrichment [144-146]. The application of these technologies will allow the mass identification of phosphorylation sites in *Giardia*, as well as quantitatively monitoring protein phosphorylation at a depth not currently explored in this parasite.

1.9.2 Acetylation

Acetylation is a common modification of lysine amino acid residues, which neutralises its positive charge [147]. The acetylation of proteins is most widely studied in the context of epigenetic gene regulation, whereby covalent histone acetylation affects chromatin condensation and decondensation, and therefore the availability of DNA for transcription [133]. The *N*-acetyl lysine group is also recognised by bromodomains found in transcription factors, and thus can also regulate gene expression by recruitment of proteins that initiate transcription [133]. The process of acetylation of histones is regulated by enzymes known as histone acetylases (HATs) and histone deacetylases (HDACs), both of which are found in the *G. duodenalis* genome [148]. In addition, α- tubulin is also known to be acetylated in *Giardia*, which is a conserved PTM of the cytoskeleton involved with the stability of microtubules [149, 150].

Epigenetic regulation of gene transcription by histone acetylation has already been demonstrated in *Giardia*, and is implicated in differentiation in the life cycle [148] as well as in antigenic switching between VSPs [151]. Recently, it has also been suggested that such epigenetic mechanisms may also play a role in drug resistance to metronidazole [120]. However, acetylation is not always confined to histone proteins [152], and in some organisms also occurs on serine and threonine residues [153]. Acetylated peptides can be enriched using immunoaffinity purification using an anti-acetyllysine antibody, which is

then coupled to LC-MS/MS for the identification of acetylated proteins and their sites [23, 152]. In addition, a combination of protein pull-down assays using immobilized biotinylated histone peptides with SILAC allows the quantitative MS analysis of histones and their modifications, as well as their interaction partners [154, 155]. There are, to date, no published global MS based analyses of acetylation in *Giardia*. The application of proteome-wide and targeted MS analyses would offer valuable information for identifying where this PTM occurs outside histone proteins in *Giardia*, and defining the nature and network of lysine acetylation in epigenetic regulation.

1.9.3 Glycosylation

Glycosylation is the most extensive PTM of proteins, involving the linking of linear or branched sugars as *N*-linked or *O*-linked glycans to specific amino acids. In *Giardia*, the GalNAc polymer is a truncated *N*-linked glycan which forms a large portion of the CW during the encystation process [156]. In addition, specific VSPs are also known to be glycosylated [157]. Glycosylation in *Giardia* is unique for a variety of reasons, particularly in that the *Giardia* lipid-linked precursor contains only two sugars (GlcNAc2) [158]. This occur as a result of the secondary loss of glycosyltransferases for adding mannose and glucose to *N*-glycans in *Giardia*, which has also occurred to varying degrees in other parasitic protists including *Plasmodium, Entamoeba*, *Theilieria* and *Trichomonas* [158, 159].

Glycosylation during the differentiation process has been analysed using wheat germ agglutinin (WGA) affinity chromatography of trophozoites and cysts [158]. A total of 91 glycopeptides were detected along with 194 secreted and membrane proteins in trophozoites, and 157 secreted and membrane proteins in cysts, of which 42 and 20 proteins were unique, respectively. Ratner *et al*. [158] also demonstrated the most abundant glycoproteins in trophozoites were associated with enzymes associated with lysosomes, N-

glycan independent quality control of protein folding in the ER, and membrane proteins lysine rich repeat or cysteine-rich membrane proteins (including VSPs). This was contrasted to the *N*-glycome of cysts, where glycine-rich repeat transmembrane proteins dominated the dataset. As the majority of the glycoproteins are membrane-associated or secreted, they are attractive targets for treatments and vaccines. It is already known that WGA arrests the cell cycle in trophozoites [160], and that it also causes trophozoite agglutination that significantly impacts attachment rates [48]. Moreover, several *Giardia* glycoproteins are known to be immunogenic [157, 161], and therefore play an important role in host immunology for parasite clearance. Therefore, further studies of glycosylation in *Giardia* have the potential to elucidate host immune and parasite interactions, as well as possible targets for chemotherapy.

1.9.4 Other Modifications

There is evidence of a range of other PTMs that occur on proteins in *Giardia* (Table 4). Disulfide bonds are known to be an important modification to the cysteine-rich VSP and HCMP families [162, 163], which are known to confer trypsin resistance and are hypothesised to alleviate oxidative stress caused by host immune molecules [115]. VSPs, aside from disulfide bonds and glycosylations, are also known to have palmitoyl lipid modifications that affect segregation into lipid rafts and antigen switching [157, 164]. In addition, members of the giardin family, which are associated with the cytoskeleton and plasma membrane, also have both myristoyl and palmitoyl lipid modifications [65].

Giardia is also considered the earliest branching eukaryote, and as such it features a reduced ubiquitination system, including a single gene which encodes the ubiquitin moiety [165, 166]. Nino *et al* [166] analysed ubiquitinated proteins during differentiation in *Giardia* using affinity enrichment by 6His-ubiquitin protein tagging [167] followed by LC-MS/MS, and identified 211 ubiquitin-protein conjugates. These ubiquitinated proteins

occurred in a wide range of functional categories including participation in the GalNAc biosynthesis pathway during CW production. Several histones and several NEK kinases were also detected as ubiquitinated. Nino *et al* [166] hypothesised that ubiquitination may play several biological roles in *Giardia*, and is not limited to proteasome-dependent protein turnover. Indeed, ubiquitin and several ubiquitin E1 enzymes were identified as up-regulated during host-parasite *in vitro* interactions [119]. These data suggest further investigation of ubiquitin-conjugates may uncover important information relevant to the pathogenesis process in *Giardia*.

1.10 Future Directions

Proteomic analyses in *Giardia* have provided biological insights into key processes and structures in this parasite, both standalone and in complement to other available genomic and transcriptomic data. However, quantitative proteomics remains an emergent field, and there are several areas identified where primary data is still required. Moreover, of those proteomic studies identified and available for *Giardia* (Table 1), less than half provide quantitative information, with the remainder reporting on protein identifications alone. As the field transitions to the larger, more complex proteomic datasets associated with shotgun proteomics, there is also an increasing importance in the data analysis of differential expression [168, 169]. The reporting of false discovery rates (FDR) at both identification and quantitation levels is necessary, as is the use of robust statistical validation of differential expression [170]. However, with accumulating genome sequence data and functional annotation resources for *Giardia*, a major barrier to potential quantitative proteomics studies has now been removed.

We believe that the utilisation of new and emerging proteomic technologies is the next frontier in *Giardia* proteomics. As previously discussed, MS-based analyses of all PTMs remain limited, with more understanding of PTMs and their interaction networks required.

For quantitative proteomics, TMT isobaric labelling currently permits the multiplexing of up to 10 channels [21], and has been shown to be applicable to sensitive quantitation of differential protein expression in *Giardia* [119]. These TMT isobaric labels have significant potential for improving protein quantitation, especially since triple-stage MS (MS3) has been shown to enhance quantitation accuracy by reducing ion fragment interference in these multiplexed tags [171]. In addition, the field of proteomics has, in recent years, embraced the idea of orthogonal validation of protein expression data generated in quantitative shotgun proteomics experiments. One such validation technique, which allows the analysis of specific peptides in particular proteins, is known as Selective Reaction Monitoring (SRM), or Multiple Reaction Monitoring (MRM). This approach allows the targeted quantification of selected peptides [172], and the experiments are typically designed using data acquired in previous quantitative shotgun proteomics experiments [173]. These are particularly relevant for *Giardia*, as there are a lack of commercially available antibodies for western blotting, which is the commonly used approach for orthogonal validation at the protein level. The design of MRM assays in *Giardia* will become increasingly important in future, as it will address current limitations on validation of differentially expressed proteins in large proteomics datasets.

Lastly, there have also been significant improvements to data-independent acquisition (DIA), using the sequential window acquisition of all theoretical mass spectra (SWATH) method [174]. The SWATH method essentially uses a shotgun proteomics approach for the DIA exploration of the total population of precursor ions present in a given sample to generate a peptide spectrum library for large scale MRM of the sample [174, 175]. SWATH combines an untargeted shotgun proteomic approach with the targeted MRM quantification of peptides in a simultaneous analysis. These new approaches to quantitative proteomics will underpin the next generation of methods in the field, and it is important step these technologies are developed for, and applied to, *Giardia*.

Systems biology has shown considerable promise in helping to provide a global expression network to navigate complex biological phenomena. For *Giardia* these include host-parasite interactions, life cycle transitions and the interactions between the whole parasite and its smaller, subcellular components. In terms of analysing the proteomic complement involved in these complex phenomena, many successful achievements have occurred, but far more data is still required. Quantitative information is essential, and increasing, but analyses of protein turnover, protein-protein interactions and PTMs are all lacking in *Giardia*. Moreover, as complementary RNA and protein abundance data emerges in the literature, more work is needed to define the nature of the relationship between expression at the RNA and the protein level. In *Giardia* there are instances of post-transcriptional regulation of genes [176], regulation by small non-coding RNA [177], epigenetic regulation though histone modification [148, 151, 178], and unique transcriptional features such as abundant production of anti-sense transcripts [179, 180]. In order to combine both transcript and protein expression data, it is important to first understand the depth of their overlap. It is likely combined analyses, rather than the independent analyses currently in the literature, will be required to evaluate this.

The dynamic changes, fluctuations and modifications that occur within the cell can be captured by quantitative proteomics. There are multiple elegant examples of this for *Giardia* in the literature, where molecular analysis demonstrates the individual components that, when combined, effect a biological process. Improvements in genome databases, particularly functional annotation, will concomitantly expand these insights into parasite pathology, virulence and diversity. Hopefully, the gaps that remain in the foundations of *Giardia* proteomics will reduce in the coming years, especially as improved technologies and multiplexed methods increase both proteome coverage and accuracy in quantitation.

1.11 Acknowledgements

SJE acknowledges funding from the Australian Government in the form of an APA scholarship, as well as financial support from Macquarie University. SJE also acknowledges ongoing support received from colleagues at Microbial Screening Technologies. PAH wishes to thank Wayne Dutschke for continued support and encouragement.

1.12 Overview of Experimental Chapters

The objectives of the experimental chapters in this thesis were to provide both (1) a quantitative proteomics method for analysing *G. duodenalis*, and (2) additional primary data which would complement the data already available. Bottom-up or shotgun proteomic techniques were utilised to achieve these objectives, and both label-free and labelled quantitative approaches have been applied. In conjunction with the data presented in each experiment we have provided robust statistical analyses of differential expression. As such, this thesis contains a variety of proteome-wide analysis of protein expression and abundance across multiple biological scenarios in *Giardia.* Multiple experimental chapters of this thesis have also been peer-reviewed and published in the literature, and where this has occurred the published versions of these experiments have been incorporated in their respective chapters.

Chapter 2 of this thesis details the optimised sample preparation and fractionation of whole trophozoite protein extracts for label-free quantitation via spectral counting. Two different sample preparations, a gel-based and an in-solution method, were applied to the evaluation of isolate diversity in two A1 subassemblage isolates. This experimental chapter aimed to explore which methodology would provide the most identifications in conjunction with accurate quantification, especially for *Giardia*-specific genes and gene-families. This chapter is presented as a research article (Publication II).

Chapter 3 of this thesis used the preferred sample preparation and fractionation methodology from the prior experimental chapter, and applied it to the analysis of seven A1 and one A2 *G. duodenalis* isolates. These eight isolates were analysed as whole trophozoite fractions and protein abundance was quantified using spectral counting. In addition, database searching was carried out against both the A1 and A2 subassemblage genomes. This was used to investigate database dependent losses based on new comparative genome evidence that supported reclassifications of A1 and A2 as cryptic species. This experiment aimed to collect and amalgamate a quantitative proteomic

baseline of assemblage A isolates. This chapter is presented as two research articles (Publication III and Publication IV).

Chapter 4 of this thesis applied label-free spectral counting to quantify differential abundances between proteins of *G. duodenalis* trophozoites and cysts. This was performed in two alternate genome isolates to complement quantitative proteomic data in the genome isolate WB C6. This allowed the identification of universal, isolate-independent markers of encystation, as well as identification of isolate-specific adaptations in the differentiation process, which may impact reinfection success and life cycle completion in subsequent hosts. This chapter is presented as a research article (Publication V).

Chapter 5 of this thesis employs a TMT labelling approach to quantify induction of protein expression during an *in vitro* host-parasite interaction model. This constitutes the first quantitative proteomic analyses of a trophozoite-IEC *in vitro* model, as well as the first application of TMT labels for protein quantitation in this parasite. This experiment aimed to evaluate the effects of host soluble products on trophozoites in a cell-free media, as well as trophozoites co-incubated with HT-29 IECs. This experiment also identified proteins specifically induced by interaction with IECs or their products, to evaluate if induction was independent and separate between *G. duodenalis* trophozoites exposed to host soluble signals alone, and trophozoites co-incubated with IECs. Although this chapter has been accepted in Scientific Reports (Publication VI), it has been reformatted it as a thesis chapter for ease of understanding, with all figures, tables and supplementary material integrated.

Experimental chapters have individual reference lists at the end of each chapter, while references for the introduction ***(Chapter 1)*** and general discussion ***(Chapter 6)*** are contained in a separate bibliography at the end of this thesis. All supplementary material for experimental chapters are available as an electronic copy on the DVD provided with this thesis.

CHAPTER 2

A comparative proteomic experiment assessing two sample preparation and fractionation methods, as applied to the characterisation of two A1 Giardia duodenalis *isolates.*

2. Proteomic analysis in *Giardia duodenalis* yields insights into strain virulence and antigenic variation

2.1 Context

Post-genomic proteomics technologies remain a largely unexplored tool for examining *Giardia duodenalis* cell biology, strain variation and disease mechanisms. This experiment utilises a comparative proteomics approach between two strains with well documented phenotypes in the literature, and which are closely related according to genetic taxonomy (A1 subassemblage). An avian and zoonotic virulent strain, BRIS/95/HEPU/2041, was compared to a control strain, BRIS/83/HEPU/106, to investigate protein expression differences that may be responsible for phenotypical differences in virulence and pathology. This constituted the first modern proteomic analysis of strain variation in *Giardia*. This experiment was also used to evaluate both a gel-based as well as in-solution based fractionation and digestion methodology, in order to assess their complementarities and their merits in the design of future proteomics experiments for *G. duodenalis*.

2.2 Contributions and Permissions

I performed 100% of all work pertaining to growing and collecting *Giardia* samples for proteomic analysis, and 90% of all lab work with assistance and guidance provided by Steve Van Sluyter. I wrote 80% of the manuscript with assistance and editing provided by Professor Paul Haynes. The paper (Publication II) is reprinted with the permission of John Wiley and Sons.

2.3 Publication II

Proteomics 2014, *14*, 2523–2534 DOI 10.1002/pmic.201400144

RESEARCH ARTICLE

Proteomic analysis in *Giardia duodenalis* yields insights into strain virulence and antigenic variation

Samantha J. Emery, Steve van Sluyter and Paul A. Haynes

Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, New South Wales, Australia

Received: April 17, 2014
Revised: August 19, 2014
Accepted: September 25, 2014

Giardia duodenalis is a protozoan parasite of the small intestine in vertebrates, including humans. Assemblage A of *G. duodenalis* is one of the two discrete subtypes that infects humans, and is considered a zoonotic assemblage. Two *G. duodenalis* Assemblage A strains BRIS/95/HEPU/2041 and BRIS/83/HEPU/106, constituting virulent and control strains respectively, were analyzed in one of the first comparative shotgun proteomic studies performed in this parasite. Protein extracts were prepared using a multiplatform approach with both an in-gel and in-solution sample preparation to enable us to assess the complementarity for future *Giardia* proteomic studies. Protein analysis revealed that BRIS/95/HEPU/2041 possessed a wider and more varied repertoire of variant surface proteins (VSPs), which are hypothesized to be involved in host adaptation, immune evasion, and virulence. A total of 35 VSPs were identified, with three common to both strains, six unique to BRIS/82/HEPU/106, and twenty-six unique to BRIS/95/HEPU/2041. Additionally, up to 25.6% of all differentially expressed proteins in BRIS/95/HEPU/2041 belonged to the VSP family, a trend not seen in the control BRIS/83/HEPU/106. Greater antigen variation in BRIS/95/HEPU/2041 may explain aspects of virulence phenotypes in *G. duodenalis*, with a highly diverse population capable of evading host immune responses.

Keywords:
Giardia duodenalis / Label-free quantitative shotgun proteomics / Variant surface proteins / Virulence

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Giardia duodenalis (syn. *G. lamblia* and *G. intestinalis*) is a gastrointestinal protozoan and major contributor toward human and veterinary diarrheal infections worldwide [1]. As a disease, *Giardiasis* is estimated to occur at a prevalence of 200 million symptomatic infections within Asia, Africa, and Latin America alone, and it has been listed on the Neglected Disease Initiative from the WHO since 2004 [2, 3]. Infections are often highly variable, ranging from asymptomatic and self-limiting to severe and chronic, lasting several months [4]. *Giardiasis* is characterized by abdominal pain, watery diarrhea, nausea, vomiting, and weight loss, with symptoms more pronounced in the immunodeficient and immunocompromized as well as in children, where "failure to thrive" is a common consequence of prolonged infection [2, 5]. The prevalence of *Giardiasis* is caused by multiple factors, including a direct life cycle where infective cysts may persist and accumulate in the environment over several months, cross-species (zoonotic) transmission across wide host-ranges, and a multitude of risk and environmental factors involved in pathogen exposure [3, 6].

Correspondence: Professor Paul A. Haynes, Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW 2109, Australia
E-mail: paul.haynes@mq.edu.au
Fax: +61-2-9850-6200

Abbreviations: **FASP**, filter-aided separation of proteins; **GPF**, gas-phase fractionation; **HCMP**, high cysteine membrane protein; **NanoLC-MS/MS**, nanoflow LC-MS/MS; **NSAF**, normalized spectral abundance factor; **SpC**, spectral count; **TFE**, 2,2,2-trifluoroethanol; **VSP**, variant surface protein

The source of variability for virulence and pathogenesis remains largely unknown for *Giardia*. Host source, geography, and longitudinal factors have long been recognized as contributors in *G. duodenalis* isolate variability [7–9], and recent advances in molecular epidemiology have indicated potential genetic diversity that may explain this variation [3, 10–12]. Taxonomic revision of *G. duodenalis* has led to the description of eight assemblages (A–H) in a possible species complex, each of which may be distinguished through multilocus sequencing and which form discrete groups phylogenetically and may be further divided into sub-assemblages (such as A1 and A2) [3, 11]. Genome sequences have been published for assemblages A1, A2, and B, which are known to infect humans, as well as the artiodactyl-specific assemblage E [13–17]. The number of protein-coding ORFs in *Giardia* genomes sequenced to date ranges from 4470 to 7477, with the A1 WB genome containing 5901 protein-coding ORFs [17]. Despite this accumulation of genetic information there are still no confirmed virulence factors or secreted toxins in *G. duodenalis*, and little is known about the roles played by different isolates, strains, and assemblages in infection [2, 3, 15, 18].

MS-based proteomic analysis has been successfully applied to studying parasites, disease, and their hosts [19, 20]. Proteomic data already exist in integrated databases for other protozoan parasites of humans [21–24], including *Plasmodium falciparum* [25], *Toxoplasma gondii* [26], and *Cryptosporidium parvum* [27]. However, published proteomic studies in *Giardia* are not as comprehensive [28]. There have been several MS-based studies recently published on the ventral disk, mitosome, and basal body proteins [29–31], but there are few published shotgun proteomics experiments aimed at exploring the entire *G. duodenalis* proteome [32]. Giardiadb.org lists only 1859 proteins for which MS evidence (fewer than one spectra) exists in *G. duodenalis*, compared to 4036 such proteins recorded for *P. falciparum* [22, 33]. There is a distinct and immediate need not only to generate information to correct this disparity, but to provide foundations for additional and ongoing proteomic analyses for *Giardia* in the postgenomic era.

Shotgun quantitative proteomics is a powerful tool for resolving biological differences among systems, states, and environments, and within strains, species, and taxa. Additionally, advances in label-free quantitative proteomics [34] have made it a versatile and reliable alternative to labeled strategies such as SILAC [35], ICAT [36], and ITRAQ [31]. Currently, only 2DE has been utilized in quantitative and comparative proteomic studies in *G. duodenalis* [37, 38]. With differences in clinical symptomology well established, and with genomes and multilocus subtyping widely available, shotgun quantitative proteomics presents a unique opportunity to analyze differences in global protein abundance between strains.

This study is among the first comparative and quantitative proteomic studies recorded in *G. duodenalis* utilizing a shotgun proteomics approach, and is also the first to analyze proteins in *G. duodenalis* that may contribute to virulence in *Giardiasis*. We examined two *G. duodenalis* strains: BRIS/83/HEPU/106 (H-106), isolated from a diarrheic child, and BRIS/95/HEPU/2041 (B-2041), which was isolated from a wild-caught cockatoo (*Cacatua galerita*) that succumbed to terminal enteritis [39]. H-106 and B-2041 constitute a lab-standard and virulent strain, respectively, with respect to several infection models. When mice were orally infected with H-106 and B-2041, the avian strain established a chronic infection with threefold higher parasite load and 20% weight impairment when compared to H-106 [39]. This was accompanied by greater disruption to the gastrointestinal mucosa in mice where, compared to H-106, B-2041 induced higher rates of villous atrophy, goblet cell hyperplasia, and cell vacuolation, as well as low concentrations of *Giardia*-specific serum [5]. These data, accompanied by experimental evidence of high zoonotic risk in B-2041 across mammals [40], and the fact this phenotype did not diminish over 9 months of continuous laboratory passage [39], were used to classify B-2041 as a virulent strain. As H-106 was used as a baseline in these models for establishing the phenotype of B-2041, it has been maintained as a control for our proteomic comparison. As both these strains are isolated from Brisbane, Australia, and have been verified as assemblage A, they will provide insight into disease mechanisms unrelated to geographical and assemblage variation [5, 41].

A combination of both gel-based and in-solution platforms and sample preparations were used in this study, with quantitation of relative protein abundance performed using spectral counting [34]. This type of multiplatform approach has been previously successfully applied to *T. gondii* [42] and *C. parvum* [27]. In this study we analyzed the proteome of *G. duodenalis* first by 1D gel LC-MS/MS, and second by combining the filter-aided separation of proteins (FASP) [43, 44] with gas-phase fractionation (GPF) [45] as an in-solution platform. We assessed the complementarity of the two techniques in context of a global protein abundance study as well as for proteome coverage.

2 Materials and methods

2.1 Parasite culture

Axenic cultures of trophozoites of H-106 and B-2041 were grown in TYI-S-33 medium supplemented with 1% bile and newborn-calf serum [46] as previously described [47]. Parasites were subcultured at end-log phase into fresh media, and grown and harvested for protein extraction within five passages of recovery from cryopreservation. Cryopreservations were selected from within the first year each strain was introduced into axenic culture. H-106 has previously been designated the lab-standard isolate based on former studies [5, 7–9, 39]. Three biological replicates generated from separate cultures were grown for each strain and harvested separately. Absence of bacterial and fungal contamination was verified using serial dilutions and streak plates to ensure no colony-forming units were detected in cultures prior to extraction.

2.2 Protein extraction

Giardia trophozoites were harvested from confluent cultures in late log phase by chilling on ice with vortexing for 15 min to detach parasites from the walls of the culture vessel. Trophozoites were harvested by centrifugation at 3000 × *g* for 10 min, and then washed twice more with ice-cold PBS to remove media traces [5]. Trophozoites were pelleted to 10^9 parasites and protein was extracted in 1 mL ice-cold SDS sample buffer containing 1 mM EDTA and 5% beta-mercaptoethanol, then reduced at 75°C for 10 min. Protein extracts were centrifuged at 0°C at 13 000 × *g* for 10 min to remove debris, and stored at −20°C. The concentration of protein in each solution was measured by BCA assay (Pierce) before fractionation and digestion.

2.3 SDS-PAGE fractionation

Aliquots of 175 μg protein were loaded into a Bio-Rad 10% Tris-HCl precast gel with Bio-Rad unstained marker run in parallel. Gels were run at 70 V for 15 min and then at 160 V for 1 h, fixed for 1 h and then stained lightly with CBB G-250 for 15 min. Each stained lane in triplicate sets was excised and cut into 16 equal fractions using a scalpel, and then further chopped and transferred to a 96-well plate. Trypsin in-gel digestion and peptide extraction were performed as previously described [48]. Peptides were stored at −20°C.

2.4 Nanoflow LC-MS/MS (nanoLC-MS/MS) for gel extracts

Each of the 16 reconstituted fractions of triplicate sets for both strains was analyzed by nanoLC-MS/MS on an LTQ-XL linear ion trap mass spectrometer (Thermo, San Jose, CA, USA). Samples were analyzed on a 150 × 0.2 mm id fused-silica column packed with Magic C18AQ (200 Å, 5 μm diameter, Michrom Bioresources, CA) connected to an Advance CaptiveSpray Source (Michrom Bioresources). An electrospray voltage of 1.8 kV was applied via a liquid junction upstream of the C18 column. Samples were injected onto the column using a Surveyor autosampler, which was followed by an initial wash step with buffer A (0.1% v/v formic acid, 1 mM ammonium formate, 0.2% v/v methanol) for 4 min at 150 μL/min. Peptides were eluted from the column with 0–45% buffer B (100% v/v ACN, 0.1% v/v formic acid) for 75 μL/min for 34 min and 100% Buffer B at 150 μL for the final 7 min. Each of the 16 fractions for each gel lane took 45 min to analyze, totaling 12 h per sample. Spectra in the positive ion mode were scanned over the range of 400–2000 amu and, using Xcalibur software (version 2.06, Thermo), automated peak recognition, dynamic exclusion, and MS/MS of the top six most intense ions at 35% normalization collision energy were performed.

2.5 FASP digestion and peptide extraction

In-solution digestion was performed using the filter-aided sample-preparation (FASP) method [43, 44], modified to solubilize protein in 2,2,2-trifluoroethanol (TFE) [49]. For each sample 500 μg of protein was extracted from SDS sample buffer using chloroform–methanol [50], with the protein pellet washed two to three times with methanol. The protein pellet was then dissolved in 300 μL 50% TFE containing 100 mM ammonium bicarbonate and 50 mM DTT, then added to a 30 K Amicon ultrafiltration device (Millipore) and centrifuged at 14 000 × *g* until retentate was less than 20 μL. Reduced disulfides were alkylated in 200 μL 50% TFE containing 100 mM ammonium bicarbonate and 0.5 M iodoacetamide in the dark for 45 min, then centrifuged at 14 000 × *g* as above. Proteins were washed five times in 200 μL 50% TFE containing 100 mM ammonium bicarbonate, centrifuging as above after each and discarding flow-through where necessary. Digestion with Lys-C was performed overnight at 30°C with 5 μL of 1 μg/μL Lys-C in 45 μL of 50% TFE containing 100 mM ammonium bicarbonate. Following digestion with Lys-C, samples were diluted with 350 μL 20% ACN containing 50 mM ammonium bicarbonate and digested with 5 μL of 1 μg/μL trypsin for 2 h at 37°C. Peptides resulting from digestion were centrifuged into fresh receptacles at 14 000 × *g* until retentate was less than 20 μL, then subsequently extracted twice with 150 μL of 50% ACN containing 2% formic acid, and centrifuged as above. Each extract was dried using a vacuum centrifuge and then reconstituted to 80 μL with 2% formic acid, 2% TFE.

2.6 Theoretically derived GPF calculations using predictive software

Optimized GPF mass ranges were calculated in silico according to Scherl et al. [45], using the Giardiadb.org 2.5 release of *G. duodenalis* strain WB genome to calculate the optimized mass ranges based on a theoretical trypsin digestion of the proteome. The charge states +2 and +3 were considered in the calculation, as well as carbamidomethyl as a modification. A total of four mass ranges were calculated covering the range from 400 to 2000 amu. The mass ranges were calculated as following: the low mass range was 400–518 amu, the low–medium mass range was 518–691 amu, the medium–high mass range was 691–988 amu, and the high mass range was 988–2000 amu.

2.7 NanoLC-MS/MS for FASP extracts

Each of the FASP protein digests for the triplicates of each strain was analyzed by LC-MS/MS on an LTQ-XL linear ion trap mass spectrometer. Samples were analyzed, as before, on a 150 × 0.2 mm id fused-silica column packed with Magic C18AQ (200 Å, 5 μm diameter) connected to an Advance

CaptiveSpray Source. The FASP protein digest for each sample was analyzed as four repeat injections, with the mass spectrometer set to scan for four 180-min runs with a low, low–medium, medium–high, and high mass range as calculated in silico using a theoretical trypsin digest. Samples were injected onto the column using a Surveyor autosampler, which was followed by an initial wash step with buffer A (0.1% v/v formic acid, 1 mM ammonium formate, 0.2% v/v methanol) for 6 min. Peptides were eluted from the column with 0–80% buffer B (100% v/v ACN, 0.1% v/v formic acid) at 150 μL/min for 167 min finished by a wash step with buffer A for 6 min at 150 μL/min. Each of the FASP protein digests required four runs of 180 min, totaling 12 h per sample. Spectra in the positive ion mode were scanned over the respective GPF ranges and, using Xcalibur software (version 2.06), automated peak recognition, dynamic exclusion, and MS/MS of the top six most intense ions at 35% normalization collision energy were performed.

2.8 Database search for protein/peptide identification

The LTQ-XL raw output files were converted into mzXML files and searched against the Giardiadb.org 2.5 release of *G. duodenalis* Assemblage A1 genome (WB Clone 6, ATCC 50803) using the global proteome machine (GPM) software (version 2.1.1) and the X!Tandem algorithm. Both the 16 fractions for the SDS-PAGE gel and the four fractions for the GPF of each replicate were processed sequentially with output files generated for each individual fraction, and a merged, nonredundant output file was produced for protein identifications with log(*e*) values <-1. Peptide identification was determined using MS and MS/MS tolerances of 4 and 0.4 Da. Carbamidomethyl was considered a complete modification, and partial modifications considered included oxidation of methionine and tryptophan. The MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [51] with the dataset identifier PDX000452.

2.9 Data processing and quantitation

The 12 lists of identified proteins for each of the three replicates of both strains across gel LC-MS/MS and GPF methods were filtered based on two criteria. Protein identifications were considered valid only when present reproducibly in all three replicates, with a total spectral count (SpC) of ≥5 [48,52]. Protein and peptide FDRs were subsequently calculated. Peptide FDR was calculated by 2 × (total number of peptides identified for reversed protein hits/total number of peptides identified for all proteins in the list) × 100, while protein FDR was calculated by (number of reversed protein hits/total number of proteins in the list) × 100 [48].

Protein abundance was calculated using normalized spectral abundance factors (NSAFs) as described previously [53]. To calculate NSAF the SpC (the total number of MS/MS spectra) identifying a protein was divided by the protein length (*L*), and normalized to the sum of SpC/L for all proteins in the experiment [53]. A spectral fraction of 0.5 was added to all SpCs, to compensate for null values and allow log transformation of the NSAF data prior to statistical analysis [53].

2.10 Statistical analysis of differentially expressed proteins

Several *t*-tests were performed to identify differentially expressed proteins between *Giardia* strains in the methods used. The *t*-tests were run on log-transformed NSAF data to determine statistical significance of differential abundance, and proteins with a *t*-test *p*-value less than 0.05 were considered to be differentially expressed.

3 Results and discussion

3.1 Gel LC-MS/MS proteomic analysis

Proteins were extracted from *G. duodenalis* trophozoites grown to late log-phase. Trophozoites from both strains were cultured from early cryopreservations of H-106 and B-2041 to reduce clinical isolate variation, and such were also limited to five passages in culture after recovery. Additionally, B-2041 grew slower than H-106 in axenic culture, which coincides with the observed variation in growth seen in the original mouse model [39]. Extracted protein from replicates of both strains was fractionated by SDS-PAGE and subjected to in-gel digestion followed by nanoLC-MS/MS. The total number of peptides initially detected at low stringency ranged between 35 223 to 44 985, with RSDs of 0.37% for B-2041 and 12.66% for H-106, indicating good reproducibility of peptide counts within strains and across experiments. A summary of the protein and peptide identification from both strains in the gel LC-MS/MS is provided in Table 1. Before filtering the data, the peptide FDR ranged from 0.88 to 1.01% across all six replicates, and after combining the triplicate data the protein FDR of the reproducibly identified proteins showed no reverse hits in both *G. duodenalis* strains. A total of 1130 reproducibly identified proteins from *G. duodenalis* were identified from 256 264 spectra using gel nanoLC-MS/MS.

3.2 Gel LC-MS/MS analysis of statistically significant differentially expressed proteins

Natural log NSAF values were calculated as a measure of abundance, and then Student's *t*-test was used to statistically evaluate differentially expressed proteins of greater abundance in either strain. A summary of the protein abundance profiles shown by label-free quantitation is given in Table 2. Protein quantitation indicated minimal variation in trophozoites between the strains examined. There were 1064

Table 1. Peptide/protein identification data of parasite proteins found in *G. duodenalis* strains (BRIS/82/HEPU/106 and BRIS/95/HEPU/2041) across all gel LC-MS/MS and GPF methods

Sample and method	Redundant count of peptides			Peptide FDR (%)			Average no. of peptides (±%RSD)	No. of proteins common to three replicates
	R1	R2	R3	R1	R2	R3		
B-2041[a]—gel LC-MS/MS	44 985	44 654	44 825	1.00	1.01	0.88	44 821 ± 0.37%	996
H-106[a]—gel LC-MS/MS	35 223	45 467	41 139	0.99	0.99	1.01	40 609 ± 12.66%	901
B-2041[a]—GPF	24 163	24 488	23 974	2.68	2.58	3.52	24 208 ± 1.07%	997
H-106[a]—GPF	23 235	22 301	22 065	3.02	3.24	3.29	22 533 ± 2.75%	1042

a) H-106 and B-2041 refer to strains BRIS/82/HEPU/106 and BRIS/95/HEPU/2041, respectively.

Table 2. Summary of protein identifications and label-free quantitation in *G. duodenalis* strains (BRIS/82/HEPU/106 and BRIS/95/HEPU/2041)

	Gel LC-MS/MS	GPF
Total no. of proteins (<1% FDR)	1130	1212
Total differentially expressed (n.r)	66	192
B-2041[a] higher abundance proteins	29	82
B-2041[a] unique proteins	2	16
H-106[a] higher abundance proteins	37	110
H-106[a] unique proteins	4	18
Unchanged common	1064	1020

a) H-106 and B-2041 refer to strains BRIS/82/HEPU/106 and BRIS/95/HEPU/2041, respectively.

proteins that showed no statistically significant difference in abundance, 37 proteins of greater abundance in the human, lab-standard H-106 strain, and 29 proteins of greater abundance in the avian, virulent strain B-2041. Interestingly, very few of the 66 differentially expressed proteins in either strain were found to be unique (total SpC = 0 in one strain), with only two unique proteins seen in B-2041 and four in H-106, with all six unique proteins of relatively low abundance (total SpC < 15).

3.3 GPF proteomic analysis

A GPF experiment was performed using an aliquot of the same protein material taken from *G. duodenalis* trophozoites for gel LC-MS/MS. Samples were fractionated using GPF, with spectra scanning ranges calculated by theoretical trypsin digest of the *G. duodenalis* reference genome from giardiadb.org [33]. A summary of the protein and peptide identification information for the GPF experiment is shown in Table 1.

For the GPF experiment a total of 1212 proteins were identified in all six replicates, and 140 398 spectra were assigned to peptides ranging from 22 065 to 24 488 in the replicates analyzed. The reproducibility of peptide counts was high, with the ±RSD range from 1.07 to 2.75%. The peptide FDR before filtering ranged from 2.58 to 3.52% and no further filtering of the dataset was needed after retaining reproducible protein identifications across triplicates, with the protein FDR at 0 and 0.09% for B-2041 and H-106, respectively.

3.4 GPF analysis of statistically significant differentially expressed proteins

Quantitation of protein abundance for proteins reproducibly identified using GPF was calculated using NSAF values as before, with Student's *t*-test used to statistically evaluate differential abundance of proteins. A summary of label-free quantitation for both sets of GPF data is shown in Table 2. In the GPF dataset, a total of 192 differentially expressed proteins were identified, with approximately 16 and 19% of these differentially expressed proteins uniquely expressed in H-106 and B-2041, respectively. It was also noted that the GPF dataset indicated a high similarity between the strains in a large pool of common proteins, with 1020 statistically unchanged proteins.

In the results of the label-free quantitation in the GPF dataset, there were 82 proteins of greater abundance in B-2041 and 110 in H-106 of which 19.5 and 28.2% of the proteins were putative uncharacterized proteins. In both strains the variable genome constituted a large proportion of these proteins of greater abundance. The variable genome represents 9% of *G. duodenalis* genes across four protein families (NEK kinase, Protein 21.1, high cysteine membrane protein (HCMP), and variant surface protein (VSP)) [15]. In H-106, 23% of proteins of greater abundance were from the variable genome, while this was as high as 33% in B-2041. In B-2041 this included an overrepresentation of the VSP family specifically, with 21 (25%) of the greater abundance proteins belonging to the VSP family. This also included 12 proteins within the 15 uniquely expressed proteins for B-2041 in the GPF datasets identified as VSPs.

3.5 Degree of overlap between methods

Between the two datasets, a nonredundant total of 1376 unique proteins were reproducibly identified, from a total

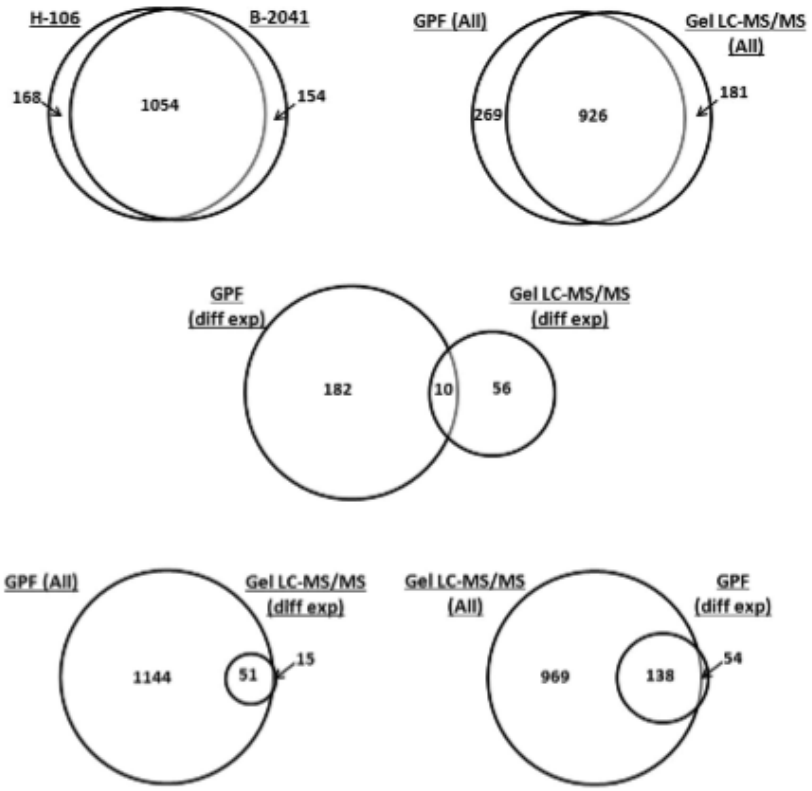


Figure 1. Proportional Venn diagrams depicting the amount of overlap in *G. duodenalis* strains (BRIS/82/HEPU/106 and BRIS/95/HEPU/2041) as well as between the methods. "GPF all" and "gel LC-MS/MS" all indicated the total number of proteins identified reproducibly at protein FDR <1%, while "GPF diff exp" and "gel LC-MS/MS diff exp" indicate the total number of nonredundant significantly differentially expressed proteins for both the GPF and SDS-PAGE gel LC-MS/MS methods.

of 413 425 spectra used in the entire experiment. The complete list of reproducibly identified proteins from the GPF and gel LC-MS/MS datasets can be found in Supporting Information Tables 1 and 2, respectively. As this is one of the first shotgun proteomics experiments to be performed using whole *Giardia* trophozoites, we aimed to assess gel-based and in-solution platforms for their suitability for further proteomic analyses in this parasite. Of the 1376 unique proteins reproducibly identified in this experiment, there were 926 (67.3%) proteins in common between the GPF datasets and gel LC-MS/MS (Fig. 1).

However, when comparing differentially expressed proteins between methods there is little overlap. Of the proteins found to be of greater abundance in H-106, only six were common with gel LC-MS/MS differentially expressed proteins, which is 5.5 and 16.2% of the differentially expressed proteins, respectively. This is also the case in B-2041, with only three proteins common between methods representing 10.3% of the gel LC-MS/MS differentially expressed proteins, and only 3.7% of the GPF differentially expressed proteins in the strain. Of the 66 proteins differentially expressed in the gel LC-MS/MS dataset, 15 proteins (22.7%) were not identified reproducibly in the GPF dataset, while of the 192 differentially expressed proteins in the GPF dataset there were 54 proteins (27.9%) not identified in the gel LC-MS/MS dataset. Finally, only ten proteins were identified as differentially expressed in both gel LC-MS/MS and GPF datasets, with many of the differentially expressed proteins seen in the gel LC-MS/MS dataset not observed as statistically significantly different in their abundance in the GPF dataset.

The GPF method identified a greater number of differentially expressed proteins than the gel-based data, as well as more proteins uniquely expressed in either strain. Moreover, a high proportion of differentially expressed proteins belonged to the variable genome in *G. duodenalis*. This was particularly due to a higher number of VSPs that were either unique or differentially expressed in B-2041, which was only seen in the GPF datasets. This increase in VSPs in the dataset is potentially due to the sample-preparation method used prior to MS analysis. The FASP method was selected as an in-solution platform, with an amended protocol that used TFE to solubilize proteins for digestion. TFE has been previously shown as effective for increasing the number of membrane proteins and proteins with lipid PTMs identified in datasets [54–56]. Solubilizing proteins in TFE may therefore increase extraction and recovery of VSP proteins, which are specifically palmitoylated on the cysteine residue in a conserved CRGKA carboxy-terminal motif located in the cytoplasmic tail [2, 57, 58]. This palmitoyl group is important in segregation of VSP into lipid rafts and has been hypothesized

as a potential mechanism of VSP release from the membrane upon cleavage [59]. TFE extraction may therefore be a more efficient sample preparation for VSPs due to this lipid modification.

This is similar to results seen in the GPF dataset for another of the variable gene families in *Giardia*, the HCMP family. The GPF dataset identified a similar number of HCMPs as the gel dataset, eight and seven HCMPs respectively, though six (86.7%) were unchanged in abundance in the gel dataset compared to only three (37.5%) in the GPF dataset, where two and three HCMPs had a higher protein abundance in H-106 and B-2041, respectively. In vegetative cells HCMPs are a large group of cysteine-rich, non-VSP, type 1 integral membrane proteins that are thought to be localized to the nuclei and nuclear envelope in vegetative cells, and are then trafficked to the cyst surface during encystation [60]. However, like VSPs, all HCMPs contain a cysteine residue in their cytoplasmic tail that may act as a substrate for the palmitoylation as in VSPs, but the presence of this lipid modification is yet to be investigated. If this modification exists, then it could possibly explain why HCMPs behave similarly in changes of abundance and differential expression between gel and GPF datasets.

The giardins are another multigene family in *Giardia* that segregate to the flagella, adhesive disk, and the plasma membrane [61]. Similar to VSPs some members may possess lipid modifications, with a dual myristoylated and pamitoylated n-terminal sequence demonstrated previously in one member [62]. Giardins are also associated with segregation to lipid rafts [62,63]. However, the extent of lipid modifications across all giardins is unknown. Though these proteins do not represent a class that is differentially expressed between strains like the VSPs or HCMPs, there is a clear bias in datasets for the giardins between methods. Of the 22 giardins reproducibly identified, 9 (40.9%) had a consistently higher NSAF value across both strains in the gel dataset and 8 (36.4%) had a higher NSAF value across both strains in the GPF dataset. While the physicochemical properties between giardins are not fully known, especially in terms of lipid modifications, our data show that some of these membrane proteins, like VSPs, are better recovered in TFE-based sample preparations when compared to detergents such as SDS. This suggests that they may contain hydrophobic PTMs.

In the SDS-PAGE datasets, highly abundant enzymes involved in anaerobic energy production, as well as cytoskeletal proteins involved in the complex organization of *Giardia* ultrastructure, dominated the results. These metabolic enzymes and cytoskeletal proteins were the most abundant in every gel fraction analyzed and likely contributed to the higher redundant peptide count, lower percentage of unique peptides, and masking of proteins of lower abundance. A higher percentage of unique peptides is especially important in identifying proteins belonging to gene families in *Giardia*, including those of the variable genome. Many of these gene family members have a high degree of protein sequence homology in conserved regions [15], and greater protein coverage ensures more confident identifications. This is particularly important as the *Giardia* gene families contain those proteins most likely to be species-, assemblage-, and isolate-specific, making them particularly relevant to *Giardia* biology. As such, SDS-PAGE was less favorable for the recovery of proteins from the VSP gene family, which were seen to be highly differentially expressed between strains using TFE in-solution digestion.

3.6 Proteomic divergence between strains

A total of 1376 unique proteins across the two methods were reproducibly identified between the two strains. When we compared our 1376 proteins to the 1859 proteins listed from the A1 genome (strain WB) with MS evidence on Giardiadb.org [33], 197 of these proteins have no spectra currently available. This would make our experiment the first to identify spectra from these proteins. There were 1054 proteins of the 1376 that were common between both strains, meaning 76.6% of proteins were present in both H-106 and B-2041. As these strains were both isolated in Brisbane, Australia, and have both been genetically classified to the A1 sub-assemblage of *G. duodenalis* [39, 41], it is not surprising that the strains share a large core of common proteins. However, B-2041 has been experimentally demonstrated to be more virulent than H-106, and this is potentially reflected in the variation between the strains in 154 unique proteins found in the avian, virulent strain and the 168 in the human, lab-standard strain (Fig. 1). Of the 154 unique proteins in B-2041, 74 (47.7%) were putative uncharacterized proteins, and a further 21 proteins (13.5%) belonged to proteins in the variable genome of *Giardia*, of which 10 were VSPs. Of the 168 unique proteins in H-106, 83 (49.1%) were putative uncharacterized proteins and 30 were from the variable genome, of which 6 were from the VSP family, 13 were from the NEK kinase family, and 8 were from the Protein 21.1 gene family.

The variable genome also played a key role in differentially expressed proteins. B-2041 contained 82 differentially expressed proteins in the GPF dataset, of which 16 proteins were expressed uniquely in the strain. For H-106 there were 110 proteins differentially expressed in the GPF dataset and of these 18 proteins were uniquely expressed. In H-106 there were 25 proteins (22.7%) of greater abundance from the four protein families that constitute the variable genome of *Giardia*, while in B-2041 there were 29 such proteins (35.4%). Further examination of the gene families show that in the virulent B-2041 the VSP family was overrepresented in proteins of greater abundance, with 21 proteins (25.6%) identified as VSPs, while NEK kinase, HCMP, and Protein 21.1 gene families only contributed 8 proteins (9.8%) of greater abundance in the strain. This is a different trend to the nonvirulent H-106 strain, where there were only 3 VSP and 22 proteins (20%) differentially expressed in the HCMP, NEK kinase, and Protein 21.1 families, in particular 13 from the Protein 21.1 family.

These results may potentially indicate that intraassemblage variation in *G. duodenalis* may be governed by only a small set of proteins, potentially those proteins specifically associated with the variable genome. Jerlstrom-Hultqvist et al. [15] compared the genomes of *G. duodenalis* assemblages A, B, and E and found that the four gene families of the variable genome, the HCMP, VSP, Protein 21.1 and NEK kinase gene families, constituted 9% of the *G. duodenalis* genome. It was also shown that these gene families were associated with synteny breaks that indicated they may evolve more rapidly than other genes [15]. VSPs and HCMPs, which code for surface proteins, were also found in other nonsyntenic regions of the genome and were often assemblage-specific genes. This led to the hypothesis that increased genome plasticity in these areas may allow isolates to diverge in their surface protein repertoire, possibly through gene duplications, insertions, and deletions [15]. The divergence of the *G. duodenalis* surface composition and antigenicity may suggest genetic adaptation to host environment, as assemblages A, B, and E have varying host specificity and zoonotic potential [3,15]. Additionally, if duplications and expansions of the variable genome could possibly be isolate-specific as well as assemblage or sub-assemblage-specific, then our results may exclude further variable genome protein identifications as these might be absent in the A1 genome used. As there is currently only one representative genome available for each of the sub-assemblages or A1 and A2, the extent of genetic diversity between isolates of the same sub-assemblage is not currently understood [17].

It is not known if the variable genome could play a role in intraassemblage variation as well as interassemblage. Host variation and specificity hve been considered a potential factor in isolate virulence in *G. duodenalis* [3], and perhaps the host species available to the parasite (whether a specific species or a broad zoonotic range) could influence the number and expression of proteins in the variable genome. As one of the main differences between B-2041 and H-106 is their host origin, it is hard to exclude host specificity as potentially responsible for the virulence phenotype itself. Indeed, B-2041 is known to be infective for birds [39], neonatal and adult mice, and also domestic kittens and lambs [40]. In contrast adult mice are found to have acquired immunity to H-106 [39] and this also suggests that B-2041 is a zoonotic strain with an ability to infect a wide range of vertebrate species. It is possible that the zoonotic phenotype may also constitute the virulence phenotype in *G. duodenalis*.

Unfortunately, excluding the VSP gene family, there is very little known about the biological function of the other proteins of the variable genome. The NEK kinases are a large gene family of which ~70% lack catalytic amino acid residues and are therefore considered inactive, while even less is known about the Protein 21.1 family, although they share homology to some kinase-like and cryptic kinase-like domains [64]. Despite a lack of functional information on the NEK kinase and Protein 21.1 family, our data suggest that proteins from these families are expressed in high numbers, especially in the proteins of greater abundance in H-106. The HCMPs, although they share similar structural features to VSPs, are regulated and expressed more like cyst wall proteins during encystation, and their role in *Giardia* biology remains unclear [60]. Until more information on the evolutionary origins, biology, and protein function is elucidated in these gene families, it will be impossible to determine their role in inter- and intraassemblage variation.

As a consequence, analysis of our data using GO was problematic, as none of the HCMP, Protein 21.1, NEK kinase, or VSP families had been assigned any GO function. Similarly, results included a large portion of putative uncharacterized proteins (19–25% of all differentially expressed proteins) with no assigned GO, which is consistent with previous genome analysis that found a very large fraction of hypothetical genes in *Giardia* with limited expression, little to no homology to other protozoa, and unknown function [15]. This meant that analysis using GO mostly showed differences in cellular and metabolic function, which correlated to the differences in growth rate and utilization of media between B-2041 and H-106, with B-2041 previously recorded as growing slower in axenic culture than the human lab-standard [5].

Though there are few virulence factors known in *G. duodenalis*, it has recently been shown that the cathepsin B family plays a role in modulating the host immune response and is upregulated in *Giardia* trophozoites exposed to gastrointestinal cells [65]. A total of seven cathepsins were identified in the GPF dataset for both strains, two cathepsin L and five cathepsin B proteins, with the two most abundant cathepsin B proteins showing at least a twofold increase in abundance in B-2041 compared to H-106. Considering B-2041 has been shown to cause prolonged and chronic infection in mice models [39], differential abundance of cathepsins between strains might be an important mechanism of virulence through host immune modulation.

3.7 VSPs and virulence

A total of 35 VSPs were reproducibly identified in both strains in the GPF dataset, which accounts for 2.9% of all proteins. Previous studies have estimated VSP coding regions to be approximately 3–5% of the *G. duodenalis* genome; 218 complete and 75 incomplete or partial VSP genes have been identified in the Assemblage A genome, strain WB [13]. However, VSP proteins constituted 75% (12) of all unique proteins and 25.6% of all differentially expressed proteins in the virulent strain, a trend not seen in the lab-standard strain. Of the 35 VSPs that were reproducibly identified, 3 were common to both strains, although H-106 expressed only 6 unique VSPs compared to 26 unique VSPs seen in virulent B-2041. This means that B-2041 has a population surface diversity and heterogeneity that far exceeds that seen in strain H-106.

VSPs are cysteine-rich surface proteins, ranging in size from 20 to 200 kDa, with a variable domain at the amino termini, a semi-conserved domain, and a conserved carboxyl

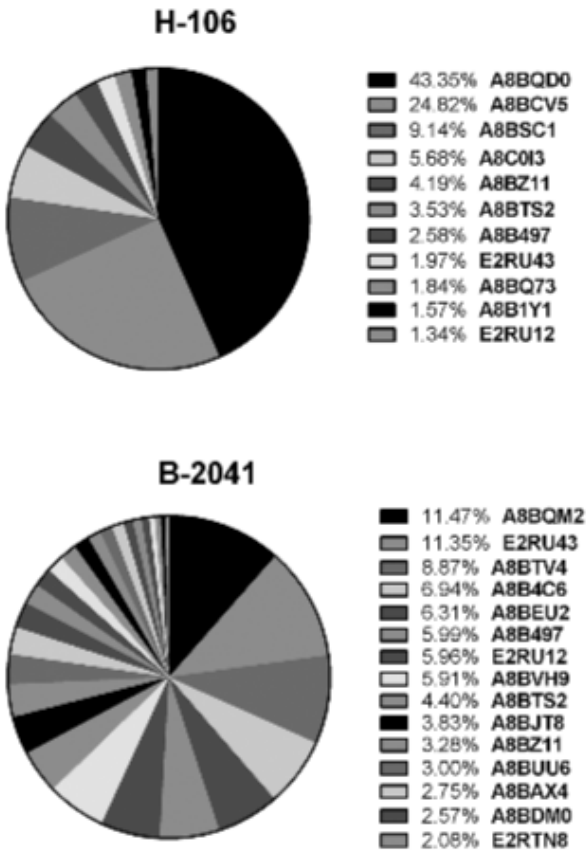


Figure 2. Pie charts depicting the relative abundance of VSPs in the trophozoite population calculated by NSAF for both strains (BRIS/82/HEPU/106 and BRIS/95/HEPU/2041) in the GPF method. Identifiers for the VSPs listed in the figure correspond to their UniProt accession numbers.

termini and cytoplasmic tail [2, 59]. The cysteine content of VSP is estimated to be around 12%, owing to multiple CXXC motifs that create disulphide bonds, and consequently render VSPs in their native state resistant to proteolysis in the small intestine [57, 66]. Although there are over 200 complete VSP genes in Assemblage A, only one VSP is expressed on the entire surface of the *Giardia* trophozoite at any given time [67]. It has been recently shown that VSP expression is controlled by posttranscriptional gene silencing (PTGS), and the multiple mRNAs generated from VSP genes are silenced so that trophozoites only accumulate the transcript of the single VSP that will be expressed on the surface [68]. VSPs are involved in antigen switching and are putative virulence factors for immune evasion [2]. Regulation of VSP switching events are still under investigation, though microRNAs (miRNAs) derived from small nucleolar RNAs (snoRNAS) may be responsible for the disappearance of one VSP on the trophozoite surface and the appearance of a new one [69]. It is currently unknown whether external or environmental factors may influence the VSP-switching event, or whether the change in one VSP to another in *Giardia* is random and unregulated. It has also been demonstrated that encystation/excystation is a universal mechanism for VSP switching [70], and that diversification of VSPs occurs throughout a population even derived from a single clone, as shown through cytotoxic MAb selection [71]. Although VSPs are characteristic of the trophozoite stage, VSPs are even detected in the cyst stage as they are expressed on the surface of the two exyzoites within the cyst structure [60]. VSPs have been shown to change in axenic culture approximately every 6–13 generations, and it is possible axenic cultures contain higher numbers of VSPs due to the absence of immune selection in a culture environment, reflective of different rates of antigen-switching events [67].

As only one VSP is expressed on the surface of *G. duodenalis* trophozoites at any given time, the presence of multiple VSPs in either strain is indicative of variant subpopulations. Surface variation within a population may help trophozoites elude host immune responses, although antigenic variation within *Giardia* still occurs in the absence of immune pressure, including within axenic culture that may be responsible for some of the diversity seen in this study [2, 59]. It has been hypothesized that VSPs might also be responsible for host specificity, with some particular VSPs predicted to be adaptive to a certain host species [15]. If this is true, then the increased number of VSPs seen in B-2041 might reflect its demonstrated ability to infect a wide range of vertebrate

species [5, 40]. In the course of an infection of a new host, increased variant populations in B-2041 would allow proliferating trophozoites to give rise to those subpopulations expressing host-adapted surface antigens. Alternatively, the higher number of VSPs might be indicative of different rates of switching events between these isolates, with B-2041 showing higher abundance in the GPF dataset of arginine deiminase, which has been experimentally shown to be involved in VSP-switching events [72].

Analysis of the VSP composition of each strain population also yielded distinct differences between the two strains beyond the number of VSPs identified. Figure 2 shows the population profile of each strain for reproducibly identified VSPs by NSAF for the GPF datasets. H-106 is dominated by two variant populations that comprise over half the trophozoite population, and after which the abundance of remaining variant populations dramatically decreases. However, there are no dominant variant populations seen in the virulent strain B-2041, instead the abundances are more evenly distributed. The most abundant VSP by NSAF value seen in H-106 is 43.4%, while the most abundant in B-2041 is only 11.5%. This means that while nearly 77.3% of the total SpCs of VSPs identified in H-106 are from the top three abundant VSP subpopulations, the top three most abundant VSPs in B-2041 only account for 31.7% of VSP SpCs in B-2041. Previously, B-2041 has been demonstrated to be more virulent, with parasite load up to 6.7 times greater than H-106 and persisting 14 days longer on average [5]. Additionally, although IgA and IgM were high in mice infected with B-2041, very few of the antibodies were specific for *Giardia* antigens [5]. Greater antigen variation with B-2041 could explain both its infection persistence as well as the lack of specific antibodies generated, with a highly diverse population capable of evading host immune responses. This is particularly interesting in light of increased protein abundance of two cathepsin B proteins in B-2041, recently shown to modulate the host immune response [65], and when considered with the greater antigen variation in B-2041 may be responsible for the persistence of the infection in the avian strain.

4 Concluding remarks

Despite recent advances in genetic taxonomy, the relative ease of axenic cultures, and the availability of several *G. duodenalis* genome sequences, proteomics in this parasite has been largely neglected. This is especially relevant for shotgun proteomics experiments, which offer a unique opportunity to generate a large amount of data for the understanding of clinically relevant parasites. In the case of *G. duodenalis*, shotgun proteomics allows the diagnosis of protein abundance trends in genetically and geographically similar strains with divergent disease phenotypes. If *Giardia* strains remain well classified in terms of assemblage, sub-assemblage, and isolate information, then proteomics has the potential to be very useful for generating protein targets for virulence, zoonoses, and host specificity.

The variable genome in *G. duodenalis* has already been implicated in genetic interassemblage diversity [15], but in this report we have shown that it might also play a role in variation between strains of the same assemblage. In particular, VSPs may play a role in virulence and host specificity, with the more virulent, zoonotic *G. duodenalis* strain having a greater number of VSPs within its population, as well as a very different VSP population structure when compared to the less virulent, lab-standard strain. Due to transcriptional gene silencing of VSP transcripts, detection of multiple VSPs in a population is problematic using genetic techniques [69, 73], and proteomics should be considered a unique tool capable of capturing a comprehensive picture of the variant VSP subpopulations in *Giardia*.

S.J.E. acknowledges support from the Australian government in the form of an APA scholarship, as well as financial support from Macquarie University. S.J.E. wishes to thank Dr. Ernest Lacey and Microbial Screening Technologies for ongoing support as well as Dr. Jacqui Upcroft for supplying the Giardia samples used. P.A.H. wishes to thank Justin Lane for continued support and encouragement.

The authors have declared no conflict of interest.

5 References

[1] Thompson, R. C., Meloni, B. P., Molecular variation in *Giardia*. *Acta Trop.* 1993, *53*, 167–184.

[2] Ankarklev, J., Jerlstrom-Hultqvist, J., Ringqvist, E., Troell, K., Svard, S. G., Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nat. Rev. Microbiol.* 2010, *8*, 413–422.

[3] Feng, Y., Xiao, L., Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. *Clin. Microbiol. Rev.* 2011, *24*, 110–140.

[4] Homan, W. L., Mank, T. G., Human giardiasis: genotype linked differences in clinical symptomatology. *Int. J. Parasitol.* 2001, *31*, 822–826.

[5] Williamson, A. L., O'Donoghue, P. J., Upcroft, J. A., Upcroft, P., Immune and pathophysiological responses to different strains of *Giardia duodenalis* in neonatal mice. *Int. J. Parasitol.* 2000, *30*, 129–136.

[6] Erickson, M. C., Ortega, Y. R., Inactivation of protozoan parasites in food, water, and environmental systems. *J. Food Prot.* 2006, *69*, 2786–2808.

[7] Capon, A. G., Upcroft, J. A., Boreham, P. F., Cottis, L. E., Bundesen, P. G., Similarities of *Giardia* antigens derived from human and animal sources. *Int. J. Parasitol.* 1989, *19*, 91–98.

[8] Upcroft, J. A., Boreham, P. F., Campbell, R. W., Shepherd, R. W., Upcroft, P., Biological and genetic analysis of a longitudinal collection of *Giardia* samples derived from humans. *Acta Trop.* 1995, *60*, 35–46.

[9] Upcroft, J. A., Boreham, P. F., Upcroft, P., Geographic variation in *Giardia* karyotypes. *Int. J. Parasitol.* 1989, *19*, 519–527.

[10] Caccio, S. M., Thompson, R. C., McLauchlin, J., Smith, H. V., Unravelling *Cryptosporidium* and *Giardia* epidemiology. *Trends Parasitol.* 2005, *21*, 430–437.

[11] Monis, P. T., Caccio, S. M., Thompson, R. C., Variation in *Giardia*: towards a taxonomic revision of the genus. *Trends Parasitol.* 2009, *25*, 93–100.

[12] Thompson, A., Human giardiasis: genotype-linked differences in clinical symptomatology. *Trends Parasitol.* 2001, *17*, 465.

[13] Adam, R. D., Nigam, A., Seshadri, V., Martens, C. A. et al., The *Giardia lamblia* vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC Genomics* 2010, *11*, 424.

[14] Franzen, O., Jerlstrom-Hultqvist, J., Castro, E., Sherwood, E. et al., Draft genome sequencing of *Giardia intestinalis* assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog* 2009, *5*, e1000560.

[15] Jerlstrom-Hultqvist, J., Franzen, O., Ankarklev, J., Xu, F. et al., Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. *BMC Genomics* 2010, *11*, 543.

[16] Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B. et al., Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 2007, *317*, 1921–1926.

[17] Adam, R. D., Dahlstrom, E. W., Martens, C. A., Bruno, D. P. et al., Genome sequencing of *Giardia lamblia* genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). *Genome Biol. Evol.* 2013, *5*, 2498–2511.

[18] Buret, A. G., Mechanisms of epithelial dysfunction in giardiasis. *Gut* 2007, *56*, 316–317.

[19] Hanash, S., Disease proteomics. *Nature* 2003, *422*, 226–232.

[20] Wastling, J. M., Armstrong, S. D., Krishna, R., Xia, D., Parasites, proteomes and systems: has Descartes' clock run out of time? *Parasitology* 2012, *139*, 1103–1118.

[21] Aurrecoechea, C., Brestelli, J., Brunk, B. P., Fischer, S. et al., EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Res.*, *38*, D415–419.

[22] Fraunholz, M. J., Roos, D. S., PlasmoDB: exploring genomics and post-genomics data of the malaria parasite, *Plasmodium falciparum*. *Redox Rep.* 2003, *8*, 317–320.

[23] Kissinger, J. C., Gajria, B., Li, L., Paulsen, I. T., Roos, D. S., ToxoDB: accessing the *Toxoplasma gondii* genome. *Nucleic Acids Res.* 2003, *31*, 234–236.

[24] Puiu, D., Enomoto, S., Buck, G. A., Abrahamsen, M. S., Kissinger, J. C., CryptoDB: the *Cryptosporidium* genome resource. *Nucleic Acids Res.* 2004, *32*, D329–331.

[25] Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M. et al., Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 2002, *419*, 537–542.

[26] Xia, D., Sanderson, S. J., Jones, A. R., Prieto, J. H. et al., The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol* 2008, *9*, R116.

[27] Sanderson, S. J., Xia, D., Prieto, H., Yates, J. et al., Determining the protein repertoire of *Cryptosporidium parvum* sporozoites. *Proteomics* 2008, *8*, 1398–1414.

[28] Steuart, R. F., Proteomic analysis of *Giardia*: studies from the pre- and post-genomic era. *Exp. Parasitol.* 2010, *124*, 26–30.

[29] Lauwaet, T., Smith, A. J., Reiner, D. S., Romijn, E. P. et al., Mining the *Giardia* genome and proteome for conserved and unique basal body proteins. *Int. J. Parasitol.* 2011, *41*, 1079–1092.

[30] Hagen, K. D., Hirakawa, M. P., House, S. A., Schwartz, C. L. et al., Novel structural components of the ventral disc and lateral crest in *Giardia intestinalis*. *PLoS Negl. Trop. Dis.* 2011, *5*, e1442.

[31] Jedelsky, P. L., Dolezal, P., Rada, P., Pyrih, J. et al., The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLoS One* 2011, *6*, e17285.

[32] Faso, C., Bischof, S., Hehl, A. B., The proteome landscape of *Giardia lamblia* encystation. *PLoS One* 2013, *8*, e83207.

[33] Aurrecoechea, C., Brestelli, J., Brunk, B. P., Carlton, J. M. et al., GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* 2009, *37*, D526–530.

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

[35] Ong, S. E., Mann, M., Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods Mol. Biol.* 2007, *359*, 37–52.

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

[37] Kim, J., Bae, S. S., Sung, M. H., Lee, K. H., Park, S. J., Comparative proteomic analysis of trophozoites versus cysts of *Giardia lamblia*. *Parasitol. Res.* 2009, *104*, 475–479.

[38] Steuart, R. F., O'Handley, R., Lipscombe, R. J., Lock, R. A., Thompson, R. C., Alpha 2 giardin is an assemblage A-specific protein of human infective *Giardia duodenalis*. *Parasitology* 2008, *135*, 1621–1627.

[39] Upcroft, J. A., McDonnell, P. A., Gallagher, A. N., Chen, N., Upcroft, P., Lethal *Giardia* from a wild-caught sulphur-crested cockatoo (*Cacatua galerita*) established in vitro chronically infects mice. *Parasitology* 1997, *114*(Pt 5), 407–412.

[40] McDonnell, P. A., Scott, K. G., Teoh, D. A., Olson, M. E. et al., Giardia duodenalis trophozoites isolated from a parrot (Cacatua galerita) colonize the small intestinal tracts of domestic kittens and lambs. *Vet. Parasitol.* 2003, *111*, 31–46.

[41] Nolan, M. J., Jex, A. R., Upcroft, J. A., Upcroft, P., Gasser, R. B., Barcoding of *Giardia duodenalis* isolates and derived lines from an established cryobank by a mutation scanning-based approach. *Electrophoresis* 2011, *32*, 2075–2090.

[42] Zhou, D. H., Yuan, Z. G., Zhao, F. R., Li, H. L. et al., Modulation of mouse macrophage proteome induced by *Toxoplasma gondii* tachyzoites in vivo. *Parasitol. Res.* 2011, *109*, 1637–1646.

[43] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* 2009, *6*, 359–362.

[44] Manza, L. L., Stamer, S. L., Ham, A. J., Codreanu, S. G., Liebler, D. C., Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* 2005, *5*, 1742–1745.

[45] Scherl, A., Shaffer, S. A., Taylor, G. K., Kulasekara, H. D. et al., Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. *Anal. Chem.* 2008, *80*, 1182–1191.

[46] Keister, D. B., Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans. R Soc. Trop. Med. Hyg.* 1983, *77*, 487–488.

[47] Boreham, P. F., Phillips, R. E., Shepherd, R. W., The activity of drugs against *Giardia intestinalis* in neonatal mice. *J. Antimicrob. Chemother.* 1986, *18*, 393–398.

[48] Gammulla, C. G., Pascovici, D., Atwell, B. J., Haynes, P. A., Differential metabolic response of cultured rice (*Oryza sativa*) cells exposed to high- and low-temperature stress. *Proteomics* 2010, *10*, 3001–3019.

[49] Chapman, B., Castellana, N., Apffel, A., Ghan, R. et al., Plant proteogenomics: from protein extraction to improved gene predictions. *Methods Mol. Biol.* 2013, *1002*, 267–294.

[50] Wessel, D., Flugge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984, *138*, 141–143.

[51] Vizcaino, J. A., Cote, R. G., Csordas, A., Dianes, J. A. et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, *41*, D1063–1069.

[52] Mirzaei, M., Soltani, N., Sarhadi, E., Pascovici, D. et al., Shotgun proteomic analysis of long-distance drought signaling in rice roots. *J. Proteome Res.* 2012, *11*, 348–358.

[53] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K. et al., Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* 2006, *5*, 2339–2347.

[54] Deshusses, J. M., Burgess, J. A., Scherl, A., Wenger, Y. et al., Exploitation of specific properties of trifluoroethanol for extraction and separation of membrane proteins. *Proteomics* 2003, *3*, 1418–1424.

[55] Ru, Q. C., Zhu, L. A., Katenhusen, R. A., Silberman, J. et al., Exploring human plasma proteome strategies: high efficiency in-solution digestion protocol for multi-dimensional protein identification technology. *J. Chromatogr. A* 2006, *1111*, 175–191.

[56] Zhang, H., Lin, Q., Ponnusamy, S., Kothandaraman, N. et al., Differential recovery of membrane proteins after extraction by aqueous methanol and trifluoroethanol. *Proteomics* 2007, *7*, 1654–1663.

[57] Papanastasiou, P., McConville, M. J., Ralton, J., Kohler, P., The variant-specific surface protein of *Giardia*, VSP4A1, is a glycosylated and palmitoylated protein. *Biochem. J.* 1997, *322*(Pt 1), 49–56.

[58] Touz, M. C., Conrad, J. T., Nash, T. E., A novel palmitoyl acyl transferase controls surface protein palmitoylation and cytotoxicity in *Giardia lamblia*. *Mol. Microbiol.* 2005, *58*, 999–1011.

[59] Prucca, C. G., Rivero, F. D., Lujan, H. D., Regulation of antigenic variation in *Giardia lamblia*. *Annu. Rev. Microbiol.* 2011, *65*, 611–630.

[60] Davids, B. J., Reiner, D. S., Birkeland, S. R., Preheim, S. P. et al., A new family of giardial cysteine-rich non-VSP protein genes and a novel cyst protein. *PLoS One* 2006, *1*, e44.

[61] Weiland, M. E., McArthur, A. G., Morrison, H. G., Sogin, M. L., Svard, S. G., Annexin-like alpha giardins: a new cytoskeletal gene family in *Giardia lamblia*. *Int. J. Parasitol.* 2005, *35*, 617–626.

[62] Saric, M., Vahrmann, A., Niebur, D., Kluempers, V. et al., Dual acylation accounts for the localization of {alpha}19-giardin in the ventral flagellum pair of *Giardia lamblia*. *Eukaryot. Cell* 2009, *8*, 1567–1574.

[63] Humen, M. A., Perez, P. F., Lievin-Le Moal, V., Lipid raft-dependent adhesion of *Giardia intestinalis* trophozoites to a cultured human enterocyte-like Caco-2/TC7 cell monolayer leads to cytoskeleton-dependent functional injuries. *Cell Microbiol.* 2011, *13*, 1683–1702.

[64] Manning, G., Reiner, D. S., Lauwaet, T., Dacre, M. et al., The minimal kinome of *Giardia lamblia* illuminates early kinase evolution and unique parasite biology. *Genome Biol.* 2011, *12*, R66.

[65] Cotton, J. A., Bhargava, A., Ferraz, J. G., Yates, R. M. et al., Giardia duodenalis cathepsin B proteases degrade intestinal epithelial interleukin-8 and attenuate interleukin-8-induced neutrophil chemotaxis. *Infect. Immun.* 2014, *82*, 2772–2787.

[66] Papanastasiou, P., Bruderer, T., Li, Y., Bommeli, C., Kohler, P., Primary structure and biochemical properties of a variant-specific surface protein of *Giardia*. *Mol. Biochem. Parasitol.* 1997, *86*, 13–27.

[67] Prucca, C. G., Lujan, H. D., Antigenic variation in *Giardia lamblia*. *Cell Microbiol.* 2009, *11*, 1706–1715.

[68] Prucca, C. G., Slavin, I., Quiroga, R., Elias, E. V. et al., Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* 2008, *456*, 750–754.

[69] Li, W., Saraiya, A. A., Wang, C. C., The profile of snoRNA-derived microRNAs that regulate expression of variant surface proteins in *Giardia lamblia*. *Cell Microbiol.* 2012, *14*, 1455–1473.

[70] Svard, S. G., Meng, T. C., Hetsko, M. L., McCaffery, J. M., Gillin, F. D., Differentiation-associated surface antigen variation in the ancient eukaryote *Giardia lamblia*. *Mol. Microbiol.* 1998, *30*, 979–989.

[71] Nash, T. E., Aggarwal, A., Adam, R. D., Conrad, J. T., Merritt, J. W., Jr., Antigenic variation in *Giardia lamblia*. *J. Immunol.* 1988, *141*, 636–641.

[72] Touz, M. C., Ropolo, A. S., Rivero, M. R., Vranych, C. V. et al., Arginine deiminase has multiple regulatory roles in the biology of *Giardia lamblia*. *J. Cell Sci.* 2008, *121*, 2930–2938.

[73] Bienz, M., Siles-Lucas, M., Wittwer, P., Muller, N., Vsp gene expression by *Giardia lamblia* clone GS/M-83-H7 during antigenic variation in vivo and in vitro. *Infect. Immun.* 2001, *69*, 5278–5285.

2.4 Supplementary Data

The following supplementary information is available for this manuscript.

Supplementary Table S1. Biological classification of differentially expressed proteins (up regulated, down regulated and unchanged) obtained from t-test analysis of BRIS/95/HEPU/2041 vs. BRIS/82/HEPU/106 in the gas phase fractionation dataset. Proteins which are differentially expressed in BRIS/95/HEPU/2041 and BRIS/82/HEPU/106 are filled in green and red respectively on the first tab. All proteins from the GPF dataset are shown in the first tab (GPF), while the 110 proteins of greater abundance in BRIS/82/HEPU/106 are shown in the second tab (H-106) and the 82 proteins of greater abundance from BRIS/95/HEPU/2041 are shown in the third tab (B-2041). **(Supplementary DVD)**

Supplementary Table S2. Biological classification of differentially expressed proteins (up regulated, down regulated and unchanged) obtained from t-test analysis of BRIS/95/HEPU/2041 vs. BRIS/82/HEPU/106 in the gel LC-MS/MS dataset. Proteins which are differentially expressed in BRIS/95/HEPU/2041 and BRIS/82/HEPU/106 are filled in green and red respectively on the first tab. All proteins from the GPF dataset are shown in the first tab (GPF), while the 37 proteins of greater abundance in BRIS/82/HEPU/106 are shown in the second tab (H-106) and the 29 proteins of greater abundance from BRIS/95/HEPU/2041 are shown in the third tab (B-2041). **(Supplementary DVD)**

CHAPTER 3

A comprehensive quantitative proteomic experiment analysing eight assemblage A Giardia duodenalis *isolates.*

3. Quantitative proteomics analysis of *Giardia duodenalis* Assemblage A – a baseline for host, assemblage and isolate variation.

3.1 Context

Giardia duodenalis is considered a species complex, and its taxonomy has been recently reclassified into eight assemblages, of which some are further classified into subassemblages. With considerable genetic data and genomes already available, this experiment was designed to provide a comprehensive dataset for a variety of isolates from the A assemblage. Eight isolates (seven A1 and one A2) with diverse origins previously characterised in the literature according to karotype, assemblage, virulence, geographic variation and drug resistance, were examined using the optimised workflow established in Chapter 2. Proteomic data was also searched against recently reclassified A1 and A2 subassemblage genomes to examine database dependent gains and losses of information. This provided a comprehensive dataset for a baseline of isolate variation, and also an evaluation of the utility of employing assemblage A genome sequence data for peptide to spectra matching in experiments involving a variety of isolates.

3.2 Contributions and Permissions

I performed 100% of all work pertaining to growing and collecting *Giardia* samples for proteomic analysis, and 100% of all lab work. I wrote 80% of the manuscript with assistance and editing provided by Professor Paul Haynes and Dr Ernest Lacey.

Publication III was published in the literature originally in the journal *Proteomics* as a continuous communication format known as a 'Dataset Brief'. Following this publication, we were invited to share our dataset with a potentially wider audience through publication of the data in a new journal called *Data in Brief,* which provides a platform to describe, share and reuse large –omics datasets.

As the *Data in Brief* publication (Publication IV) is focused mainly on the methods and data, while the *Proteomics* publication (Publication III) contains a general discussion of the results and some consideration of their biological implications, we have chosen to include the *Data in Brief* article first, followed by the *Proteomics* article in this thesis.

3.5 Publication IV

Data in Brief 5 (2015) 23–27

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data from a proteomic baseline study of Assemblage A in *Giardia duodenalis*

Samantha J. Emery [a], Ernest Lacey [b], Paul A. Haynes [a,*]

[a] *Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW 2109, Australia*

[b] *Microbial Screening Technologies Pty Ltd, Smithfield, NSW 2165, Australia*

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form
4 August 2015
Accepted 9 August 2015
Available online 19 August 2015

Keywords:
Assemblage A
Giardia duodenalis
Label-free quantitative shotgun proteomics
Variant Surface Protein
Variable genome
Parasite proteomics

ABSTRACT

Eight Assemblage A strains from the protozoan parasite *Giardia duodenalis* were analysed using label-free quantitative shotgun proteomics, to evaluate inter- and intra-assemblage variation and complement available genetic and transcriptomic data. Isolates were grown in biological triplicate in axenic culture, and protein extracts were subjected to in-solution digest and online fractionation using Gas Phase Fractionation (GPF). Recent reclassification of genome databases for subassemblages was evaluated for database-dependent loss of information, and proteome composition of different isolates was analysed for biologically relevant assemblage-independent variation. The data from this study are related to the research article "*Quantitative proteomics analysis of Giardia duodenalis Assemblage A – a baseline for host, assemblage and isolate variation*" published in Proteomics (Emery et al., 2015 [1]).

Specifications table

Subject area	Biology

* Corresponding author.
E-mail address: paul.haynes@mq.edu.au (P.A. Haynes).

http://dx.doi.org/10.1016/j.dib.2015.08.003

More specific subject area	Quantitative proteomic data of 8 *Giardia duodenalis* Assemblage A isolates using gas phase fractionation and normalised spectral abundance factors (NSAF).
Type of data	Table, Figure, Supplementary Tables
How data was acquired	Protein extracts from biological triplicates were digested in solution, and fractionated online using GPF with mass range fraction optimised for the *G. duodenalis* A1 subassemblage genome. Data was acquired on a LTQ-XL Linear Ion Trap (Thermo).
Data format	Raw data, reproducibly identified proteins.
Experimental factors	8 *G. duodenalis* strains grown in Axenic culture from animal and human hosts, covering both subassemblage A1 and A2 to analyse isolate variation. Data was searched against both A1 subassemblage genome database and recently released A2 subassemblage database to compare database-specific losses.
Experimental features	Sample triplicates were combined to produce reproducibly identified proteins and spectral counts of each protein were used to calculate NSAF values for each protein.
Data source location	Sydney, NSW, Australia
Data accessibility	Data is available from http://www.ebi.ac.uk/pride/archive/projects/PXD001272 and will also be made available through the giardiadb.org website later in 2015.

Value of the data

- First proteomic baseline for taxonomy and isolate variation in Assemblage A strains.
- Provides proteome coverage of isolates from animal and human hosts, both A1 and A2 subassemblages, with an emphasis on Australian isolates.
- Evaluates database-dependent losses based on new genome reclassifications and releases in Assemblage A.
- Identifies sources of inter- and intra-assemblage A isolate variation and its impacts.

1. Experimental design, materials and methods

1.1. Isolate selection, axenic culture, protein extraction and digestion

Eight Assemblage A strains [1], including the A1 genome strain, were assembled from animal and human infections, previously characterised in the literature according to karotype [2,3], subassemblage [4], virulence [2], geographic variation [5,6] and drug resistance [7]. The full description of strains can be seen in Table 1.

G. duodenalis strains were cultured in triplicate axenically in TYI-S33 media supplemented with 10% newborn calf serum and 1% bile as previously described [8] and harvested from confluent cultures in late log-phase. Trophozoites were harvested by centrifugation, washed twice in ice-cold PBS to remove media traces [9] and pellets of 10^8 trophozoites were extracted into 1 mL ice-cold SDS sample buffer containing 1 mM EDTA and 5% beta-mercaptoethanol, then disulphides were reduced at 75 °C

Table 1
Classification information for the eight *G. duodenalis* strains used in this study including subassemblage, geographic origin, and the host species the strain was isolated from. Strain identification coincides with those previously published in the literature.

Strain	Assemblage	Origin	Host source
BRIS/83/HEPU 106	A1	Brisbane, Australia	Human
BRIS87/HEPU/713	A1	Brisbane, Australia	Human
OAS1	A1	Canada	Sheep (*Ovis aries*)
Bac2	A1	Australia	Cat (*Felis catus*)
BRIS/95/HEPU/2041	A1	Victoria, Australia	Cockatoo (*Cacatua galerita*)
BRIS/89/HEPU/1065	A1	Brisbane, Australia	Human
WB*	A1	Afghanistan	Human
BRIS/89/HEPU/1003	A2	Brisbane, Australia	Human

* Assemblage A1 genome strain (ATCC 50803).

for 10 min. Trophozoite protein extracts were centrifuged at 0 °C at 13,000 × g for 10 min to remove debris, and protein concentration was measured by BCA assay (Pierce). A 500 μg protein pellet was extracted using methanol–chloroform precipitation [10] and in-solution digestion was performed using a modified filter aided sample preparation (FASP) [11]. After peptide extraction all samples were dried using a vacuum centrifuge and reconstituted to 60 μL with 2% formic acid, 2% 2,2,2-trifluorethanol (TFE).

1.2. Nanoflow LC-MS/MS using gas phase fractionation

Optimised gas phase fractionation (GPF) mass ranges were calculated using the 2.5 release of the *G. duodenalis* WB genome for Assemblage A from giardiaDB.org [12]. Charge states +2 and +3 were considered as well as carbamidomethyl as a cysteine modification, and 4 mass ranges were calculated over 400–2000 amu. The mass ranges were as following: the low mass range was 400–518 amu, the low-medium mass range was 518–691 amu, the medium-high mass range was 691–988 amu and the high mass range was 988–2000 amu. Each FASP protein digest for the triplicates of each strain were analysed by nanoLC-MS/MS on an LTQ-XL linear ion trap mass spectrometer (Thermo, San Jose, CA). Peptides were separated on a 150 × 0.2 mm I.D fused-silica column packed with Magic C18AQ (200 Å, 5 μm diameter, Michrom Bioresources, California) connected to an Advance CaptiveSpray Source (Michrom Bioresources, California). Each FASP protein digest was analysed as 4 repeat injections, with the mass spectrometer scanning for 180 min runs for each of the four calculated mass ranges. Samples were injected onto the column using a Surveyor autosampler, followed by an initial wash step with buffer A (0.1% v/v formic acid, 1 mM ammonium formate, 0.2% v/v methanol) for 4 min followed by 150 μL/min for 2 min. Peptides were eluted from the column with 0–80% buffer B (100% v/v ACN, 0.1% v/v formic acid) at 150 μL/min for 167 min finished by a wash step with buffer A for 6 min at 150 μL/min. Spectra in the positive ion mode were scanned over the respective GPF ranges and, using Xcalibur software (Version 2.06, Thermo), automated peak recognition, dynamic exclusion and MS/MS of the top six most-intense ions at 35% normalisation collision energy were performed.

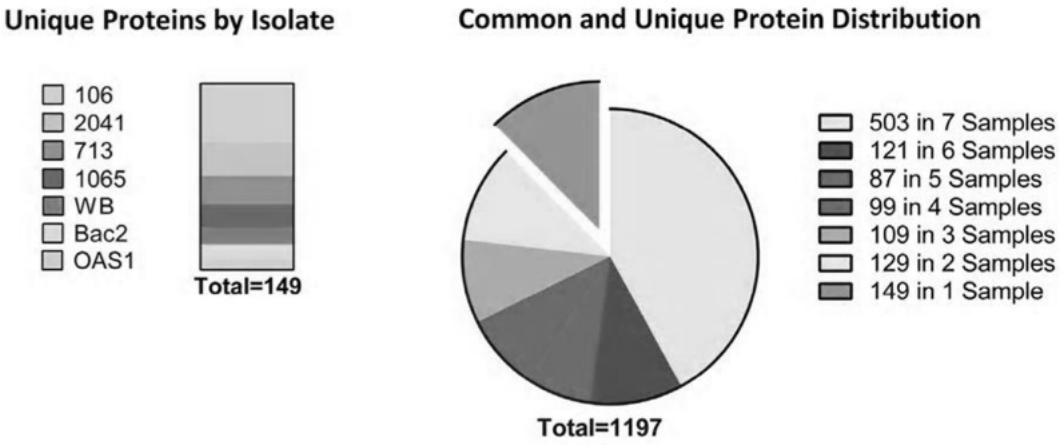


Fig. 1. Distribution of shared and unique proteins in the A1 subassemblage between the 1197 non-redundant proteins identified within the seven isolates analysed. The 1197 proteins were reproducibly identified in at least one isolate, with 149 (12.4%) of these proteins identified within only one isolate, and therefore considered to be uniquely expressed. Part A (left) shows the distribution of these 149 uniquely expressed proteins by isolate in the seven A1 isolates analysed in this study. Part B (right) shows the distribution of the shared proteins between the seven subassemblage A1 isolates. A total of 503 (42%) proteins were identified in all seven isolates examined in this study, and are considered common between isolates of the A1 subassemblage. The remaining segments indicates proteins common within decreasing numbers of isolates, while the final elevated segment indicates the 149 isolate-unique proteins.

1.3. Database searching for protein/peptide information

The LTQ-XL raw output files were converted into mzXML files and searched against the Giardiadb.org 4.0 release of *G. duodenalis* strain Assemblage A1 and A2 genome using the global proteome machine (GPM) software (version 2.1.1) and the X!Tandem algorithm. The 4 fractions for the GPF of each replicate were processed sequentially with output files generated for each individual fraction, and a merged, non-redundant output file for protein identifications with $\log(e)$ values < -1. Peptide identification was determined using MS and MS/MS tolerances of $+2$ Da and $+0.4$ Da. Carbamidomethyl was considered a complete modification, and partial modifications considered included oxidation of methionine and tryptophan.

1.4. Data processing and quantitation

The output from the GPM software (version 2.1.1) [13,14] constituted low stringency protein and peptide identifications, and was used to assess experimental consistency. These data were further processed using the Scrappy software package [15], which combines biological triplicates into a single list of reproducibly identified proteins, which we define in this study as those proteins present reproducibly in all three replicates of at least one strain, with a total spectral count (SpC) of ≥ 5 [15]. Reversed database searching was used for calculating peptide and protein false discovery rates (FDRs) as previously described [15]. Complete protein and peptide data for replicates, including database-dependent losses are shown in Supplementary data 1, Table 1 and in *Giardia* specific gene-families in Supplementary data, Table 2. Protein abundance was calculated using NSAF values [16]. Distribution of reproducibly identified proteins by strain can be viewed in Fig. 1. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [17] via the PRIDE partner repository with the dataset identifier PXD001272.

2. Direct link to deposited data

Data is available through the PRIDE proteomics database through the following link http://www.ebi.ac.uk/pride/archive/projects/PXD001272 and will also be made available through the giardiadb.org website later in 2015.

3. Conflict of interest

The authors declare that there is no conflict of interest on any work published in this paper.

Acknowledgements

SJE acknowledges funding from the Australian Government in the form of an APA scholarship, as well as financial support from Macquarie University. SJE wishes to thank Dr Jacqui Upcroft for supplying the Giardia samples used and for the ongoing support received from colleagues at Microbial Screening Technologies. PAH wishes to thank Justin Lane for continued support and encouragement.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2015.08.003.

References

[1] S.J. Emery, E. Lacey, P.A. Haynes, Quantitative proteomics analysis of *Giardia duodenalis* Assemblage A – a baseline for host, assemblage and isolate variation, Proteomics 15 (2015) 2281–2285.
[2] A.L. Williamson, P.J. O'Donoghue, J.A. Upcroft, P. Upcroft, Immune and pathophysiological responses to different strains of *Giardia duodenalis* in neonatal mice, Int. J. Parasitol. 30 (2000) 129–136.
[3] J.A. Upcroft, N. Chen, P. Upcroft, Mapping variation in chromosome homologues of different *Giardia* strains, Mol. Biochem. Parasitol. 76 (1996) 135–143.
[4] M.J. Nolan, A.R. Jex, J.A. Upcroft, P. Upcroft, R.B. Gasser, Barcoding of *Giardia duodenalis* isolates and derived lines from an established cryobank by a mutation scanning-based approach, Electrophoresis 32 (2011) 2075–2090.
[5] J.A. Upcroft, P.F. Boreham, R.W. Campbell, R.W. Shepherd, P. Upcroft, Biological and genetic analysis of a longitudinal collection of *Giardia* samples derived from humans, Acta Trop. 60 (1995) 35–46.
[6] J.A. Upcroft, P.F. Boreham, P. Upcroft, Geographic variation in *Giardia* karyotypes, Int. J. Parasitol. 19 (1989) 519–527.
[7] J.A. Upcroft, A. Healey, D.G. Murray, P.F. Boreham, P. Upcroft, A gene associated with cell division and drug resistance in *Giardia duodenalis*, Parasitology 104 (Pt 3) (1992) 397–405.
[8] D.B. Keister, Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile, Trans. R. Soc. Trop. Med. Hyg. 77 (1983) 487–488.
[9] L.A. Dunn, J.A. Upcroft, E.V. Fowler, B.S. Matthews, P. Upcroft, Orally administered *Giardia duodenalis* extracts enhance an antigen-specific antibody response, Infect. Immun. 69 (2001) 6503–6510.
[10] D. Wessel, U.I. Flugge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, Anal. Biochem. 138 (1984) 141–143.
[11] B. Chapman, N. Castellana, A. Apffel, R. Ghan, et al., Plant proteogenomics: from protein extraction to improved gene predictions, Methods Mol. Biol. *1002* (2013) 267–294.
[12] A. Scherl, S.A. Shaffer, G.K. Taylor, H.D. Kulasekara, et al., Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides, Anal. Chem. 80 (2008) 1182–1191.
[13] D. Fenyo, R.C. Beavis, A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes, Anal. Chem. 75 (2003) 768–774.
[14] R. Craig, R.C. Beavis, TANDEM: matching proteins with tandem mass spectra, Bioinformatics 20 (2004) 1466–1467.
[15] K.A. Neilson, I.S. George, S.J. Emery, S. Muralidharan, et al., Analysis of rice proteins using SDS-PAGE shotgun proteomics, Methods Mol. Biol. 1072 (2014) 289–302.
[16] B. Zybailov, A.L. Mosley, M.E. Sardiu, M.K. Coleman, et al., Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*, J. Proteome Res. 5 (2006) 2339–2347.
[17] J.A. Vizcaino, R.G. Cote, A. Csordas, J.A. Dianes, et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013, Nucleic Acids Res. 41 (2013) D1063–1069.

3.3 Publication III

Proteomics 2015, *00*, 1–5 DOI 10.1002/pmic.201400434

DATASET BRIEF

Quantitative proteomic analysis of *Giardia duodenalis* assemblage A: A baseline for host, assemblage, and isolate variation

Samantha J. Emery[1], *Ernest Lacey*[2] *and Paul A. Haynes*[1]

[1] Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW, Australia
[2] Microbial Screening Technologies Pty Ltd, Smithfield, NSW, Australia

Received: September 11, 2014
Revised: December 9, 2014
Accepted: February 24, 2015

Giardia duodenalis is a gastrointestinal protozoan parasite of vertebrates and is a species complex comprised of eight assemblages, with the zoonotic assemblage A one of two subtypes infective for humans. With increasing genomic and transcriptomic data publicly available through the centralized giardiaDB.org, we have quantitatively analyzed the proteomes of eight *G. duodenalis* assemblage A strains (seven A1 and one A2) to provide a proteomic baseline to complement the available data. A nonredundant total of 1197 subassemblage A1 proteins and 719 subassemblage A2 proteins were identified with an average of 770 proteins in each strain. The eight strains were also searched against both assemblage A genome sequences (subassemblage A1 and A2 genomes) and demonstrated subassemblage specific differences in protein identifications, especially for variable gene families. Substantial differences were observed in the numbers and abundance in the variable surface protein family, and two different variable surface protein expression profiles that were independent of host origin, subassemblage, or geographic origin. We hypothesize that this variation in surface antigen switching events may be related to karotype and chromosomal variation, which would indicate an assemblage-independent mechanism of diversity generation in *G. duodenalis*. All MS data have been deposited in the ProteomeXchange with identifier PXD001272 (http://proteomecentral.proteomexchange.org/dataset/PXD001272).

Keywords:
Assemblage A / *Giardia duodenalis* / Label-free quantitative shotgun proteomics / Microbiology / Variable genome / Variable surface protein

Additional supporting information may be found in the online version of this article at the publisher's web-site

Giardia duodenalis is a protozoan parasite of vertebrates with both human and veterinary significance, and a major contributor to diarrhoea and gastroenteritis worldwide [1]. *Giardia* is transmitted through the fecal-oral route and has a simple, two-stage life cycle: the environmentally resistant and infective cyst, and the vegetative trophozoite. *Giardiasis* is sometimes clinically asymptomatic, though the disease is usually self-limiting and characterized by diarrhoea, malabsorption, abdominal cramps, nausea, and weight loss along with failure-to-thrive or ill thrift in children and animals [2].

Molecular epidemiology has reclassified *G. duodenalis* as a species complex consisting of eight genetic assemblages (A–H), which segregate phylogenetically and according to host species and zoonoses [1]. Assemblages A and B are infective in humans; assemblage A has a broad host specificity and infects humans, livestock, companion, and wild animals, and hence is considered zoonotic [1]. Intra-assemblage variation between strains classified as closest genetically remains unclear, as mice studies comparing assemblage A strains have

Correspondence: Professor Paul A. Haynes, Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW 2109, Australia
E-mail: paul.haynes@mq.edu.au
Fax: +61-2-9850-8313

Abbreviations: **HCMP**, high cysteine membrane protein; **NSAF**, normalized spectral abundance factor; **SpC**, spectral count; **VSP**, variable surface protein

Table 1. Summary of peptide and protein identification data of *G. duodenalis* proteins across the eight strains analyzed

Sample and subassemblage		Low stringency peptide count			Average number of peptide (±%RSD)	Number of R.I. proteins common to three replicates	R.I. protein FDR (%)	R.I. peptide FDR (%)
		Replicate 1	Replicate 2	Replicate 3				
BRIS/83/HEPU 106	A1	23 016	21 198	22 837	22 350 ± 4.48	895	0.45	0.08
BRIS87/HEPU/713	A1	20 558	19 945	18 893	19 799 ± 4.25	798	0.50	0.12
OAS1	A1	21 111	21 209	18 288	20 203 ± 8.21	716	0.42	0.06
Bac2	A1	20 097	19 285	19 629	19 670 ± 1.99	701	0.29	0.06
BRIS/95/HEPU/2041	A1	21 724	20 807	21 635	21 389 ± 2.36	836	0.72	0.15
BRIS/89/HEPU/1065	A1	18 635	20 740	20 547	19 974 ± 5.83	728	0.27	0.04
WB	A1	21 227	20 699	21 003	20 976 ± 1.26	769	0.39	0.90
BRIS/89/HEPU/1003	A2	19 126	21 036	21 189	20 450 ± 5.62	713	0.42	0.08

Of the eight strains analyzed, all but the A2 strain BRIS/80/HEPU/1003, were searched against the A1 subassemblage reference genome. R.I.: reproducibly identified.

demonstrated differences in pathology and infection intensity [3]. Assemblage A is comprised of the subassemblages A1 and A2, and genome comparisons show high similarity in chromosome synteny and phylogeny, with 5901 and 6724 protein coding ORFs in the A1 and A2 genomes, respectively [4].

Although a considerable amount of genetic information is available for *G. duodenalis*, more baseline data is required at the proteomic level for a systems biology approach to be pursued. In this study, we analyzed eight assemblage A strains of diverse origins previously characterized in the literature according to karotype [3, 5], assemblage [6], virulence [3], geographic variation [7, 8], and drug resistance [9]. Of these eight strains, three are from animals (one each of wild, companion, and livestock animal) and five were isolated from human infections, including one A2 strain as well as the subassemblage A1 genome strain (full strain classification information can be found in Supporting Information Table 1). These strains include host and intra-assemblage variation that will provide a diverse proteomic baseline currently lacking for assemblage A, and will complement current genetic and transcriptomic data available through the parasite database giardiaDB.org [10].

A full description of *G. duodenalis* axenic culture, sample preparation and MS analysis can be found in the Supporting Information. The output from the GPM software (version 2.1.1) [11, 12] constituted low stringency protein and peptide identifications, and was used to assess experimental consistency. These data were further processed using the Scrappy software package [13], which combines biological triplicates into a single list of reproducibly identified proteins, which we define as those proteins present reproducibly in all three replicates of at least one strain, with a total spectral count (SpC) of ≥5 [13]. Reversed database searching was used for calculating peptide and protein false discovery rates (FDRs) as previously described [13]. Protein abundance was calculated using normalized spectral abundance factors (NSAF) [14]. The MS proteomics data have been deposited to the ProteomeXchange Consortium [15] via the PRIDE partner repository with the dataset identifier PXD001272.

A total of 5 376 925 spectra were collected from the eight strains for searching against databases. In total, 1197 nonredundant proteins were identified across the seven subassemblage A1 strains when searched against the A1 genome, with an average of 778 proteins identified in each strain, and ranging from 716 in an A1 animal strain (OAS1) to 895 in an A1 human strain (BRIS/83/HEPU/106). Protein and peptide information for all strains can be viewed in Table 1. Protein and peptide FDR of the reproducibly identified proteins were <1 and <0.15%, respectively, across the eight strains. A total of 503 proteins were common to all seven A1 strains, which accounts for at least 56.1% and up to 75.3% of proteins identified in each strain, strongly supporting the idea of a core set of proteins within *G. duodenalis*. Supporting Information Fig. 1 shows the distribution of the 1197 nonredundant proteins shared and unique between the seven A1 strains. While 90% of *G. duodenalis* genes represent a common or "housekeeping" genome, 10% represents four large genome families, which is referred to as the "variable" genome [16]. These four gene families are integral for *Giardia* biology and are the variable surface proteins (VSPs), high cysteine membrane proteins (HCMPs), NEK kinases, and protein 21.1/ankyrin repeat proteins. Proteins from the variable genome ranged from 10.3 to 16.1% identified in each strain, with VSPs in each strain varying from 0.1 to 5.3% while percent compositions of other variable gene families remained similar. Supporting Information Table 2 shows variable genome proteins identified by strain.

Until recently, the WB strain was the only assemblage A genome of *G. duodenalis* fully sequenced, but the release of a second assemblage A genome sequence has led to a reclassification in the databases. Strain WB is now designated the A1 subassemblage reference genome, while the newly sequenced DH isolate is considered the A2 subassemblage reference genome [4]. Of the eight strains examined in this report, seven have been subtyped as A1 and one as A2 (BRIS/89/HEPU/1003), therefore we searched all eight strains against both A1 and A2 genome sequences to examine database-dependent differences in protein identifications.

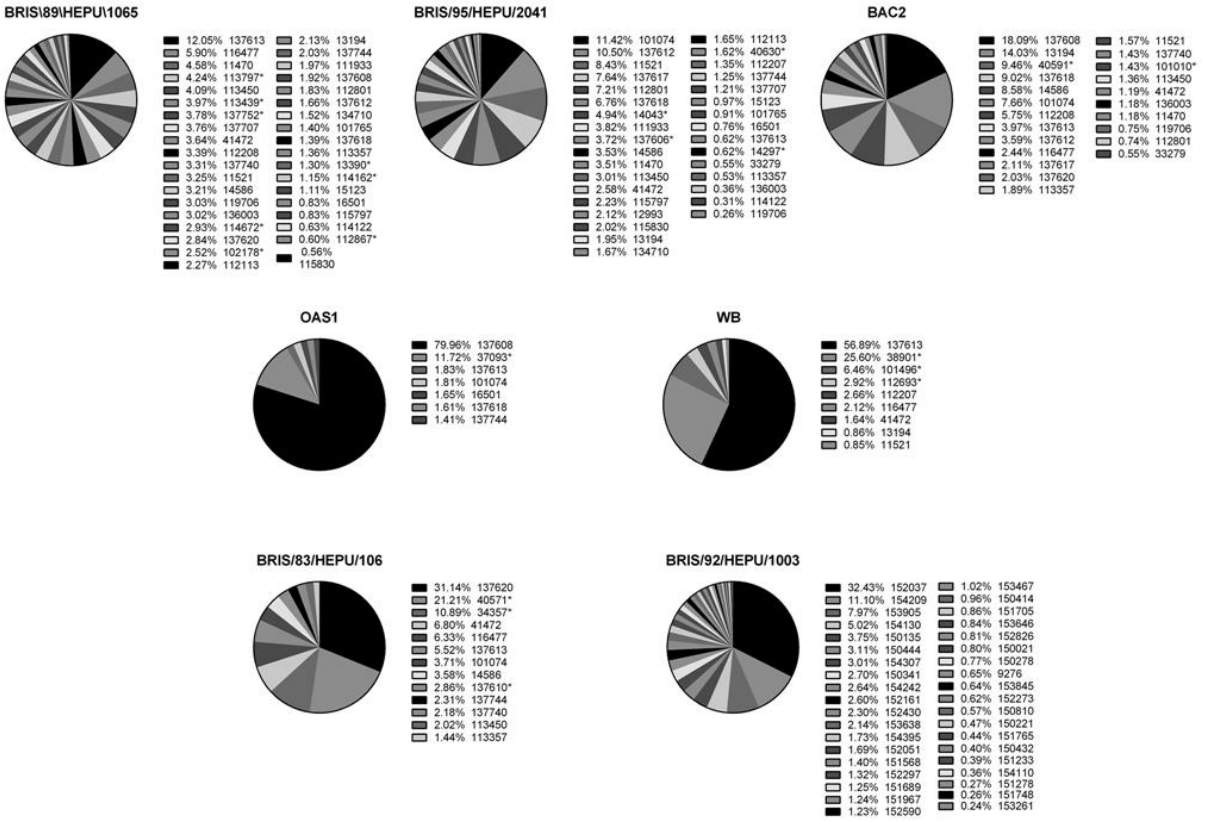


Figure 1. Pie charts demonstrating the distribution of reproducibly identified VSP subpopulations in the eight *Giardia* strains using the triplicate average NSAF. Each VSP is designated using its Giardiadb.org accession and a "*" next to the identifier is indicative that the VSP was identified in that strain only. As BRIS/80/HEPU/1003 is the single A2 strain in this study all its VSPs are considered unique. BRIS87/HEPU/713 is not shown as only one VSP was reproducibly identified.

A detailed comparison of protein and peptide information for results from searches against both genome sequences can be viewed in Supporting Information Table 2. When BRIS/89/HEPU/1003 was searched against the A2 genome sequence, the average peptide count increased by 925 (4.7%) and the number of reproducibly identified proteins increased from 668 to 713. For the seven A1 *G. duodenalis* strains the trend observed was similar; when searched against the subassemblage A1 genome, an average of 928 more peptides per replicate and 32 more reproducibly identified proteins in total were observed than when the same data was searched against subassemblage A2 genome. The variable genome was particularly sensitive regarding which subassemblage gene sequence was used for searching.

For BRIS/89/HEPU/1003, searching against the congruent subassemblage genome sequence caused an increase from 15 to 38 reproducibly identified VSPs, a decrease from 9 to 0 HCMPs, a decrease from 27 to 5 NEK kinases, and an increase from 40 to 55 protein 21.1/ankyrin repeat proteins. Most of these changes are reflective of the differences in variable gene family compositions between the A1 and A2 genome sequences, including a decrease of annotated NEK kinases from 179 in the A1 genome to 32 in the A2 genome, an increase from 243 annotated protein 21.1/ankyrin repeat genes to 340 in the A2 genome, and a loss of the majority of the HCMP family from 59 to 2 in the A1 to A2 genomes, respectively. Interestingly, although there was a decrease from 186 VSP genes in the A1 genome to 121 in the A2 genome, searching BRIS/89/HEPU/1003 against the A2 genome saw reproducibly identified VSPs more than double. In summary, our results clearly support the reclassification of genome sequences into the A1 and A2 subassemblages, since peptide-to-spectrum matching is significantly improved in important gene families (full information in Supporting Information Table 3).

In *Giardia*, VSPs are a family of cysteine-rich proteins responsible for surface antigen variation and are also putative virulence factors [17]. All VSP genes are transcribed simultaneously from the nucleus, and all but one transcript is silenced through RNA interference, and a single transcript goes on to be expressed on the entirety of the trophozoite surface [17]. However, previous genetic and antibody studies have

struggled to detect the full extent of VSP population diversity, because detection of multiple VSPs in the trophozoite population at these levels is confounded by the most abundant VSP transcript or protein in the population [17]. Our analytical approach has overcome this limitation and was able to identify multiple VSPs within the trophozoite population, and to simultaneously measure their relative abundance.

Analysis of the VSP composition indicated two population profiles in the eight strains studied, which we have designated as "dominant distribution" and "variable distribution." Three strains (Bac2, BRIS/89/HEPU/1065, and BRIS/95/HEPU/2041) had a VSP distribution characterized by multiple smaller subpopulations, while five strains (WB, OAS1, BRIS/89/HEPU/1003, BRIS87/HEPU/713, BRIS/83/HEPU/106) were dominated by only one or two VSP populations (Fig. 1). As axenic culture is independent of host immune selection, this population heterogeneity is a potential indication of variation in rates of antigenic switching. Indeed, BRIS/95/HEPU/2041, a virulent strain in several animal models [3], contained the third highest number of VSPs identified, and significantly higher abundance of arginine deiminase, a protein previously implicated in increasing the rate of surface antigen switching [18]. It has previously been shown that strains from different assemblages vary in rates of antigenic switching [19], but our study indicates that this may also occur between strains within assemblages.

The trends in total number of VSPs as well as VSP subpopulation distribution did not appear to coincide with host, subassemblage, geographic origin, or time since introduction to axenic culture. However, chromosomal rearrangements and duplications [9,20] have been previously demonstrated across *G. duodenalis* strains including several strains used in this study. Many VSPs in the *Giardia* genome are organized into inverted or tandem pairs and linear arrays, implicating gene duplication and transposition events in the evolution and expansion of the VSP repertoire [21]. The distribution of VSPs by chromosome can be seen in Table 2 for the seven A1 strains (no physical map is available for the A2 genome). Strains displaying a variable distribution expressed at least twice as many VSPs on chromosomes 3 and 4 as those strains displaying a dominant distribution. Previously, partial chromosome 3/4 duplication had been demonstrated in BRIS/83/HEPU/106 and BRIS/95/HEPU/2041, and chromosome 3/4 rearrangements were seen in Bac2 and BRIS/89/HEPU/1065, while no rearrangements or duplications were observed in OAS1 or WB [3, 20]. All strains expressing VSPs from all five chromosomes, including higher numbers from chromosome 3/4, possessed these rearrangements or duplications (also shown in Table 2). Increasing evidence suggests that the VSP repertoire is under positive evolutionary selection. As coverage of *G. duodenalis* genomes expands to multiple assemblages and subassemblages [4,16,22], genome comparisons indicate gene duplication and recombination are responsible for expansion of the VSP repertoire. Analyses show that some VSP genes exhibit significant recombination between each other,

Table 2. Chromosome distribution of reproducibly identified VSPs across the seven A1 *G. duodenalis* strains

Strain	C1	C2	C3	C4	C5	N/A	Total
BRIS/83/HEPU 106	1	1	3	4	3	1	13
BRIS87/HEPU/713	0	0	0	1	0	0	1
OAS1	1	0	2	3	1	0	7
Bac2	3	4	4	6	5	1	23
BRIS/95/HEPU/2041	4	4	6	11	6	2	33
BRIS/89/HEPU/1065	5	4	9	10	6	3	37
WB	2	1	0	0	5	1	9
No. of VSP on chromosome	12	20	18	81	65		

"N/A" stands for chromosomes that have not been able to be localized for the *G. duodenalis* Assemblage A physical map. The bottom table shows the total number of VSP genes that have been localized to each of the 5 *G. duodenalis* chromosomes. Rows that have been shaded are indicative of strains with previously reported chromosome 3/4 duplications/rearrangements.

as well as evidence of high mutation frequency regions separate from low mutation frequency regions [21]. Additionally, many VSPs are organized into tandem or inverted pairs and linear arrays, further indicating the importance of past gene duplication and translocation in emergence of an expanded VSP repertoire [21]. As the assemblages and subassemblages sequenced so far display unique repertoires, it is possible that pathogenesis and ecological niches may have been the driver for diversification of this genetic reservoir for surface variation. Considering that VSP genes occur in nonsyntenic regions of the genome prone to duplication events, and variation is observed between VSP gene numbers in currently sequenced genomes [4,21], we hypothesize that genome plasticity, including karyotype, directly influences VSP expression. This appears to be an assemblage-independent mechanism of isolate variation in *G. duodenalis*.

P.A.H. acknowledges support from the ARC Training Centre for Molecular Technologies in the Food Industry, and wishes to thank Justin Lane for continued support and encouragement. S.J.E. acknowledges funding from the Australian Government in the form of an APA scholarship, as well as financial support from Macquarie University. S.J.E. wishes to thank Dr. Jacqui Upcroft for supplying the Giardia samples used and for the ongoing support received from colleagues at Microbial Screening Technologies. P.A.H. wishes to thank Justin Lane for continued support and encouragement. The MS proteomics data have been deposited to the ProteomeXchange Consortium [15] via the PRIDE partner repository with the dataset identifier PXD001272.

The authors have declared no conflict of interest.

References

[1] Feng, Y., Xiao, L., Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. *Clin. Microbiol. Rev.* 2011, *24*, 110–140.

[2] Ankarklev, J., Jerlstrom-Hultqvist, J., Ringqvist, E., Troell, K., Svard, S. G., Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nat. Rev. Microbiol.* 2010, *8*, 413–422.

[3] Williamson, A. L., O'Donoghue, P. J., Upcroft, J. A., Upcroft, P., Immune and pathophysiological responses to different strains of *Giardia duodenalis* in neonatal mice. *Int. J. Parasitol.* 2000, *30*, 129–136.

[4] Adam, R. D., Dahlstrom, E. W., Martens, C. A., Bruno, D. P. et al., Genome sequencing of *Giardia lamblia* genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). *Genome Biol. Evol.* 2013, *5*, 2498–2511.

[5] Upcroft, J. A., Chen, N., Upcroft, P., Mapping variation in chromosome homologues of different *Giardia* strains. *Mol. Biochem. Parasitol.* 1996, *76*, 135–143.

[6] Nolan, M. J., Jex, A. R., Upcroft, J. A., Upcroft, P., Gasser, R. B., Barcoding of *Giardia duodenalis* isolates and derived lines from an established cryobank by a mutation scanning-based approach. *Electrophoresis* 2011, *32*, 2075–2090.

[7] Upcroft, J. A., Boreham, P. F., Campbell, R. W., Shepherd, R. W., Upcroft, P., Biological and genetic analysis of a longitudinal collection of *Giardia* samples derived from humans. *Acta Trop.* 1995, *60*, 35–46.

[8] Upcroft, J. A., Boreham, P. F., Upcroft, P., Geographic variation in *Giardia* karyotypes. *Int. J. Parasitol.* 1989, *19*, 519–527.

[9] Upcroft, J. A., Healey, A., Murray, D. G., Boreham, P. F., Upcroft, P., A gene associated with cell division and drug resistance in *Giardia duodenalis*. *Parasitology* 1992, *104*, 397–405.

[10] Aurrecoechea, C., Brestelli, J., Brunk, B. P., Carlton, J. M. et al., GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* 2009, *37*, D526–D530.

[11] Fenyo, D., Beavis, R. C., A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal. Chem.* 2003, *75*, 768–774.

[12] Craig, R., Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, *20*, 1466–1467.

[13] Neilson, K. A., George, I. S., Emery, S. J., Muralidharan, S. et al., Analysis of rice proteins using SDS-PAGE shotgun proteomics. *Methods Mol. Biol.* 2014, *1072*, 289–302.

[14] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K. et al., Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* 2006, *5*, 2339–2347.

[15] Vizcaino, J. A., Cote, R. G., Csordas, A., Dianes, J. A. et al., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, *41*, D1063–D1069.

[16] Jerlstrom-Hultqvist, J., Franzen, O., Ankarklev, J., Xu, F. et al., Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. *BMC Genomics* 2010, *11*, 543.

[17] Prucca, C. G., Lujan, H. D., Antigenic variation in *Giardia lamblia*. *Cell. Microbiol.* 2009, *11*, 1706–1715.

[18] Touz, M. C., Ropolo, A. S., Rivero, M. R., Vranych, C. V. et al., Arginine deiminase has multiple regulatory roles in the biology of *Giardia lamblia*. *J. Cell. Sci* 2008, *121*, 2930–2938.

[19] Nash, T. E., Conrad, J. T., Merritt, J. W., Jr., Variant specific epitopes of *Giardia lamblia*. *Mol. Biochem. Parasitol.* 1990, *42*, 125–132.

[20] Upcroft, J. A., Healey, A., Upcroft, P., Chromosomal duplication in *Giardia duodenalis*. *Int. J. Parasitol.* 1993, *23*, 609–616.

[21] Adam, R. D., Nigam, A., Seshadri, V., Martens, C. A. et al., The *Giardia lamblia* vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC Genomics* 2010, *11*, 424.

[22] Franzen, O., Jerlstrom-Hultqvist, J., Castro, E., Sherwood, E. et al., Draft genome sequencing of *Giardia intestinalis* assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog.* 2009, *5*, e1000560.

3.4 Publication III Supplementary Data

The following supplementary information is available for this manuscript.

Giardia Axenic Culture, Sample Preparation and Mass Spectrometry Analysis

G. duodenalis strains were cultured in triplicate axenically in TYI-S33 media supplemented with 10% newborn calf serum and 1% bile as previously described [1] and harvested from confluent cultures in late log-phase. Trophozoites were harvested by centrifugation, washed twice in ice-cold PBS to remove media traces [2] and pellets of 10^8 trophozoites were extracted into 1mL ice-cold SDS sample buffer containing 1mM EDTA and 5% beta-mercaptoethanol, then reduced at 75°C for 10 min. Trophozoite protein extracts were centrifuged at 0°C at 13 000 x *g* for 10 min to remove debris, and protein concentration was measured by BCA assay (Pierce). A 500 µg protein pellet was extracted using methanol-chloroform precipitation [3] and in-solution digestion was performed using a modified filter aided sample preparation (FASP) [4]. After peptide extraction all samples were dried using a vacuum centrifuge and reconstituted to 60 µL with 2% formic acid, 2% 2,2,2-trifluorethanol (TFE) .

Optimised GPF mass ranges were calculated using the 2.5 release of the *G. duodenalis* WB genome for Assemblage A from giardiaDB.org [5]. Charge states +2 and +3 were considered as well as carbamidomethyl as a cysteine modification, and 4 mass ranges were calculated over 400-2000amu. The mass ranges were as following: the low mass range was 400-518amu, the low-medium mass range was 518-691amu, the medium-high mass range was 691-988amu and the high mass range was 988-2000amu. Each FASP protein digest for the triplicates of each strain were analysed by nanoLC-MS/MS on an LTQ-XL linear ion trap mass spectrometer (Thermo, San Jose CA). Peptides were separated on a 150 x 0.2mm I.D fused-silica column packed with Magic C18AQ (200Å, 5µm diameter,

Michrom Bioresources, California) connected to an Advance CaptiveSpray Source (Michrom Bioresources, California). Each FASP protein digest was analysed as 4 repeat injections, with the mass spectrometer scanning for 180 minute runs for each of the four calculated mass ranges. Samples were injected onto the column using a Surveyor autosampler, followed by an initial wash step with buffer A (0.1% v/v formic acid, 1mM ammonium formate, 0.2% v/v methanol) for 4 minutes followed by 150µL/min for 2 minutes. Peptides were eluted from the column with 0-80% buffer B (100% v/v ACN, 0.1% v/v formic acid) at 150µL/min for 167 minutes finished by a wash step with buffer A for 6 minutes at 150µL/min. Spectra in the positive ion mode were scanned over the respective GPF ranges and, using Xcalibur software (Version 2.06, Thermo), automated peak recognition, dynamic exclusion and MS/MS of the top six most-intense ions at 35% normalisation collision energy were performed.

The LTQ-XL raw output files were converted into MzMXL files and searched against the Giardiadb.org 4.0 release of *G. duodenalis* strain Assemblage A1 and A2 genome using the global proteome machine (GPM) software (version 2.1.1) and the X!Tandem algorithm. The 4 fractions for the GPF of each replicate were processed sequentially with output files generated for each individual fraction, and a merged, non-redundant output file for protein identifications with log(*e*) values < -1. Peptide identification was determined using a MS and MS/MS tolerances of $\pm$4 Da and $\pm$0.4 Da. Carbamidomethyl was considered a complete modification, and partial modifications considered included oxidation of methionine and tryptophan.

Supplementary Data References

[1] Keister, D. B., Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans R Soc Trop Med Hyg* 1983, *77*, 487-488.

[2] Dunn, L. A., Upcroft, J. A., Fowler, E. V., Matthews, B. S., Upcroft, P., Orally administered *Giardia duodenalis* extracts enhance an antigen-specific antibody response. *Infect Immun* 2001, *69*, 6503-6510.

[3] Wessel, D., Flugge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* 1984, *138*, 141-143.

[4] Chapman, B., Castellana, N., Apffel, A., Ghan, R.*, et al.*, Plant proteogenomics: from protein extraction to improved gene predictions. *Methods in molecular biology* 2013, *1002*, 267-294.

[5] Scherl, A., Shaffer, S. A., Taylor, G. K., Kulasekara, H. D.*, et al.*, Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. *Anal Chem* 2008, *80*, 1182-1191.

Supplementary Tables and Figures:

Supplementary Table S1. Classification information for the eight *G. duodenalis* strains used in this study including subassemblage, geographic origin, and the host species the strain was isolated from. Names of strains coincide with those previously published in the literature.

Supplementary Table S2. Complete summary of peptide and protein identification data of *G. duodenalis* proteins across the eight strains analysed for both the subassemblage A1 and A2 reference genome. For the two tables A) shows the protein and peptide summary when searched against the subassemblage A1 genome sequence while B) shows the protein and peptide summary against the subassemblage A2 genome sequence.

Supplementary Table S3. Numbers of reproducibly identified proteins by family in each strain from the *G. duodenalis* variable genome. Final column shows the total, both by number and as a percentage of all reproducibly identified proteins for the strain. For the two tables A) shows the numbers of reproducibly identified proteins by family when searched against the subassemblage A1 genome sequence while B) shows the numbers of reproducibly identified proteins by family when searched against the subassemblage A2 genome sequence. The A2 strain BRIS/89/HEPU/1003 is distinguished with a '*' next to its identifier.

Supplementary Table S4. Excel spreadsheet showing the biological classification of proteins reproducibly identified (found in all three replicates with SpC ≥ 5) for the *Giardia* strains in this study when searched against the A1 subassemblage reference genome. The first sheet shows the 1197 non-redundant proteins identified across the seven A1 *G. duodenalis* strains in this study, including the spectral counts in each triplicate in each strain. For a protein to be included in the non-redundant total it had to meet the criteria above for reproducibly identified proteins in at least one strain. The second tab shows the 895 reproducibly identified proteins and their replicate SpC & NSAF in BRIS/82/HEPU/106. The third tab shows the 836 reproducibly identified proteins and their

replicate SpC & NSAF in BRIS/95/HEPU/2041. The fourth tab shows the 798 reproducibly identified and their replicate SpC & NSAF proteins in BRIS/87/HEPU/713. The fifth tab shows the 701 reproducibly identified proteins and their replicate SpC & NSAF proteins in Bac2. The sixth tab shows the 716 reproducibly identified proteins and their replicate SpC & NSAF in OAS1. The seventh tab shows the 728 reproducibly identified proteins and their replicate SpC & NSAF proteins in BRIS/89/HEPU/1065. The eighth tab shows the 769 reproducibly identified and their replicate SpC & NSAF proteins proteins in WB. The ninth tab shows the 668 reproducibly identified proteins and their replicate SpC & NSAF in BRIS/89/HEPU/1003, which is the single A2 subassemblage strain. **(Supplementary DVD)**

Supplementary Table S5. Excel spreadsheet showing the biological classification of proteins reproducibly identified (found in all three replicates with SpC ≥ 5) for the *Giardia* strains in this study when searched against the A2 subassemblage reference genome. For a protein to be included in the non-redundant total it had to meet the criteria above for reproducibly identified proteins in at least one strain. The first tab shows the 719 reproducibly identified proteins and their replicate SpC & NSAF in BRIS/89/HEPU/1003 which is the single A2 subassemblage strain. The second tab shows the 864 reproducibly identified proteins and their replicate SpC & NSAF in BRIS/82/HEPU/106. The third tab shows the 795 reproducibly identified proteins and their replicate SpC & NSAF in BRIS/95/HEPU/2041. The fourth tab shows the 775 reproducibly identified and their replicate SpC & NSAF proteins in BRIS/87/HEPU/713. The fifth tab shows the 658 reproducibly identified proteins and their replicate SpC & NSAF proteins in Bac2. The sixth tab shows the 681 reproducibly identified proteins and their replicate SpC & NSAF in OAS1. The seventh tab shows the 677 reproducibly identified proteins and their replicate SpC & NSAF proteins in BRIS/89/HEPU/1065. The eighth tab shows the 769 reproducibly identified and their replicate SpC & NSAF proteins in WB. **(Supplementary DVD)**

Supplementary Figure S1. Pie chart showing the distribution of shared and unique proteins of the 1197 non-redundant proteins identified across the seven subassemblage A1 *G. duodenalis* strains. The elevated segment representing the 149 proteins identified in 1 strain is further broken down in the upper box which shows the distribution of proteins identified uniquely to each strain.

Supplementary Table S1

Strain	Assemblage	Origin	Host Source
BRIS/83/HEPU 106	A1	Brisbane, Australia	Human
BRIS87/HEPU/713	A1	Brisbane, Australia	Human
OAS1	A1	Canada	Sheep (*Ovis aries*)
Bac2	A1	Australia	Cat (*Felis catus*)
BRIS/95/HEPU/2041	A1	Victoria, Australia	Cockatoo (*Cacatua galerita*)
BRIS/89/HEPU/1065	A1	Brisbane, Australia	Human
WB*	A1	Afghanistan	Human
BRIS/89/HEPU/1003	A2	Brisbane, Australia	Human

*WB is the Assemblage A1 genome strain

Supplementary Table S2

A) A1 Genome

Sample & Subassemblage		Low Stringency Peptide Count			Average No. Peptide (±%RSD)	No. R.I. Proteins Common to 3 Replicates	R.I. Protein FDR (%)	R.I. Peptide FDR(%)
		Replicate 1	Replicate 2	Replicate 3				
BRIS/83/HEPU 106	A1	23016	21198	22837	22350 ± 4.48%	895	0.45	0.08
BRIS87/HEPU/713	A1	20558	19945	18893	19799 ± 4.25%	798	0.50	0.12
OAS1	A1	21111	21209	18288	20203 ± 8.21%	716	0.42	0.06
Bac2	A1	20097	19285	19629	19670 ± 1.99%	701	0.29	0.06
BRIS/95/HEPU/2041	A1	21724	20807	21635	21389 ± 2.36%	836	0.72	0.15
BRIS/89/HEPU/1065	A1	18635	20740	20547	19974 ± 5.83%	728	0.27	0.04
WB	A1	21227	20699	21003	20976 ± 1.26%	769	0.39	0.90
BRIS/89/HEPU/1003	A2	18296	20103	20179	19526 ± 5.46%	668	0.30	0.07

B) A2 Genome

Sample & Subassemblage		Low Stringency Peptide Count			Average No. Peptide (±%RSD)	No. R.I. Proteins Common to 3 Replicates	R.I. Protein FDR (%)	R.I. Peptide FDR(%)
		Replicate 1	Replicate 2	Replicate 3				
BRIS/83/HEPU 106	A1	21772	21111	23043	21975 ± 4.47%	864	0.81	0.17
BRIS87/HEPU/713	A1	19487	18916	18168	18857 ± 3.51%	775	0.90	0.19
OAS1	A1	19135	19476	17052	18554 ± 7.07%	681	0.29	0.05
Bac2	A1	18906	18127	18278	18437 ± 2.24%	658	0.61	0.14
BRIS/95/HEPU/2041	A1	20364	19511	20208	20028 ± 2.27%	795	0.50	0.12
BRIS/89/HEPU/1065	A1	18708	20812	20647	20055 ± 5.83%	677	0.74	0.13
WB	A1	20227	19844	19804	19958 ± 1.17%	724	0.39	0.09
BRIS/89/HEPU/1003	A2	19126	21036	21189	20450± 5.62%	713	0.42	0.08

***R.I. stands for 'Reproducibly Identified'.**

Supplementary Table S3

A) A1 Genome

Strain	VSP	HCMP	Nek Kinase	Protein 21.1	TOTAL:
BRIS/83/HEPU 106	13	4	39	53	109 (12.2%)
BRIS87/HEPU/713	1	2	35	44	82 (10.3%)
OAS1	7	5	29	39	80 (11.2%)
Bac2	23	4	28	36	91 (13.0%)
BRIS/95/HEPU/2041	33	2	38	52	125 (15.0%)
BRIS/89/HEPU/1065	37	7	33	40	117 (16.1%)
WB	9	7	31	47	94 (12.2%)
BRIS/89/HEPU/1003*	15	7	27	40	89 (13.3%)

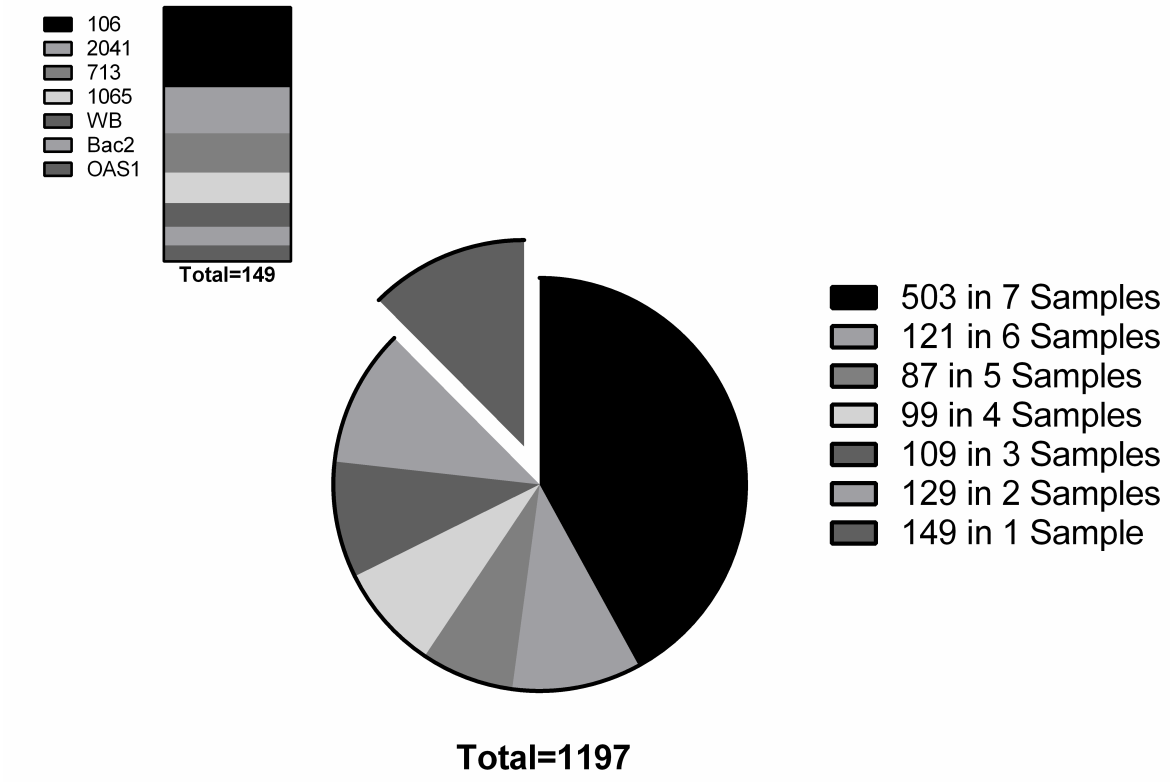
Genome Total:	186	59	179	243

B) A2 Genome

Strain	VSP	HCMP	Nek Kinase	Protein 21.1	TOTAL:
BRIS/83/HEPU 106	15	0	11	65	91 (10.5%)
BRIS87/HEPU/713	2	0	8	53	63 (8.1%)
OAS1	7	1	7	47	62 (9.1%)
Bac2	13	0	7	38	58 (8.8%)
BRIS/95/HEPU/2041	25	0	2	63	90 (11.3%)
BRIS/89/HEPU/1065	29	1	7	40	77 (11.4%)
WB	11	0	8	48	67 (9.3%)
BRIS/89/HEPU/1003*	38	0	5	52	95 (13.3%)

Genome Total:	121	2	32	340

Supplementary Figure S1



CHAPTER 4

Quantitative and comparative proteomics of G. duodenalis *trophozoites and cysts from two genome-alternate A1 isolates.*

4. The generation gap: proteome changes and strain variation during encystation in *Giardia duodenalis*

4.1 Context

The prevalence of *Giardia* is partially attributed to its direct lifecycle, where cysts accumulate and remain infective in the environment. The formation of this environmentally resistant infective form (ERIF) is a complex process, which has been comprehensively studied using transcriptomics but requires further data for understanding the proteomic complement. In addition, the majority of current information is based around isolate WB C6 (ATCC 50803), and does not account for isolate differences in the encystation process that may affect infection outcome in the next generation. In this study we compared the cyst and trophozoite stages of two, genome-alternate isolates, BRIS/95/HEPU/2041 and BRIS/83/HEPU/106, to verify the presence of encystation protein markers that are isolate- and method- independent. In addition, we also assessed differences between isolates that might influence completion of the life cycle in a new host. This study provided further insight into isolate variation established in the previous chapters by extending the investigations into biological processes central to the reproductive and infective success of *Giardia*.

4.2 Contributions and Permissions

I performed 100% of all work pertaining to growing and collecting *Giardia* samples for proteomic analysis, and 100% of all lab work. I wrote 80% of the manuscript with data analysis assisted by Dana Pascovici and editing provided by Professor Paul Haynes and Dr Ernest Lacey. The paper (Publication V) is reprinted with the permission of Elsevier.

4.3 Publication V

Molecular & Biochemical Parasitology 201 (2015) 47–56

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

The generation gap: Proteome changes and strain variation during encystation in *Giardia duodenalis*

Samantha J. Emery[a], Dana Pascovi[b], Ernest Lacey[c], Paul A. Haynes[a,*]

[a] Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW 2109, Australia
[b] Australian Proteome Analysis Facility (APAF), Macquarie University, North Ryde, NSW, 2109, Australia
[c] Microbial Screening Technologies Pty Ltd, Smithfield, NSW 2165, Australia

ARTICLE INFO

Article history:
Received 22 April 2015
Received in revised form 26 May 2015
Accepted 28 May 2015
Available online 1 June 2015

Keywords:
Encystation
VSP
Protein 21.1
Assemblage A
Proteomics

ABSTRACT

The prevalence of *Giardia duodenalis* in humans is partly owed to its direct and simple life cycle, as well as the formation of the environmentally resistant and infective cysts. Proteomic and transcriptomic studies have previously analysed the encystation process using the well-characterised laboratory genomic strain, WB C6. This study presents the first quantitative study of encystation using pathogenically relevant and alternative assemblage A strains: the human-derived BRIS/82/HEPU/106 (H-106)and avian-derived BRIS/95/HEPU/2041 (B-2041). We utilised tandem MS/MS with a label-free quantitative approach to compare cysts and trophozoite life stages for strain variation, as well as confirm universal encystation markers of assemblage A. A total of 1061 non-redundant proteins were identified from both strains, including trophozoite- and cyst-specific proteomes and life-stage differentially expressed proteins. Additionally, 24 proteins previously classified in the literature as encystation-specific were confirmed as strain-independent markers of encystation. Functional cluster analysis of differentially expressed proteins saw significant overlap between strains, including protein trafficking and localisation in cysts, NEK kinase function, and carbohydrate metabolism in trophozoites. Two significant points of strain specific adaptations in cysts were also identified. B-2041 possessed major up-regulation of the ankyrin repeat protein 21.1 family compared to H-106. Furthermore, cysts of B-2041 retained near-complete VSP variant diversity between cysts and trophozoites, while H-106 lost 45% of its VSP variant diversity between life cycle stages, a constriction previously observed in studies of WB C6. This is the first report of strain variation in the cyst stage in *G. duodenalis*, and highlights cyst variation and its impacts on reinfection and life cycle success.

1. Introduction

Giardia duodenalis is a flagellated protozoan responsible for a global pandemic afflicting 250 million of the world's population at any one time, making it the most common diarrheal parasite. *G. duodenalis* is the perfectly adapted parasite associated with short-term morbidity rather than mortality, as it is adapted to a wide range of host species and displays zoonoses overlapping those of other closely related *Giardia* species [1]. At the heart of the evolutionary success of *G. duodenalis* is the process of cyst formation. *G. duodenalis* has a simple, direct life cycle comprising discrete parasitic trophozoites and cyst stages with transmission predominantly mediated by contaminated waterborne routes. The cyst is the environmentally resistant infective form (ERIF), formed in the jejunum and shed in the faeces, making it essential for life cycle completion. Cysts are able to accumulate and persist in the environment for months at a time awaiting ingestion, even resisting some disinfectants, until primed by host gastric signals which prompt trophozoite emergence [2].

Encystation is a complex process involving significant amounts of cell structure remodelling and metabolic adjustment. Tear-dropped trophozoites, which are flagellated and binucleate (4N), become increasingly rounded, lose the ability to attach to host cells,

Abbreviations: ANK, ankyrin; CW, cyst wall; CWP, cyst wall protein; ES, enrichment score; ESV, encystation-specific vesicle; FDR, false discovery rate; GalNAc, β(1-3)-*N*-acetyl-D-galactosamine; GO, gene ontology; GPF, gas phase fractionation; HCMP, high cysteine membrane protein; Nano LC–MS/MS, nanoflow liquid chromatography tandem mass spectrometry; NSAF, normalised spectral abundance factor; ORF, open reading frame; PDI, protein disulfide isomerase; PDX, peroxiredoxin; P.I, post induction; PV, peripheral vesicle; SAGE, serial analysis of gene expression; SpC, spectral count; TCP-1, chaperonin T complex 1; TMT, tandem mass tag; VSP, variable surface protein.

* Corresponding author. Tel.: +61 2 9850 6258; fax: +61 2 9850 6200.
E-mail address: paul.haynes@mq.edu.au (P.A. Haynes).

http://dx.doi.org/10.1016/j.molbiopara.2015.05.007
0166-6851/

internalise their flagella, disassemble median bodies and become immotile cysts with 4 nuclei and 16N ploidy [3,4]. Trophozoites bound for encystation are shunted out of the cell-cycle at the G_2-M transmission to begin the process, which takes approximately 20–24 h [5]. Early encystation is characterised by cyst wall (CW) formation and the presence of encystation-specific vesicles (ESV) within the first 8–10 h post induction (p.i). The CW is comprised of three cyst wall proteins (CWP), CWP1-3, which are synthesised, concentrated and transported in the ESV to be complexed with β(1-3)-*N*-acetyl-D-galactosamine (GalNAc) polymer (~60% of the CW) to produce the 300 nm thick CW [4,6]. Encystation is triggered *in vitro* by simulating conditions of the mid-lower portion of the jejunum. *In vitro* encystations predominately utilise high bile concentrations and high pH, or cholesterol starvation, but may also involve addition of lactic acid, primary bile salts and deprivation of bile prior to induction [5,8,9]. Currently, the specific factors that induce encystation in bile, an unfractionated non-standardised media addition, remain unknown. Additionally, *in vitro* encystation is not universally induced under such conditions, as some laboratory passaged strains and clones are non-responsive to *in vitro* signals and fail to encyst [7]. The induction of encystation in the host, where host immune pressures may contribute to, or influence, the cyst formation process, is yet uncharacterised.

To date, our understanding of regulation of *Giardia* encystation through proteomic and transcriptomic studies is limited to the *G. duodenalis* laboratory strain WB strain first isolated in 1979, or the laboratory optimal and clonally derived WB C6 (ATCC 50803) generated in 1983. So far, microarray analyses of cysts and trophozoites have revealed a small set of 18 encystation genes up-regulated independent of encystation method used [5], while further microarray analysis revealed a limited transcriptome within cysts, implying metabolic dormancy [8]. Recently, Serial Analysis of Gene Expression (SAGE) was used to monitor changes in mRNA abundance across multiple time-points during both encystation and excystation, allowing a global snapshot of the changes throughout key life cycle processes [9]. Using mass-spectrometry based quantitative proteomics, cysts and trophozoites have been previously compared using 2DE gel electrophoresis [10], as well LC–MS/MS proteomic analysis of encystations across 14 h using label-free quantitation [6]. The increasing number of post-genomic technologies quantitating gene and protein expression are amalgamating a reproducible core set of stage-specific markers throughout the multiple stages of encystation, but only using WB C6 as the best characterised laboratory strain [6]. While this provides pivotal information it fails to provide an adequate understanding of genomic, transcriptomic and proteomic transition in alternative and pathogenically relevant strains. Furthermore, with wide variations in host range and zoonoses, as well as experimental evidence of varying cyst infectivity [11], it is highly likely that strain variations during encystation occur to impact life cycle completion and zoonotic potential in subsequent generations.

The present study reports the first comparative and quantitative proteomic analysis of cysts and trophozoites from two alternate assemblage A strains, specifically derived from pathogenic strains of diverse host specificity. We have utilised two Australian *G. duodenalis* strains from the A1 sub-assemblage, BRIS/82/HEPU/106 (H-106), isolated from a diarrhoeic child, and BRIS/95/HEPU/2041 (B-2041), which was isolated from a wild-caught cockatoo (*Cacatua galerita*) who succumbed to terminal enteritis [12]. Both these strains have well characterised phenotypes and infection models in the literature, with B-2041 showing a more intense pathogenicity and pathophysiology compared to H-106 [13–15]. This makes this study the first to consider encystation processes outside the genomic strain WB C6, in the search for reproducible and pathogenically relevant markers of the *Giardia* life cycle.

2. Methods

2.1. Parasite culture and encystation

Axenic cultures of trophozoites of H-106 and B-2041 were grown in TYI-S-33 medium supplemented with 1% bile and newborn calf serum [16] as previously described [17]. Parasites were subcultured at end-log phase into fresh media, and grown and harvested for protein extraction within 5 passages of recovery from cryopreservation. Three biological replicates generated from separate cultures were grown for each strain and encysted and harvested separately. Absence of bacterial and fungal contamination was verified using serial dilutions and streak plates to ensure no colony forming units were detected in cultures prior to encystation and extraction.

Half of each triplicate culture was harvested to extract the trophozoite proteins, and the remaining half progressed through to encystation. *Giardia* trophozoites were harvested from pre-confluent cultures in by chilling on ice with vortexing for 15 min to detach parasites from the walls of the culture vessel. Trophozoites were harvested by centrifugation at 3000 × *g* for 10 min, then washed twice more with ice-cold PBS to remove media traces [14]. Encystation conditions were optimised to maximise cyst yield as previously described for each strain [7,18]. Maximum Type 1 cysts for H-106 were achieved using a 24 h/24 h encystation to growth media method [7], and 40 h in encystation media was used for the slower growing B-2041 [18]. Both strains were incubated in encystation media containing high bile, lactic acid and increased pH of 7.8 [18]. After encystation, cysts were enriched using hypotonic lysis, washed twice in dH_2O to remove cell debris and collected using sedimentation and centrifugation.

2.2. Protein extraction, digestion and peptide extraction

Trophozoites and cysts were extracted in ice-cold SDS sample buffer containing 1 mM EDTA and 5% beta-mercaptoethanol, and then reduced at 75 °C for 10 min. Protein extracts were centrifuged at 0 °C at 13,000 × *g* for 10 min to remove debris, and stored at −20 °C. The concentration of protein in each solution was measured by BCA assay (Pierce) before fractionation and digestion. For each sample 250 μg of protein was extracted from SDS sample buffer using chloroform-methanol [19], with the protein pellet washed 2–3 times with methanol. In-solution digestion was performed using the filter aided sample preparation (FASP) method [20,21], modified to solubilise protein in 2,2,2-trifluoroethanol (TFE) [22] as previously described [23]. Each extract was reconstituted to 60 μL with 2% formic acid, 2% TFE and stored at −20 °C until analysis by LC–MS/MS.

2.3. Theoretically derived GPF calculations using predictive software

Optimised GPF mass ranges were calculated using theoretical trypsin digest according to Scherl et al. [24], with the Giardiadb.org 2.5 release of the A1 genome (Strain WB) with charge states +2 and +3 considered, as well as carbamidomethyl modifications. The mass ranges were calculated as following: the low mass range was 400–520 amu, the low-medium mass range was 515–690 amu, the medium-high mass range was 685–990 amu and the high mass range was 985–2000 amu.

2.4. Nanoflow LC–MS/MS

Each FASP protein digest was analysed as 4 repeat injections across the four mass ranges as calculated in section 2.3, on a LTQ Velos Pro linear ion trap mass spectrometer (Thermo, San Jose CA).

Samples were analysed on a 150 × 0.1 mm I.D fused-silica column packed in-house with Magic C18AQ resin (200 Å, 5 μm diameter, Michrom Bioresources, California) in a fused silica capillary with an integrated electrospray tip, coupled to a 300 × 0.1 mm pre-column packed with PS-DVB resin (100 μM I.D., Agilent). A 1.8 kV electrospray voltage was applied *via* a liquid junction upstream of the C18 column and samples were injected using an Easy-nLC II (Thermo). Peptides were initially washed with Buffer A (2% v/v ACN, 0.1% v/v formic acid) at 500 nl/min for 7 min and eluted with 0-50% Buffer B (99.9% v/v ACN, 0.1% v/v formic acid) at 500 nl/min for 168 min, followed by awash step with 95% Buffer B at 500 nl/min for 5 min. Each of the FASP protein digests required 410 μL runs of 180 min, totalling 12 h per sample. Automated peak recognition, MS/MS of the top nine most intense precursor ions at 35% normalisation collision energy and dynamic exclusion duration of 90 s were performed using Xcalibur software (Version 2.06, Thermo).

2.5. Database search for protein/peptide identification

The Giardiadb.org 4.0 release of the *G. duodenalis* strain assemblage A1 (strain WB) genome was used as a database for searching with the global proteome machine (GPM) software (version 2.1.1) [25,26] and the X!Tandem algorithm. For each replicate in each sample the 4 fractions were searched sequentially and processed with output files for each individual fraction, and a merged, non-redundant output file for protein identifications with log(*e*) values <−1. Peptide identification was determined using MS and MS/MS tolerances of ±4 Da and ±0.4 Da. Carbamidomethyl was considered a complete modification, and partial modifications considered included oxidation of methionine and tryptophan. Point mutations were also considered. Complete protein and peptide data for triplicates of both samples and strains have been deposited to the ProteomeXchange Consortium [27] *via* the PRIDE partner repository and can viewed under the identifier PXD002002.

2.6. Data processing and quantitation

The 9 lists of identified proteins from the merged output of the GPM constituted low-stringency protein and peptide information, which was transformed into lists of high stringency reproducibly identified proteins using the Scrappy software package [28,29]. Proteins were defined as reproducibly identified only when present in all three replicates with a total spectral count (SpC) of ≥5 [29]. Reversed database searching was used for calculating peptide and protein false discovery rates (FDRs) as previously described [29] while protein abundance was calculated using normalised spectral abundance factors (NSAF) [30]. To calculate NSAF values the spectral count (the total number of MS/MS spectra) was divided by the protein length (L), and normalized to the sum of SpC/L for all proteins in the experiment. To compensate for null values and to allow log transformation prior to statistical analysis, a spectral fraction of 0.5 was added to all SpC values [28].

2.7. Statistical analysis of differentially expressed proteins

Statistical analysis was performed between cysts and trophozoites of each strain, as well as cysts from both strains using the Scrappy software package [28]. Student's *t* tests were performed on log-transformed NSAF data to determine statistical significance of differential abundance, and proteins with a *t* test *p*-value less than 0.05 were considered to be differentially expressed.

2.8. Gene Ontology information and functional annotation

Functional annotation analysis was performed using the DAVID bioinformatics resource [30] as previously described [6]. Proteins statistically significantly up- and down-regulated were converted from giardial ORF codes from the giardiadb.org genome release to gene identification numbers using the NCBI Batch Entrez tool (http://www.ncbi.nlm.nih.gov/sites/batchentrez). These gene identification codes were then submitted as gene lists to DAVID for functional annotation clustering. Clusters were analysed using high stringency, with both gene ontology (GO) annotations and interpro annotations considered, and only clusters with enrichment scores (ES) ≥1 were selected for further analysis.

3. Results and Discussion

3.1. Proteome analysis of reproducibly identified proteins

In this first comparative proteomic analysis of alternate strains from the *G. duodenalis* A assemblage, a total of 1061 non-redundant proteins were reproducibly identified from both strains, with a non-redundant total of 672 and 972 proteins reproducibly identified in H-106 and B-2041. A full summary of peptide counts and NSAF values for reproducibly identified proteins in both strains can be found in Supplementary Table 1. After combining triplicate data for H-106, there were a total of 586 reproducibly identified proteins in trophozoites and 488 reproducibly identified proteins in cysts. The protein and peptide FDR for both cysts and trophozoites was very low as no reverse hits were detected. For B-2041 there were 816 reproducibly identified protein in trophozoites, and a total of 688 reproducibly identified proteins in cysts. The protein and peptide FDR was low at 0.04% and 0.07%, respectively. A summary of protein and peptide data can be found in Table 1.

For both strains, analysis of cysts and trophozoites revealed a common core proteome between both life cycle stages. This was achieved by cyst enrichment using hypotonic lysis to remove non-viable cysts and unencysted trophozoites, and avoid sub-sampling a mixed population [7]. This allowed increased resolution of cyst- and trophozoite-specific proteomes. The degree of overlap between life stages in strains can be viewed in Fig. 1. For H-106, 402 (59.82%) of all reproducibly identified proteins were common, while 531 (54.57%) were common between cysts and trophozoites in B-2041. A similar proportion of trophozoite-unique and cyst-unique proteins were seen in the strains: 184 (27.38%) and 285 (29.29%) proteins identified uniquely in H-106 and B-2041 trophozoites, respectively. For both H-106 and B-2041, despite consistent protein loading, less reproducibly identified proteins were found in cysts, which suggest a simpler proteome composition for this life stage. More importantly, comparisons indicate the existence of a unique cyst proteome for both strains, with 86 (12.80%) reproducibly proteins uniquely identified in H-106 cysts and 157 (16.14%) reproducibly identified proteins found only in B-2041 cysts. This data strongly reinforces that a robust core proteome exists across strains of the *Giardia* life cycle, supplemented by a subset of stage-specific proteins.

3.2. Analysis of differentially expressed proteins between cysts and trophozoites

In H-106, a total of 155 proteins were up-regulated in cysts and 129 proteins down-regulated, while in B-2041 a total of 204 and 181 proteins were up-regulated and down-regulated in cysts, respectively. The full dataset for differentially expressed proteins in both strains can be found in Supplementary Table 2. Fig. 2 shows the distribution of up- and down- regulated proteins, and illustrates that both strains possess similar proportions of differentially expressed and unchanged proteins.

To validate the encystation process in our strains, we compared markers derived from previous proteomic and transcript

Table 1
Peptide/protein identification data of parasite proteins found in *G. duodenalis* strains (H-106 and B-2041) in both trophozoite and cyst life cycle stages.

Sample	Average Low Stringency Peptide Count (±%RSD)	No. Proteins Common to 3 Replicates	Protein FDR	Peptide FDR
H-106 Trophozoites	17224 ± 17.9%	586	NRH[a]	NRH[a]
H-106Cysts	10867 ± 7.59%	488	NRH[a]	NRH[a]
B-2041 Trophozoites	22601 ± 3.08%	816	0.49%	0.04%
B-2041Cysts	15625 ± 19.85%	688	0.58%	0.07%

[a] NRH = no reverse hits.

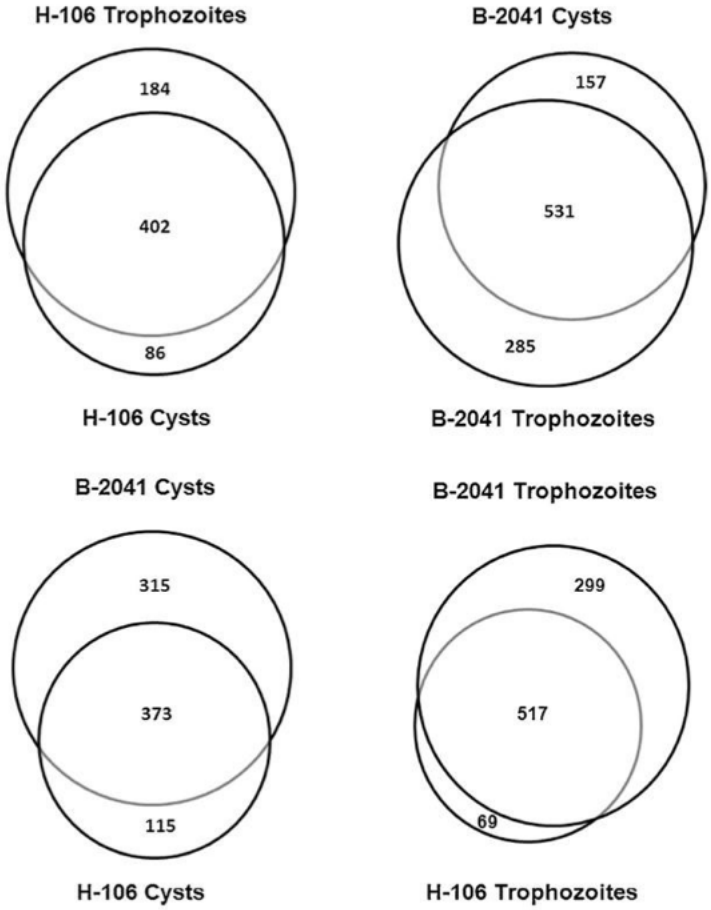


Fig. 1. Unique and Common Proteomes. Proportional Venn diagrams depicting the amount of overlap for proteins reproducible identified (found in all three replicates with SpC ≥ 5) in *G. duodenalis* life cycle stages cysts and trophozoites in respective strains H-106 and B-2041. Also shown is the degree of similarity for proteins identified in the cyst stage and the trophozoite stage for the two strains H-106 and B-2041.

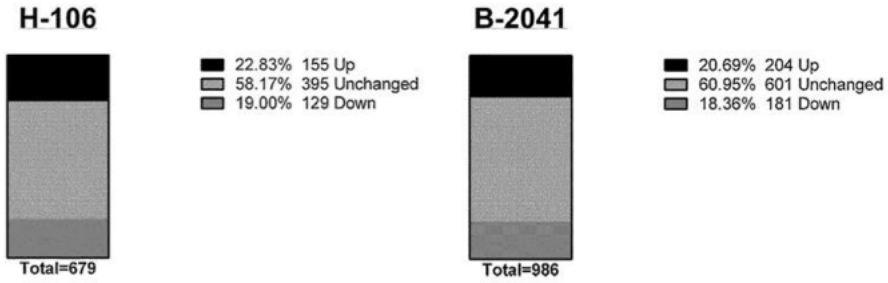


Fig. 2. Distribution of differentially expressed proteins in cysts. For both H-106 ($n = 679$) and B-2041 ($n = 986$) despite differences in total number of proteins, proportions of up- and down-regulated proteins in cysts remain similar between both strains.

Table 2
G. duodenalis proteins significantly upregulated in cysts of one or both H-106 and B-2041 and previously identified as an encystation marker [5,6,9,10]. An ORF designated an encystation marker in an earlier study is indicated by a 'Y'. The '↑' indicates the ORF was upregulated while '↓' indicates downregulation. Where there is a '–' the ORF was unchanged in abundance while an 'X' indicates it was not reproducibly identified. A shaded cell specifies the ORF was upregulated in both H-106 and B-2041.

ORF	Annotation	H-106	B-2041	Protein studies		Transcript studies	
				2D PAGE [10]	Tandem MS [6]	Microarray [5]	SAGE [9]
GL50803_103676	Alpha-tubulin	–	↑	Y	Y		
GL50803_88765	Cytosolic HSP70	↑	↓	Y	Y		
GL50803_98054	Heat shock protein HSP 90-alpha	↑	–	Y	Y		
GL50803_4812	Beta-giardin	–	↑	Y	Y		
GL50803_102813	Protein 21.1	↑	↑		Y	Y	
GL50803_5638	Cyst wall protein 1	↑	↑			Y	Y
GL50803_5435	Cyst wall protein 2	X	↑		Y	Y	Y
GL50803_8245	Glucosamine-6-phosphate deaminase	↑	↑		Y		Y
GL50803_7982	UDP-glucose 4-epimerase	↑	↑		Y		Y
GL50803_2897	Furin precursor putative serine protease	↑	↑		Y		Y
GL50803_16424	Hypothetical protein	↑	↑		Y		Y
GL50803_16217	UDP-*N*-acetylglucosaminepyrophosphorylase	↑	X		Y		
GL50803_14626	Oxidoreductase, short chain dehydrogenase	↑	X		Y	Y	
GL50803_8377	Hypothetical protein	–	↑		Y		y
GL50803_103676	Alpha-tubulin	–	↑	Y			
GL50803_136021	Beta-tubulin	–	↑	Y			
GL50803_14626	Oxioreductase	↑	X			Y	
GL50803_92729	Fatty acid elongase 1	↑	X			Y	
GL50803_88581	Synaptic glycoprotein SC2	↑	↑			Y	Y
GL50803_7260	Aldose reductase	↑	–		Y	Y	
GL50803_102813	Protein 21.1	–	↑			Y	
GL50803_8987	Hypothetical protein	X	↑			Y	Y
GL50803_7374	Hypothetical protein	↑	X				Y
GL50803_10885	4-Alpha-glucanotransferase, amylo-alpha-1,6-glucosidase	↑	↑		Y		

experiments to differentially expressed cyst proteins in our dataset. We compiled data from studies involving proteins up-regulated from tandem MS/MS analysis of encysting trophozoites across 14 h [6], 2DE-gel electrophoresis of trophozoites and cysts [10], SAGE analysis of encysting trophozoites and cysts [9] and microarray analysis of encystation [5]. Table 2 shows proteins significantly up-regulated in cysts of one or both strains in this study, which have also been previously identified as an encystation marker. There are a total of 24 such proteins identified in our cyst up-regulated dataset. Of these 24 proteins, 7 proteins were reproducibly identified and up-regulated in both B-2041 and H-106. Of these 7 proteins, 3 are involved in the GalNAc biosynthesis pathway: 4-alpha-glucanotransferase, amylo-alpha-1,6-glucosidase (GL50803_10885), and UDP-glucose 4-epimerase (GL50803_7982). Additionally, Glucosamine-6-phosphate deaminase (GL50803_8245) and UDP-*N*-acetylglucosamine pyrophosphorylase (GL50803_16217) were also identified as up-regulated only in H-106. These correlate with the GalNAc biosynthesis pathway as encystation-regulated [31], and various members of the pathway have already been identified as markers in earlier studies [5,6,9]. Also up-regulated in both H-106 and B-2041 cysts were CWP 1 (GL50803_5638), synaptic glycoprotein SC2 (GL50803_88581), a hypothetical protein (GL50803_8377) and the furin precursor putative serine protease (GL50803_2897), the last of which is a known late-encystation marker [6,9]. A further 17 previously identified markers were also up-regulated, 8 from H-106 and 9 from B-2041, indicating a consistency between our experiment and previous experiments [5,6,8–10]. It also further reinforces the occurrence of reproducible encystation markers and pathways as universal among isolates.

Analysis of up-regulated proteins in cysts also revealed a subset of proteins exclusively expressed in cysts with no spectra identified in any trophozoite biological replicates. In H-106, of the 155 proteins of higher abundance in cysts, 21 (13.55%) were unique (SpC = 0 in trophozoites) while of the 129 proteins of higher abundance in trophozoites 39 (30.23%) were unique (SpC = 0 in cysts). Of these 21 proteins uniquely expressed in H-106 cysts, 5 proteins are previously identified encystation markers and 7 (35.0%) were only annotated as 'hypothetical proteins' (GL50803_113303, GL50803_7374, GL50803_16794, GL50803_17262, GL50803_15594, GL50803_10808, GL50803_6982). Similarly in B-2041 there were 22 (10.78%) of the 204 and 61 (33.70%) of the 181 proteins considered unique in up-regulated and down-regulated proteins in cysts, respectively. Again, of these 22 unique proteins, a total of 6 consisted of previously identified encystation markers and 14 (63.66%) were hypothetical proteins (GL50803_10892, GL50803_11604, GL50803_17803, GL50803_2692, GL50803_8789, GL50803_7035, GL50803_17062, GL50803_29796, GL50803_4692, GL50803_8987, GL50803_10522, GL50803_9296, GL50803_17174, GL50803_12109). This firstly confirms that expression of several reproducible encystation markers is confined exclusively within the cyst life stage. It also reinforces the need for ongoing work to annotate the *Giardia* genome to increase our understanding of biological processes that may aid designation of new functional markers.

3.3. Functional analysis of differentially expressed proteins

To extrapolate biological and metabolic processes and relate these back to encystation, functional annotation and clustering of differentially expressed proteins was performed using DAVID [30]. For the analysis, *Giardia* ORF identifiers were converted into gene identification numbers through NCBI, which resulted in only minor losses. For B-2041 a total of 201/204 up-regulated and 177/181 down-regulated proteins in cysts were submitted, while a total of 148/155 up-regulated and 127/129 down-regulated proteins were analysed from H-106. Across both strains and life stages, a non-redundant total of 15 functional annotation clusters were detected with an enrichment score (ES) of ≥1 as seen in Fig. 3. The total output of the DAVID functional clustering analysis can be seen in Supplementary Table 3.

Of the proteins significantly down-regulated in cysts, there were two common functional clusters between H-106 and B-2041; 'carbohydrate metabolism' and 'thioredoxin/cell redox homeostasis'. The 'carbohydrate metabolism' cluster involved 10 proteins in each H-106 and B-2041, of which 7 were com-

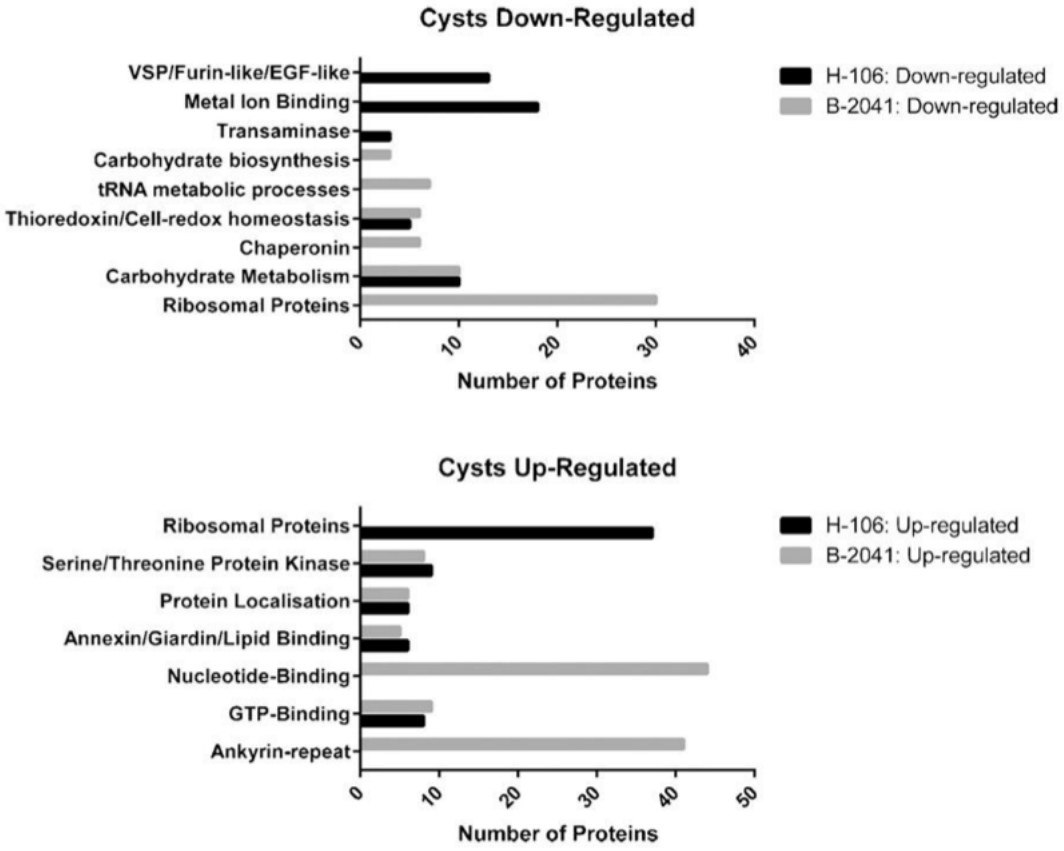


Fig. 3. Functional annotation clusters. Differentially expressed proteins for H-106 and B-2041 were analysed using DAVID. Proteins considered as up- and down-regulated in cysts of both strains assigned to a total of 15 non-redundant clusters which had enrichment scores of ≥1. The first graph (above) shows the functional annotation clusters considered down-regulated in cysts in H-106 and B-2041. The second graph (below) shows functional annotation clusters from proteins considered to be differentially expressed and of higher abundance in the cysts life stage in H-106 and B-2041.

mon. The majority of these proteins were from the glycolysis pathway including enolase (GL50803_11118), phosphoenolpyruvate carboxykinase (GL50803_10623), fructose-bisphosphate adolase (GL50803_11043), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (GL50803_8822), glucokinase (GL50803_8826), glucose-6-phosphate isomerise (GL50803_9115) and glyceraldehyde 3-phosphate dehydrogenase (GL50803_6687), with phosphoglycerate kinase (GL50803_3206) down-regulated in only B-2041 and pyrophosphate-fructose 6-phosphate 1-phosphotransferase alpha subunit (GL50803_14993) down-regulated only in H-106. This cumulative down-regulation of 9 members of the glycolytic pathway cluster is consistent with previous proteomic data during encystation [6] and the process of vegetative trophozoites becoming dormant cysts [8]. In the 'thioredoxin/cell redox homeostasis' cluster there were 6 and 5 proteins in B-2041 and H-106 respectively, of which 3 were common, and all feature the thioredoxin protein domain (GL50803_ 8394, GL50803_6289, GL50803_3910). Within the B-2041 cluster were another 2 proteins including a protein disulfide isomerise (PDI) (GL50803_14670), glutaredoxin-related protein (GL50803_2013) and another hypothetical protein with the thioredoxin domain (GL50803_9355), while two peroxiredoxins (PRX) (GL50803_14521, GL50803_15383) were clustered with H-106. Both PDIs and PRXs have been shown to be up-regulated in early encystation [6] and PRXs GL50803_14521 and GL50803_15383 were identified in SAGE of trophozoites and encyzoites [9]. It has been suggested that increased PRX expression is found in trophozoites *in vitro* anticipating *in vivo* changing oxygen concentrations through the large intestines during early encystation [6].

A further four functional clusters were found down-regulated in cysts in B-2041: 'ribosomal family proteins', 'chaperonin T-complex', 'aminoacyl tRNA synthetase' and 'carbohydrate biosynthesis'. In H-106 three functional clusters were also down-regulated in cysts including 'transaminase', 'ion binding/metal ion binding' and 'VSP/Furin-like/EGF-like', the last of which will be further discussed in context of strain variation in Section 3.4. A total of seven non-redundant functional clusters with an ES ≥1 were identified as up-regulated in cysts between both H-106 and B-2041. A total of 4 clusters were common between both strains: 'Giardins', 'protein localisation', 'serine/threonine protein kinase' and 'GTP-binding'. For B-2041 there were two additional functional clusters: 'nucleotide-binding' and 'ANK-repeat'. The 'ANK-repeat' cluster was comprised of mostly members of the multigene 21.1 protein family, and will be further discussed in Section 3.4 as a point of variation between H-106 and B-2041 cysts. Clusters which are common between cysts are further discussed as universal markers and strain-independent pathways in encystation, which offer insight into the reproducible aspects of the differentiation process as well as potential drug targets for blocking transmission.

The 'protein localisation' cluster featured proteins involved in the secretory pathway and ESV trafficking in both strains. Of these proteins, Rab2a (GL50803_15567) and Rab11 (GL50803_1695) were common as up-regulated in cysts of both strains. Rab11 has been previously shown to be up-regulated in early and late

encystation, responsible for ESV trafficking and specifically the transport of CWP-1 [32]. Clathrin heavy-chain (GL50803_102108) was also up-regulated in B-2041, which is implicated in protein trafficking and localises to peripheral vesicles (PV) [33,34], and previously observed as up-regulated during encystation [6]. Similarly, coatomer beta subunit (GL50803_88082) was up-regulated in H-106 cysts, which is associated with the Sec61 translocon and early secretory pathway during encystation [6]. Furthermore, the signal recognition particle receptor (GL50803_14856) was also up-regulated in B-2041, which has been localised to the wall of the ER and PV [35]. Sec23 (GL50803_9376) was also up-regulated in B-2041 cysts, and is implicated in ER protein trafficking and secretory pathways, in particular during differentiation due to its role in maintenance of ER exit sites [36]. Another associated protein, Sar1 (GL50803_7569), was also up-regulated in H-106 cysts and is a GTPase responsible for recruitment for COPII transport and closely associated with Sec23 rescruitment to ER exit sites [36]. The 'GTP-binding' cluster also featured proteins relating to protein trafficking during encystation, reinforcing the importance of secretory pathways and protein trafficking within differentiation in *Giardia* [37]. Both H-106 and B-2041 had up-regulated dynamin expression (GL50803_14373) in cysts, as dynamin-negative mutants possess impaired endocytic function and arrested encystation due to its role in CWP trafficking and ESV homeostasis [38]. Our results match those already quantitatively shown using tandem MS/MS proteomics [6], that reinforce protein trafficking and localisation are indispensible to complete encystation *Giardia* and as such are universal among strains.

Giardins are a family of annexin related proteins, which are calcium-ion binding and phospholipid-binding proteins that are often associated with the membrane and flagella-related structyres [39]. The giardins are comprised of several multigene families, of which alpha-giardin has multiple representatives, and are involved in immunogenicity, the cytoskeleton, flagella motility, membrane stability, and attachment, as well as possibly assemblage-specific [40–44]. A total of 6 (33.33%) and 5 (27.78%) alpha-giardins were clustered in H-106 and B-2041. Additionally, gamma giardin (GL50803_17230) was also up-regulated in cysts in both strains, and beta-giardin (GL50803_4812) and delta giardin (GL50803_86676) were up-regulated in B-2041 cysts. Across the literature, there are multiple studies which document changes in expression within the giardin family across encystation [9,45,46]. Some of these changes may be related to cytoskeletal remodelling, such as disassembly of flagella [4,45], which may cause changes in protein abundance of associated *Giardins* which localise to such structures. Overall, there is significant evidence that during encystation and differentiation the abundance of giardins fluctuates, and this correlates with our results for both strains.

The 'serine/threonine protein kinase' functional cluster in up-regulated cyst proteins contained 9 kinases in H-106 and 8 in B-106. Of the 9 kinases in H-106, 3 were NEK kinases, (GL50803_101534, GL50803_11311, GL50803_95593) while 6 NEK kinases were clustered in B-2041 (GL50803_11311, GL50803_113553, GL50803_5375, GL50803_8152, GL50803_8445, GL50803_95593). Two NEK kinases were commonly up-regulated in cysts between strains (GL50803_11311, GL50803_95593). NEK kinases form an interesting family within *Giardia*, constituting 71% of all kinase encoding genes, with an estimated 70% of this multigene kinase family lacking catalytic residues for phosphorylating targets [47]. NEK 1 (GL50803_92498) and NEK 2 (GL50803_5375) have been shown to regulate the cell cycle in *Giardia*, with NEK 2 up-regulated in early encystation [6], and observed as up-regulated in B-2041 cysts. Several other NEK kinases have been identified as up-regulated during encystation in other studies [5,6,9]. Localisation data for a few NEK kinase members is also available, and when compared to the proteins seen to be up-regulated in our study are associated with the flagella (GL50803_5375, GL50803_101534) [47], and may be involved in signalling during remodelling and disassembly of flagella. More data regarding the function, localisation and targets of NEK kinases in *G. duodenalis* is needed, but is highly likely that more members thought in this family than previously play a role in differentiation.

3.4. Strain-specific adaptations during encystation

The current literature has analysed the well-characterised genome strain WB, specifically WB C6 (ATCC 50803), to provide global analysis of differentiation in *Giardia* [5,6,8–10]. This has provided a comprehensive foundation of consistent metabolic processes and reproducible encystation markers, but cannot identify strain-dependent variation. Differences in infectivity of cysts between isolates have been demonstrated in animal models [11], which adds credence to the possibility of strain variation that impacts reinfection. Analysis of protein abundance differences between cysts of H-106 and B-2041 revealed 143 proteins of greater abundance in H-106 cysts while 294 proteins had increased abundance in B-2041 cysts. While our data has highlighted the presence of strain-universal processes within differentiation, H-106 and B-2041 also emerged with strain-specific adaptations after encystation.

In a previous study, we examined strain variation between H-106 and B-2041 trophozoites from axenic culture, and found that surface antigen proteins from the VSP family varied in both number and subpopulation distribution between strains [23]. In this study we show that other proteins that we previously identified as differentially expressed between H-106 and B-2041 trophozoites are also differentially expressed in the corresponding cysts. This includes increased abundance in B-2041 of cathepsin B precursors (GL50803_16160 and GL50803_16779), which may further correlate with the more virulent phenotype of B-2041 seen in animal models [14,48]. Intra-assemblage variation within *G. duodenalis* is yet to be comprehensively analysed since genomes have become available, although it has been shown that *Giardia*-specific gene families of the variable genome differ between assemblages [49,50]. Protein expression of VSP variants also varies between cultured trophozoites from multiple strains of the same subassemblage [51]. Between the 488 and the 688 reproducibly identified proteins in H-106 and B-2041 cysts, 373 proteins were common, which corresponds to 76.4% of cyst proteins from H-106 and 54.2% of proteins from B-2041 cysts (Fig. 1). However, 517 proteins were common between trophozoites, equalling 89.0% and 63.3% of proteins in H-106 and B-2041. This data suggests that greater strain variation occurs at the protein level in the cyst life cycle stage compared to trophozoites (Fig. 1).

The functional analysis of differentially expressed proteins in Section 3.3 highlighted that H-106 and B-2041 are overall remarkably similar metabolically and structurally in axenic culture, and that proteomic variation in cysts appears to be generated from the variable genome, as we have seen previously in trophozoites [23]. Of the 15 functional clusters identified, six were common between strains in the same life cycle stage (Fig. 3). Several clusters which were of higher abundance in trophozoites in only one strain, such as 'metal ion binding', 'transaminase' and 'chaperonin', in fact had multiple common members differentially expressed, but failed to meet enrichment score criteria in the other strain. This was the same in the 'nucleotide Binding' cluster in B-2041 cysts, which was below the enrichment cut-off in H-106. In the 'ribosomal proteins' cluster in H-106 trophozoites and B-2041 cysts, each featured a unique group of proteins of which the majority were still differentially expressed in the other strain in the same life cycle stage, and therefore not a true source of strain variation. Although 2 functional clusters, 'carbohydrate biosynthesis' and

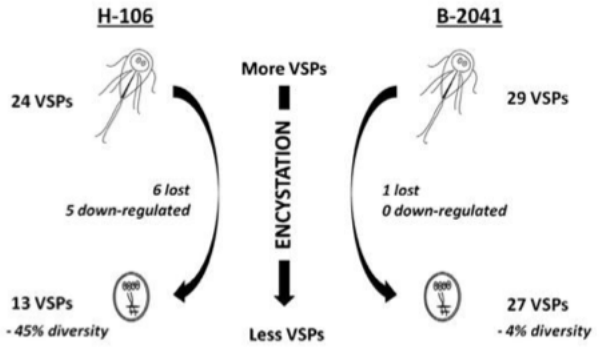


Fig. 4. Strain-specific VSP variant losses. Flow chart demonstrating VSP variant expression for both strains H-106 and B-2041 through the encystation process. Encystation is considered a universal switch for VSP turnover and a hypothesised bottleneck for loss of VSP diversity. The figure shows the number of VSP subpopulations identified in original trophozoite populations prior to encystations, followed by how many VSP variants were considered statistically significantly lower in abundance, and the final number of VSP variants with comparable or higher abundance in cyst populations. The figure shows that H-106 loses 45% of its original VSP variant diversity, compared to only 4% in B-2041. H-106 follows a similar pattern of VSP variant loss as WB C6 seen in tandem MS/MS analysis [6].

'tRNA metabolic processes', were up-regulated in B-2041 trophozoites, these were relatively small clusters of 3 and 7 proteins, respectively. In total, there were 2 large functional clusters that varied between strains and life cycle stages; 'VSP/Furin-like/EGF-like' and 'ankyrin-repeat'. Both these clusters feature proteins from the *Giardia* variable genome already implicated in inter- and intra-assemblage variation [23,49,51]. These data reinforce the hypothesis that the diversity seen in such families at the genome level translates to crucial differences at the level of the expressed proteome between strains at both life cycle stages.

Functional annotation clustering analysis (Section 3.3) revealed a trend of VSP down-regulation in H-106, which was absent from B-2041 (Fig. 3). Of the 24 VSPs identified in H-106 trophozoites, 11 VSPs were statistically significantly down-regulated in abundance in H-106 cysts. Meanwhile in B-2041, of the 29 VSPs identified in trophozoites, only 1 VSP was found to be down-regulated in B-2041 cysts, with 28 VSPs at comparable or greater abundance. This data suggests significant differences in VSP diversity retention between *G. duodenalis* strains, as depicted in Fig. 4. A similar pattern of VSP loss was observed in WB C6 during encystation using tandem MS/MS [6]. In H-106 VSP diversity dropped a striking 45% between trophozoites and cysts, similar to the reduction of 26.6% in WB C6 in the study by Faso et al. [6]. This is compared to retention of 96% of VSP subpopulation diversity in B-2041 cysts at abundances comparable to the preceding generation. Faso et al. hypothesised that this encystation-induced bottleneck for VSP diversity could be from depletion of subpopulations carrying VSP antigens incompatible with survival and proliferation in the encystation media [6]. However, the 11 VSP variants significantly down-regulated in H-106 remain at comparable abundance between B-2041 trophozoites and cysts, with both strains induced in the same high-bile media. In the case of the three strains, all were capable of generating VSP variant heterogeneity in culture, though only B-2041 was able to retain diversity through the differentiation process. This suggests differences in VSP diversity loss are a strain-dependent phenomenon separate to encystation culture conditions.

VSPs comprise a gene family of ~200 members in the A1 assemblage and form a protein coat on the trophozoite surface that is responsible for immune evasion and hypothesised as involved in virulence as well as host and environment adaptation [2,52,53]. In *G. duodenalis* post-transcriptional regulations means only a single VSP is expressed on the entirety of the parasite surface [54], and in axenic culture the absence of immune selection allows multiple VSP variants in the population to proliferate, which are detected in proteomics [6,23,51]. Though microRNAs (miRNAs) derived from small nucleolar RNAs (snoRNAS) play a role in VSP-switching [55] it is not confirmed experimentally if exogenous signals from environment, media or host affect the proliferation and survival of some VSP variants. B-2041 is a well-established zoonotic isolate in animal models [12,15] while H-106 is human-derived. Furthermore, B-2041 has already shown differences in VSP expression patterns compares to H-106 in earlier proteomic studies [23], and it does appear that this heterogeneity is not depleted through the life cycle and the process of differentiation. We hypothesise that if different VSPs grant advantages for specific host species, that broad VSP variant diversity in cysts is selected for zoonotic advantage. It would also correlate well with VSP retention seen in the zoonotic B-2041, as well as the reduction of VSP variation seen in human-derived H-106 and WB C6 [6].

Another significant point of divergence between H-106 and B-2041 emerged in functional cluster analysis, where ankyrin (ANK) repeat containing proteins clustered in the up-regulated proteins in B-2041 cysts. Of the 41 proteins assigned to this cluster, 27 (65.85%) were protein 21.1 members (Figure 3, Supplementary Table 3). Of the 36 protein 21.1 members in H-106, 8 were down-regulated and 1 was up-regulated in cysts, while 27 remained at the same abundance in both life stages. However, in B-2041 of the 61 protein 21.1 identified, 34 were unchanged in abundance between cysts and trophozoites but 27 (44.26%) were at higher abundance in cyst and clustered in function annotation (Fig. 3). The biological functions of this protein family in *Giardia* are unknown. While the protein 21.1 family bears a significant homology to the NEK kinases they lack kinase domains [47], and their most significant feature remains the presence and number of ANK repeats (such that in the recently sequenced A2 genome they are labelled 'ANK repeat protein' [49]). Interestingly, ANK repeat proteins are considered at least an order of magnitude higher in abundance in the *G. duodenalis* genome than in other eukaryotes [56]. Though some of these ANK repeats in protein 21.1 are paired with coiled-coil domains that localise to the flagella and axonemes [56], this only constitutes a subset of all protein 21.1 genes. In biology, proteins containing ANK repeats do not have enzymatic function, but rather mediate a range of protein-protein interactions that modulate diverse cellular pathways in all areas of cell functioning, across all phyla [57]. As ANK repeat proteins mediate protein-protein interactions in such a wide range of cell processes, of which countless pathways occur during encystation, it is difficult to extrapolate if protein 21.1 up-regulation in B-2041 contribute to a single or many biological phenomena. Nonetheless, though the biological significance of differential protein 21.1 expression between H-106 and B-2041 remains unclear, it is a point of considerable strain variation. It remains to be investigated what the targets and the functions of this large class of proteins are in *G. duodenalis*, of which many members are also constitutively expressed [6,23,51].

4. Concluding remarks

For the simple life cycle of *Giardia*, encystation is the essential process that generates cysts for reinfection and completes the life cycle. Post-genomic technologies have permitted identification and quantification of encystation-dependent pathways and markers by utilising the laboratory and genome standard WB C6. In this study we have expanded this foundation from the perspective of pathogenic relevance as well as providing complementary data from alternative strains. It is evident between H-106, B-2041 and studies using WB C6 that there are distinct processes and *Giardia* gene families uniquely and principally encystation-regulated, and

whose expression is also distinctly trophozoite- or cyst-specific. Not only do multiple functional clusters of proteins overlap within cysts and trophozoites of both strains, but many correlate with observation in the literature for encystation. Overall, the strains adhere to consistent metabolic and structural pathways that are central and indispensible for differentiation, making these universal among strains and optimal as drug targets for blocking transmission.

Differences in strain infectivity, host range and pathophysiological intensity are all intrinsic in *G. duodenalis*, and this study has demonstrated the encystation process is tailored to produce strain specific variation. Utilising alternate strains has confirmed universal encystation markers but demonstrated pronounced strain adaptations. VSPs are an important group of proteins at the forefront of parasite virulence and immune evasion, and the retention of VSP diversity throughout encystation in the zoonotic B-2041, but not the human derived H-106 or WB C6, may be central to its wide host infectivity. Furthermore, though the absence of protein 21.1 functional data makes it difficult to understand their biological impact in encystation, significant differences in expression of this gene family indicates a divergent role in differentiation between strains. This is the first evidence that during encystation *G. duodenalis* strains are consistent in metabolism and cell re-structuring, but also produce an unexplored diversity for possible advantages in their subsequent generations. Analyses must extend beyond the foundations from laboratory standards to avoid excluding the variation inherent in this successful and exquisitely adapted parasite. A recently sequenced A2 genome has shown that inter-assemblage variation exists at the genetic level [49], and we await confirmation of intra-assemblage variation in isolates of the same subassemblage for further insight into strain specific adaptation.

Acknowledgements

SJE acknowledges funding from the Australian Government in the form of an APA scholarship, as well as financial support from Macquarie University. SJE wishes to thank Dr Jacqui Upcroft for supplying the *Giardia* samples used and for the ongoing support received from colleagues at Microbial Screening Technologies. PAH wishes to thank Justin Lane for continued support and encouragement. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [27] *via* the PRIDE partner repository with the dataset identifier PXD002002.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.molbiopara.2015.05.007

References

[1] Feng Y, Xiao L. Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. Clin Microbiol Rev 2011;24:110–40.
[2] Ankarklev J, Jerlstrom-Hultqvist J, Ringqvist E, Troell K, Svard SG. Behind the smile: cell biology and disease mechanisms of *Giardia* species. Nat Rev Microbiol 2010;8:413–22.
[3] Bernander R, Palm JE, Svard SG. Genome ploidy in different stages of the *Giardia lamblia* life cycle. Cell Microbiol 2001;3:55–62.
[4] Lauwaet T, Davids BJ, Reiner DS, Gillin FD. Encystation of *Giardia lamblia*: a model for other parasites. Curr Opin Microbiol 2007;10:554–9.
[5] Morf L, Spycher C, Rehrauer H, Fournier CA, Morrison HG, Hehl AB. The transcriptional response to encystation stimuli in *Giardia lamblia* is restricted to a small set of genes. Eukaryotic cell 2010;9:1566–76.
[6] Faso C, Bischof S, Hehl AB. The proteome landscape of Giardia lamblia encystation. PLoS One 2013;8:e83207.
[7] Kane AV, Ward HD, Keusch GT, Pereira ME. In vitro encystation of *Giardia lamblia*: large-scale production of in vitro cysts and strain and clone differences in encystation efficiency. J Parasitol 1991;77:974–81.
[8] Faghiri Z, Widmer G. A comparison of the *Giardia lamblia* trophozoite and cyst transcriptome using microarrays. BMC Microbiol 2011;11:91.
[9] Birkeland SR, Preheim SP, Davids BJ, Cipriano MJ, Palm D, Reiner DS, et al. Transcriptome analyses of the *Giardia lamblia* life cycle. Mol Biochem Parasitol 2010;174:62–5.
[10] Kim J, Bae SS, Sung MH, Lee KH, Park SJ. Comparative proteomic analysis of trophozoites versus cysts of *Giardia lamblia*. Parasitol Res 2009;104:475–9.
[11] Visvesvara GS, Dickerson JW, Healy GR. Variable infectivity of human-derived *Giardia lamblia* cysts for Mongolian gerbils (*Meriones unguiculatus*). J Clin Microbiol 1988;26:837–41.
[12] Upcroft JA, McDonnell PA, Upcroft P. Virulent avian *Giardia duodenalis* pathogenic for mice. Parasitol Today 1998;14:281–4.
[13] Upcroft JA, McDonnell PA, Gallagher AN, Chen N, Upcroft P. Lethal *Giardia* from a wild-caught sulphur-crested cockatoo (*Cacatua galerita*) established in vitro chronically infects mice. Parasitology 1997;114(Pt 5):407–12.
[14] Williamson AL, O'Donoghue PJ, Upcroft JA, Upcroft P. Immune and pathophysiological responses to different strains of *Giardia duodenalis* in neonatal mice. Int J Parasitol 2000;30:129–36.
[15] McDonnell PA, Scott KG, Teoh DA, Olson ME, Upcroft JA, Upcroft P, et al. *Giardia duodenalis* trophozoites isolated from a parrot (*Cacatua galerita*) colonize the small intestinal tracts of domestic kittens and lambs. Vet Parasitol 2003;111:31–46.
[16] Keister DB. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. Trans R Soc Trop Med Hyg 1983;77:487–8.
[17] Boreham PF, Phillips RE, Shepherd RW. The activity of drugs against *Giardia intestinalis* in neonatal mice. J Antimicrob Chemother 1986;18:393–8.
[18] Davids B, Gillin F. Methods for *Giardia* culture, cryopreservation, encystation, and excystation in vitro. In: Luján H, Svärd S, editors. Giardia. Vienna: Springer; 2011. p. 381–94.
[19] Wessel D, Flugge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem 1984;138:141–3.
[20] Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods 2009;6:359–62.
[21] Manza LL, Stamer SL, Ham AJ, Codreanu SG, Liebler DC. Sample preparation and digestion for proteomic analyses using spin filters. Proteomics 2005;5(7):1742–5.
[22] Chapman B, Castellana N, Apffel A, Ghan R, Cramer GR, Bellgard M, et al. Plant proteogenomics: from protein extraction to improved gene predictions. Methods Mol Biol 2013;1002:267–94.
[23] Emery SJ, van Sluyter S, Haynes PA. Proteomic analysis in *Giardia duodenalis* yields insights into strain virulence and antigenic variation. Proteomics 2014;14:2523–34.
[24] Scherl A, Shaffer SA, Taylor GK, Kulasekara HD, Miller SI, Goodlett DR. Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. Anal Chem 2008;80:1182–91.
[25] Fenyo D, Beavis RC. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal Chem 2003;75:768–74.
[26] Craig R, Beavis RC. TANDEM matching proteins with tandem mass spectra. Bioinformatics 2004;20:1466–7.
[27] Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The proteomics identifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res 2013;41:D1063–9.
[28] Neilson KA, Keighley T, Pascovici D, Cooke B, Haynes PA. Label-free quantitative shotgun proteomics using normalised spectral abundance factors. Methods Mol Biol 2013;1002:205–22.
[29] Neilson KA, George IS, Emery SJ, Muralidharan S, Mirzaei M, Haynes PA. Analysis of rice proteins using SDS-PAGE shotgun proteomics. Methods Mol Biol 2014;1072:289–302.
[30] Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L. Washburn MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J Proteome Res 2006;5:2339–47.
[31] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
[32] Castillo-Romero A, Leon-Avila G, Wang CC, Perez Rangel A, Camacho Nuez M, Garcia Tovar C, et al. Rab11 and actin cytoskeleton participate in *Giardia lamblia* encystation, guiding the specific vesicles to the cyst wall. PLoS Negl Trop Dis 2010;4:e697.
[33] Rivero MR, Vranych CV, Bisbal M, Maletto BA, Ropolo AS, Touz MC. Adaptor protein 2 regulates receptor-mediated endocytosis and cyst formation in *Giardia lamblia*. Biochem J 2010;428:33–45.
[34] Hernandez Y, Castillo C, Roychowdhury S, Hehl A, Aley SB, Das S. Clathrin-dependent pathways and the cytoskeleton network are involved in ceramide endocytosis by a parasitic protozoan, *Giardia lamblia*. Int J Parasitol 2007;37:21–32.
[35] Svard SG, Rafferty C, McCaffery JM, Smith MW, Reiner DS, Gillin FD. A signal recognition particle receptor gene from the early-diverging eukaryote, *Giardia lamblia*. Mol Biochem Parasitol 1999;98:253–64.
[36] Faso C, Konrad C, Schraner EM, Hehl AB. Export of cyst wall material and Golgi organelle neogenesis in *Giardia lamblia* depend on endoplasmic reticulum exit sites. Cell Microbiol 2013;15.
[37] Langford TD, Silberman JD, Weiland ME, Svard SG, McCaffery JM, Sogin ML, et al. *Giardia lamblia*: identification and characterization of Rab and GDI

proteins in a genome survey of the ER to Golgi endomembrane system. Exp Parasitol 2002;101:13–24.
[38] Gaechter V, Schraner E, Wild P, Hehl AB. The single dynamin family protein in the primitive protozoan *Giardia lamblia* is essential for stage conversion and endocytic transport. Traffic 2008;9:57–71.
[39] Weiland ME, McArthur AG, Morrison HG, Sogin ML, Svard SG. Annexin-like alpha giardins: a new cytoskeletal gene family in *Giardia lamblia*. Int J Parasitol 2005;35:617–26.
[40] Weiland ME, Palm JE, Griffiths WJ, McCaffery JM, Svard SG. Characterisation of alpha-1 giardin: an immunodominant *Giardia lamblia* annexin with glycosaminoglycan-binding activity. Int J Parasitol 2003;33:1341–51.
[41] Weiland ME, McArthur AG, Morrison HG, Sogin ML, Svard SG. Annexin-like alpha giardins: a new cytoskeletal gene family in *Giardia lamblia*. Int J Parasitol 2005;35:617–26.
[42] Steuart RF, O'Handley R, Lipscombe RJ, Lock RA, Thompson RC. Alpha 2 giardin is an assemblage A-specific protein of human infective *Giardia duodenalis*. Parasitology 2008;135:1621–7.
[43] Vahrmann A, Saric M, Koebsch I, Scholze H. Alpha14-Giardin (annexin E1) is associated with tubulin in trophozoites of *Giardia lamblia* and forms local slubs in the flagella. Parasitol Res 2008;102:321–6.
[44] Feliziani C, Merino MC, Rivero MR, Hellman U, Pistoresi-Palencia MC, Ropolo AS. Immunodominant proteins alpha-1 giardin and beta-giardin are expressed in both assemblages A and B of *Giardia lamblia*. BMC Microbiol 2011;11:233.
[45] Jenkins MC, O'Brien CN, Macarisin D, Miska K, Fetterer R, Fayer R. Analysis of giardin expression during encystation of *Giardia lamblia*. J Parasitol 2012;98:1266–70.
[46] Kim J, Lee HY, Lee MA, Yong TS, Lee KH, Park SJ. Identification of alpha-11 giardin as a flagellar and surface component of *Giardia lamblia*. Exp Parasitol 2013;135:227–33.
[47] Manning G, Reiner DS, Lauwaet T, Dacre M, Smith A, Zhai Y, et al. The minimal kinome of *Giardia lamblia* illuminates early kinase evolution and unique parasite biology. Genome Biol 2011;12:R66.
[48] Cotton JA, Bhargava A, Ferraz JG, Yates RM, Beck PL, Buret AG. *Giardia duodenalis* Cathepsin B. Proteases degrade intestinal epithelial interleukin-8 and attenuate interleukin-8-induced neutrophil chemotaxis. Infect Immun 2014;82:2772–87.
[49] Adam RD, Dahlstrom EW, Martens CA, Bruno DP, Barbian KD, Ricklefs SM, et al. Genome sequencing of *Giardia lamblia* genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). Genome Biol Evol 2013;5:2498–511.
[50] Jerlstrom-Hultqvist J, Franzen O, Ankarklev J, Xu F, Nohynkova E, Andersson JO, et al. Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. BMC Genomics 2010;11:543.
[51] **Emery SJ, Lacey E, Haynes PA. Quantitative proteomics analysis of *Giardia duodenalis* assemblage A – a baseline for host, assemblage and isolate variation. Proteomics 2015,** http://dx.doi.org/10.1002/pmic.201400434 **[Epub ahead of print].**
[52] Adam RD, Nigam A, Seshadri V, Martens CA, Farneth GA, Morrison HG, et al. The *Giardia lamblia* vsp gene repertoire: characteristics, genomic organization, and evolution. BMC Genomics 2010;11:424.
[53] Prucca CG, Lujan HD. Antigenic variation in *Giardia lamblia*. Cell Microbiol 2009;11:1706–15.
[54] Prucca CG, Slavin I, Quiroga R, Elias EV, Rivero FD, Saura A, et al. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. Nature 2008;456:750–4.
[55] Li W, Saraiya AA, Wang CC. The profile of snoRNA-derived microRNAs that regulate expression of variant surface proteins in *Giardia lamblia*. Cell Microbiol 2012;14:1455–73.
[56] Elmendorf HG, Rohrer SC, Khoury RS, Bouttenot RE, Nash TE. Examination of a novel head-stalk protein family in *Giardia lamblia* characterised by the pairing of ankyrin repeats and coiled-coil domains. Int J Parasitol 2005;35:1001–11.
[57] Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY. The ankyrin repeat as molecular architecture for protein recognition. Protein Sci 2004;13:1435–48.

4.4 Supplementary Data

The following supporting information is available for this manuscript.

Supplementary Table S1. Excel spreadsheet showing the biological classification of proteins reproducibly identified (found in all three replicates with SpC ≥ 5) for the two *Giardia* strains in this study for each of their life cycle stages. The first tab shows the 586 reproducibly identified proteins from H-106 Trophozoites and their replicate SpC & NSAF. The second tab shows the 488 reproducibly identified proteins from H-106 Cysts and their replicate SpC & NSAF. The third tab shows the 816 reproducibly identified proteins from B-2041 Trophozoites and their replicate SpC & NSAF. The fourth tab shows the 688 reproducibly identified proteins from B-2041 Cysts and their replicate SpC & NSAF. **(Supplementary DVD)**

Supplementary Table S2. Excel spreadsheet showing the biological classification of differentially expressed proteins (up regulated, down regulated and unchanged) obtained from t-test analysis between *Giardia* life cycle stages. Proteins which are differentially expressed between H-106 Cysts and H-106 Trophozoites are filled in green and red respectively on the first tab. Proteins which are differentially expressed between B-2041 Cysts and B-2041 Trophozoites are filled in green and red respectively on the second tab. Proteins which are differentially expressed between B-2041 Cysts and H-106 Trophozoites are filled in green and red respectively on the third tab. **(Supplementary DVD)**

Supplementary Table S3. Excel spreadsheet showing the functional annotation clustering analysis using DAVID for statistically significant differentially expressed proteins in H-106 Cysts and B-2041 Cysts. Only functional annotation clusters with an enrichment score of ≥ 1 were considered in the paper, and these are shown in the spreadsheet. The spreadsheet shows the name of the cluster in one column, and the GO annotations and Interpro annotations in the following columns that were grouped under the name of the cluster. The final columns contain the multiple Giardiadb.orf ORF numbers of the genes assigned to each cluster. **(Supplementary DVD)**

CHAPTER 5

Quantitative proteomics of G. duodenalis *during* in vitro *host-parasite interactions between host-cell attached, and host-secretion exposed trophozoites.*

5. Differential stimulation of *Giardia duodenalis* trophozoites between host soluble signals and host cell attachment during *in vitro* interactions

5.1 Context

Giardiasis is a multi-factorial disease, where symptoms are caused by both parasite factors and the host immune response. *Giardia duodenalis* induces disease in the absence of cell invasion, with no known secreted toxins and with relatively little inflammation. Recent research also indicates several mechanisms of host immunomodulation by *Giardia* also occur. Host-parasite models utilising intestinal epithelial cells (IEC) have been used to investigate disease induction in *Giardia*, which until now has only been characterised using transcriptomics. This investigation is the first to adopt a proteomic approach to investigate changes in parasite expression, as well as the first to use tandem mass tag (TMT) labelling for protein quantitation. Trophozoites were separately exposed to IEC host monolayers or host soluble factors (HSF) to observe differential induction of virulence driven by host signals, as compared to parasite attachment to IECs. This study provides insight into key process and signals regulating virulence and disease induction in *Giardia*.

5.2 Contributions and Permissions

I performed 80% of all work pertaining to growth of *Giardia* and IEC cells, with assistance provided by Daniel Vuong for growth and maintenance of HT-29 cells. I performed 80% of all lab work for processing protein extracts and analysis of peptides, with guidance for methodology for TMT labelling and experimental design provided by Mehdi Mirzaei. I wrote 80% of the manuscript with assistance in data analysis provided by Dana Pascovi and editing provided by Professor Paul Haynes, Mehdi Mirzaei and Dr Ernest Lacey.

This paper was submitted to *Nature: Scientific Reports* on 29/7/15 (SREP-15-21979-T) and accepted for publication on 04/01/16. We have chosen to reformat the journal submission of this paper (Publication VI) as a thesis chapter for ease of understanding. This consists of two major changes to the format: firstly, the methods have been moved from the end of the manuscript to in between the introduction and results; and, secondly, we have incorporated the substantial supplementary information into the chapter. The supplementary methods, supplementary tables and supplementary figures, as well as the main figures and tables associated with the data, have been embedded in the text near where they are first referred to.

5.3 Manuscript Information

Induction of virulence factors in *Giardia duodenalis* independent of host attachment

Samantha J. Emery[1], Mehdi Mirzaei[1], Daniel Vuong[2], Dana Pascovici[3], Joel M. Chick[4], Ernest Lacey[2], Paul A. Haynes[1*]

[1] Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, NSW 2109, Australia

[2] Microbial Screening Technologies, Pty, Ltd, Smithfield, NSW 2165, Australia

[3] Australian Proteome Analysis Facility (APAF), Macquarie University, North Ryde, NSW, 2109, Australia

[4] Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, USA

* To whom correspondence should be addressed

Reviewer login details for PRIDE data submission:

http://www.ebi.ac.uk/pride/archive/login

Email: reviewer72542@ebi.ac.uk

Password: HxIK9rtW

(This data remains confidential during the review process, and will be deposited into ProteomeXchange upon acceptance)

Submitted to Nature: Scientific Reports

1. 29/7/15: SREP-15-21979-T (Original)
2. 16/11/15: SREP-15-21979A (Resubmission)
3. 11/12/15: SREP-15-21979B (Resubmission)
4. 04/01/16: Accepted for Publication

5.4 Abstract

Giardia duodenalis is responsible for the majority of parasitic gastroenteritis in humans worldwide. Host-parasite interaction models *in vitro* provide insights into disease and virulence and help us to understand pathogenesis. Using HT-29 intestinal epithelial cells (IEC) as a model we have demonstrated that initial sensitisation by host secretions reduce proclivity for trophozoite attachment, while inducing virulence factors. Host soluble factors triggered up-regulation of membrane and secreted proteins, including Tenascins, Cathepsin-B precursor, cystatin, and numerous Variant-specific Surface Proteins (VSPs). By comparison, host-cell attached trophozoites up-regulated intracellular pathways for ubiquitination, reactive oxygen species (ROS) detoxification and pyridoxal phosphate (PLP). We reason that these results demonstrate early pathogenesis in *Giardia* involves two independent host-parasite interactions. Motile trophozoites respond to soluble secreted signals, which deter attachment and induce expression of virulence factors. Trophozoites attached to host cells, in contrast, respond by up-regulating intracellular pathways involved in clearance of ROS, thus anticipating the host defence response.

5.5 Introduction

Our understanding of how *Giardia* causes disease is incomplete, particularly concerning the early stages of trophozoite pathogenesis [1]. *Giardia* trophozoites attach strongly to the intestinal epithelial cells via a ventral adhesive disc and cause significant damage and disruption to gastroepithelial cells in the absence of cell invasion, secreted toxins and overt inflammation [2]. The interplay between the host and the parasite on establishment is a gap in our knowledge. Recently, host-parasite interaction models with human intestinal epithelial cells (IEC) *in vitro* have provided a foundation for understanding disease induction by *Giardia* trophozoites. Results indicate that these interaction models are stimulatory, inducing expression of parasite factors which have limited or no expression in axenic culture alone [3]. Additional studies have addressed gene expression and transcriptional changes in *Giardia* trophozoites co-incubated with Caco-2 and HCT-8 cells[4], and HT-29 cells[5], and analysed the secreted proteomes [3]. There have also been complementary studies of transcripts from IECs exposed to *Giardia* trophozoites [6, 7]. Together, these studies indicate the efficacy of *in vitro* models to explore the induction of giardiasis.

Proteomics is one of the few exploratory tools available to understand parasite biology at a physiological level [8]. Currently there are a limited number of proteomic studies performed on *Giardia* [9, 10, 11, 12, 13, 14, 15], but these do not focus on the host-parasite interplay at key events in pathogenesis. The present study addresses the dynamics of the origin of the infective cycle of attachment during the early stages of pathogenesis. This was achieved using tandem mass tag (TMT) labelling of trophozoite proteins after *in vitro* exposure to host cells during co-incubation (CI) and host secretions (host soluble factors (HSF)) (Figure 1, Part A). TMT labelling is a quantitative proteomics technique that uses multiplexed isobaric tags which allow greater parallelisation without increasing analysis complexity [16] (Figure 1, Part B). This is the first instance of TMT labelling in *Giardia* and demonstrates its sensitivity for protein

quantitation for parasite proteomics, even for the subtle changes in protein expression which can occur during short incubation periods.

In this study we have utilised a cell-free incubation, with only soluble products from host target cells, which has facilitated discovery of the very early, attachment independent, stage of *Giardial* pathogenesis. Using HT-29 cells as an *in vitro* model, we have demonstrated that preceding host attachment, trophozoites are actively responding to secreted soluble host signals, and activating manifestly different mechanisms to those involved with attaching to the host. Our data supports the hypothesis that the early stages of Giardial pathogenesis involve a distinctive biphasic process which involves induction of virulence factors in the trophozoites, independent of attachment to the host cells.

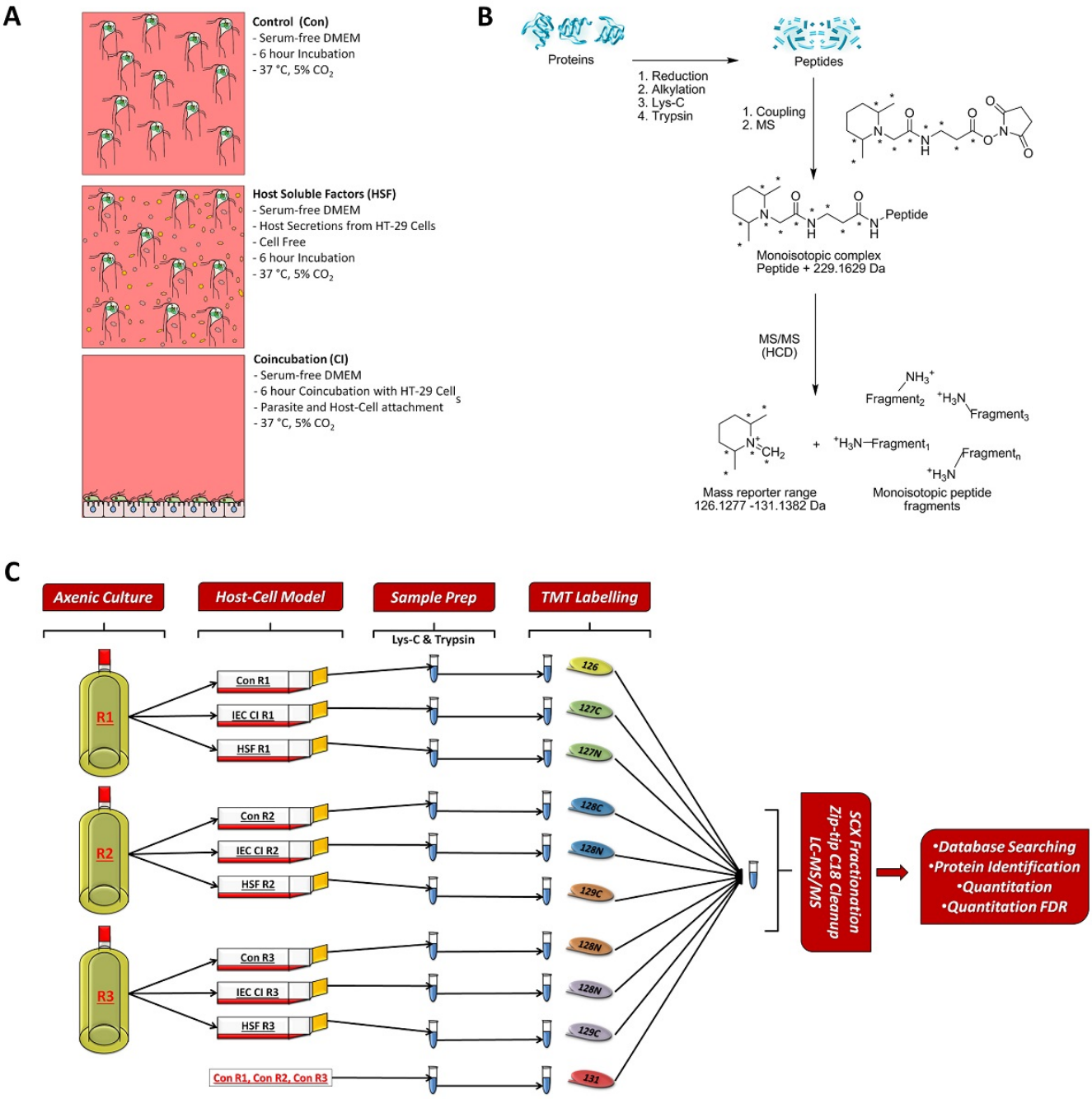


Figure 1: Experimental design and TMT labelling workflow for the experiment. (A) Summary of the experiment conditions of the control, host soluble factor and co-incubation treatments. (B) Explanation of the TMT-labelling strategy utilised in the experiments, peptides from the triplicates of the 3 conditions and a pooled control were labelled with one of each the TMT 10plex reagents. These reagents are observed as a monoisotopic complex in the first round of MS analysis on a high resolution mass spectrometer. During MS/MS and HCD based fragmentation, the TMT labels are fragmented to produce 10 reporter ions with distinguishable masses in the low m/z range, which allow relative protein quantitation (C) Overarching experimental design and workflow. Biological triplicates of BRIS/95/HEPU/2041 were grown to confluence in parasite culture in TYI-S-33 medium, before replicates were split into a replicate of each control for cell culture conditions (Con), co-incubation with IEC monolayers (CI IEC) and incubation in host soluble factors generated by IECs (HSF). Proteins were extracted from the 9 replicates, and a pooled control was generated from equal aliquots of protein from the control triplicates. After proteolytic digestion, samples were labelled in a 10 plex TMT reaction and then pooled. The combined sample was fractionated by SCX chromatography and desalted using a C18 ZipTip prior to LC-MS/MS on a Q-Exactive Orbitrap.

5.6 Methods

5.6.1 Cell and Giardia cultures

Human intestinal epithelial cell line HT-29 were grown in high glucose DMEM containing GlutaMAX™ (Gibco, Life Technologies) supplemented with 10% Foetal Bovine Serum (FBS) (Gibco, Life Technologies) and 1% Penicillin/Streptomycin (5000 U/mL) (Gibco, Life Techologies). HT-29 cells were maintained in 75cm^2 flasks (Corning Incorporated, New York) and sub-cultured 3 times a week in an incubator at 5% CO_2 at 37°C. *Giardia* trophozoites of BRIS/95/HEPU/2041 were grown in TYI-S-33 medium supplemented with 10% newborn calf serum and 1% unfractionated bovine bile [17]. Parasites were subcultured at end-log phase into fresh media and interaction studies were carried within 5 passages from recovery from cryopreservation. Absence of bacterial and fungal contamination was verified using serial dilutions and nutrient agar Petri plates to ensure no colony forming units were detected in either human or parasite cultures prior to interaction and protein extraction.

5.6.2 In vitro interaction and co-incubation

HT-29 cells were grown in 75cm^2 flasks to confluence prior to interaction studies, and washed twice with 37°C PBS to remove media and serum traces. In order to generate the HSF fraction for interaction, confluent monolayers of HT-29 were incubated for 20 hours in serum-free, DMEM media. Cell viability and monolayer integrity was monitored throughout the 20 hour incubation and final cell viability measured through trypan blue dye exclusion (Sigma Aldrich). Co-incubated media was decanted from the monolayer, centrifuged to remove any whole cells or cellular debris and filtered through a 0.22µm pore filter (Merck Millipore). To normalise the HSF fraction for flask variation, all DMEM generated from confluent monolayers was pooled.

Figure 1 shows the experimental design and workflow of the TMT experiment. *Giardia* trophozoites were grown to mid-log phase in triplicate in ‘inside-out’ custom roller bottles [18].

Trophozoites were washed twice with 37°C PBS to remove media and serum traces. Trophozoites were both motile and viable at the beginning of host-parasite interactions, and twice the volume of serum-free DMEM used in normal cell culture was used to reduce oxygen tension in all treatments as previously described [4]. HT-29 monolayers, grown in 175cm^2 flasks to 100% confluence, were washed with PBS prior to interaction to remove HSF to minimise overlap between interaction studies. For co-incubation with the HT-29 monolayer, cells were incubated in a 3:1 ratio in serum-free DMEM. For incubation with HSF, trophozoites were incubated in the filtered media from confluent HT-29 cells and for the control *Giardia* trophozoites were incubated in serum-free DMEM. Incubation and interactions were performed in triplicate for 6 hours at 5% CO_2 at 37°C. Cell and parasite viability was monitored throughout the 6 hours as well as the viability of the HT-29 monolayer and final cell viability quantified through exclusion dye assay using erythrosin B (Sigma Aldrich) and trypan blue for trophozoite and HT-29 cells, respectively, with viability >95% considered acceptable. Trophozoite adherence was also monitored in co-incubation replicates, as was trophozoite motility monitored throughout all treatments. The integrity of the HT-20 monolayer was also observed throughout the co-incubation period.

5.6.3 In vitro attachment assays

To measure rates of *Giardia*-host cell attachment, HT-29 cells were grown in 75cm^2 flasks to confluence, and *Giardia* trophozoites were grown to mid-log phase. Co-incubation of trophozoites and HT-29 cells were performed in triplicate as in *in vitro* interaction studies, with serum-free DMEM at twice normal volume with a 3:1 trophozoite to cell ratio. A control for adherence consisted of trophozoites incubated in triplicate in 75cm^2 flasks, in serum free DMEM. The assay was performed over 6 hours at 5% CO_2 at 37°C, with attachment monitored at hourly time points from T_0 to T_{360} minutes. Media was sub-sampled and total number of free,

unattached trophozoites in the flask counted by haemocytometer in both co-incubation and control triplicate flasks. The total number of free trophozoites was expressed as a percentage of trophozoites at T_0. Difference in rates of attachment was assessed for statistical significance using an unpaired t-test at each hourly time point, with a p-value ≥ 0.05 considered significant.

A second assay was performed to measure the impact of host soluble factors on both trophozoite adherence to flasks, as well as trophozoite host-cell attachment. The assay conditions were run in triplicate with the same volumes, ratios and media as previously specified. HSF fractions were generated as before. The assay was run over 12 hours, with the first incubation in flasks without cells from T_0 to T_{360}, followed by a second round of co-incubation with HT-29 cells from T_{360} to T_{720}. The experimental design of the attachment assay using HSF-exposed trophozoites is shown in Figure 2. A total of 3 treatments were analysed; trophozoites in serum-free DMEM in the first round and then co-incubated with HT-29 cells in the second round (Con/CI), trophozoites exposed to HSF in serum-free DMEM in the first round and then co-incubated with HT-29 in the presence of HSF in the second round (HSF/CI). Lastly, a control was run in serum free DMEM without HT-29 monolayers in both rounds of the assay (Con/Con). The same trophozoites were used in both rounds of the assay. Trophozoites were detached from the flask after the first round incubation, and transferred with the media into flasks containing HT-29 for subsequent host-cell co-incubation. The control triplicates from the first round were detached and transferred into fresh flasks for the second round of the assay. Trophozoites were counted every 2 hours in the first round, and then hourly during co-incubations with IEC monolayers. As before, the total number of free trophozoites was expressed as a percentage of trophozoites at T_0. Difference in rates of attachment was assessed for statistical significance using an unpaired t-test at each hourly time point, with a p-value ≥ 0.05 considered significant.

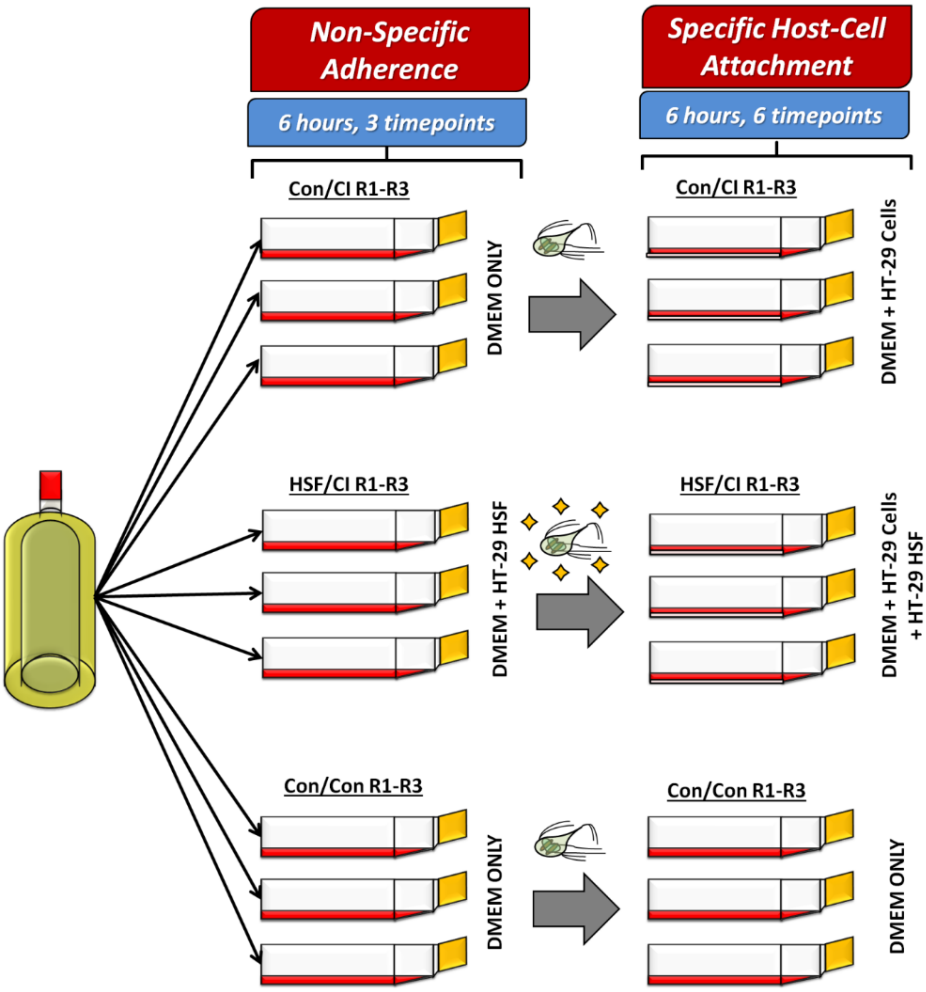


Figure 2: Design of host-soluble factor exposed non-specific adherence and specific host-cell attachment assay across 2 rounds of treatments. Trophozoites were raised in a single axenic culture and split between triplicates in three treatments. The first were trophozoites incubated in serum-free DMEM for 6 hours in the first round, and then co-incubated in the second round with confluent HT-29 (Con/CI). The second treatment exposed trophozoites to host soluble factors in serum-free DMEM for the 6 hours in the first round, and then co-incubated in the second round with confluent HT-29 also in the presence of host soluble factors (HSF/CI). The last treatment was a control for both rounds, and trophozoites were incubated in serum-free DMEM in both rounds of the assay (Con/Con). Non-specific adherence was measured in three timepoints (120 min, 240 min and 360 min) and specific host-cell attachment was measured hourly from 0-6 hours. Trophozoites were detached from the flasks used in the first round of the assay for non-specific adherence, and the same population was used to measure specific host-cell attachment.

5.6.4 Protein extraction and digestion

Trophozoites were detached from the HT-29 monolayer or flasks by briefly chilling the culture flask, collected by centrifugation and washed once with ice-cold PBS. Microscopy of detached trophozoites post host-cell interaction was performed to ensure absence of IEC contamination. Trophozoites were extracted in ice-cold SDS sample buffer containing 1mM EDTA and 5%

beta-mercaptoethanol, and then reduced at 75°C for 10 min. Protein extracts were centrifuged at 0°C at 13 000 x *g* for 10 min to remove debris, and stored at -20°C.

Protein extracts were reduced with 5mM dithiothreitol and then alkylated with 10mM iodoacetamide. Alkylation was quenched with 5mM dithiothreitol. Protein extracts were precipitated using methanol/chloroform [19], followed by resuspension in 8M Urea in 50mM Tris (pH 8.8). The concentration of protein in each triplicate sample was measured by BCA assay (Pierce) before fractionation and digestion, initially with Lys-C overnight at 30°C at concentrations of 1µg enzyme to 100µg protein. Lys-C digestion was followed with a sequential digestion with Trypsin at 37°C with 1µg enzyme for 100µg protein. Digestion with trypsin was performed at 37°C for 6 hours. Samples were acidified with trifluoroacetic acid, and then desalted on a 200mg C18 SepPak (Waters, Massachusetts). Protein extracts were dried down using a vacuum centrifuge, resuspended and peptide concentration determined using micro BCA (Pierce).

5.6.5 TMT labelling

Samples for TMT labelling were resuspended in 200mM HEPES (pH 8) and a total of 70mg from each triplicate for each of the 3 treatments, as well as a pooled control, were labelled in a 10plex TMT reaction (Thermo, San Jose, CA) using 0.14mg of each reagent. Labelling was performed for 1 hour at room temperature and then quenched with 5% hydroxylamine. Each of the 10 samples were then combined, dried down using a vacuum centrifuge, reconstituted in 1% formic acid, and desalted on a 200mg C18 SepPak (Waters, Massachusetts). The combined sample of TMT labelled peptides were dried down again and reconstituted in 1% formic acid prior to fractionation by strong cation exchange (SCX) high pressure liquid chromotography (HPLC) using a PolyLC PolySulfoethyl A (200 mm × 2.1 mm × 5 µm, 200 Å) column and UV detection at 210nm. Samples were resuspended and initially loaded with buffer A (5mM

KH_2PO_4, pH 2.7, 25% ACN), and fractionated with a linear gradient of 10-45% buffer B (5mM KH_2PO_4, pH 2.72, 350mM KCl, 25% ACN) for 70 minutes, which was rapidly increased from 45-100% buffer B for 10 minutes at a flow rate of 300µl/min. A total of 36 fractions of varying volumes were collected and dried down by vacuum centrifugation, before being combined to 12 fractions based on peptide content. These 12 fractions were desalted using C18 OMIX® tips (Agilent), dried down using a vacuum centrifuge and reconstituted in 1% formic acid in preparation for nanoflow liquid chromatography tandem mass spectrometry (NanoLC-MS/MS).

5.6.6 Nanoflow LC-MS/MS for TMT labelling

Samples were analysed on a Q Exactive Orbitrap mass spectrometer (Thermo Scientific) coupled to an EASY-nLC1000 (Thermo Scientific). Reversed-phase chromatographic separation was carried out on a 75 µm id. × 100 mm, C18 HALO column, 2.7 µm bead size, 160 Å pore size. A linear gradient of 1-30% solvent B (99.9% ACN/0.1% FA) was run over 170 minutes. The mass spectrometer was operated in the data-dependent mode to automatically switch between Orbitrap MS and ion trap MS/MS acquisition. Survey full scan MS spectra (from *m*/*z* 350 to 1850) were acquired with a resolution of 70,000 at *m*/*z* 400 and an AGC (Automatic Gain Control) target value of 1×10^6 ions. For identification of TMT labelled peptides, the ten most abundant ions were selected for higher energy collisional dissociation (HCD) fragmentation. HCD normalised collision energy was set to 35% and fragmentation ions were detected in the Orbitrap at a resolution of 70.000. Target ions that had been selected for MS/MS were dynamically excluded for 90 sec. For accurate mass measurement, the lock mass option was enabled using the polydimethylcyclosiloxane ion (*m*/*z* 445.12003) as an internal calibrant.

5.6.7 Database searching for protein/peptide identification

For peptide identification, raw data files produced in Xcalibur software (Thermo Scientific) were processed in Proteome Discoverer V1.3 (Thermo Scientific) prior to Mascot searching against the Giardiadb.org 5.0 release for Assemblage A1, isolate WB (ATCC 50803). For searching, the MS tolerance was set to ±10 ppm and the MS/MS tolerance to 0.1 Da. One missed cleavage was allowed and carbamidomethylation of cysteines was set as a static modification. TMT 10plex modification of peptide N-termini and lysine residues, methionine oxidation, and deamidation of asparagine and glutamine were set as variable modifications. Search result filters were selected as follows; only peptides with a score >15 and below the Mascot significance threshold filter of $p = 0.05$ were included and single peptide identifications required a score equal to or above the Mascot identity threshold. Protein grouping was enabled such that when a set of peptides in one protein were equal to, or completely contained, within the set of peptides of another protein, the two proteins were contained together in a protein group. Quantitative information calculated from reporter ion intensities was only accepted for peptides with scores equal to or above the Mascot homology score, and the median value was taken to compare protein ratios.

5.6.7 Analysis of differentially expressed proteins

The relative quantitation for host-cell interactions were derived by the ratio of TMT labels for each of the treatments over their respective replicate control (i.e. HSF R1 against Con R1 and CI IEC R1 against Con R1). A total of three expression ratios were derived for both HSF and CI IEC biological replicates, and an average fold change was calculated for each protein identified. In addition to TMT ratios for differential expression, proteins were analysed statistically via one-sample t-test to evaluate significance of observed protein expression changes. Proteins were only considered differentially expressed if they met both fold change

criteria as well as ≥ 0.05 p-value significance. Functional annotation of proteins was performed using Uniprot to assign GO function, subcellular localisation, Interpro protein domains and structure annotations where available.

Prediction of secreted proteins was analysed using a series of bioinformatics tools to assess subcellular localisation [20], presence of signal peptides [21], transmembrane helices [22] and nuclear localisation [23]. Proteins were submitted to TargetP v1.1 using the default settings for the algorithm for non-plant sequences, with a reliability score ≤ 3 selected for cutoff (with 1 being the highest reliability score). For analysis of signal peptides, proteins were submitted to SignalP v4.01, with the default settings for eukaryotic sequences. The presence of transmembrane helices was predicted using THMHH Server v2.0. Finally, as an exclusionary tool, proteins were submitted to NucPred to assess nuclear localisation signals with proteins above a prediction reliability score of ≥ 0.90 considered significant.

The mass spectrometry raw data files, database search results and TMT labelling protein quantitation results have all been deposited to the ProteomeXchange Consortium [24] via the PRIDE partner repository with the dataset identifier PXD002398.

5.6.7 Statistical analysis of dataset

Several additional statistical analyses of the TMT dataset were performed to evaluate the variability of the dataset, and establish the presence of an underlying biological difference between HSF and CI treatments compared to control. These include an assessment of sample variability based on control/control ratios, an unsupervised multivariate principal component analysis (PCA), and an analysis of the p-value distribution using paired t-tests between triplicates of HSF/Control and CI/Control ratios.

Estimates of the variability were generated of the control replicates using the log2-transformed ratios to the pooled control sample available from the TMT labelling experiment as in Song *et al* [25] (specifically Figure 2 in Song *et al*). Meanwhile, the unsupervised multivariate PCA analysis was performed using the log-transformed ratios of all samples to the control pooled sample (label channel 131), to visualise the treatment replicate samples for the whole dataset. Finally, p-value histograms of the two treatments (HSF vs Con and CI vs Con) were generated according to Pounds *et al* [26] to investigate the distribution of p-values to demonstrate whether the null hypothesis registered a true effect of differential expression

5.7 Results

5.7.1 In vitro co-incubation is an active model for trophozoite-host attachment

Comparisons of the rates of adherence of trophozoites to either empty flasks or HT-29 cell monolayers can be viewed in Figure 3, and the complete dataset can be found in Supplementary Data S1. High rates of adherence occur in the first two hours, with no significant difference in the number of free trophozoites occurring between monolayer co-incubation and control flasks. In the first hour, 78.2% and 72.4% of trophozoites remain free in the media in co-incubation and control flasks, respectively, and this decreases to 56.7% and 61.7% free trophozoites by the second hour (Figure 3A). After 2 hours adherence plateaus, with only 10% more trophozoites adhering to the flasks in the control between 2-6 hours. However, during co-incubation with HT-29 IEC, attachment increases significantly after 2 hours, with 70% of trophozoites attached after 6 hours. Differences between free and attached trophozoites between co-incubation and control flasks are statistically significant ($p<0.05$) beyond 2 hours to assay completion at 6 hours (Figure 3A). This highlights the viability of *in vitro* host-parasite models, which prompt active attachment that is distinct and separate from adherence.

Further evidence of active interaction during *in vitro* co-incubation is exemplified by changes induced in HT-29 cell morphology over the 6 hour co-incubation. Figure 3B shows that HT-29 cells become enlarged and amorphous by 2 hours, and at later time points detachment from the monolayer has begun to occur (denoted by arrows). Over the 6 hours, increased cellular debris from damaged host cells also accumulated, providing further evidence of pathogenic effects from the interaction with trophozoites.

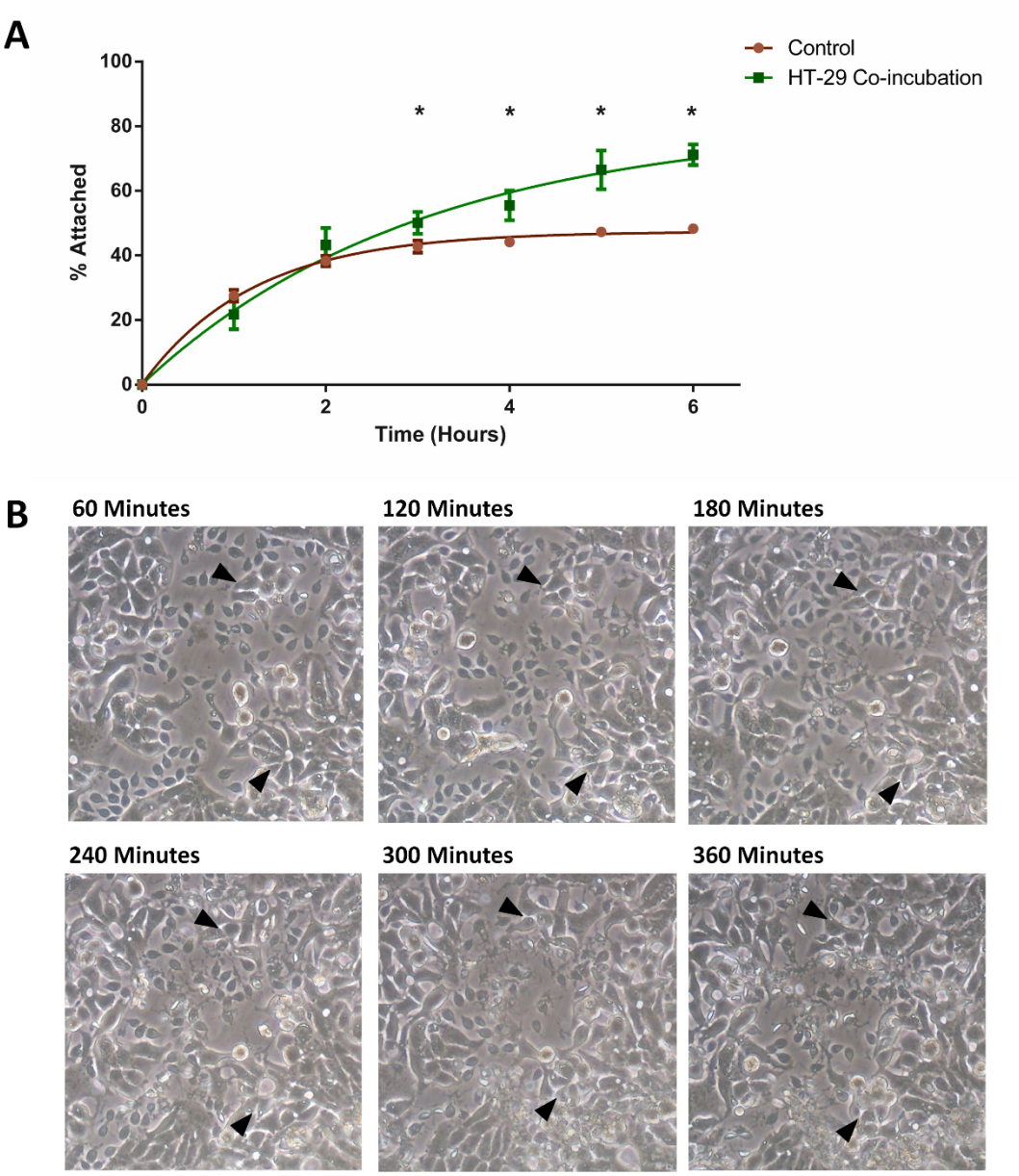


Figure 3: Results of the *in vitro* host-cell attachment versus non-specific adherence. (A) Rates of attachment between *Giardia* trophozoites incubated with HT-29 cells over 6 hours against a control for adherence (T75 flasks with media only). A '*' indicates a significant difference in % attached trophozoites compared to control (designated by p-value ≥ 0.05). (B) Changes in HT-29 cell morphology induced during co-incubation with *Giardia* trophozoites. The arrows (▲) highlight 2 regions of affected cells throughout the 6 hour co-incubation.

5.7.2 Host secretions trigger a non-attaching trophozoite phenotype

The complete results for rates of adherence and host-cell attachment between control and HSF-exposed trophozoites are shown in Figure 4, with the full dataset in Supplementary Data S1. Trophozoites exposed to HSF showed remarkably reduced rates of both adherence to T75cm^2

flasks (Figure 4A), as well as attachment to HT-29 cells during co-incubation (Figure 4C). The differences in the numbers of adhered/attached trophozoites were significantly lower ($p<0.05$) at every time point for both conditions, indicating the switch in phenotype was immediate and sustained. Trophozoites exposed to HSF during conditions for adherence had a consistent average reduction of 45.7% (standard deviation ± 1.4%) compared to controls across all hourly timepoints. Trophozoites in control flasks had an average adherence of 49.7% at 6 hours compared to 26.6% in HSF-exposed trophozoites, meaning 46.5% less trophozoites had adhered. During the second round of co-incubation, trophozoites in control flasks of only DMEM had reduced viability, likely due to oxygen tension on parasites in the absence of IECs [4]. This manifested as a 7.6% decrease in adherence between 5 and 6 hours in controls (Figure 4C). Regardless, differences between adhered controls and host-cell attached trophozoites exposed to HSF also peaked at an average of 20.3% between 3-5 hours in the second round. HSF-exposed and unexposed trophozoites co-incubated with HT-29 showed similar trends of increasing attachment, with similar linear slopes distinct from the trend of exponential plateau observed in the control flasks (Figure 4C). This suggests HSF-exposed trophozoites are still capable of host-cell attachment, albeit at lower rates. Co-incubations in the presence of HSF reduced host-cell attachment between 9.0% and 24.4% after 1 to 6 hours respectively, with 47.2% less trophozoites attached in co-incubations with HSF after 6 hours. This indicates that HSF produce similar reductions in both adherent and host-attaching populations.

Reduction in the rate of adherence in the first 6 hours was readily observed microscopically (Figure 4B). The trophozoites that adhered in the presence of HSF were also semi-motile on the flask wall compared to trophozoites in controls, which maintained their position after settling and adhering (Supplementary Media 1).

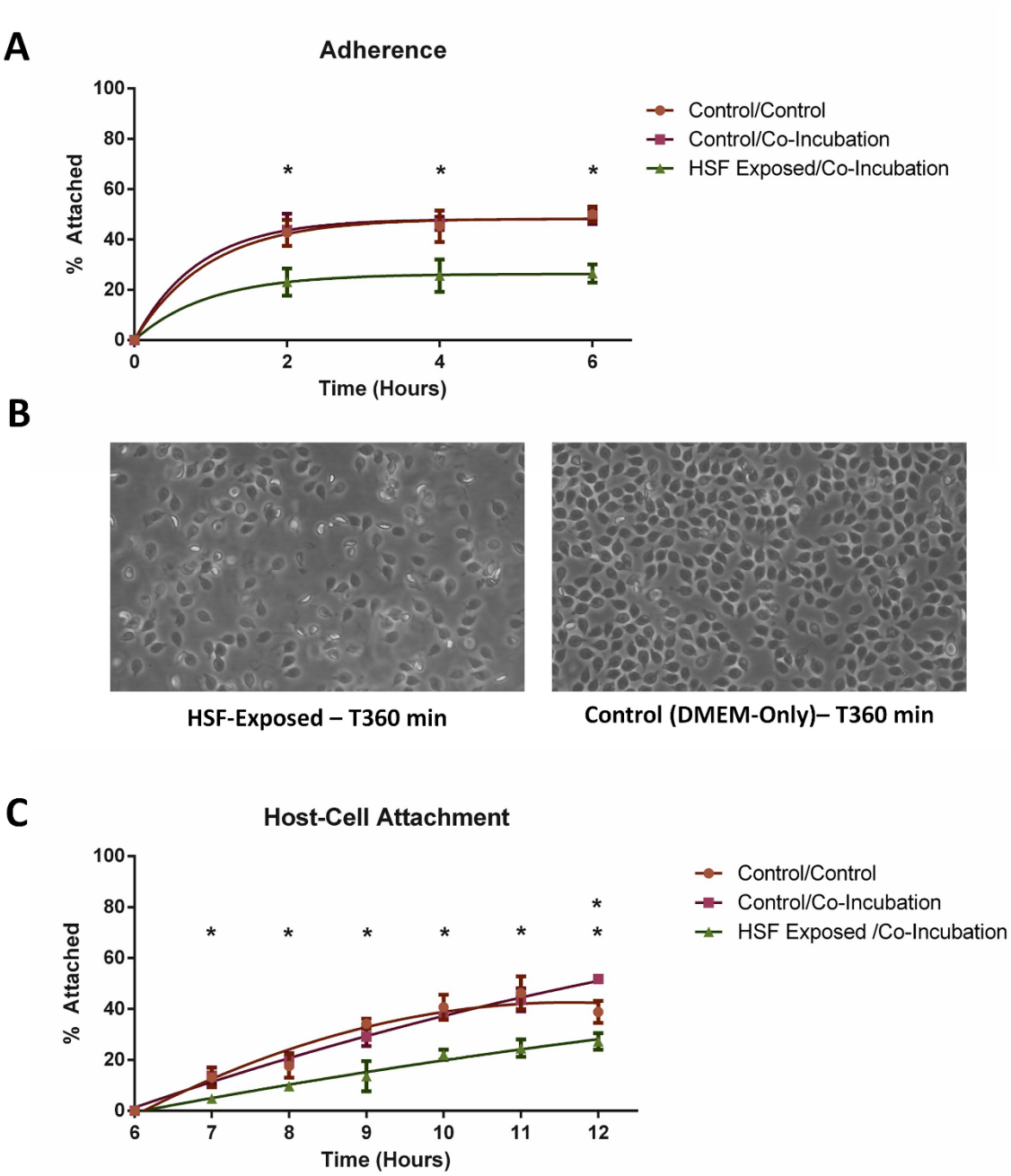


Figure 4: Results for the effects of HSF on adherence and host-cell attachment during co-incubation. The 3 treatments are as follows: trophozoites in serum-free DMEM in the first round followed by co-incubation with HT-29 (Con/CI), trophozoites exposed to HSF in serum-free DMEM and then co-incubated with HT-29 in the presence of HSF in the second round (HSF/CI) and a control of serum free DMEM in both rounds of the assay (Con/Con). Trophozoites were transferred from the first round of adherence to the second round of co-incubation. (A) Rates for adherence during the first 6 hours between trophozoites co-incubated in T75 flasks containing serum-free DMEM (Con/CI, Con/Con) and trophozoites incubated in the presence of HSF (HSF/CI). A '*' indicates a significant difference in % attached trophozoites (designated by p-value ≥ 0.05). Rates of adherence in HSF-exposed trophozoites were statistically significantly lower at all 3 timepoints compared to unexposed trophozoites. (B) Images depicting the differences in density of adhered trophozoites in HSF-exposed and HSF-free flasks after 6 hours of incubation. Flasks shown were seeded with the same number of trophozoites. (C) Rates of host-cell attachment in the second 6 hours of the assay. Rates of host-cell attachment between trophozoites exposed to HT-29 monolayers (Con/CI) are compared to trophozoites incubated with HT-29 monolayers in the presence of HSF (HSF/CI). A control for adherence was also performed in triplicate in serum DMEM without HT-29 cells. A '*' indicates a significant difference in % attached trophozoites compared to control (**cont).**

(cont.) (designated by p-value ≥ 5). Rates of attachment in HSF-exposed cells was significantly lower at all time points between both the control for adherence, and the rate for host-cell attachment in unexposed trophozoites. The number of attached trophozoites co-incubated with HT-29 monolayers without HSF was statistically significantly higher compared to the control for trophozoite adherence in flasks only after 6 hours incubation.

This indicates a reduced proclivity for adherence, further supported by observations during detachment from flasks, with controls containing only DMEM requiring twice the chilling time on ice as well as additional vortexing to liberate trophozoites before the second round of the assay. Trophozoites exposed to HSF showed no reduction in viability and remained motile throughout the entire assay.

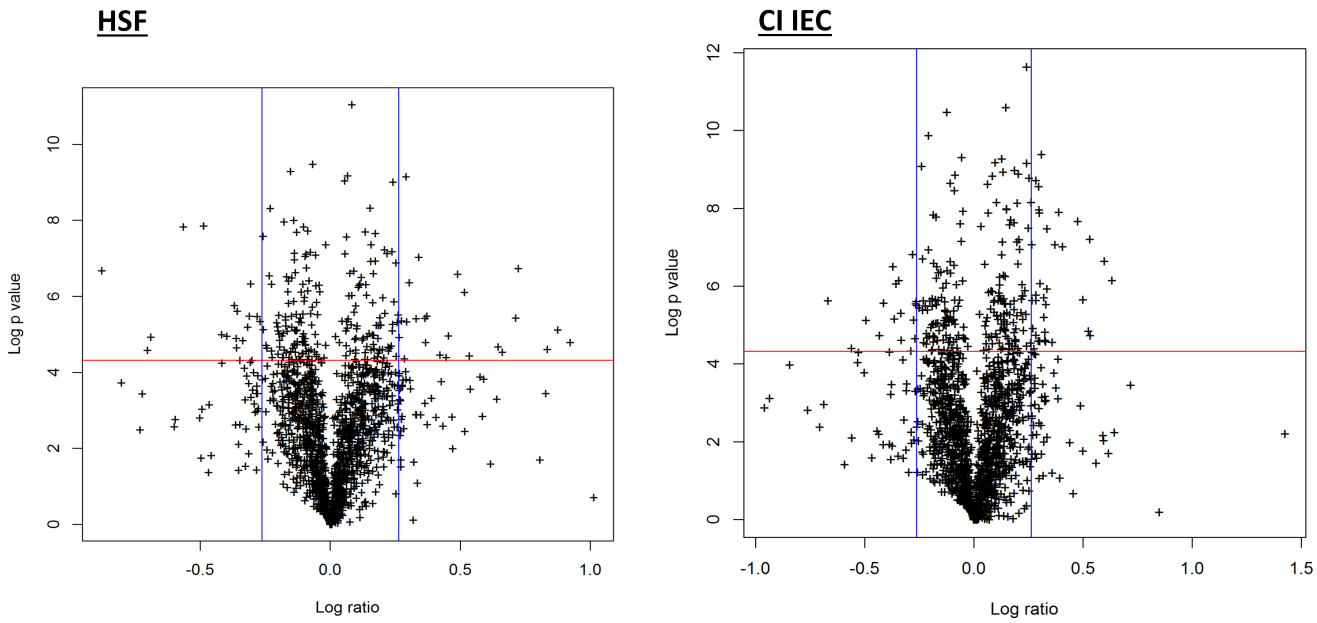
5.7.3 Quantitative proteomics

The complete TMT dataset, including protein identifications, label ratios and peptide information can be viewed in Supplementary Data S2. A non-redundant total of 1664 proteins from *G. duodenalis* isolate BRIS/95/HEPU/2041 were identified from a non-redundant total of 13 465 peptides (Figure 5A). Peptide to spectrum matching was performed using the Assemblage A1 genome sequence of isolate WB C6 (ATCC 50803). Previous quantitative proteomic analyses have demonstrated no significant difference in peptide numbers identified in subassemblage A1 isolates when using the WB C6 genome sequence as a database [27]. Similarly, recent comparative genomics analysis demonstrated that the A1 subassemblage is more conserved than the A2 subassemblage, with approximately 7.5 single nucleotide polymorphisms (SNPs) per 100 000 between genome sequences of two A1 isolates [28], compared to 350 SNPs per 100 000 between two newly sequenced A2 isolates [29]. Both genomic and proteomic data therefore indicates the A1 subassemblage is sufficiently conserved for the WB C6 genome sequence to provide a reference database for quantitative proteomics experiments within this taxonomic group.

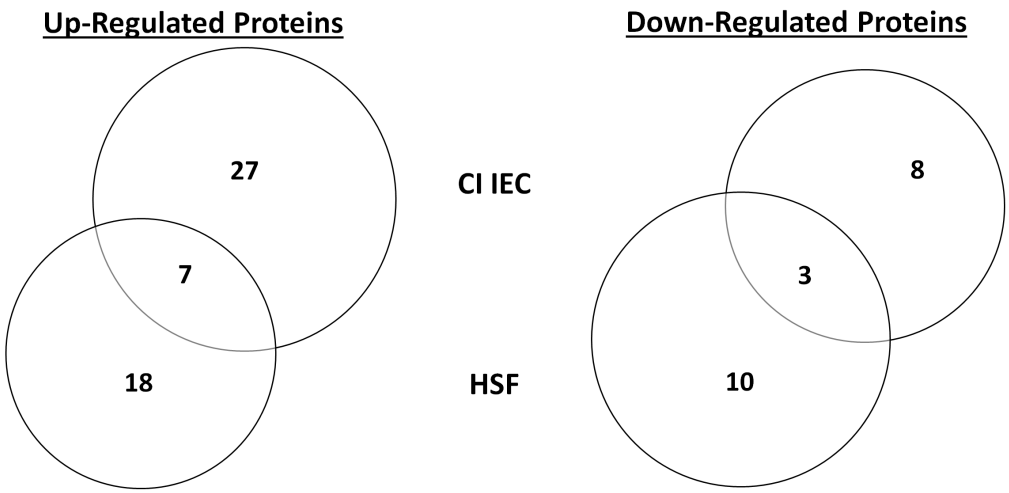
A

Dataset Summary	
Proteins Identified	1664
Non-redundant peptides	13 465
Proteins Up-regulated *(Ratio ≥ 1.2, p-value ≤ 0.05)*	34 (IEC CI), 25 (HSF)
Proteins Down-regulated *(Ratio ≤ 0.8, p-value ≤ 0.05)*	11 (IEC CI), 13 (HSF)

B



C



D

Figure 5: Protein identification and protein quantitation summary from TMT labelling of trophozoites co-incubated with the IEC monolayer (CI IEC) and with host soluble factors alone (HSF). (A) Outline of protein identification, differentially expressed proteins and protein quantitation FDR for the dataset. (B) Volcano plots illustrating the dual criteria for differentially expressed proteins. The x-axis represents log fold change with the vertical blue lines indicating 1.2 and 0.8 ratio, while the log p-value is plotted on the y-axis with proteins above the red horizontal line indicating significance ≤ 0.05. Each data point represents a single identified protein. Proteins within the upper and outer quadrants meet both the fold change and p-value cut-off, and are therefore considered as differentially expressed. (C) Proportional venn diagrams showing overlap between up-regulated and down-regulated proteins in trophozoites between CI IEC and HSF treatments.

Quantitation ratios were calculated as biological triplicate values from host-parasite interaction replicates relative to their respective control replicates (Figure 1C). Proteins were considered differentially expressed above a fold change of 1.2 and below a fold change of 0.8 in addition to a significant p-value of ≥ 0.05 [30] (Figure 5B). Using a two-stage criterion for differentially expressed proteins greatly improved the statistical confidence, as single-paired t-tests eliminated proteins with high ratio variability between replicates.

Statistical evaluation of the dataset by comparing individual control replicates against the pooled control indicated that levels of variability in the entire dataset were very low (Figure 6), which is consistent of labelled experiments such as iTRAQ and TMT [31]. Approximately 95% of the proteins have a standard deviation below 0.2, consequently in such cases a fold change of 1.2 ($\log_2(1.2) = 0.26$) would correspond to a z-score greater than 1 (Figure 6). The standard deviations are, naturally, considerably lower than those determined in the evaluation of Song *et al* [25], since the experiments fits into a single TMT 10-plex labelling experiment.

The Principal Component Analysis (PCA) also indicated that the control replicates were both clustered together, and were clearly discriminated from HSF and CI treatments along the first principal component (Figure 7). When considering the top 5% proteins with the highest loadings for PC1, we find amongst them several of the differentially expressed proteins identified in the paired analyses of HSF vs Con and CI vs Con (GL50803_3910, GL50803_27918, GL50803_13390, GL50803_10358, GL50803_6430, GL50803_14567, GL50803_9779, GL50803_17163, GL50803_42357). The principal component scores and loadings from this analysis can be found in Supplementary Data S3.

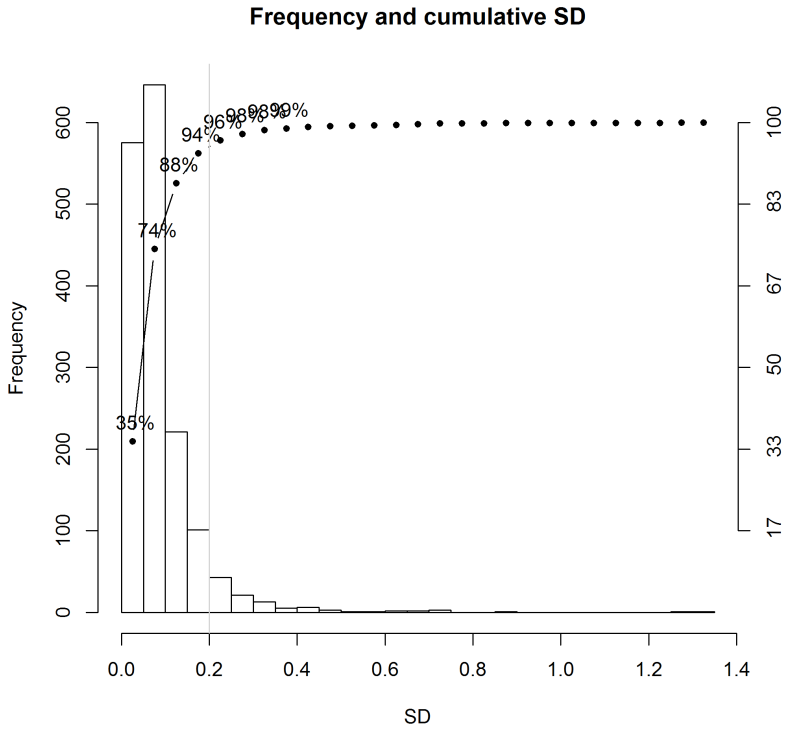


Figure 6: Histogram of standard deviations generated from triplicate ratios of $\log_2$(Control/PooledControl), based on all quantitated proteins in the dataset. Overlayed are the cumulative percentages of ratios with standard deviations lower than the respective bin. Approximately 95% of quantitated proteins have standard deviations less than 0.2.

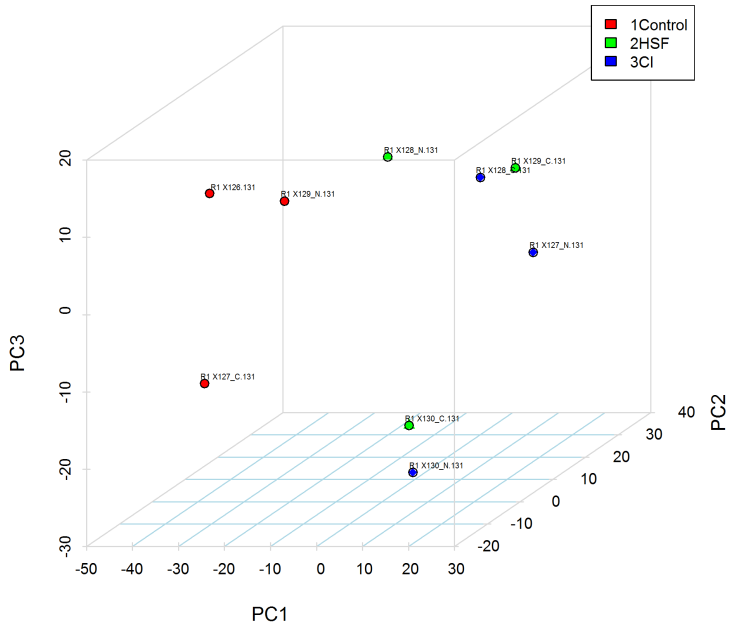


Figure 7: Principal component scores plot in the space of the first three principal components generated for the whole dataset of $\log_2$ ratios of all samples with respect to the pooled control (label 131). The plot shows the control ratios well separated from the rest along the first principal component.

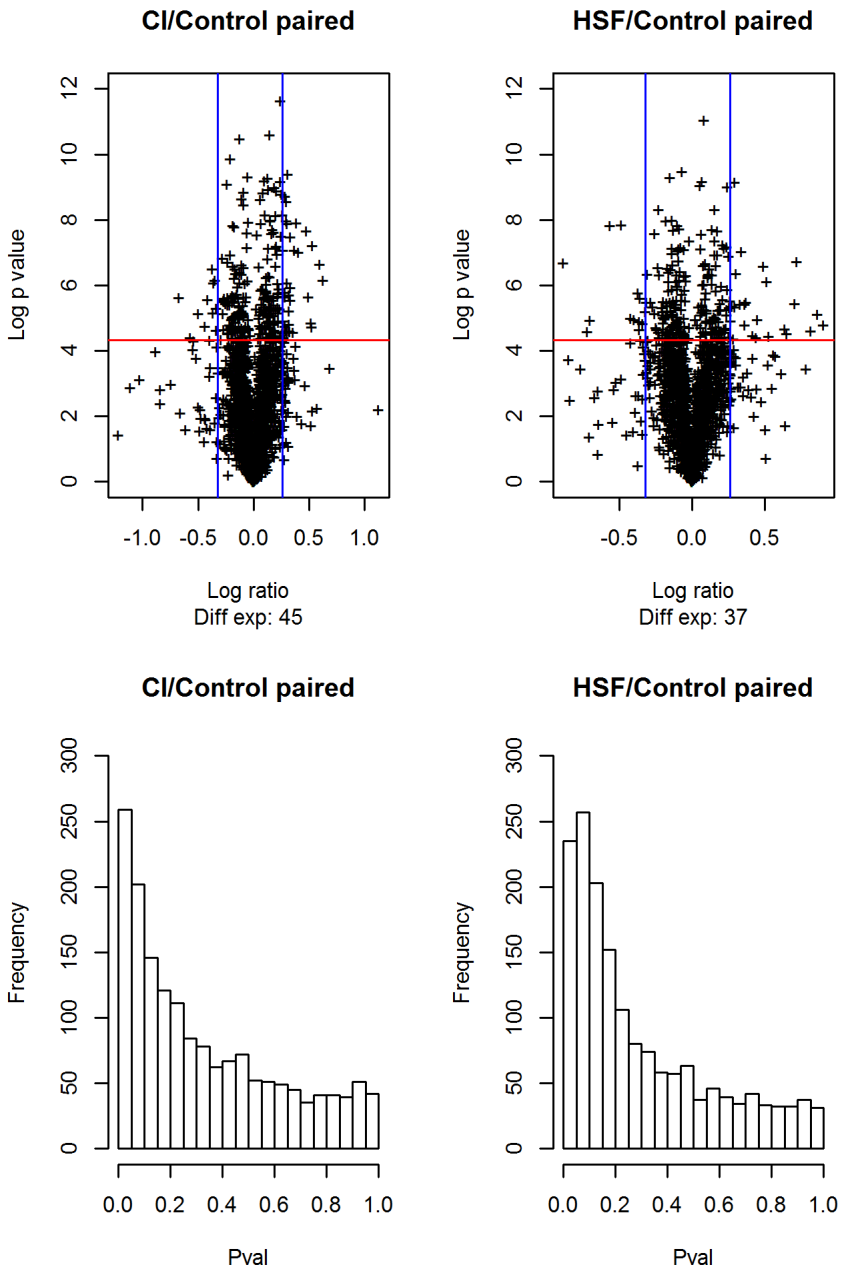


Figure 8: Histograms showing the distribution of p-values resulting from the paired t-tests comparing the Control samples respectively to the CI and HSF samples, underneath the volcano plots showing the p-values and fold changes. The p-value histograms have a peak corresponding to a larger number of low p-values, which is indicative of a real underlying effect; a random or noisy dataset is expected to generate a uniform distribution of p-values and hence a flat histogram.

In addition, the distribution of the p-values indicates an underlying signal of differential protein expression when comparing the HSF and CI treatments to the control replicates (Figure 8). Pounds *et al* [26] shows that understanding the distribution of p-values is crucial to understanding whether the null hypothesis holds – in the case of a repeated test undertaken on an essentially

random distribution, the p-value histogram is expected to be essentially flat, and the p-values are expected to be uniformly distributed; in contrast, where a real effect exists, the histogram will show a peak at the low end corresponding to lower p-values arising from real effects. In the case of this experiment, the p-value histograms obtained when comparing the two experimental states against the control are consistent with the existence of an underlying real effect.

A total of 68 differentially expressed proteins were identified in *Giardia* trophozoites during host-parasite interactions (Figure 5A), with 45 proteins differentially expressed in trophozoites co-incubated with IECs (Table 1), and 38 proteins in HSF-exposed trophozoites (Table 2). These included several up-regulated proteins previously identified the transcript level during co-incubation of *Giardia* trophozoites with IECs [4, 5], including tenascin proteins, cathepsin B precursor, uridine phosphorylase 1 (UPL-1) and thioredoxin. A total of 30 variant-specific surface proteins (VSPs) were also identified in the whole dataset, with eight and four variants up-regulated in HSF and CI treatments, respectively (Table 3). This up-regulation of a large number of VSP variants in HSF-exposed trophozoites has not been previously reported for *in vitro* host-parasite models in *Giardia*. The *G. duodenalis* VSP gene family contains both conserved, homologous regions of gene sequence as well as unique regions [32]. Of the eight differentially expressed VSP variants in HSF-exposed trophozoites, five of the protein-level identifications are from non-homologous peptides. Though the remaining three VSP variants have homologous peptides associated with the identification, multiple peptides have been detected for these identifications, and protein-level identification was assigned based on the most likely candidate given the total and composition of peptides matched.

Gene ID	Description	Fold Change
GL50803_16693 *	hypothetical protein	1.55
GL50803_113415	hypothetical protein	1.51
GL50803_27918 *	Cystatin homologue	1.45
GL50803_16188	SMC1 beta-like protein	1.45
GL50803_96264	hypothetical protein	1.44
GL50803_10529	Bardet-Biedl syndrome 4 protein-like protein	1.42
GL50803_3345	hypothetical protein	1.39
GL50803_13390 *	Variant-Specific Surface Protein 127 (VSP-127)	1.33
GL50803_480 *	Translation initiation inhibitor	1.31
GL50803_14045	hypothetical protein	1.29
GL50803_5810	hypothetical protein	1.28
GL50803_40630	Variant-Specific Surface Protein 70 (VSP-70)	1.26
GL50803_3171	UBCE14	1.26
GL50803_2267	hypothetical protein	1.26
GL50803_8329 *	hypothetical protein (VPS25(EAP20))	1.26
GL50803_14567	hypothetical protein	1.25
GL50803_3042	Hybrid cluster protein lateral transfer candidate	1.25
GL50803_9506	hypothetical protein	1.25
GL50803_9779	UPL-1	1.25
GL50803_114609	Pyruvate-flavodoxin oxidoreductase	1.24
GL50803_13194	Variant-Specific Surface Protein 38 (VSP-38)	1.24
GL50803_15039	hypothetical protein	1.23
GL50803_7110 *	Ubiquitin	1.23
GL50803_10358	A-type flavoprotein lateral transfer candidate	1.23
GL50803_29078	hypothetical protein	1.23
GL50803_114674	Hypothetical protein	1.23
GL50803_12941	hypothetical protein	1.22
GL50803_6430 *	14-3-3 protein	1.22
GL50803_15252	Ubiquitin-conjugating enzyme E2-17 kDa 3	1.22
GL50803_17163	Peptidyl-prolyl cis-trans isomerase B precursor	1.22
GL50803_4946	Peptide methionine sulfoxide reductase msrA	1.21
GL50803_15089	hypothetical protein	1.21
GL50803_11043	Fructose-bisphosphate aldolase	1.20
GL50803_14392	Metallo-beta-lactamase superfamily protein	1.20
GL50803_17062	hypothetical protein	0.79
GL50803_7593	hypothetical protein	0.79
GL50803_17375	hypothetical protein	0.77
GL50803_10875	hypothetical protein	0.78
GL50803_15137	hypothetical protein	0.68
GL50803_14457	Ripening regulated protein DDTFR19	0.79
GL50803_11311 *	Kinase, NEK	0.78
GL50803_12229 *	hypothetical protein	0.74
GL50803_5785	Qb-SNARE 4	0.75
GL50803_10522 *	hypothetical protein	0.71
GL50803_5881	Protein 21.1	0.63

Table 1: Differentially expressed proteins in *G. duodenalis* trophozoites co-incubated with HT-29 cells (IEC CI) for 6hr in serum-free media. Up-regulated proteins are designated by a ratio ≥ 1.2 also accompanied by a p-value ≤ 0.05. Down-regulated proteins are indicated with grey shading, and were designated based on a ratio of ≤ 0.8 which was accompanied by a p-value ≤ 0.05. Gene identifiers marked with a '*' indicate a protein that was common between trophozoites co-incubated with the IEC monolayer and trophozoites incubated with HSF.

Gene ID	Description	Fold Change
GL50803_16693 *	hypothetical protein	1.90
GL50803_137618	Variant-Specific Surface Protein 8 (VSP-8)	1.83
GL50803_14069	hypothetical protein	1.79
GL50803_41472	Variant-Specific Surface Protein 49 (VSP-49)	1.65
GL50803_8687	Tenascin precursor	1.64
GL50803_16779	Cathepsin B precursor	1.58
GL50803_3910	hypothetical protein	1.57
GL50803_14573	Tenascin-X	1.45
GL50803_3581	hypothetical protein	1.43
GL50803_101832	High cysteine protein	1.41
GL50803_27918 *	Cystatin homologue	1.37
GL50803_115830	Variant-Specific Surface Protein 1.1 (VSP-1.1)	1.36
GL50803_98861	Surface protein	1.34
GL50803_112208	Variant-Specific Surface Protein 98.1 (VSP-98.1)	1.29
GL50803_13390 *	Variant-Specific Surface Protein 127 (VSP-127)	1.29
GL50803_112867	Variant-Specific Surface Protein 16 (VSP-16)	1.29
GL50803_14278	hypothetical protein	1.27
GL50803_17340	hypothetical protein	1.26
GL50803_6430 *	14-3-3 protein	1.24
GL50803_7110 *	Ubiquitin	1.23
GL50803_480 *	Translation initiation inhibitor	1.22
GL50803_8329 *	hypothetical protein (VPS25(EAP20))	1.22
GL50803_3755	hypothetical protein	1.21
GL50803_2012	hypothetical protein	1.20
GL50803_32890	Variant-Specific Surface Protein 10 (VSP-10)	1.20
GL50803_16070	hypothetical protein	0.79
GL50803_113876	ABC transporter, ATP-binding protein, putative	0.79
GL50803_42357	hypothetical protein	0.78
GL50803_14539	Protein 21.1	0.78
GL50803_11311 *	Kinase, NEK	0.78
GL50803_14569	hypothetical protein	0.78
GL50803_7513	hypothetical protein	0.76
GL50803_12229 *	hypothetical protein	0.75
GL50803_22291	hypothetical protein	0.71
GL50803_10522 *	hypothetical protein	0.68
GL50803_12109	hypothetical protein	0.62
GL50803_17404	hypothetical protein	0.61
GL50803_16587	Kinase	0.54

Table 2: Differentially expressed proteins in *G. duodenalis* trophozoites incubated with host soluble factors generated from HT-29 cells for 6hr in serum-free media. Up-regulated proteins are designated by a ratio ≥ 1.2 also accompanied by a p-value ≤ 0.05. Down-regulated proteins are indicated with grey shading, and were designated based on a ratio of ≤ 0.8 which was accompanied by a p-value ≤ 0.05. Gene identifiers marked with a '*' indicate a protein that was common between trophozoites co-incubated with the IEC monolayer and trophozoites incubated with HSF.

Gene ORF	Description	Fold Change		Addition Protein Domains
		6hr IEC CI	6hr HSF	
137618	VSP 8	**1.65**	**1.83**	
13194	VSP 38	**1.24**	**1.78**	
41472	VSP 49	**1.53**	**1.65**	
16472	VSP 52	0.91	**1.53**	
136003	VSP 7.1	0.77	**1.51**	
113163	VSP 29	1.18	**1.50**	
115830	VSP 1.1	**1.21**	**1.36**	
112208	VSP 98.1	**1.22**	**1.29**	
13390	VSP 127	**1.33**	**1.29**	
112867	VSP 16	1.02	**1.29**	BmKX domain (IPR015215)
115742	VSP 31	**1.41**	**1.26**	
34196	VSP 193	1.07	**1.21**	
32890	VSP 10	1.31	**1.20**	Peptidase M8, leishmanolysin (IPR001577)
40630	VSP 70	**1.26**	1.19	Peptidase M8, leishmanolysin (IPR001577)
37093	VSP 25	1.19	1.16	
13727	VSP 183	1.13	1.15	
113439	VSP 45	1.13	1.13	
101074	VSP 88	1.11	1.10	
11521	VSP 126.1	1.16	1.10	
113450	VSP 44	1.01	1.10	
98861	Surface protein	1.13	**1.34**	Peptidase M8, leishmanolysin (IPR001577)
119706	VSP 168.2	1.14	1.04	
16158	VSP, putative	1.04	1.01	
137723	VSP 26.1	**1.25**	1.01	
137714	VSP 53.2	0.99	0.97	
115797	VSP 54	0.99	0.95	
101765	VSP 116	1.00	0.90	
33279	VSP 100	0.92	0.85	
113357	VSP 122	0.94	0.84	
114674	Hypothetical protein	**1.23**	1.18	

Table 3: Variant surface proteins (VSPs) identified across the TMT experiment, including their gene identifier, descriptor and fold change. TMT ratios above the 1.2 threshold have been bolded, while a shaded cell indicates a p-value ≥ 0.05. Only VSPs that were above the threshold for ratio and below the p-value for significance were considered differentially expressed. Interpro domains other than the 4 consistent VSP protein domains (Giardia variant-specific surface protein (IPR005127), Insulin-like growth factor binding protein, n-terminal (IPR009030), EGF-like (IPR000742) and Furin-like repeat (IPR006212) domains) were considered for additional functional protein information, and listed where appropriate.

Overall, a non-redundant total of 47 proteins were up-regulated, while 21 were down-regulated during *in vitro* host-parasite interactions. This smaller down-regulated protein response suggests the response of *Giardia* trophozoites to host signals *in vitro* may be inductive rather than repressive. The dataset size of differentially expressed proteins is similar to those seen previously with HT-29 cells over 6 hours incubation in RNA studies [5]. Although some common differentially expressed proteins were observed, the majority of proteins were expressed uniquely between trophozoites stimulated by host secretions compared to attachment (Figure 5C). The process of washing IEC-monolayers prior to co-incubation, as well as introducing *Giardia* trophozoites to fresh, serum-free interaction media, ensured secreted factors from host IECs were removed at the beginning of co-incubation, and did not establish to sufficient levels during the 6 hour time course. Therefore, there was not the same induction of responses observed in trophozoites incubated solely in the presence of pre-established levels of HSF, and only minimal overlap between the treatments. Expression intensity also varied, with HSF driving a higher fold change than those in trophozoites interacting with the IEC monolayer (Figure 5B). Lower fold changes in co-incubation are likely due to the few hours delay in host-attachment by trophozoites, as evidenced during the co-incubation attachment assay (Figure 3A). These results indicate during *in vitro* interactions *Giardia* trophozoites respond at the very early stages to host secretions prior to attachment. These soluble signals from the host stimulate distinct and independent pathways for establishing disease.

5.7.4 Host secretions induce virulence factors in a motile population

Trophozoites incubated in HSF up-regulated production of 25 proteins (Table 2), the majority of which were membrane-associated or secreted. A total of 11 up-regulated proteins contained evidence of secretion or transmembrane helices in one or more predictive algorithms (Figure 9). These 11 proteins consisted of VSP variants, tenascins, cathepsin B and a high cysteine

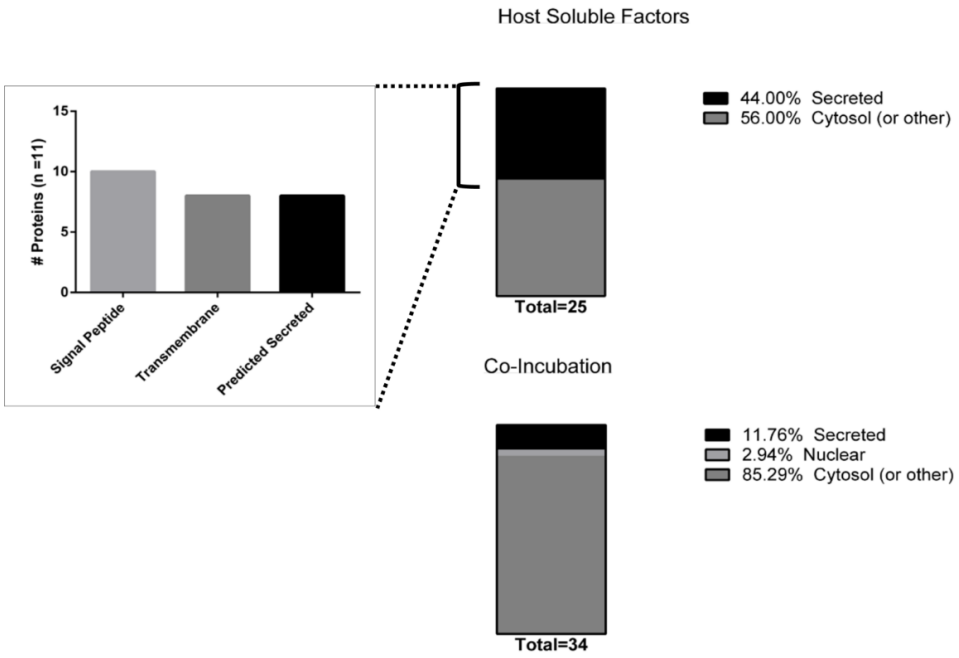


Figure 9: Bioinformatic predictions of membrane and secreted protein analysis for up-regulated proteins. Distribution of secreted proteins in trophozoites incubated in HSF and trophozoites co-incubated with HT-29 IECs. Proteins were considered exported to the membrane and/or secreted if they were positive for 1 of the 3 bioinformatics tools (TargetP, THMHH, SignalP) and negative for nuclear localisation (NucPred). The breakdown of the 11 proteins in HSF incubated trophozoites is further broken down in the column graph insert in the right. Complete summary of the predictive bioinformatics analysis can be viewed in Supplementary Data S4.

membrane protein (HCMP) (Supplementary Data S5). A total of 8 VSPs were up-regulated in the 6 hours, which constitute 32% of all up-regulated proteins in trophozoites incubated in HSF, and 26.7% of all VSP variants identified. While multiple HCMP variants have been previously reported during *in vitro* host-parasite interactions at the transcript level [4, 5], the up-regulation of multiple VSPs has not been previously observed in RNA studies. There were 4 VSP variants identified in trophozoites co-incubated with IECs, but lower fold changes in these VSPs suggest that this is likely due to partial re-establishment of host-soluble signals throughout the experimental timecourse. Five of the 8 VSPs were predicted as secretory as well as possessing transmembrane helices, while 2 lacked prediction for secretion but contained transmembrane helices, and one met no predicted secreted or membrane criteria (GL50803_98861). Expression of both tenascin precursor (GL50803_8687) and tenascin X (GL50803_14573) was increased in trophozoites exposed to HSF, which is consistent with

previous observations of increased expression in tenascin gene transcripts during *in vitro* host-parasite interactions [5]. Tenascin proteins share similar domains to the membrane-associated VSPs, but lack transmembrane helices (Supplementary Data S4 and S5). Tenascin X also possesses an additional EGF-like extracellular domain (IPR013111). In light of functional and predictive bioinformatics, it is highly likely that these proteins are secreted. Exposure of trophozoites to HSF also resulted in up-regulation of the cysteine protease cathepsin B (GL50803_16779), which is consistent with previous RNA [4, 5]. The *G. duodenalis* cystatin homologue [33] was up-regulated in trophozoites exposed to HSF and during co-incubation, indicating sensitivity to multiple host signals. This cystatin (GL50803_27918) is currently unannotated, with a single ortholog in all *G. duodenalis* genomes, and lacks GO and Interpro annotations whilst displaying low sequence homology to other parasite and eukaryotic cystatins. However, sequence analysis indicates conserved crucial residues, including a glycine in the N-terminal region, and hydrophobic residues in both first and second binding loops which forms the wedge that inserts and deactivates the active site of cathepsin proteases [34]. A putative thioredoxin (GL50803_3910) was also up-regulated, consistent with potential oxygen stress in cell culture conditions, and data from previous studies [4, 5]. The single *Giardia* 14-3-3 homologue (GL50803_6430) was also up-regulated, which is involved in binding to signal proteins involved in phosphorylation cascades [35]. There were five up-regulated proteins which were only annotated as 'hypothetical proteins' (GL50803_16693, GL50803_14278, GL50803_17340, GL50803_3755, GL50803_2012), and for which there are no current GO annotations or Interpro protein structure and fold information. Interestingly, none of these proteins have been previously reported as differentially expressed in transcriptomic *G. duodenalis in vitro* host-parasite interaction models

5.7.5 Host-cell attachment prompts intracellular anticipation of host defences

A total of 34 proteins were up-regulated in trophozoites co-incubated with the IEC monolayer (Table 1). In contrast to trophozoites incubated with soluble host signals, there were only four secreted and membrane proteins, all of which were VSPs, indicating no new secreted or membrane protein classes were detected (Supplementary Data S5). A single protein (GL50803_16188), with annotations associated with chromosome organisation, was localised to the nucleus, meaning 29 (85%) of the up-regulated proteins are likely to be localised to the cytosol (Figure 9). Of these up-regulated proteins, five were annotated with oxidoreductase functions (GL50803_5810, GL50803_3042, GL50803_114609, GL50803_10358, GL50803_4946), which include functions for oxidative defence and reactive oxygen species (ROS) detoxification, as well as production of pyridoxal phosphate (PLP). These five proteins constitute 9.8% of all annotated oxidoreductases in the *G. duodenalis* A1 genome. GL50803_5810 is currently unannotated but is most likely to be the *Giardia* homolog for pyridoxamine-phosphate oxidase, the enzyme responsible for the rate-limiting reaction in the production of the active form of coenzyme vitamin B_6, due to the singular presence of these functional protein domains in the genome. Concordantly, GL50803_480 was also upregulated, which has gene ontology (GO) and Interpro annotations associated with deaminase activity in converting reactive enamine/imine intermediates in PLP-dependent enzyme reactions. Furthermore, GL50803_29708 is a hypothetical protein with annotations for pyridoxal-phosphate dependent aminotransferase. Additionally, several proteins in the ubiquination pathway were also up-regulated, including ubiquitin (GL50803_7110), an ubiquitin carrier enzyme (UBCE) (GL50803_3171) and ubiquitin-conjugating enzyme E2-17 (GL50803_15252). The cystatin homologue (Gl50802_27918) was up-regulated, as also observed during exposure to host soluble factors. UPL- 1was also up-regulated, which has been previously been reported in trophozoites co-incubated with HT-29 cells [5]. Again, the *Giardia* 14-3-3 homologue (GL50803_6430) was up-regulated as seen in HSF-exposed trophozoites,

along with another similar tetracopeptide (TPR) containing protein (GL50803_10529). There were seven hypothetical proteins that lacked GO and Interpro functional and structural information (GL50803_16693, GL50803_113415, GL50803_3345, GL50803_2267, GL50803_14567, GL50803_9506, GL50803_15039), none of which have been previously reported in earlier transcriptomic studies of *in vitro* host-parasite interaction models for *G. duodenalis.*

5.8 Discussion

In vitro host-parasite models in *G. duodenalis* are designed to replicate trophozoite attachment to host cells and thereby induce expression of virulence and disease factors. Previous studies have demonstrated co-incubation induces secreted proteins [3] and expression of gene transcripts demonstrably different from constitutive expression in culture [4, 5]. We have shown co-incubation with cell monolayers permits attachment of trophozoites to host cells, which is significantly different from trophozoite adherence to flasks alone (Figure 10). Further, our results demonstrate that *Giardia* induces proteins upon exposure to either host cell products, or upon attachment to host cells, and the proteins expressed in these responses are independent and distinct (Figure 10). We have also demonstrated a unique sensitivity in trophozoites to secreted products from host cells, and that HSF-exposed trophozoites switch to a non-attaching motile phenotype (Figure 10). Significantly, exposure to these host secretions was sufficient for induction of virulence factors in trophozoites. To our knowledge, this is the first demonstration of an interactive, biphasic process during early pathogenesis which shows a clear difference between motile trophozoites responding to host soluble signals, and trophozoites attached to host-cells.

We have provided experimental evidence that exposure to HSF results in up-regulation of membrane and secreted proteins prior to attachment in *Giardia* trophozoites (Figure 10). Several of these induced proteins are also virulence factors, and we have demonstrated their sensitivity to host secretions in the absence of host cells. VSPs are a gene family and collective virulence factor responsible for immune evasion and antigen variation [32]. A single VSP is expressed on the trophozoite surface at any given time, though multiple variants accumulate in culture in the absence of immune selection. On exposure to HSF a quarter of all expressed VSP variants were up-regulated, which constituted over a third of observed up-regulated proteins.

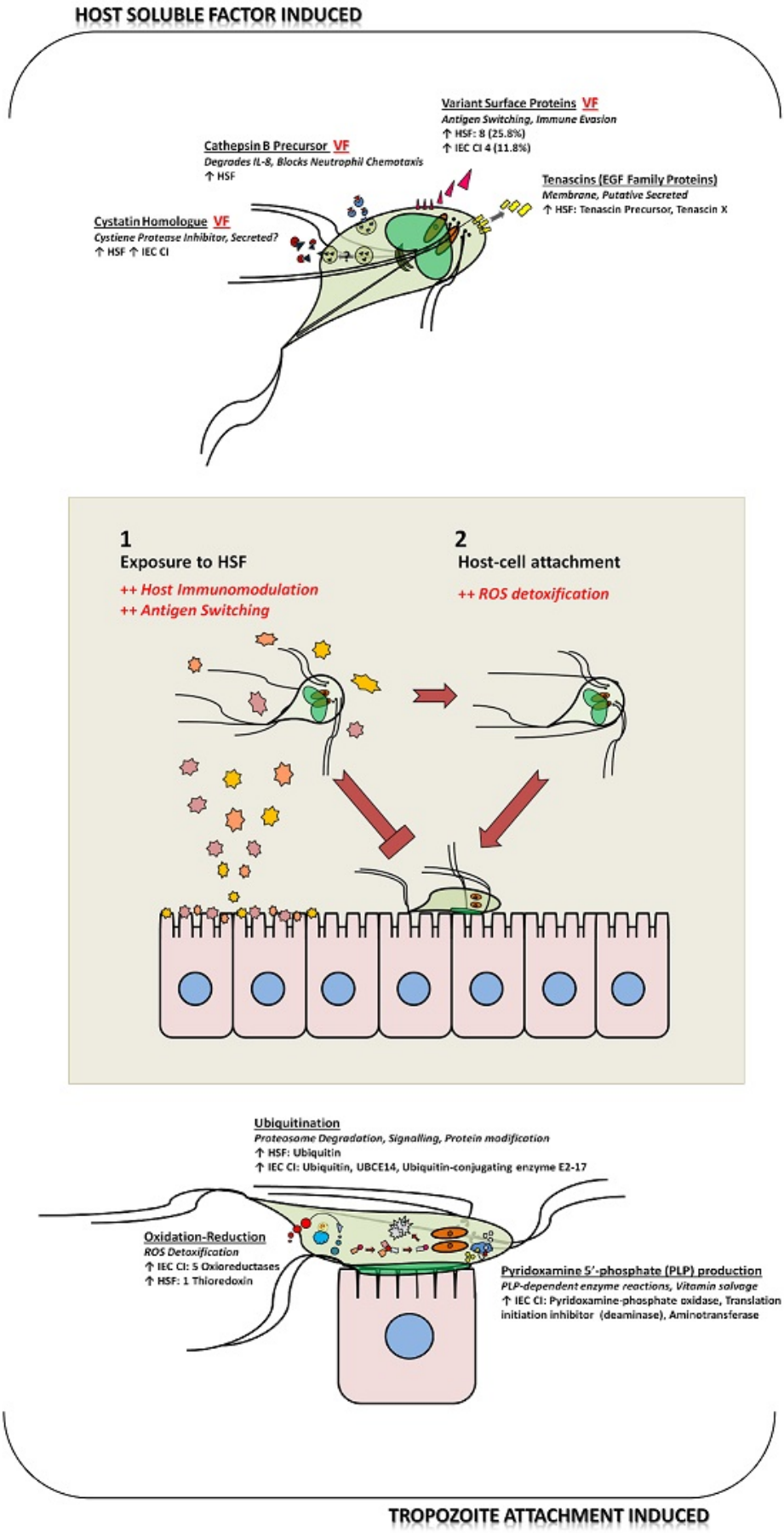


Figure 10: Figure depicting the biphasic model of interaction between *G. duodenalis* trophozoites **(cont.)**

(cont.) and host-cells proposed in our paper. Proteins trends, families and pathways induced during host-parasite interactions in *Giardia* are distinguished between HSF-induced in non-attached, motile trophozoites (above) separate to cascades induced in host-cell attached trophozoites (below). 'VF' indicates induced protein groups related to known or putative virulence factors in *Giardia*, which were induced by host secretions. The middle of the figure shows the two distinct stages observed in early pathogenesis, where host-soluble factors lead to a switch to a non-attaching, motile population phenotype. Motile trophozoites migrate further through the gastrointestinal tract, where in the absence of host soluble factors and more optimum conditions, *Giardia* attaches to host cells. These two stages of host-parasite interactions induce distinct and independently protein responses.

These changes in VSP expression likely represent selection of favourable variants for host pathogenesis or virulence. Interestingly, four up-regulated VSPs possessed additional functional domains to the core VSP protein domains (Table 3). These included the metallopeptidase domain homologous to the virulence factor leishmanolysin from the parasite *Leishmania* [36, 37], and the BmKX domain found in scorpion toxins [38]. This suggests individual VSP variants may act as independent virulence factors beyond their role collectively in immune evasion [39]. The magnitude of differential expression in VSPs has not been reported previously in *Giardia* host-parasite interactions. This may be due to the absence of host secretions in interaction media, or possibly the virulent phenotype of BRIS/95/HEPU/2041 [40, 41], which features diverse VSP variant repertoires in culture [12, 27]. In addition to previous studies showing VSP switching elicited by specific monoclonal antibodies [42, 43], we present here experimental evidence suggesting soluble host factors may drive antigen switching events.

Cathepsin B is a confirmed virulence factor and secreted protein which degrades IL-8 and inhibits neutrophil chemotaxis [44, 45]. Exposure to secreted host factors up-regulated cathepsin B as well as cystatin, a protease inhibitor of cathepsins. When secreted, cystatins potentially modulate host immune response, as observed in parasitic nematodes [46, 47], but may also internally regulate parasite cathepsins [48]. The *Giardia* cystatin is phylogenetically basal to eukaryotic cystatins and more closely related to bacterial equivalents [33, 37], which makes it difficult to extrapolate its internal or external targets [47].

Tenascins share similar 'EGF-like' (IPR000742) domains to VSPs, and are glycoproteins involved in cell-to-cell adhesion in mammals and chordates, with unknown functions in parasites and early eukaryotes [49, 50]. Interestingly, research in mammals demonstrates that tenascins bind and interact with lectin domains [51, 52], which may have implications for parasite interactions with molecules involved in mammalian innate immunity. Significantly, both transcript and protein data reproducibly identify tenascins in host-*Giardia* interactions, and our bioinformatics analysis demonstrated for the first time that these are possibly secreted proteins during early pathogenesis.

Trophozoites exposed to host secretions displayed a non-attaching phenotype (Figure 4) and a distinct protein response from co-incubated trophozoites (Figure 5). In *Giardia*, attachment occurs via its ventral disk through cytoskeletal mechanisms related to microtubules, but also via *Giardial* lectins [53]. Earlier attachment assays demonstrate several factors that decrease attachment, including temperature, acidification, osmolality and tonicity, incubation with lectins, exposure to lectin-binding carbohydrates and, most prominently, interference with contractile filaments [53, 54, 55]. None of these factors account for the reduction in attachment observed in HSF-exposed trophozoites during our experiments. Previous investigation of host factors is limited to trypsin, which produced minor reductions in attachment [53], or exogenous or host lectins, which are associated with trophozoite agglutination[55], which did not occur during our assay. Our results indicate a unique factor may be responsible for non-attachment in trophozoites exposed to host secretions. In addition, our results also support the hypothesis that host secretions promote an active switch to a motile population, potentially to continue migration through the gut to a less 'hostile' environment, as has been previously observed in gastrointestinal nematodes which relocate in response to localised host immune responses [56]. If one assumes that host secretions contain immune molecules involved in parasite clearance, such as cytokines [57], up-regulating VSPs, for immune evasion, and cathepsin B and cystatins,

for immunomodulation [39], are consistent with the hypothesis that host secretions induce proteins to counteract host immunity in a motile *Giardia* population seeking ideal conditions for attachment (Figure 10).

In contrast, trophozoites co-incubated with IECs induced intracellular pathways in anticipation of non-specific immune responses to infection. The lower intensity of observed protein expression fold changes coincides with the delay observed between adherence to control flasks and adherence to host cell monolayers (Figure 2A). Co-incubation induced up-regulation of proteins for ubiquitination, including two E2 carrier/conjugating enzymes as well as the ubiquitin moiety. Though ubiquitination pathway and enzymes are simpler in *Giardia* compared to mammals [58], it has been shown to play an important role during differentiation, with modifications on a diverse range of proteins [59]. In our results, an increase in ubiquitin-modified proteins is likely to indicate increased proteasome activity in response to rapid transition from axenic culture to host-parasite interaction.

Parasite co-incubation with host cells prompted up-regulation of a diverse range of oxidoreductases highlighting the importance of ROS detoxification in parasite establishment and survival. Trophozoites exposed to host-soluble signals up-regulated thioredoxin for oxidative defense in increased oxygen during cell-culture conditions, but not oxidoreductases. Co-incubation with host IECs in aerobic cell culture conditions provides some protection against oxidative stress in *Giardia*, but defence genes are still up-regulated at the transcriptomic level [4, 5]. As trophozoites incubated in cell-free interactions have greater environmental oxygen stress, this up-regulation of oxidoreductases is likely to be a specific response to interaction with host cells which are known to produce exogenous ROS in response to gastroepithelial parasites [60]. Similarly, production of reactive nitrogen species (RNS) negatively impacts *Giardia* growth, differentiation and viability. Therefore, trophozoites inhibit RNS production by outcompeting host cells for substrates by rapid arginine consumption [61], highlighting the

importance of this pathway in pathogenicity and virulence [7, 62]. Earlier transcriptomic studies have confirmed IECs co-incubated with *Giardia* up-regulate genes associated with ROS and RNS production [6, 7], and our results reinforce the necessity for *Giardia* to counteract these host defences. A flavo-diiron protein (GL50803_10358) with very low nitric oxide (NO) reductase activity but remarkably high O_2 detoxification activity, which has been shown experimentally to promote parasite survival in the small intestine [63], was up-regulated in our experiments. This was accompanied by a cumulative up-regulation of three other oxidoreductases with functions in removal of ROS, indicating that co-incubation either prompts, or trophozoite attachment anticipates, ROS production in host cells [6].

Pyruvate flavodoxin oxidoreductase (PFOR) converts pyruvate to acetyl coenzyme A during anaerobic energy production [64] but is also up-regulated during oxidative stress and involved in the antioxidant system [60, 65]. Similarly, peptide methionine sulphoxide reductase (Msr) (GL50803_4946) is likely to alleviate oxidative stress by reversing oxidation of critical methionine residues which might otherwise cause protein inactivation [66]. Following this trend, an iron-sulphur (FeS) containing hybrid cluster protein (GL50803_3042) was also up-regulated, which is a member of a protein class implicated in defence against oxidative stress in bacteria, Archaea and protozoans including *Trichomonas vaginalis*, and *Entamoeba histolytica* [67]. Collectively, these reinforce the necessity of maintaining redox homeostasis in the face of ROS produced by host defences, with multiple oxidoreductases induced by trophozoites soon after first contact with IECs. The remaining up regulated oxidoreductase is the *Giardial* pyridoxamine-phosphate oxidase homolog, which is the only protein annotated with this functional domain (GO:0004733) with a single orthologue in all sequenced *G. duodenalis* genomes.

Two PLP-associated enzymes were also up-regulated, suggesting PLP-regulated enzymes and pathways play a role in disease induction. The production of PLP, its biological role in *Giardia*,

and its antioxidant properties are all unknown, although PLP-dependent enzymes have been suggested as drug candidates for multiple protozoans [68]. Importantly, crucial PLP-dependent cysteine desulfurases in *Giardia* possess conserved PLP-binding residues and are involved in FeS cluster biosysnthesis [69]. Additionally, PLP-dependent enzymes are involved in polyamine biosynthesis in *Giardia*, and it has been demonstrated that specific inhibition of ornithine decarboxylase results in interruption of this biosynthesis and eventuates in parasite death [70]. PLP-dependent pathways and enzymes have been wholly unexplored in *Giardia* for therapeutics, and provide a potentially novel pathway for blocking disease induction.

In *Giardia*, analysis of pathogenesis has emphasised the importance of trophozoite attachment to host-cells. However, we have demonstrated for the first time that trophozoites respond independently to host soluble signals early in pathogenesis, and initial exposure to these secretions prompts a switch to a motile population phenotype. We hypothesise protein expression induced by host secretions is aimed at counteracting host immune defences while a motile population migrates to an optimum environment for attachment. In this biphasic model, trophozoites are either initially sensitised to host soluble signals, or undergo host attachment and induce proteins in advance of host defences (Figure 10). Anticipating the production of molecules from the host cells which are designed to clear parasite infection, trophozoites attached to host cells produce a wide range of oxidoreductases for neutralising exogenous ROS. These trophozoites also up-regulate proteins associated with PLP production, while increasing ubiquitin/proteasome mediated protein turnover for production of disease related proteins. This dual combination of distinct responses to either host soluble factors, or host attachment, indicates early pathogenesis involves multiple and distinct levels of crosstalk between host and parasite. These are independently regulated and do not require attachment to host-cells for induction of virulence factors. Thus, these host secreted signals are sufficient to induce

virulence factor expression in *Giardia* cells, which occurs independent of parasite attachment to host cells.

5.9 Acknowledgements

SJE acknowledges funding from the Australian Government in the form of an APA scholarship, as well as financial support from Macquarie University. We gratefully acknowledge the staff and support from the Australian Proteome Analysis Facility (APAF). SJE wishes to thank Dr Jacqui Upcroft for supplying the *Giardia* samples used and for the ongoing support received from colleagues at Microbial Screening Technologies, including Alistair Lacey for support going beyond the call of duty. PAH wishes to thank Justin Lane for continued support and encouragement.

5.10 Supplementary Data

The following supplementary information is available as supplementary data files associated with this manuscript:

Supplementary Media 1: video showing trophozoites in control flasks (DMEM-only) compared to HSF-exposed trophozoites during the first 6-hour incubation in T75cm2 flasks. Dark, tear-shaped trophozoites are adhered to the flask at low density, and continue to appear semi-motile against the flask wall during exposure to HSF. In contrast, control trophozoites in DMEM only are stationary on the flask wall at high density. **(Supplementary DVD)**

Supplementary Data S1: complete dataset for the attachment assay for trophozoites and HT-29 cells, including mean, standard deviations and p-values between co-incubation and control triplicates. Tab one contains the data for the adherence versus host-cell attachment assay, while the second tab contains the data for the HSF-exposure attachment assay. **(Supplementary DVD)**

Supplementary Data S2: Excel spreadsheet showing the complete protein identification and quantitation information for the TMT labelling experiment. **(Supplementary DVD)**

Supplementary Data S3: spreadsheet containing principal component scores (first tab) and loadings (second tab) resulting from a PCA analysis of the log2 (ratios) of all samples to the control pool, using all available quantitated proteins. The loadings were sorted in decreasing order of the first principal component, which shows a good separation of the control ratios from the rest; the top 5% highest loadings were highlighted. **(Supplementary DVD)**

Supplementary Data S4: Excel spreadsheet showing the complete functional annotation information for proteins up- and down-regulated during host-cell interaction analysis. Tables show the Giardiadb.org ORF number, protein description, fold change as well as GO annotation, subcellular localisation information and interpro protein domain/fold information.

Up-regulated proteins in CI IEC and HSF are shown on tab 1 and 2, respectively, while the down-regulated proteins in CI IEC interaction are on tab 3 and down-regulated proteins in HSF are on tab 4. **(Supplementary DVD)**

Supplementary Data S5: Excel spreadsheet showing the output from bioinformatics analysis of secreted proteins. Proteins were submitted to Target P for subcellular localisation, with a RI of ≤ 3 considered. Signal P was used to assess presence of a signal peptide, with a cutoff ≥ 0.5 considered for confident identifications and TMHMM used to detect presence of transmembrane helices. Finally, NucPred was used as an exclusionary tool for false positives in the secretory pathway, and scores ≥0.9 considered a hit for nuclear localisation. Output from analyses for up-regulated proteins in CI IEC and HSF are shown on tab 1 and 2, respectively, while the down-regulated proteins in CI IEC interaction are on tab 3 and down-regulated proteins in HSF are on tab 4. **(Supplementary DVD)**

5.11 References

1. Cotton JA, Beatty JK, Buret AG. Host parasite interactions and pathophysiology in *Giardia* infections. *Int J Parasitol* **41**, 925-933 (2011).

2. Buret AG. Pathophysiology of enteric infections with *Giardia duodenalis*. *Parasite* **15**, 261-265 (2008).

3. Ringqvist E*, et al.* Release of metabolic enzymes by *Giardia* in response to interaction with intestinal epithelial cells. *Mol Biochem Parasitol* **159**, 85-91 (2008).

4. Ringqvist E, Avesson L, Soderbom F, Svard SG. Transcriptional changes in *Giardia* during host-parasite interactions. *International journal for parasitology* **41**, 277-285 (2011).

5. Ferella M*, et al.* Gene expression changes during *Giardia*-host cell interactions in serum-free medium. *Mol Biochem Parasitol* **197**, 21-23 (2014).

6. Roxstrom-Lindquist K, Ringqvist E, Palm D, Svard S. *Giardia lamblia*-induced changes in gene expression in differentiated Caco-2 human intestinal epithelial cells. *Infect Immun* **73**, 8204-8208 (2005).

7. Stadelmann B*, et al.* The role of arginine and arginine-metabolizing enzymes during *Giardia* - host cell interactions in vitro. *BMC Microbiol* **13**, 256 (2013).

8. Wastling JM, Armstrong SD, Krishna R, Xia D. Parasites, proteomes and systems: has Descartes' clock run out of time? *Parasitology* **139**, 1103-1118 (2012).

9. Faso C, Bischof S, Hehl AB. The proteome landscape of *Giardia lamblia* encystation. *PloS one* **8**, e83207 (2013).

10. Lauwaet T*, et al.* Mining the *Giardia* genome and proteome for conserved and unique basal body proteins. *Int J Parasitol* **41**, 1079-1092 (2011).

11. Jedelsky PL*, et al.* The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLoS One* **6**, e17285 (2011).

12. Emery SJ, van Sluyter S, Haynes PA. Proteomic analysis in *Giardia duodenalis* yields insights into strain virulence and antigenic variation. *Proteomics* **14**, 2523-2534 (2014).

13. Wampfler PB*, et al.* Proteomics of secretory and endocytic organelles in *Giardia lamblia*. *PLoS One* **9**, e94089 (2014).

14. Stefanic S, Palm D, Svard SG, Hehl AB. Organelle proteomics reveals cargo maturation mechanisms associated with Golgi-like encystation vesicles in the early-diverged protozoan *Giardia lamblia*. *J Biol Chem* **281**, 7595-7604 (2006).

15. Emery SJ, Pascovici D, Lacey E, Haynes PA. The generation gap: Proteome changes and strain variation during encystation in *Giardia duodenalis*. *Mol Biochem Parasitol* **201**, 47-56 (2015).

16. McAlister GC*, et al.* Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem* **84**, 7469-7478 (2012).

17. Keister DB. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans R Soc Trop Med Hyg* **77**, 487-488 (1983).

18. Farthing MJ, Pereira ME, Keusch GT. Giardia lamblia: evaluation of roller bottle cultivation. *Exp Parasitol* **54**, 410-415 (1982).

19. Wessel D, Flugge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* **138**, 141-143 (1984).

20. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**, 1005-1016 (2000).

21. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783-795 (2004).

22. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953-971 (2007).

23. Brameier M, Krings A, MacCallum RM. NucPred--predicting nuclear localization of proteins. *Bioinformatics* **23**, 1159-1160 (2007).

24. Vizcaino JA*, et al.* The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* **41**, D1063-1069 (2013).

25. Song X*, et al.* iTRAQ experimental design for plasma biomarker discovery. *J Proteome Res* **7**, 2952-2958 (2008).

26. Pounds SB. Estimation and control of multiple testing error rates for microarray studies. *Brief Bioinform* **7**, 25-36 (2006).

27. Emery SJ, Lacey E, Haynes PA. Quantitative proteomics analysis of *Giardia duodenalis* Assemblage A - a baseline for host, assemblage and isolate variation. *Proteomics*, (2015).

28. Nageshan RK, Roy N, Hehl AB, Tatu U. Post-transcriptional repair of a split heat shock protein 90 gene by mRNA trans-splicing. *J Biol Chem* **286**, 7116-7122 (2011).

29. Ankarklev J*, et al.* Comparative genomic analyses of freshly isolated *Giardia intestinalis* assemblage A isolates. *BMC genomics* **16**, 697 (2015).

30. Mahoney DW*, et al.* Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J Proteome Res* **10**, 4325-4333 (2011).

31. Karp NA*, et al.* Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics* **9**, 1885-1897 (2010).

32. Adam RD*, et al.* The *Giardia lamblia* vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC genomics* **11**, 424 (2010).

33. Kordis D, Turk V. Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. *BMC evolutionary biology* **9**, 266 (2009)

34. Pavlova A, Bjork I. Grafting of features of cystatins C or B into the N-terminal region or second binding loop of cystatin A (stefin A) substantially enhances inhibition of cysteine proteinases. *Biochemistry* **42**, 11326-11333 (2003).

35. Lalle M*, et al.* Interaction network of the 14-3-3 protein in the ancient protozoan parasite *Giardia duodenalis*. *J Proteome Res* **11**, 2666-2683 (2012).

36. Joshi PB*, et al.* Targeted gene deletion in *Leishmania major* identifies leishmanolysin (GP63) as a virulence factor. *Mol Biochem Parasitol* **120**, 33-40 (2002).

37. Sanderson SJ*, et al.* Functional conservation of a natural cysteine peptidase inhibitor in protozoan and bacterial pathogens. *FEBS Lett* **542**, 12-16 (2003).

38. Wang CG*, et al.* A novel short-chain peptide BmKX from the Chinese scorpion Buthus martensi Karsch, sequencing, gene cloning and structure determination. *Toxicon* **45**, 309-319 (2005).

39. Ankarklev J*, et al.* Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nat Rev Microbiol* **8**, 413-422 (2010).

40. Upcroft JA*, et al.* Lethal *Giardia* from a wild-caught sulphur-crested cockatoo (*Cacatua galerita*) established in vitro chronically infects mice. *Parasitology* **114 (Pt 5)**, 407-412 (1997).

41. Williamson AL, O'Donoghue PJ, Upcroft JA, Upcroft P. Immune and pathophysiological responses to different strains of *Giardia duodenalis* in neonatal mice. *Int J Parasitol* **30**, 129-136 (2000).

42. Nash TE. Antigenic variation in *Giardia lamblia* and the host's immune response. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **352**, 1369-1375 (1997).

43. Nash TE*, et al.* Antigenic variation in *Giardia lamblia*. *J Immunol* **141**, 636-641 (1988).

44. Cotton JA*, et al.* Giardia duodenalis cathepsin B proteases degrade intestinal epithelial interleukin-8 and attenuate interleukin-8-induced neutrophil chemotaxis. *Infect Immun* **82**, 2772-2787 (2014).

45. Cotton JA*, et al.* Giardia duodenalis infection reduces granulocyte infiltration in an in vivo model of bacterial toxin-induced colitis and attenuates inflammation in human intestinal tissue. *PloS one* **9**, e109087 (2014).

46. Hartmann S, Lucius R. Modulation of host immune responses by nematode cystatins. *Int J Parasitol* **33**, 1291-1302 (2003).

47. Vray B, Hartmann S, Hoebeke J. Immunomodulatory properties of cystatins. *Cell Mol Life Sci* **59**, 1503-1512 (2002).

48. Puente-Rivera J*, et al.* Trichocystatin-2 (TC-2): an endogenous inhibitor of cysteine proteinases in Trichomonas vaginalis is associated with TvCP39. *The international journal of biochemistry & cell biology* **54**, 255-265 (2014).

49. Chiu PW*, et al.* A novel family of cyst proteins with epidermal growth factor repeats in *Giardia lamblia*. *PLoS Negl Trop Dis* **4**, e677 (2010).

50. Bornstein P. Matricellular proteins: an overview. *J Cell Commun Signal* **3**, 163-165 (2009).

51. Aspberg A, Binkert C, Ruoslahti E. The versican C-type lectin domain recognizes the adhesion protein tenascin-R. *Proc Natl Acad Sci U S A* **92**, 10590-10594 (1995).

52. Aspberg A*, et al.* The C-type lectin domains of lecticans, a family of aggregating chondroitin sulfate proteoglycans, bind tenascin-R by protein-protein interactions independent of carbohydrate moiety. *Proc Natl Acad Sci U S A* **94**, 10116-10121 (1997).

53. Katelaris PH, Naeem A, Farthing MJ. Attachment of *Giardia lamblia* trophozoites to a cultured human intestinal cell line. *Gut* **37**, 512-518 (1995).

54. Inge PM, Edson CM, Farthing MJ. Attachment of *Giardia lamblia* to rat intestinal epithelial cells. *Gut* **29**, 795-801 (1988).

55. Magne D*, et al.* Role of cytoskeleton and surface lectins in *Giardia duodenalis* attachment to Caco2 cells. *Parasitol Res* **77**, 659-662 (1991).

56. Wagland BM, Emery DL, McClure SJ. Studies on the host-parasite relationship between *Trichostrongylus colubriformis* and susceptible and resistant sheep. *Int J Parasitol* **26**, 1279-1286 (1996).

57. Roxstrom-Lindquist K*, et al. Giardia* immunity--an update. *Trends Parasitol* **22**, 26-31 (2006).

58. Gallego E, Alvarado M, Wasserman M. Identification and expression of the protein ubiquitination system in *Giardia intestinalis*. *Parasitol Res* **101**, 1-7 (2007).

59. Nino CA*, et al.* Ubiquitination dynamics in the early-branching eukaryote *Giardia intestinalis*. *Microbiologyopen* **2**, 525-539 (2013).

60. Ansell BR*, et al.* Drug resistance in *Giardia duodenalis*. *Biotechnol Adv* **33**, 888-901 (2015).

61. Eckmann L*, et al.* Nitric oxide production by human intestinal epithelial cells and competition for arginine as potential determinants of host defense against the lumen-dwelling pathogen *Giardia lamblia*. *J Immunol* **164**, 1478-1487 (2000).

62. Stadelmann B, Merino MC, Persson L, Svard SG. Arginine consumption by the intestinal parasite *Giardia intestinalis* reduces proliferation of intestinal epithelial cells. *PLoS One* **7**, e45325 (2012).

63. Di Matteo A*, et al.* The O2-scavenging flavodiiron protein in the human parasite *Giardia intestinalis*. *J Biol Chem* **283**, 4061-4068 (2008).

64. Adam RD. Biology of *Giardia lamblia*. *Clin Microbiol Rev* **14**, 447-475 (2001).

65. Raj D*, et al.* Differential gene expression in *Giardia lamblia* under oxidative stress: significance in eukaryotic evolution. *Gene* **535**, 131-139 (2014).

66. Weissbach H*, et al.* Peptide methionine sulfoxide reductase: structure, mechanism of action, and biological function. *Arch Biochem Biophys* **397**, 172-178 (2002).

67. Almeida CC*, et al.* The role of the hybrid cluster protein in oxidative stress defense. *J Biol Chem* **281**, 32445-32450 (2006).

68. Kappes B, Tews I, Binter A, Macheroux P. PLP-dependent enzymes as potential drug targets for protozoan diseases. *Biochim Biophys Acta* **1814**, 1567-1576 (2011).

69. Tachezy J, Sanchez LB, Muller M. Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. *Mol Biol Evol* **18**, 1919-1928 (2001).

70. Gillin FD, Reiner DS, McCann PP. Inhibition of growth of *Giardia lamblia* by difluoromethylornithine, a specific inhibitor of polyamine biosynthesis. *J Protozool* **31**, 161-163 (1984).

CHAPTER 6

A general discussion of the work presented in this thesis, including implications for the field and further research directions.

6. General Discussion

6.1 Thesis Outcomes

Chapter 1 of this thesis provided a comprehensive literature review of the proteomics field in *G. duodenalis*. Within this review it was acknowledged that there are several gaps in which proteomic data is incomplete or absent for multiple biological scenarios. In addition, the final section (*Subheading 1.10 – Future Directions*) provided commentary on the necessity of utilising newer quantitative technologies. The review in ***Chapter 1*** also included the publications resulting from the experimental chapters of this thesis (Publications II-VI), and contextualised them in light of their technical and biological contributions. The overall aim of the literature review of this thesis was to craft a complete summary of post-genomic proteomics in *Giardia*, including the largely unexplored range of post-translational modifications of core proteins. We believe that this is an important step in order to direct and streamline the next generation of proteomic data, specifically quantitative data. Accordingly, we believe that the publication which will arise from ***Chapter 1*** will provide an important resource which is currently unavailable in the literature.

In ***Chapter 2*** of this thesis, we successfully compared a gel-based and an in-solution sample preparation and fractionation method for quantitative proteomics in *Giardia* using label-free spectral counting. We evaluated these methods by characterising two phenotypically distinct A1 isolates, in order to demonstrate the performance of the different methods when applied in biologically meaningful comparisons. This optimised protocol for profiling combined an improved FASP method using TFE [188, 189] with online fractionation during MS using GPF [190]. Our data indicated that this method improved identification of proteins with lipid modifications and provided increased peptide coverage for the proteins identified. This was compared to the 1-DE SDS-PAGE approach, which was

problematic due to highly abundant cytoskeletal and metabolic enzymes that could not be resolved to discrete regions of the gel, and as such dominated the peptide results. Overall, the combined FASP-GPF approach allowed more identifications with improved coverage and fewer peptides, which has been documented in similar studies with *Vitis riparia* (riverbank grape) [191].

The optimised methodology described above in ***Chapter 2*** was then applied to several investigations in ***Chapters 2-4.*** In ***Chapter 2*** the proteomes of two A1 isolates were compared directly, and then a similar but broader approach for investigating isolate diversity was carried out in ***Chapter 3***. In ***Chapter 2*** comparative proteomics revealed differences in protein abundance in known virulence factors, including immunomodulatory proteases such as cathepsin B [100], as well as differences in the total number and subpopulation diversity of VSPs between isolates [101]. Differences in VSP distribution, abundance and diversity were also observed between the eight isolates in ***Chapter 3***. In ***Chapter 3*** this was shown to be independent of subassemblage, and not correlated to host, geographic origins, or time since introduction to culture.

The large scale comparative proteomics of eight isolates of assemblage A in ***Chapter 3*** also demonstrated that peptide tospectrum matching was improved when spectra were matched to genome sequence from the correct subassemblage. Database searching of spectra revealed that quantitation of multi-gene families in *Giardia* was greatly affected by which genome sequence database was used in this process, which coincides with comparative genomic analyses between assemblages [32], between subassemblages [28] and within subassemblages [33]. These studies indicate diversity, expansion and recombination occurs at high frequency in multigene families. We believe that the data generated from this study provides a comprehensive proteomic foundation that will remain a useful resource in both parasite biology and taxonomy.

In ***Chapter 4,*** isolate diversity was investigated in the biological context of differentiation. As in ***Chapter 2*** and ***Chapter 3,*** a label-free spectral-counting approach was applied, and the optimised methodology from ***Chapter 2*** was also utilised. A sound proteomic baseline was available for the genome isolate WB C6 [39, 106], as well as several complementary transcriptomic studies [107-109]. In ***Chapter 4*** two genome-alternate isolates was used, which confirmed that the metabolic and cytoskeletal restructuring during differentiation provides universal markers for encystation that are isolate- and method-independent. However, two functional clusters relating to the VSP subpopulation retention between trophozoites to cysts, and ankyrin-repeat/protein 21.1 up-regulation were observed, which may constitute isolate-specific adaptations that affect reinfection success in the next generation. We also believe that this experiment in ***Chapter 4*** provided an important precedent to utilise multiple isolates to investigate biological processes in *Giardia*. With the demonstration of broad differences in host range and zoonoses [192], phenotypes [106] and cyst infectivity [9], it is likely that studies of single isolates will not illuminate those factors which contribute to *Giardia's* inherent isolate diversity.

The final experiment in this thesis, discussed in ***Chapter 5*****,** provided several important technical and biological foundations in *Giardia* proteomics. This chapter constitutes the first experiment to use TMT labels for quantitative proteomics in *Giardia*. Two studies have previously utilised iTRAQ tags, though these have been for filtering of contaminants rather than quantitation [19], and in less complex protein samples [37]. As such, we believe this is an important study as we have demonstrated the feasibility of using isobaric tags for sensitive and dynamic quantification of complex, whole-trophozoite lysates.

The experiment in ***Chapter 5*** also constituted the first proteomic analysis of *Giardia* in an *in vitro* IEC-trophozoite interaction model, complementing transcriptomic predecessors [115-117]. This study, also for the first time, independently investigated disease induction

by soluble factors released by HT-29 IECs compared to co-incubation, the latter of which permits trophozoite attachment. When proteomic data was combined with results from attachment assays, this experiment demonstrated for the first time a biphasic model in early pathogenesis. The data from this experiment indicated that motile trophozoites respond very rapidly to soluble signals from the host, and that these are enough to induce expression of many secreted and membrane associated proteins, including virulence factors. In contrast, trophozoite attachment induced proteins associated with the anticipation of host defences, in particular ROS. We believe that this constitutes an exciting and novel observation in disease induction in *Giardia*, and that host-parasite interactions are occurring on multiple levels from the very early stages of pathogenesis.

As acknowledged in ***Chapter 1*** (*Subheading 2 – Genomes and Annotations*), the recent availability of multiple genomes has made quantitative proteomics possible for *G. duodenalis*. In this vein, this thesis has presented four independent, post-genomic, quantitative proteomics experiments in *G. duodenalis*. The culmination of these studies provides a cohesive proteomics pipeline using both labelled and label-free proteomics approaches. These methodologies validate how quantifying changes in protein abundance correlate with isolate diversity (***Chapter 2*** and ***Chapter 3***), reflect changes during differentiation (***Chapter 4***), and demonstrate induction of virulence and disease factors during host-parasite interactions (***Chapter 5***). We believe these studies validate important approaches and resources to functionally evaluate parasite biology, either using proteomics alone, or as part of a collective effort to combine protein, gene and transcript data in a systems biology approach [10, 193]. In addition, we believe that each of the studies in this thesis will provide important technical counsel for sample preparation and fractionation, protein identification and quantitation technologies, and data and statistical analysis.

6.2 Reflections and Future Research

Although ***Chapter 1*** covers the technical and biological shortcomings for the overall field of *Giardia* proteomics, there are a number of possible future extensions which could follow on from the specific research in this thesis. While we believe that the information in this thesis constitutes a valuable technical and molecular resource, there are several limitations highlighted, as well as further biological questions linked to the outcomes of experiments that will be discussed.

As indicated in ***Chapter 1*** (*Subheading 1.3 – Genomes and Annotations, Table 1*) there is still considerable work required to fully annotate the *Giardia* genome sequence. For the WB C6 genome, which was the genome predominantly used throughout this thesis, half the genes are still annotated at 'hypothetical proteins'. The absence of annotations diminishes the primary capacity of quantitative proteomics – to relate differentially expressed proteins to their molecular functions and therefore bestow biological significance on results. Though proteins may be differentially expressed in a particular experimental context, the absence of functional annotation makes it impossible to extrapolate what this change in expression is correlated with at a functional level. This is frustrating when the remainder of differentially expressed proteins are functionally annotated and clearly related to biological outcomes, such as in ***Chapter 5***. In the study of proteins during *in-vitro* host-parasite interactions, multiple known virulence factors were induced, indicating that IEC soluble signals, or co-incubation, were indeed simulating disease processes. However, multiple hypothetical proteins without functional annotations were also detected, and though it is likely these also correlate to important functions in early pathogenesis, drawing biological conclusions for these proteins is not possible at this stage. It is therefore critical that researchers deposit their quantitative proteomics information in databases to await further annotation. In particular, submitting to the centralised *Giardiadb.org* [26], which

combines genome, transcript and protein data, may connect earlier and current proteomics data to future improvements in annotation.

Another limitation of quantitative proteomics in *Giardia* is the validation of differentially expressed proteins in large datasets. Though inclusion of western blot assays is traditionally championed as the validation tool of shotgun proteomics, there are now several MS alternatives which are gaining favour as technologies improve [193, 194]. This may prove beneficial for *Giardia* in particular, as there is a paucity of commercially available antibodies for use, which has made western blotting significantly costly in time and money. Targeted MS proteomic techniques such as SRM and MRM assays therefore allow the routine quantification of proteins for which no affinity reagent or antibody is available. In future, these will continue to improve in accuracy and sensitivity, in parallel to MS instrumentation and software. We believe that designing such SRM assays for validation of shotgun proteomics experiments in *Giardia* will therefore be a significant and broadly applicable advance in quantitative proteomics, and is a natural progression of the studies in this thesis.

However, there are additional measures of confidence which can be applied to proteomics datasets prior to validation. Protein and peptide FDR at the identification levels are applied in ***Chapters 2-4***. These FDR are calculated based on reverse searching against decoy datasets [195] and are used to assess the quality of the datasets, specifically in conjunction to filtering. For label-free spectral counting in ***Chapters 2-4,*** differentially expressed proteins are only considered when they are reproducibly identified in all biological triplicates of at least one condition. This filtering is shown to greatly reduce, and in some case eliminate, reverse hits in the spectral counting datasets presented in this thesis. In addition, in ***Chapter 5,*** TMT labelling allowed the multiplexing of samples, which permits each peptide to be detected and relatively quantified for all 10 TMT channels in the second

MS scan. In addition to the fold changes calculated in TMT ratios, we also provided a secondary criteria of a significant p-value, which is known to improve confidence in quantitating differences between proteins with smaller fold changes, due to the increased variability that occurs at lower abundance [170]. A forward-looking approach we are working on relates to the analysis of control versus control experiments for label-free spectral counting. By comparing control technical or biological replicates an estimation of the 'noise' level in quantification can be experimentally determined. By comparing this to the protein fold-changes observed between control and treatments, and ensuring it exceeds the calculated level of 'noise', quantification can also be validated for label-free experiments. We believe that these FDR and statistical analysis methods are essential for all quantitative proteomics experiments, and their importance needs to be emphasised so such standards in data quality and analysis are a requirement in the field for *Giardia*.

As covered in ***Chapter 1*** there are still many gaps in the biology of *Giardia* that require initial proteomic data. We believe that the methods developed in this thesis should be applied to generate this missing data in the near future. This thesis provides comprehensive analyses of protein expression in multiple biological scenarios for the A assemblage, particularly A1. There is currently a deficit of proteomic data in Assemblage B, which is also infective for humans and linked to more damaging clinical symptomology [98]. Comparative genomics of assemblage A and B support the hypothesis that these are in fact two distinct species [28, 32, 196], and therefore divergent mechanisms for virulence, pathology and disease induction may occur, that cannot be extrapolated from studies of assemblage A. We believe that, given that an improved genome sequence for assemblage B is now available [28], generating proteomic data in this assemblage should be prioritised.

This thesis has provided evidence that proteomics is an exquisite tool for elucidating alterations in protein expression in *Giardia*. More importantly, we feel that the changes in

protein expression described throughout the chapters in the thesis are true reflections of parasite diversity and biology. However, further and novel proteomic investigations are required, and these hopefully will utilise new quantitative technologies. Nonetheless, we believe that this thesis constitutes a thematically and technically cohesive foundation for the field, and we hope that it will support ongoing efforts for quantitative proteomics in *Giardia.*

Bibliography

*Reference list for **Chapter 1** and **Chapter 6**.*

Bibliography

[1] Adam, R. D., Biology of *Giardia lamblia*. *Clin Microbiol Rev* 2001, *14*, 447-475.

[2] Feng, Y., Xiao, L., Zoonotic potential and molecular epidemiology of *Giardia* species and giardiasis. *Clin Microbiol Rev* 2011, *24*, 110-140.

[3] Halliez, M. C., Buret, A. G., Extra-intestinal and long term consequences of *Giardia duodenalis* infections. *World J Gastroenterol* 2013, *19*, 8974-8985.

[4] Buret, A. G., Pathophysiology of enteric infections with *Giardia duodenalis*. *Parasite* 2008, *15*, 261-265.

[5] Hanevik, K., Kristoffersen, E., Svard, S., Bruserud, O.*, et al.*, Human cellular immune response against *Giardia lamblia* 5 years after acute giardiasis. *J Infect Dis* 2011, *204*, 1779-1786.

[6] Cotton, J. A., Beatty, J. K., Buret, A. G., Host parasite interactions and pathophysiology in *Giardia* infections. *Int J Parasitol* 2011, *41*, 925-933.

[7] Gerbaba, T. K., Gupta, P., Rioux, K., Hansen, D., Buret, A. G., *Giardia duodenalis* - induced alterations of commensal bacteria kill *Caenorhabditis elegans*: A new model to study microbial - microbial interactions in the gut. *Am J Physiol Gastrointest Liver Physiol* 2015, ajpgi 00335 02014.

[8] Williamson, A. L., O'Donoghue, P. J., Upcroft, J. A., Upcroft, P., Immune and pathophysiological responses to different strains of *Giardia duodenalis* in neonatal mice. *Int J Parasitol* 2000, *30*, 129-136.

[9] Visvesvara, G. S., Dickerson, J. W., Healy, G. R., Variable infectivity of human-derived Giardia lamblia cysts for Mongolian gerbils (Meriones unguiculatus). *J Clin Microbiol* 1988, *26*, 837-841.

[10] Wastling, J. M., Armstrong, S. D., Krishna, R., Xia, D., Parasites, proteomes and systems: has Descartes' clock run out of time? *Parasitology* 2012, *139*, 1103-1118.

[11] Jex, A. R., Koehler, A. V., Ansell, B. R., Baker, L.*, et al.*, Getting to the guts of the matter: the status and potential of 'omics' research of parasitic protists of the human gastrointestinal system. *Int J Parasitol* 2013, *43*, 971-982.

[12] Steuart, R. F., Proteomic analysis of *Giardia*: studies from the pre- and post-genomic era. *Exp Parasitol* 2010, *124*, 26-30.

[13] Ong, S. E., Foster, L. J., Mann, M., Mass spectrometric-based approaches in quantitative proteomics. *Methods* 2003, *29*, 124-130.

[14] Ong, S. E., Mann, M., Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* 2005, *1*, 252-262.

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

[16] Wang, H., Hanash, S., Multi-dimensional liquid phase based separations in proteomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003, *787*, 11-18.

[17] Ong, S. E., Mann, M., Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods Mol Biol* 2007, *359*, 37-52.

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

[19] Jedelsky, P. L., Dolezal, P., Rada, P., Pyrih, J.*, et al.*, The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLoS One* 2011, *6*, e17285.

[20] Thompson, A., Schafer, J., Kuhn, K., Kienle, S.*, et al.*, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 2003, *75*, 1895-1904.

[21] McAlister, G. C., Huttlin, E. L., Haas, W., Ting, L.*, et al.*, Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem* 2012, *84*, 7469-7478.

[22] Mann, M., Jensen, O. N., Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003, *21*, 255-261.

[23] Witze, E. S., Old, W. M., Resing, K. A., Ahn, N. G., Mapping protein post-translational modifications with mass spectrometry. *Nat Methods* 2007, *4*, 798-806.

[24] Caccio, S. M., Ryan, U., Molecular epidemiology of giardiasis. *Mol Biochem Parasitol* 2008, *160*, 75-80.

[25] Monis, P. T., Caccio, S. M., Thompson, R. C., Variation in *Giardia*: towards a taxonomic revision of the genus. *Trends Parasitol* 2009, *25*, 93-100.

[26] Aurrecoechea, C., Brestelli, J., Brunk, B. P., Carlton, J. M.*, et al.*, GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res* 2009, *37*, D526-530.

[27] Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B.*, et al.*, Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science* 2007, *317*, 1921-1926.

[28] Adam, R. D., Dahlstrom, E. W., Martens, C. A., Bruno, D. P.*, et al.*, Genome sequencing of *Giardia lamblia* genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). *Genome Biol Evol* 2013, *5*, 2498-2511.

[29] Perry, D. A., Morrison, H. G., Adam, R. D., Optical map of the genotype A1 WB C6 *Giardia lamblia* genome isolate. *Mol Biochem Parasitol* 2011, *180*, 112-114.

[30] Krauer, K. G., Burgess, A. G., Dunn, L. A., Upcroft, P., Upcroft, J. A., Sequence map of the 2 Mb *Giardia lamblia* assemblage A chromosome. *J Parasitol* 2010, *96*, 660-662.

[31] Franzen, O., Jerlstrom-Hultqvist, J., Castro, E., Sherwood, E.*, et al.*, Draft genome sequencing of *Giardia intestinalis* assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS pathogens* 2009, *5*, e1000560.

[32] Jerlstrom-Hultqvist, J., Franzen, O., Ankarklev, J., Xu, F*., et al.*, Genome analysis and comparative genomics of a *Giardia intestinalis* assemblage E isolate. *BMC genomics* 2010, *11*, 543.

[33] Ankarklev, J., Franzen, O., Peirasmaki, D., Jerlstrom-Hultqvist, J*., et al.*, Comparative genomic analyses of freshly isolated *Giardia intestinalis* assemblage A isolates. *BMC genomics* 2015, *16*, 697.

[34] Nageshan, R. K., Roy, N., Hehl, A. B., Tatu, U., Post-transcriptional repair of a split heat shock protein 90 gene by mRNA trans-splicing. *J Biol Chem* 2011, *286*, 7116-7122.

[35] Capon, A. G., Upcroft, J. A., Boreham, P. F., Cottis, L. E., Bundesen, P. G., Similarities of *Giardia* antigens derived from human and animal sources. *Int J Parasitol* 1989, *19*, 91-98.

[36] Steuart, R. F., O'Handley, R., Lipscombe, R. J., Lock, R. A., Thompson, R. C., Alpha 2 giardin is an assemblage A-specific protein of human infective *Giardia duodenalis*. *Parasitology* 2008, *135*, 1621-1627.

[37] Lingdan, L., Pengtao, G., Wenchao, L., Jianhua, L*., et al.*, Differential dissolved protein expression throughout the life cycle of *Giardia lamblia*. *Exp Parasitol* 2012, *132*, 465-469.

[38] Franzen, O., Jerlstrom-Hultqvist, J., Einarsson, E., Ankarklev, J*., et al.*, Transcriptome profiling of *Giardia intestinalis* using strand-specific RNA-seq. *PLoS Comput Biol* 2013, *9*, e1003000.

[39] Faso, C., Bischof, S., Hehl, A. B., The proteome landscape of *Giardia lamblia* encystation. *PloS one* 2013, *8*, e83207.

[40] Emery, S. J., Pascovi, D., Lacey, E., Haynes, P. A., The generation gap: Proteome changes and strain variation during encystation in *Giardia duodenalis*. *Mol Biochem Parasitol* 2015.

[41] Yang, J., Zhang, Y., I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* 2015.

[42] Yang, J., Yan, R., Roy, A., Xu, D*., et al.*, The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 2015, *12*, 7-8.

[43] Wang, S., et al., RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res*, 2016. 44(W1): p. W430-5.

[44] Kelley, L.A., et al., The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 2015. 10(6): p. 845-58.

[45] Soding, J., A. Biegert, and A.N. Lupas, The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, 2005. **33**(Web Server issue): p. W244-8.

[46] Davids, B., Gillin, F., in: Luján, H., Svärd, S. (Eds.), *Giardia*, Springer Vienna 2011, pp. 381-394.

[47] Katelaris, P. H., Naeem, A., Farthing, M. J., Attachment of *Giardia lamblia* trophozoites to a cultured human intestinal cell line. *Gut* 1995, *37*, 512-518.

[48] Magne, D., Favennec, L., Chochillon, C., Gorenflot, A.*, et al.*, Role of cytoskeleton and surface lectins in *Giardia duodenalis* attachment to Caco2 cells. *Parasitol Res* 1991, *77*, 659-662.

[49] Townson, S. M., Laqua, H., Upcroft, P., Boreham, P. F., Upcroft, J. A., Induction of metronidazole and furazolidone resistance in *Giardia*. *Trans R Soc Trop Med Hyg* 1992, *86*, 521-522.

[50] Upcroft, J. A., Campbell, R. W., Upcroft, P., Quinacrine-resistant *Giardia duodenalis*. *Parasitology* 1996, *112 (Pt 3)*, 309-313.

[51] Upcroft, J., Mitchell, R., Chen, N., Upcroft, P., Albendazole resistance in *Giardia* is correlated with cytoskeletal changes but not with a mutation at amino acid 200 in beta-tubulin. *Microb Drug Resist* 1996, *2*, 303-308.

[52] Ankarklev, J., Jerlstrom-Hultqvist, J., Ringqvist, E., Troell, K., Svard, S. G., Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nat Rev Microbiol* 2010, *8*, 413-422.

[53] Svard, S. G., Hagblom, P., Palm, J. E., *Giardia lamblia* -- a model organism for eukaryotic cell differentiation. *FEMS Microbiol Lett* 2003, *218*, 3-7.

[54] Dreger, M., Proteome analysis at the level of subcellular structures. *Eur J Biochem* 2003, *270*, 589-599.

[55] Dreger, M., Subcellular proteomics. *Mass Spectrom Rev* 2003, *22*, 27-56.

[56] Elmendorf, H. G., Dawson, S. C., McCaffery, J. M., The cytoskeleton of *Giardia lamblia*. *Int J Parasitol* 2003, *33*, 3-28.

[57] Martincova, E., Voleman, L., Pyrih, J., Zarsky, V.*, et al.*, Probing the biology of *Giardia intestinalis* mitosomes using in vivo enzymatic tagging. *Mol Cell Biol* 2015.

[58] Wampfler, P. B., Tosevski, V., Nanni, P., Spycher, C., Hehl, A. B., Proteomics of secretory and endocytic organelles in *Giardia lamblia*. *PLoS One* 2014, *9*, e94089.

[59] Jerlstrom-Hultqvist, J., Stadelmann, B., Birkestedt, S., Hellman, U., Svard, S. G., Plasmid vectors for proteomic analyses in *Giardia*: purification of virulence factors and analysis of the proteasome. *Eukaryot Cell* 2012, *11*, 864-873.

[60] Hagen, K. D., Hirakawa, M. P., House, S. A., Schwartz, C. L.*, et al.*, Novel structural components of the ventral disc and lateral crest in *Giardia intestinalis*. *PLoS Negl Trop Dis* 2011, *5*, e1442.

[61] Lauwaet, T., Smith, A. J., Reiner, D. S., Romijn, E. P.*, et al.*, Mining the *Giardia* genome and proteome for conserved and unique basal body proteins. *Int J Parasitol* 2011, *41*, 1079-1092.

[62] Weiland, M. E., McArthur, A. G., Morrison, H. G., Sogin, M. L., Svard, S. G., Annexin-like alpha giardins: a new cytoskeletal gene family in *Giardia lamblia*. *Int J Parasitol* 2005, *35*, 617-626.

[63] Jenkins, M. C., O'Brien, C. N., Murphy, C., Schwarz, R.*, et al.*, Antibodies to the ventral disc protein delta-giardin prevent in vitro binding of *Giardia lamblia* trophozoites. *J Parasitol* 2009, *95*, 895-899.

[64] Macarisin, D., O'Brien, C., Fayer, R., Bauchan, G., Jenkins, M., Immunolocalization of beta- and delta-giardin within the ventral disk in trophozoites of *Giardia duodenalis* using multiplex laser scanning confocal microscopy. *Parasitol Res* 2012, *111*, 241-248.

[65] Saric, M., Vahrmann, A., Niebur, D., Kluempers, V.*, et al.*, Dual acylation accounts for the localization of {alpha}19-giardin in the ventral flagellum pair of *Giardia lamblia*. *Eukaryot Cell* 2009, *8*, 1567-1574.

[66] Palm, D., Weiland, M., McArthur, A. G., Winiecka-Krusnell, J.*, et al.*, Developmental changes in the adhesive disk during *Giardia* differentiation. *Mol Biochem Parasitol* 2005, *141*, 199-207.

[67] Woessner, D. J., Dawson, S. C., The *Giardia* median body protein is a ventral disc protein that is critical for maintaining a domed disc conformation during attachment. *Eukaryot Cell* 2012, *11*, 292-301.

[68] Manning, G., Reiner, D. S., Lauwaet, T., Dacre, M.*, et al.*, The minimal kinome of *Giardia lamblia* illuminates early kinase evolution and unique parasite biology. *Genome Biol* 2011, *12*, R66.

[69] Lourenco, D., Andrade Ida, S., Terra, L. L., Guimaraes, P. R.*, et al.*, Proteomic analysis of the ventral disc of *Giardia lamblia*. *BMC Res Notes* 2012, *5*, 41.

[70] Tumova, P., Kulda, J., Nohynkova, E., Cell division of *Giardia intestinalis*: assembly and disassembly of the adhesive disc, and the cytokinesis. *Cell Motil Cytoskeleton* 2007, *64*, 288-298.

[71] Davids, B. J., Williams, S., Lauwaet, T., Palanca, T., Gillin, F. D., *Giardia lamblia* aurora kinase: a regulator of mitosis in a binucleate parasite. *Int J Parasitol* 2008, *38*, 353-369.

[72] Lauwaet, T., Davids, B. J., Reiner, D. S., Gillin, F. D., Encystation of Giardia lamblia: a model for other parasites. *Curr Opin Microbiol* 2007, *10*, 554-559.

[73] Nohynkova, E., Draber, P., Reischig, J., Kulda, J., Localization of gamma-tubulin in interphase and mitotic cells of a unicellular eukaryote, *Giardia intestinalis*. *Eur J Cell Biol* 2000, *79*, 438-445.

[74] Correa, G., Morgado-Diaz, J. A., Benchimol, M., Centrin in *Giardia lamblia* - ultrastructural localization. *FEMS Microbiol Lett* 2004, *233*, 91-96.

[75] Paredez, A. R., Nayeri, A., Xu, J. W., Krtkova, J., Cande, W. Z., Identification of obscure yet conserved actin-associated proteins in *Giardia lamblia*. *Eukaryot Cell* 2014, *13*, 776-784.

[76] Dagley, M. J., Dolezal, P., Likic, V. A., Smid, O.*, et al.*, The protein import channel in the outer mitosomal membrane of *Giardia intestinalis*. *Mol Biol Evol* 2009, *26*, 1941-1947.

[77] Tovar, J., Leon-Avila, G., Sanchez, L. B., Sutak, R.*, et al.*, Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* 2003, *426*, 172-176.

[78] Tachezy, J., Sanchez, L. B., Muller, M., Mitochondrial type iron-sulfur cluster assembly in the amitochondriate eukaryotes *Trichomonas vaginalis* and *Giardia intestinalis*, as indicated by the phylogeny of IscS. *Mol Biol Evol* 2001, *18*, 1919-1928.

[79] Stefanic, S., Palm, D., Svard, S. G., Hehl, A. B., Organelle proteomics reveals cargo maturation mechanisms associated with Golgi-like encystation vesicles in the early-diverged protozoan *Giardia lamblia*. *J Biol Chem* 2006, *281*, 7595-7604.

[80] Buret, A., Amat, C., Manko, A., Beatty, J.*, et al.*, *Giardia duodenalis*: new research developments in pathophysiology, pathogenesis, and virulence factors. *Curr Trop Med Rep* 2015, 1-9.

[81] Ringqvist, E., Palm, J. E., Skarin, H., Hehl, A. B.*, et al.*, Release of metabolic enzymes by *Giardia* in response to interaction with intestinal epithelial cells. *Mol Biochem Parasitol* 2008, *159*, 85-91.

[82] Eckmann, L., Laurent, F., Langford, T. D., Hetsko, M. L.*, et al.*, Nitric oxide production by human intestinal epithelial cells and competition for arginine as potential determinants of host defense against the lumen-dwelling pathogen *Giardia lamblia*. *J Immunol* 2000, *164*, 1478-1487.

[83] Stadelmann, B., Merino, M. C., Persson, L., Svard, S. G., Arginine consumption by the intestinal parasite *Giardia intestinalis* reduces proliferation of intestinal epithelial cells. *PLoS One* 2012, *7*, e45325.

[84] Cotton, J. A., Bhargava, A., Ferraz, J. G., Yates, R. M.*, et al.*, *Giardia duodenalis* cathepsin B proteases degrade intestinal epithelial interleukin-8 and attenuate interleukin-8-induced neutrophil chemotaxis. *Infect Immun* 2014, *82*, 2772-2787.

[85] de Carvalho, T. B., David, E. B., Coradi, S. T., Guimaraes, S., Protease activity in extracellular products secreted in vitro by trophozoites of *Giardia duodenalis*. *Parasitol Res* 2008, *104*, 185-190.

[86] Rodriguez-Fuentes, G. B., Cedillo-Rivera, R., Fonseca-Linan, R., Arguello-Garcia, R.*, et al.*, *Giardia duodenalis*: analysis of secreted proteases upon trophozoite-epithelial cell interaction in vitro. *Mem Inst Oswaldo Cruz* 2006, *101*, 693-696.

[87] Mantel, P. Y., Marti, M., The role of extracellular vesicles in *Plasmodium* and other protozoan parasites. *Cell Microbiol* 2014, *16*, 344-354.

[88] Twu, O., de Miguel, N., Lustig, G., Stevens, G. C.*, et al.*, *Trichomonas vaginalis* exosomes deliver cargo to host cells and mediate hostratioparasite interactions. *PLoS Pathog* 2013, *9*, e1003482.

[89] Cronemberger-Andrade, A., Aragao-Franca, L., de Araujo, C. F., Rocha, V. J.*, et al.*, Extracellular vesicles from *Leishmania*-infected macrophages confer an anti-infection cytokine-production profile to naive macrophages. *PLoS Negl Trop Dis* 2014, *8*, e3161.

[90] Hassani, K., Olivier, M., Immunomodulatory impact of *Leishmania*-induced macrophage exosomes: a comparative proteomic and functional analysis. *PLoS Negl Trop Dis* 2013, *7*, e2185.

[91] Figuera, L., Acosta, H., Gomez-Arreaza, A., Davila-Vera, D.*, et al.*, Plasminogen binding proteins in secreted membrane vesicles of *Leishmania mexicana*. *Mol Biochem Parasitol* 2013, *187*, 14-20.

[92] Silverman, J. M., Reiner, N. E., *Leishmania* exosomes deliver preemptive strikes to create an environment permissive for early infection. *Front Cell Infect Microbiol* 2011, *1*, 26.

[93] Garcia-Silva, M. R., das Neves, R. F., Cabrera-Cabrera, F., Sanguinetti, J.*, et al.*, Extracellular vesicles shed by *Trypanosoma cruzi* are linked to small RNA pathways, life cycle regulation, and susceptibility to infection of mammalian cells. *Parasitol Res* 2014, *113*, 285-304.

[94] Bayer-Santos, E., Lima, F. M., Ruiz, J. C., Almeida, I. C., da Silveira, J. F., Characterization of the small RNA content ofI extracellular vesicles. *Mol Biochem Parasitol* 2014, *193*, 71-74.

[95] Bayer-Santos, E., Aguilar-Bonavides, C., Rodrigues, S. P., Cordero, E. M.*, et al.*, Proteomic analysis of *Trypanosoma cruzi* secretome: characterization of two populations of extracellular vesicles and soluble proteins. *J Proteome Res* 2013, *12*, 883-897.

[96] Montaner, S., Galiano, A., Trelis, M., Martin-Jaular, L.*, et al.*, The role of extracellular vesicles in modulating the host immune response during parasitic infections. *Front Immunol* 2014, *5*, 433.

[97] Benchimol, M., The release of secretory vesicle in encysting *Giardia lamblia*. *FEMS Microbiol Lett* 2004, *235*, 81-87.

[98] Homan, W. L., Mank, T. G., Human giardiasis: genotype linked differences in clinical symptomatology. *Int J Parasitol* 2001, *31*, 822-826.

[99] Upcroft, J. A., McDonnell, P. A., Upcroft, P., Virulent avian *Giardia duodenalis* pathogenic for mice. *Parasitology today* 1998, *14*, 281-284.

[100] Emery, S. J., Lacey, E., Haynes, P. A., Quantitative proteomics analysis of *Giardia duodenalis* Assemblage A - a baseline for host, assemblage and isolate variation. *Proteomics* 2015.

[101] Emery, S. J., van Sluyter, S., Haynes, P. A., Proteomic analysis in *Giardia duodenalis* yields insights into strain virulence and antigenic variation. *Proteomics* 2014, *14*, 2523-2534.

[102] Williamson, A. L., O'Donoghue, P. J., Upcroft, J. A., Upcroft, P., Immune and pathophysiological responses to different strains of *Giardia duodenalis* in neonatal mice. *Int J Parasitol* 2000, *30*, 129-136.

[103] McDonnell, P. A., Scott, K. G., Teoh, D. A., Olson, M. E.*, et al.*, *Giardia duodenalis* trophozoites isolated from a parrot (*Cacatua galerita*) colonize the small intestinal tracts of domestic kittens and lambs. *Vet Parasitol* 2003, *111*, 31-46.

[104] Nolan, M. J., Jex, A. R., Upcroft, J. A., Upcroft, P., Gasser, R. B., Barcoding of *Giardia duodenalis* isolates and derived lines from an established cryobank by a mutation scanning-based approach. *Electrophoresis* 2011, *32*, 2075-2090.

[105] Upcroft, J. A., McDonnell, P. A., Gallagher, A. N., Chen, N., Upcroft, P., Lethal *Giardia* from a wild-caught sulphur-crested cockatoo (*Cacatua galerita*) established in vitro chronically infects mice. *Parasitology* 1997, *114 (Pt 5)*, 407-412.

[106] Kim, J., Bae, S. S., Sung, M. H., Lee, K. H., Park, S. J., Comparative proteomic analysis of trophozoites versus cysts of *Giardia lamblia*. *Parasitol Res* 2009, *104*, 475-479.

[107] Birkeland, S. R., Preheim, S. P., Davids, B. J., Cipriano, M. J.*, et al.*, Transcriptome analyses of the *Giardia lamblia* life cycle. *Mol Biochem Parasitol* 2010, *174*, 62-65.

[108] Morf, L., Spycher, C., Rehrauer, H., Fournier, C. A.*, et al.*, The transcriptional response to encystation stimuli in *Giardia lamblia* is restricted to a small set of genes. *Eukaryot Cell* 2010, *9*, 1566-1576.

[109] Faghiri, Z., Widmer, G., A comparison of the *Giardia lamblia* trophozoite and cyst transcriptome using microarrays. *BMC Microbiol* 2011, *11*, 91.

[110] Bernander, R., Palm, J. E., Svard, S. G., Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cell Microbiol* 2001, *3*, 55-62.

[111] Faso, C., Konrad, C., Schraner, E. M., Hehl, A. B., Export of cyst wall material and Golgi organelle neogenesis in *Giardia lamblia* depend on endoplasmic reticulum exit sites. *Cell Microbiol* 2013, *15*, 537-553.

[112] Elmendorf, H. G., Rohrer, S. C., Khoury, R. S., Bouttenot, R. E., Nash, T. E., Examination of a novel head-stalk protein family in *Giardia lamblia* characterised by the pairing of ankyrin repeats and coiled-coil domains. *Int J Parasitol* 2005, *35*, 1001-1011.

[113] Jeelani, G., et al., Metabolic profiling of the protozoan parasite Entamoeba invadens revealed activation of unpredicted pathway during encystation. *PLoS One*, 2012. 7(5): p. e37740.

[114] De Cadiz, A.E., et al., Transcriptome analysis of encystation in Entamoeba invadens. *PLoS One*, 2013. 8(9): p. e74840.

[115] Ringqvist, E., Avesson, L., Soderbom, F., Svard, S. G., Transcriptional changes in *Giardia* during host-parasite interactions. *Int J Parasitol* 2011, *41*, 277-285.

[116] Ferella, M., Davids, B. J., Cipriano, M. J., Birkeland, S. R.*, et al.*, Gene expression changes during *Giardia*-host cell interactions in serum-free medium. *Mol Biochem Parasitol* 2014, *197*, 21-23.

[117] Roxstrom-Lindquist, K., Ringqvist, E., Palm, D., Svard, S., *Giardia lamblia*-induced changes in gene expression in differentiated Caco-2 human intestinal epithelial cells. *Infect Immun* 2005, *73*, 8204-8208.

[118] Stadelmann, B., Hanevik, K., Andersson, M. K., Bruserud, O., Svard, S. G., The role of arginine and arginine-metabolizing enzymes during *Giardia* - host cell interactions in vitro. *BMC Microbiol* 2013, *13*, 256.

[119] Emery, S.J., et al., Induction of virulence factors in Giardia duodenalis independent of host attachment. *Sci Rep*, 2016. 6: p. 20765.

[120] Ansell, B. R., McConville, M. J., Ma'ayeh, S. Y., Dagley, M. J.*, et al.*, Drug resistance in *Giardia duodenalis*. *Biotechnol Adv* 2015, *33*, 888-901.

[121] Adam, R. D., Nigam, A., Seshadri, V., Martens, C. A.*, et al.*, The *Giardia lamblia* vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC genomics* 2010, *11*, 424.

[122] Solaymani-Mohammadi, S., Genkinger, J. M., Loffredo, C. A., Singer, S. M., A meta-analysis of the effectiveness of albendazole compared with metronidazole as treatments for infections with *Giardia duodenalis*. *PLoS Negl Trop Dis* 2010, *4*, e682.

[123] Lemee, V., Zaharia, I., Nevez, G., Rabodonirina, M.*, et al.*, Metronidazole and albendazole susceptibility of 11 clinical isolates of *Giardia duodenalis* from France. *J Antimicrob Chemother* 2000, *46*, 819-821.

[124] Paz-Maldonado, M. T., Arguello-Garcia, R., Cruz-Soto, M., Mendoza-Hernandez, G., Ortega-Pierres, G., Proteomic and transcriptional analyses of genes differentially expressed in *Giardia duodenalis* clones resistant to albendazole. *Infect Genet Evol* 2013, *15*, 10-17.

[125] Leitsch, D., Schlosser, S., Burgess, A., Duchene, M., Nitroimidazole drugs vary in their mode of action in the human parasite *Giardia lamblia*. *Int J Parasitol Drugs Drug Resist* 2012, *2*, 166-170.

[126] Leitsch, D., Burgess, A. G., Dunn, L. A., Krauer, K. G.*, et al.*, Pyruvate:ferredoxin oxidoreductase and thioredoxin reductase are involved in 5-nitroimidazole activation while flavin metabolism is linked to 5-nitroimidazole resistance in *Giardia lamblia*. *J Antimicrob Chemother* 2011, *66*, 1756-1765.

[127] Liu, S. M., Brown, D. M., O'Donoghue, P., Upcroft, P., Upcroft, J. A., Ferredoxin involvement in metronidazole resistance of *Giardia duodenalis*. *Mol Biochem Parasitol* 2000, *108*, 137-140.

[128] Muller, J., Schildknecht, P., Muller, N., Metabolism of nitro drugs metronidazole and nitazoxanide in *Giardia lamblia:* characterization of a novel nitroreductase (GlNR2). *J Antimicrob Chemother* 2013, *68*, 1781-1789.

[129] Muller, J., Ley, S., Felger, I., Hemphill, A., Muller, N., Identification of differentially expressed genes in a *Giardia lamblia* WB C6 clone resistant to nitazoxanide and metronidazole. *J Antimicrob Chemother* 2008, *62*, 72-82.

[130] Muller, J., Sterk, M., Hemphill, A., Muller, N., Characterization of *Giardia lamblia* WB C6 clones resistant to nitazoxanide and to metronidazole. *J Antimicrob Chemother* 2007, *60*, 280-287.

[131] Tejman-Yarden, N., Millman, M., Lauwaet, T., Davids, B. J.*, et al.*, Impaired parasite attachment as fitness cost of metronidazole resistance in *Giardia lamblia*. *Antimicrob Agents Chemother* 2011, *55*, 4643-4651.

[132] Zhao, Y., Jensen, O. N., Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* 2009, *9*, 4632-4641.

[133] Walsh, C. T., Garneau-Tsodikova, S., Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 2005, *44*, 7342-7372.

[134] Gibson, C., Schanen, B., Chakrabarti, D., Chakrabarti, R., Functional characterisation of the regulatory subunit of cyclic AMP-dependent protein kinase A homologue of *Giardia lamblia*: Differential expression of the regulatory and catalytic subunits during encystation. *Int J Parasitol* 2006, *36*, 791-799.

[135] Kim, K. T., Mok, M. T., Edwards, M. R., Protein kinase B from *Giardia intestinalis*. *Biochem Biophys Res Commun* 2005, *334*, 333-341.

[136] Ellis, J. G. t., Davila, M., Chakrabarti, R., Potential involvement of extracellular signal-regulated kinase 1 and 2 in encystation of a primitive eukaryote, *Giardia lamblia*. Stage-specific activation and intracellular localization. *J Biol Chem* 2003, *278*, 1936-1945.

[137] Abel, E. S., Davids, B. J., Robles, L. D., Loflin, C. E.*, et al.*, Possible roles of protein kinase A in cell motility and excystation of the early diverging eukaryote *Giardia lamblia*. *J Biol Chem* 2001, *276*, 10320-10329.

[138] Alvarado, M. E., Wasserman, M., Analysis of phosphorylated proteins and inhibition of kinase activity during *Giardia intestinalis* excystation. *Parasitol Int* 2010, *59*, 54-61.

[139] Lalle, M., Salzano, A. M., Crescenzi, M., Pozio, E., The *Giardia duodenalis* 14-3-3 protein is post-translationally modified by phosphorylation and polyglycylation of the C-terminal tail. *J Biol Chem* 2006, *281*, 5137-5148.

[140] Lalle, M., Camerini, S., Cecchetti, S., Sayadi, A.*, et al.*, Interaction network of the 14-3-3 protein in the ancient protozoan parasite *Giardia duodenalis*. *J Proteome Res* 2012, *11*, 2666-2683.

[141] Villen, J., Gygi, S. P., The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* 2008, *3*, 1630-1638.

[142] Thingholm, T. E., Jorgensen, T. J., Jensen, O. N., Larsen, M. R., Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat Protoc* 2006, *1*, 1929-1935.

[143] Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., Jorgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* 2005, *4*, 873-886.

[144] Zhou, H., Ye, M., Dong, J., Corradini, E.*, et al.*, Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* 2013, *8*, 461-480.

[145] Zhou, H., Ye, M., Dong, J., Han, G.*, et al.*, Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J Proteome Res* 2008, *7*, 3957-3967.

[146] de Graaf, E. L., Giansanti, P., Altelaar, A. F., Heck, A. J., Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. *Mol Cell Proteomics* 2014, *13*, 2426-2434.

[147] Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L.*, et al.*, Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 2009, *325*, 834-840.

[148] Sonda, S., Morf, L., Bottova, I., Baetschmann, H.*, et al.*, Epigenetic mechanisms regulate stage differentiation in the minimized protozoan *Giardia lamblia*. *Mol Microbiol* 2010, *76*, 48-67.

[149] Weber, K., Schneider, A., Westermann, S., Muller, N., Plessmann, U., Posttranslational modifications of alpha- and beta-tubulin in *Giardia lamblia*, an ancient eukaryote. *FEBS Lett* 1997, *419*, 87-91.

[150] Soltys, B. J., Gupta, R. S., Immunoelectron microscopy of *Giardia lamblia* cytoskeleton using antibody to acetylated alpha-tubulin. *J Eukaryot Microbiol* 1994, *41*, 625-632.

[151] Kulakova, L., Singer, S. M., Conrad, J., Nash, T. E., Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol Microbiol* 2006, *61*, 1533-1542.

[152] Kim, S. C., Sprung, R., Chen, Y., Xu, Y.*, et al.*, Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* 2006, *23*, 607-618.

[153] Mukherjee, S., Hao, Y. H., Orth, K., A newly discovered post-translational modification--the acetylation of serine and threonine residues. *Trends Biochem Sci* 2007, *32*, 210-216.

[154] Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H.*, et al.*, Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 2010, *142*, 967-980.

[155] Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W.*, et al.*, Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 2007, *131*, 58-69.

[156] Gerwig, G. J., van Kuik, J. A., Leeflang, B. R., Kamerling, J. P.*, et al.*, The *Giardia intestinalis* filamentous cyst wall contains a novel beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study. *Glycobiology* 2002, *12*, 499-505.

[157] Papanastasiou, P., McConville, M. J., Ralton, J., Kohler, P., The variant-specific surface protein of *Giardia*, VSP4A1, is a glycosylated and palmitoylated protein. *Biochem J* 1997, *322 (Pt 1)*, 49-56.

[158] Ratner, D. M., Cui, J., Steffen, M., Moore, L. L.*, et al.*, Changes in the N-glycome, glycoproteins with Asn-linked glycans, of *Giardia lamblia* with differentiation from trophozoites to cysts. *Eukaryot Cell* 2008, *7*, 1930-1940.

[159] Bushkin, G. G., Ratner, D. M., Cui, J., Banerjee, S.*, et al.*, Suggestive evidence for Darwinian Selection against asparagine-linked glycans of *Plasmodium falciparum* and *Toxoplasma gondii*. *Eukaryot Cell* 2010, *9*, 228-241.

[160] Ortega-Barria, E., Ward, H. D., Keusch, G. T., Pereira, M. E., Growth inhibition of the intestinal parasite *Giardia lamblia* by a dietary lectin is associated with arrest of the cell cycle. *J Clin Invest* 1994, *94*, 2283-2288.

[161] Jimenez, J. C., Morelle, W., Michalsky, J. C., Dei-Cas, E., Excreted/secreted glycoproteins of *G. intestinalis* play an essential role in the antibody response. *Parasitol Res* 2007, *100*, 715-720.

[162] Papanastasiou, P., Bruderer, T., Li, Y., Bommeli, C., Kohler, P., Primary structure and biochemical properties of a variant-specific surface protein of *Giardia*. *Mol Biochem Parasitol* 1997, *86*, 13-27.

[163] Davids, B. J., Reiner, D. S., Birkeland, S. R., Preheim, S. P.*, et al.*, A new family of giardial cysteine-rich non-VSP protein genes and a novel cyst protein. *PLoS One* 2006, *1*, e44.

[164] Touz, M. C., Conrad, J. T., Nash, T. E., A novel palmitoyl acyl transferase controls surface protein palmitoylation and cytotoxicity in *Giardia lamblia*. *Mol Microbiol* 2005, *58*, 999-1011.

[165] Gallego, E., Alvarado, M., Wasserman, M., Identification and expression of the protein ubiquitination system in *Giardia intestinalis*. *Parasitol Res* 2007, *101*, 1-7.

[166] Nino, C. A., Chaparro, J., Soffientini, P., Polo, S., Wasserman, M., Ubiquitination dynamics in the early-branching eukaryote *Giardia intestinalis*. *Microbiologyopen* 2013, *2*, 525-539.

[167] Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C.*, et al.*, A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* 2003, *21*, 921-926.

[168] Nesvizhskii, A. I., Vitek, O., Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* 2007, *4*, 787-797.

[169] Nesvizhskii, A. I., Aebersold, R., Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. *Drug Discov Today* 2004, *9*, 173-181.

[170] Mahoney, D. W., Therneau, T. M., Heppelmann, C. J., Higgins, L.*, et al.*, Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J Proteome Res* 2011, *10*, 4325-4333.

[171] Ting, L., Rad, R., Gygi, S. P., Haas, W., MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* 2011, *8*, 937-940.

[172] Picotti, P., Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 2012, *9*, 555-566.

[173] Prakash, A., Tomazela, D. M., Frewen, B., Maclean, B.*, et al.*, Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. *J Proteome Res* 2009, *8*, 2733-2739.

[174] Gillet, L. C., Navarro, P., Tate, S., Rost, H.*, et al.*, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 2012, *11*, O111 016717.

[175] Rost, H. L., Rosenberger, G., Navarro, P., Gillet, L.*, et al.*, OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 2014, *32*, 219-223.

[176] Prucca, C. G., Slavin, I., Quiroga, R., Elias, E. V.*, et al.*, Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* 2008, *456*, 750-754.

[177] Chen, X. S., Penny, D., Collins, L. J., Characterization of RNase MRP RNA and novel snoRNAs from *Giardia intestinalis* and *Trichomonas vaginalis*. *BMC genomics* 2011, *12*, 550.

[178] Vranych, C. V., Rivero, M. R., Merino, M. C., Mayol, G. F.*, et al.*, SUMOylation and deimination of proteins: two epigenetic modifications involved in *Giardia* encystation. *Biochim Biophys Acta* 2014, *1843*, 1805-1817.

[179] Militello, K. T., Refour, P., Comeaux, C. A., Duraisingh, M. T., Antisense RNA and RNAi in protozoan parasites: working hard or hardly working? *Mol Biochem Parasitol* 2008, *157*, 117-126.

[180] Teodorovic, S., Walls, C. D., Elmendorf, H. G., Bidirectional transcription is an inherent feature of *Giardia lamblia* promoters and contributes to an abundance of sterile antisense transcripts throughout the genome. *Nucleic Acids Res* 2007, *35*, 2544-2553.

[181] Merino, M. C., Zamponi, N., Vranych, C. V., Touz, M. C., Ropolo, A. S., Identification of *Giardia lamblia* DHHC proteins and the role of protein S-palmitoylation in the encystation process. *PLoS Negl Trop Dis* 2014, *8*, e2997.

[182] Saric, M., Vahrmann, A., Niebur, D., Kluempers, V.*, et al.*, Dual acylation accounts for the localization of {alpha}19-giardin in the ventral flagellum pair of *Giardia lamblia*. *Eukaryot Cell* 2009, *8*, 1567-1574.

[183] Sinha, A., Datta, S. P., Ray, A., Sarkar, S., A reduced VWA domain-containing proteasomal ubiquitin receptor of *Giardia lamblia* localizes to the flagellar pore regions in microtubule-dependent manner. *Parasit Vectors* 2015, *8*, 737.

[184] Lopez, A. B., Hossain, M. T., van Keulen, H., *Giardia intestinalis* glucosamine 6-phosphate isomerase: the key enzyme to encystment appears to be controlled by ubiquitin attachment. *J Eukaryot Microbiol* 2002, *49*, 134-136.

[185] Ellis, J. E., Wyder, M. A., Jarroll, E. L., Kaneshiro, E. S., Changes in lipid composition during in vitro encystation and fatty acid desaturase activity of *Giardia lamblia*. *Mol Biochem Parasitol* 1996, *81*, 13-25.

[186] Lujan, H. D., Mowatt, M. R., Chen, G. Z., Nash, T. E., Isoprenylation of proteins in the protozoan *Giardia lamblia*. *Mol Biochem Parasitol* 1995, *72*, 121-127.

[187] Knodler, L. A., Noiva, R., Mehta, K., McCaffery, J. M.*, et al.*, Novel protein-disulfide isomerases from the early-diverging protist *Giardia lamblia*. *J Biol Chem* 2000, *275*, 28339.

[188] Appfel, A., Tom-Moy, M., Smart, B., *Human Proteome World Congress*, Sydney 2009, p. 238.

[189] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* 2009, *6*, 359-362.

[190] Scherl, A., Shaffer, S. A., Taylor, G. K., Kulasekara, H. D.*, et al.*, Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. *Anal Chem* 2008, *80*, 1182-1191.

[191] George, I. S., Fennell, A. Y., Haynes, P. A., Protein identification and quantification from riverbank grape, Vitis riparia: Comparing SDS-PAGE and FASP-GPF techniques for shotgun proteomic analysis. *Proteomics* 2015, *15*, 3061-3065.

[192] Thompson, R. C., The zoonotic significance and molecular epidemiology of *Giardia* and giardiasis. *Vet Parasitol* 2004, *126*, 15-35.

[193] Aebersold, R., Burlingame, A. L., Bradshaw, R. A., Western blots versus selected reaction monitoring assays: time to turn the tables? *Mol Cell Proteomics* 2013, *12*, 2381-2382.

[194] Mann, M., Can proteomics retire the western blot? *J Proteome Res* 2008, *7*, 3065.

[195] Gupta, N., Pevzner, P. A., False discovery rates of protein identifications: a strike against the two-peptide rule. *J Proteome Res* 2009, *8*, 4173-4181.

[196] Jerlstrom-Hultqvist, J., Ankarklev, J., Svard, S. G., Is human giardiasis caused by two different Giardia species? *Gut microbes* 2010, *1*, 379-382.