# PROTEOMIC AND TRANSCRIPTOMIC ANALYSIS OF THE PROTOZOAN PARASITE *NEOSPORA CANINUM*

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by Sarah J Vermont

September 2012

## AUTHOR'S DECLARATION

Apart from help and advice acknowledged, this thesis represents the unaided work of the author

.....

Sarah J Vermont

September 2012

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Professor Jonathan Wastling and Dr. Andy Jones, not only for accepting me as a student, but for all their expertise, advice and support along the way. They have not been the only people to whom I have turned for help: to everyone (past and present) in Infection Biology; Nadine, Sanya, Beccy, Sophie, Stu, Dong, Andy, Hos, Sarah T, Corrado, Mariwan and Poom; a massive thank you. Similarly, to members of the PFG group and especially Duncan, thank you for your patience and understanding in mass spectrometry assistance! Without Gianluca, Achuchuthan, Kat, Ritesh and the rest of the Bioinformatics group I could never had understood my data. A special thank you is due to Dr. Iain Young, not only for inspiring me to attempt a PhD in the first place, but also for being a fantastic and supportive boss in my three years as a halls tutor.

My PhD has involved collaborations on a number of projects: Professor Arnab Pain, Adam, Peter, Professor Andrew Hemphill, Thierry; thank you all for your involvement and fruitful discussions. To Steffi and Professor Jonathan Howard, thank you for making me welcome in your labs in Cologne and for all your help with the ROP18 work.

Thank you also to the BBSRC and Intervet for providing the funding that allowed me to carry out this work and to the University of Liverpool for 4 more brilliant years.

My partners in crime, Emma and Jenna, along with Alex and Monika, have made studying for my PhD especially enjoyable and entertaining; Maz and Luke have been so generous in letting me invade their home these last few weeks - thank you all.

Finally, my parents have provided unending support and encouragement and have provided a home to my continually expanding family of horses and other animals, without all of whom to distract me at times and provide escapism; I would surely never have managed. So, Mum and Dad, I can never thank you enough.

# ABSTRACT

*Neospora caninum* is an economically significant parasitic protozoan causing the disease neosporosis in cattle and dogs. Although a close relative of the zoonotic apicomplexan *Toxoplasma gondii*, the two organisms exhibit differing host ranges and infection dynamics. *T. gondii* is a model organism that has been much studied, and a great deal is known about the genes and proteins involved when it invades a host cell. This thesis explores protein expression in the proliferative and invasive tachyzoite stage of *N. caninum*, in particular the expression of proteins pertaining to the apical complex of organelles; those responsible for entry and establishment within a host cell. Almost 20 % of the predicted proteome has been identified by this analysis to be expressed in the tachyzoite stage, with approximately 50 % of the predicted repertoire of apical proteins being detected.

The discovery of differences between these two parasites' highly syntenic genomes could lead to a better understanding of the process by which *T. gondii* is able to cause disease in humans, while *N. caninum* has not been observed to do so. One finding of the recent genome sequencing and annotation project in *N. caninum* was that a key *T. gondii* virulence determinant, rhoptry gene 18 (*ROP18*) was pseudogenised in *N. caninum*. This finding was investigated further in this thesis to demonstrate that the pseudogenisation of *ROP18* was conserved across a range of *N. caninum* isolates and that *in vitro*, *N. caninum* was not able to subvert the murine interferon-gamma (IFN-γ) immune response using ROP18 in the way that virulent *T. gondii* tachyzoites do.

The tissue-dwelling Coccidia have a multi-stage life cycle which includes a latent tissue cyst-encapsulated stage called the bradyzoite. Tachyzoites convert to this more quiescent form when induced by cellular stress, and are able to remain as such for long periods, even years. At times of weakened host immunity, bradyzoites can recrudesce to produce an active infection, which can cross the placenta in a pregnant animal to infect the foetus. This a major route by which *N. caninum* infection is maintained within cattle herds, therefore the biology of stage conversion from tachyzoite to bradyzoite and vice-versa is of interest to researchers. An RNA-Seq analysis of cultured tachyzoites and bradyzoites identified a marked reduction in rhoptry gene expression, and differing expression profiles of other invasion-related genes from the micronemes and dense granules.

Overall, these data identify proteins released from the apical organelles in *N. caninum* and give an insight into the different repertoires expressed by the tachyzoite and bradyzoite life stages. Furthermore, a comparison between *N. caninum* and *T. gondii* predicted apical proteomes indicates that although most genes are shared in a one-to-one orthologous relationship between the two organisms, there are a small number of differences which may turn out to be important to the biology of the parasite, as in the case of *ROP18*.

# CONTENTS

CHAPTER 7: DISCUSSION

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| 1-DE | One-dimensional electrophoresis |
| 2-DE | Two-dimensional electrophoresis |
| AMA | Apical membrane antigen |
| APS | Ammonium persulphate |
| BLAST | Basic Local Alignment Search Tool |
| bp | Base pairs |
| BSA | Bovine serum albumin |
| cDNA | Complementary DNA |
| CID | Collision-induced dissociation |
| D.E. | Differentially expressed |
| DAPI | 4′,6-Diamidine-2′-phenylindole dihydrochloride |
| $ddH_20$ | Milli-Q grade water, resistivity 18.2 Ω.cm |
| DNA | Deoxyribonucleic acid |
| DO media | Dulbecco's modified Eagle's medium, 2 Mm glutamine, 10 mM HEPES |
| DTT | Dithiothrietol |
| EDTA | Ethylenediaminetetraacetic acid |
| ESA | Excretory/secretory antigen sample |
| ESI | Electrospray ionisation |
| EST | Expressed sequence tag |
| EtOH | Ethanol |
| EuPathDB | Eukaryotic Pathogen database |
| FIKK | Serine/threonine protein kinase in *Plasmodium spp* |
| FPKM | Fragments per kilobase of exon per million fragments mapped |
| GO | Gene Ontology |
| GRA | Dense granule protein |
| GTPase | Guanosine-triphosphatase |
| HCl | Hydrochloric acid |
| HMM | Hidden Markov model |

| | |
|---|---|
| HPLC | High-performance liquid chromatography |
| IAA | Iodoacetemide |
| IFN-γ | Interferon-gamma |
| IL-10 | Interleukin-ten |
| IL-12 | Interleukin-twelve |
| IMDM | Iscove's-modified Dulbecco's medium |
| IRG | Immunity-related GTPase |
| LAMP | Liverpool Library of Apicomplexan Metabolic Pathways |
| LC | Liquid chromatography |
| LTQ | Linear-trap quadrupole |
| m.o.i. | Multiplicity of infection |
| *m/z* | Mass-to-charge ratio |
| M2AP | MIC2-associated protein |
| MALDI | Matrix-assisted laser desorption-ionisation |
| MAPK | Mitogen-activated protein kinase |
| MAR | Microneme adhesive repeat |
| MCP | Microneme adhesive repeat-containing protein |
| MEF | Mouse embryonic fibroblast |
| $MgCl_2$ | Magnesium chloride |
| MIC | Microneme protein |
| MIPS | Munich Information Center for Protein Sequences |
| MOPS | 3-(N-morpholino)propanesulfonic acid buffer |
| mRNA | Messenger RNA |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| MudPIT | Multidimensional protein-identification technology |
| NTPase | Nucleosidetriphosphatase |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate-buffered saline |
| PCR | Polymerase chain reaction |

| | |
|---|---|
| PlasmoDB | *Plasmodium* database, member of EuPathDB |
| Poly(A) RNA | Polyadenylated RNA |
| PV | Parasitophorous vacuole |
| PVM | Parasitophorous vacuole membrane |
| QTL | Quantitative trait loci |
| R | Rhoptry sample |
| R/DG | Rhoptry/Dense granule sample |
| RNA | Ribonucleic acid |
| RNA-Seq | RNA sequencing |
| RON | Rhoptry neck protein |
| ROP | Rhoptry protein |
| ROPK | Rhoptry kinase |
| RP | Reverse phase |
| RPKM | Reads per kilobase of exon per million reads mapped |
| SCX | Strong cation-exchange |
| S/T domain | Serine/threonine domain |
| SAG | Surface antigen |
| SDS | Sodium dodecyl suphate |
| SNP | Single nucleotide polymorphism |
| SRS | SAG-related surface protein |
| SUB | Subtilisin protease |
| TCA cycle | Tricarboxcylic acid cycle |
| TEMED | Tetramethyladenamine |
| TFA | Trifluoroacetic acid |
| TOF | Time-of-flight |
| ToxoDB | *Toxoplasma* and *Neospora* database, member of EuPathDB |
| UTR | Untranslated region |

# CHAPTER 1: INTRODUCTION

## 1.1 *Neospora caninum*

*Neospora caninum* is a parasitic protozoan which invades and replicates in mammalian host cells, causing the important veterinary disease neosporosis. It is a member of the eukaryotic phylum Apicomplexa, along with some important medical and veterinary pathogens worldwide: such as *Plasmodium* (the causative agent of malaria), *Toxoplasma* (toxoplasmosis), *Cryptosporidium* (cryptosporidiosis) and *Eimeria* (coccidiosis) among others. The Apicomplexa are named with reference to their specialised 'apical complex' – a collection of organelles with roles in the process of host-cell invasion (Dubremetz *et al.* 1998a). These parasites are obligately intracellular for most of their life cycle, so an understanding of the process of invasion could lead to the development of drugs and vaccines to interfere with this vital step.

There are two species of *Neospora*: *N. caninum* and *N. hughesi;* which, alongside *Toxoplasma gondii,* two species of *Hammondia* and six species of *Besnoitia*, make up the subfamily Toxoplasmatinae, of the Sarcosystidae (Hemphill & Gottstein 2000; Mugridge *et al.* 1999; Tenter & Johnson 1997). *N. caninum* has a canid definitive host (Lindsay, Dubey & Duncan 1999; McAllister *et al.* 1998) and can infect a range of mammals as intermediate hosts, although it has the most significant impact as a pathogen of cattle. In addition to dogs (*Canis familiaris*), several sylvatic definitive hosts may exist; coyotes (*Canis latrans*) and grey wolves (*Canis lupus*) (Dubey *et al.* 2011) have been confirmed to shed oocysts (Gondim *et al.* 2004), while foxes remain unconfirmed as definitive hosts (Almeria *et al.* 2002).

## 1.2 Lifecycle and epidemiology

The coccidian life cycle is complex and involves multiple life stages. The oocyst, or egg, sporulates and excysts as sporozoites to infect the intestinal cells. The tachyzoite is the highly proliferative form that disseminates around the host. A latent phase, the bradyzoite, can sit quiescent, encysted in the host tissues until such time as the host's immunity is weakened and the infection reactivates. Sexual recombination takes place in the gut of the definitive host, whilst clonal population expansion occurs in the tissues of an intermediate host. Dogs can become infected by a number of routes: by ingesting oocysts or tissue that contains bradyzoite cysts (Lindsay, Dubey & Duncan 1999; McAllister *et al.* 1998); probably by ingestion of tachyzoites (Dijkstra *et al.* 2001), or from transplacental transmission of tachyzoites to

developing foetuses (Barber & Trees 1998; Dubey *et al.* 1988b; Dubey, Koestner & Piper 1990; Dubey & Lindsay 1989; Dubey, Schares & Ortega-Mora 2007). The exact process of sexual reproduction is not yet proven, but it is thought that oocysts are produced from gametogeny and syngamy in the intestinal epithelial cells (Williams *et al.* 2009). They are then shed via the faeces into the environment (McAllister *et al.* 1998), where they sporulate to produce two sporocysts, each containing four sporozoites. When intermediate hosts ingest material contaminated by oocysts (horizontal transmission), the sporozoites hatch and differentiate into tachyzoites, which likely disseminate via the circulatory system by invading mononuclear phagocytic cells (Gibney *et al.* 2008; Williams *et al.* 2009). Tachyzoites can invade a range of cell types but appear to have a predilection for central nervous system (CNS) and placental tissues (Dubey 2003). They replicate within parasitophorous vacuoles (PVs) by endodyogeny (Dubey *et al.* 2004; Speer & Dubey 1989), a form of internal budding resulting in two daughter cells (Roberts, Schmidt & Janovy 2009). Figure 1.1 shows a representation of the *N. caninum* life cycle, in cattle, the intermediate host of major veterinary importance.

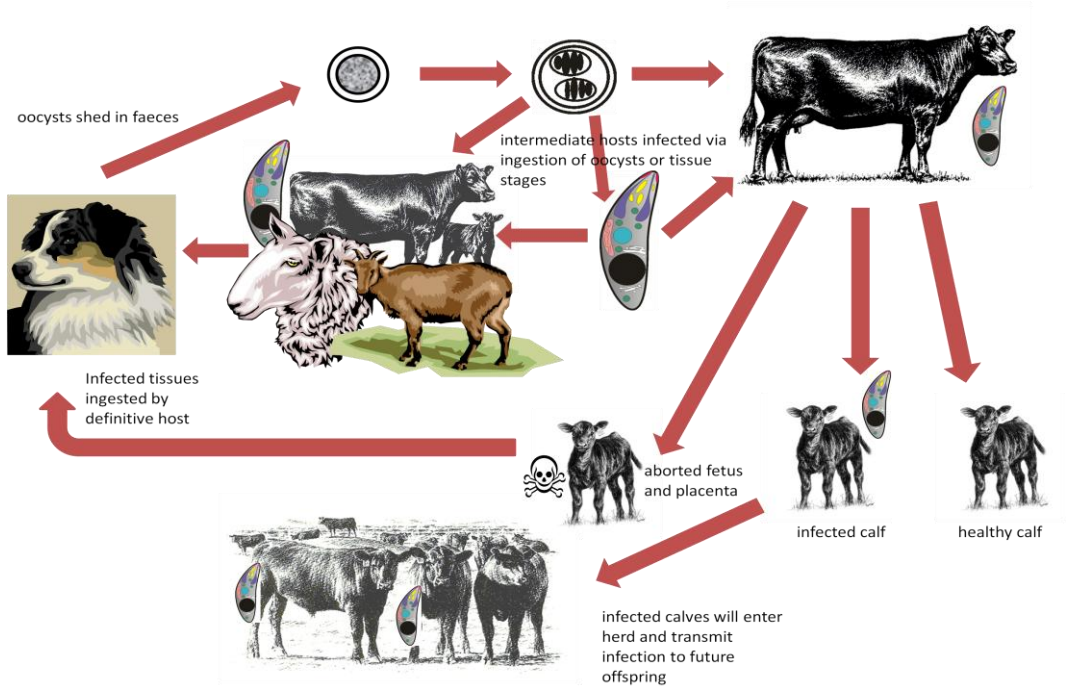



**Figure 1.1:** Lifecycle of *Neospora caninum*. Information collated from (Dubey 1999a; Dubey 1999b; Intervet website). An oocyst, after being shed by a canid definitive host, sporulates and is ingested by an intermediate host; where sporozoites are released, invade cells lining the gut and form tachyzoites which then migrate to tissues. Infection transmission can then occur horizontally by a definitive host ingesting infected tissue and the cycle proceeding through the sexual stages, or vertically via the placenta to successive generations of intermediate host.

Vertical transmission of an infection, passed down to subsequent generations via the placenta, can be from a newly acquired exogenous infection or recrudescence of an endogenous infection (Trees & Williams 2005). The relative importance of the different routes of transmission has been under scrutiny but it appears that both horizontal and vertical routes play a role in maintaining infection levels and therefore ensuring the continuation of the parasite population *(*Williams *et al*. 2009). Endogenous vertical transmission efficiencies in dairy herds have been reported to be as high as 78-95% (Davison, Otter & Trees 1999a; Pare, Thurmond & Hietala 1997) however and this is considered to be the principle route by which *N. caninum* transmission occurs (Dubey, Buxton & Wouda 2006; Dubey, Schares & Ortega-Mora 2007; Williams *et al.* 2009). As a result, a vaccination that could prevent recrudescence of bradyzoites in cows and heifers would have a significant impact on the population of this parasite.

### 1.3 Economic impact of neosporosis

*Neospora caninum* causes the disease neosporosis, which manifests most significantly in cattle, especially from an economic perspective, but also causes pathology in dogs.

Canine neosporosis manifests as a debilitating neuromuscular dysfunction, mainly affecting puppies and young dogs under six months, although clinical disease can occur in older dogs (Dubey 2005). The infection can be obtained from ingesting infected tissues (such as an intermediate host) as well as congenitally transmitted, and tissue cysts can remain even if there is clinical recovery (Dubey *et al.* 2007).

Bovine neosporosis impacts upon both the dairy and beef industries worldwide. Neosporosis is the UK's most frequently diagnosed cause of abortion in dairy cattle; estimated at 12.5 %, or 6000 abortions per year (Davison, Otter & Trees 1999b; Davison, Otter & Trees 1999c). This figure will be somewhat lower than the prevalence of *N. caninum* as there are many cases where presence of the parasite does not result in abortion, or when an alternative abortifacient may also be identified (Hemphill & Gottstein 2000; Thurmond, Hietala & Blanchard 1997, 1999; Wouda *et al*. 1997). A recent UK investigation into prevalence in cattle herds found that over 90 % of herds had at least one cow seropositive for *N. caninum*; the median seroprevalence in positive herds was 10%, but ranged from 0.4 - 58.8 % (Woodbine *et al.* 2008). Considerable economic losses result (Trees *et al.* 1999) from a combination of abortion and stillbirths, reduced milk yield (Hobson *et al*. 2002), reduced weight gain (Barling *et al.* 2000), reduced value of breeding stock (Trees *et al.* 1999) and culling (Thurmond & Hietala

1996, 1997). With dairy farmers under enormous strain in today's market to make any profit at all, it is of the essence that a vaccine be developed to effectively control the spread of this disease.

## 1.4 Host cell invasion and the apical complex

The apical complex of organelles comprises the micronemes, rhoptries and dense granules (Figure 1.2). Rhoptries are large, club shaped secretory organelles with tapered necks which act as ducts to discharge their contents through (Bannister *et al.* 2003; Bradley et al. 2005). Clustered around these are the smaller, but more numerous, micronemes, which can number up to 150 in *N. caninum* (Dubey & Lindsay 1996). Dense granules can be found in the posterior end of the parasite as well as the apical end, and are more rounded than micronemes (Bannister *et al.* 2003; Speer *et al.* 1999). The conoid is a fibrous spiral-shaped structure whose extrusion plays a role in invasion (Del Carmen *et al.* 2009) and the apicoplast is a relict plastid unique and essential to most Apicomplexa (not present in *Cryptosporidium spp* (Zhu, Marchewka & Keithly 2000)), but with exact function unknown (Foth & McFadden 2003). Also present in the fully formed tachyzoite, are 22 subpellicular microtubules, a nucleus with nucleolus, centrioles, a polar ring, two apical rings, up to three mitochondria, ribosomes, golgi complex, rough and smooth endoplasmic reticulum, lipid bodies and a posterior pore, all contained within a pellicle comprised of a plasmalemma and a single inner membrane (Speer & Dubey 1989). A diagram of a tachyzoite, with organelles labelled, can be seen in Figure 1.2.

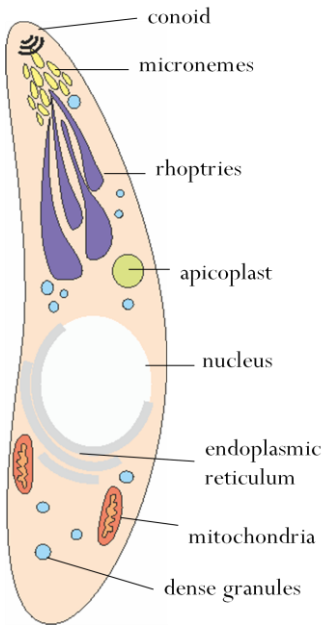



**Figure 1.2:** Diagram of a *Neospora caninum* tachyzoite. Organelles are labelled, diagram not to scale: a tachyzoite measures approximately 7.5 x 2 µm (Speer *et al*. 2009).

Tachyzoite invasion into a host cell is a key stage towards implementing the pathologies associated with apicomplexan parasites. It is fairly well conserved within the phylum and consists of a complex sequence of events (Mital & Ward 2008) summarized in Figure 1.3.

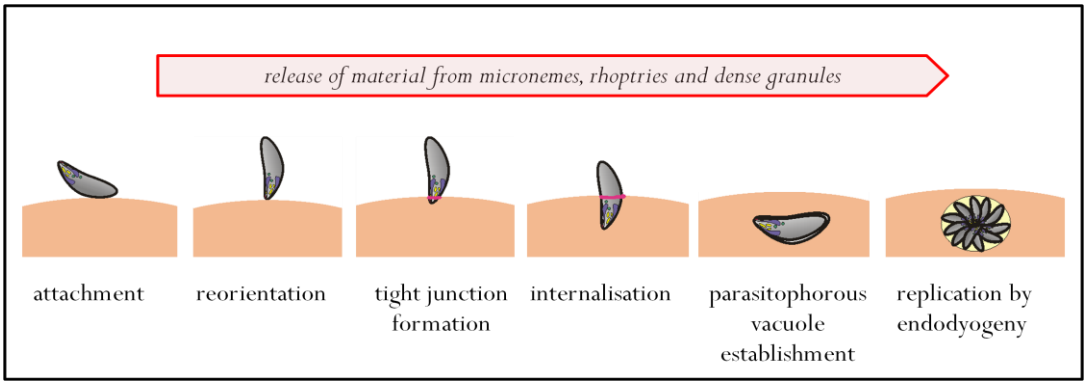



**Figure 1.3:** Summary of the invasion process in Apicomplexa. The parasite attaches to a host cell and reorientates so that the apical tip is aligned with the host cell surface. Entry to the host cell is an active process reliant on parasite proteins. Once internalised, a parasitophorous vacuole is established within which replication occurs.

To begin with, the tachyzoite moves over the surface of the cell by a process termed gliding motility (reviewed in (Kappe *et al. 2004)*) which is an active process driven by an internal actin-myosin motor, not dependent upon flagella or pseudopodia (Dobrowolski & Sibley 1996). To invade, the parasite must first orientate itself so that it is perpendicular to the host

cell surface (Dubremetz *et al*. 1998b; Nichols, Chiappino & O'Connor 1983) before forming an intimate 'moving junction' with the host cell membrane, dependent upon secreted proteins from the micronemes and rhoptries (Besteiro *et al.* 2009). The whole process of invasion is rapid, lasting only a few seconds, but the organelles are thought to discharge their contents sequentially; micronemes are quickly followed by rhoptries, then dense granules, whose proteins are involved in establishment of the host-parasite interaction once the zoite has internalised within the parasitophorous vacuole (PV) (reviewed in (Dubremetz *et al.* 1998b)). The parasitophorous vacuole membrane (PVM) is comprised mostly of host cell plasmalemma with integrated parasite proteins (Beckers *et al.* 1994), host transmembrane proteins are excluded so that the vacuole is unable to fuse with lysosomes (Mordue *et al.* 1999a; Mordue *et al.* 1999b) and is hence protected from host immune defences. Host organelles, including mitochondria and endoplasmic reticulum, accumulate around the PV (de Melo, de Carvalho & de Souza 1992), with parasite PVM-host organelle interactions mediated by parasite proteins (Sinai & Joiner 2001). Whilst the invasion process has been well characterised in *T. gondii* and other Apicomplexa such as *Plasmodium spp*, less is known about whether *N. caninum* behaves in exactly the same way.

### 1.4.1 Micronemes

The micronemes release their contents upon contact with a host cell (Carruthers & Sibley 1997), this has been shown to be controlled by intracellular calcium (Carruthers & Sibley 1999) which is essential for host cell invasion in *T. gondii* (Carruthers, Giddings & Sibley 1999). Microneme proteins in *T. gondii* and *N. caninum* are mostly known as MICs. A number of MICs have been implicated in the invasion process in *T. gondii*. Many MICs, including MIC1, have adhesive domains facilitating attachment to host cells (Garnett *et al.* 2009). MIC 1 associates with MIC4 and MIC6 to form a complex that is critical to invasion (Reiss *et al.* 2001; Saouros *et al.* 2005). A subset of MICs, called MCPs (microneme adhesive repeat containing proteins), have been found to bind sialic acid which is a component of glycoproteins and glycolipids in host cell membranes, and is often a target for host cell entry by pathogens (Friedrich *et al.* 2010). Different arrangement of the MAR type I and type II domains in *T. gondii, N. caninum and E. tenella* result in different binding affinities with a range of sialic acid types (Cowper, Matthews & Tomley 2012). *T. gondii* MIC2 is released at the junction between the host and the parasite, where it gets translocated along the parasite from the apical to posterior end, where it is then shed as the parasite internalises (Carruthers, Giddings & Sibley 1999). MIC2 binds the MIC2-associated protein, M2AP to form a complex that is required for

efficient host cell entry (Jewett & Sibley 2004). The subtilisin protease *T. gondii* SUB1 is required for host cell surface processing of MICs upon their release from the micronemes and a deficiency in SUB1 results in poor cell attachment and gliding motility (Lagal *et al.* 2010).

Cowper *et al.* (2012) predict that the MICs are likely to be responsible for the range of tissue type and host cell tropisms exhibited by different coccidian Apicomplexa and suggest that an HxT motif, lacking in the *N. caninum* orthologue of *T. gondii* MIC13, could potentially be a basis for the difference in zoonotic capability between the two organisms, as it mediates binding of α2,9-disialyl which has been identified on some human cells.

### 1.4.2 Rhoptries

The rhoptries are not only essential to host cell invasion but a number of rhoptry kinases have recently been implicated as determinants of virulence in *T. gondii*. Rhoptry proteins primarily comprise the RONs, or rhoptry neck proteins, which localise to the long neck-like part of the organelle under immunostaining (Bradley *et al.* 2005) and the ROPs, which are secreted from the bulbous rhoptry body. RON2, RON4, RON5 and RON8 form a complex in *T. gondii* with the microneme protein Apical membrane antigen 1 (AMA1) that localises to the moving junction, forming a tight contact between the parasite and host membranes (Alexander *et al.* 2005; Besteiro, Dubremetz & Lebrun 2011; Besteiro *et al.* 2009; Tyler & Boothroyd 2011).

The ROP proteins tend to be released, and function, slightly later in the invasion process than the RONs, and many possess kinase domains. ROP1 is secreted into the PV (Saffer *et al.* 1992) and ROP2 into the PVM, where it mediates attachment of host mitochondria, probably to enable acquisition of nutrients (Beckers *et al.* 1994; Sinai & Joiner 2001). ROP2 is actually present in triplicate in the *T. gondii* genome, its three paralagous genes are named ROP2a, ROP2b and ROP8 (Beckers, Wakefield & Joiner 1997). Pernas & Boothroyd (2010) called into question the role of ROP2/ROP8 in associating host mitochondria with the PVM and demonstrated that knock out mutants were able to retain this function. Interestingly, *N. caninum* has been observed to recruit host mitochondria to its PVM at a much lower frequency than *T. gondii* and there is uncertainty around the existence of such an event in *N. caninum* (Lindsay et al. 1993; Pernas & Boothroyd 2010; Speer et al. 1999). The role of ROP2/ROP8 remains unclear, as the triple-knock out strains demonstrated no observable phenotype, so it may be that the wider ROP2 family allows for a certain level of redundancy (Pernas & Boothroyd 2010).

ROP5, ROP16 and ROP18 have all been shown to confer virulence in different strains of *T. gondii:* ROP5 exists in different copy numbers between different strains (Reese *et al.* 2011; Reese & Boothroyd 2011), ROP16, which traffics to the host cell nucleus, phosphorylates STAT6 and in this way modulates host interleukin-12, a proinflammatory cytokine (Denkers *et al.* 2012; Saeij *et al.* 2007). ROP18 subverts the host interferon-gamma (IFN-γ) immune response by phosphorylating an immunity related GTPase that would otherwise cause rupture of the PV. ROP38 is expressed at higher levels in less virulent *T. gondii* strains and down regulates the MAPK (Mitogen-activated protein kinase) pathway of the host cell to interfere with apoptosis and cell proliferation (Peixoto *et al.* 2010). Most of these findings have yet to be confirmed in *N. caninum*, though it is hypothesised that host cell invasion is largely analogous to that observed in *T. gondii*. An analysis of protein expression in *N. caninum* tachyzoites will aid understanding of the similarities and differences in the invasion process between the two organisms.

### 1.4.3 Dense Granules

The GRA family of dense granule proteins (reviewed in Nam (2009), regarding *T. gondii*) are largely secretory proteins that also contain transmembrane domains. GRA1 is secreted into the PV lumen (Sibley *et al.* 1995) while GRA2, GRA4 and GRA6 are associated with an intravacuolar membranous network (Labruyere *et al.* 1999). GRA3, GRA5 and GRA10 are secreted into the PVM for direct interaction with the host cytoplasm (Ahn *et al.* 2006). *T. gondii* GRA7 gets phosphorylated during host cell invasion and associates with ROP2 and ROP4, probably on the PVM (Dunn *et al.* 2008). These proteins, along with ROP1, GRA3 and GRA1, form punctate strands ('beads-on-a-string' (Ravindran & Boothroyd 2008)) extending into the host cell, to and from the nascent PV, that are observed to partially overlap with each other by immunostaining (Dunn *et al.* 2008). These strands may be organised along microtubules and are likely to be involved in intracellular trafficking of proteins (Ravindran & Boothroyd 2008). The dense granule proteins are therefore involved with establishing and maintaining PV during and after host cell invasion and provide a mechanism for molecular interaction between the parasite and the host.

## 1.5 *Neospora* and *Toxoplasma* – similarities and differences

*N. caninum* was only differentiated from *T. gondii* in 1984 (Bjerkas, Mohn & Presthus), and described and characterised in 1988 (Dubey *et al.*); hence current understanding of its biology is somewhat behind that of *T. gondii,* which has been recognised as a protozoan parasite since

the beginning of the 20th century (Hemphill & Gottstein 2000). However, due to their close relatedness and the many genetic, physiological and morphological similarities, much research in *N. caninum* has benefitted from prior knowledge of *T. gondii* biology.

However, despite the close synteny of their genomes (DeBarry & Kissinger 2011; Reid *et al.* 2012), there are notable differences between *T. gondii* and *N. caninum*, which has led to them being placed in separate genera (Dubey *et al.* 1988a). One example is their host specificity. While *T. gondii* appears to be able to infect almost any warm blooded cell type, *N. caninum* exhibits more specificity: it has so far not been demonstrated to be zoonotic (Dubey, Schares & Ortega-Mora 2007; McCann *et al.* 2008), despite readily infecting human cells *in vitro*. Elucidating the molecular basis of this difference could be important towards discovering ways of preventing infection.

Although apparently identical by light microscopy, there are ultrastructural differences between the two organisms (Speer & Dubey 1989; Speer *et al.* 1999). *N. caninum* tachyzoites are slightly larger than *T. gondii*; at 7.5 x 2 μm; and numbers of organelles vary between the two. Micronemes are more numerous in *N. caninum*, while dense granules are numerous in both (Speer *et al.* 1999). *N. caninum* possess six to sixteen rhoptries compared to four to ten in *T. gondii* (Speer *et al.* 1999). Compared to tachyzoites, bradyzoites of both species contain fewer rhoptries. Interestingly, the electron-dense rhoptries of *N. caninum* tachyzoites are similar to those of the mature bradyzoites of both species, whereas those of *T. gondii* are spongy and electron-lucent (Speer *et al*. 1999). *T. gondii* cyst walls are smooth and less than 0.5 μm thick: *N. caninum* tissue cysts have an irregular, thick cyst wall (≤ 4 μm) but are smaller when compared to *T.gondii* tissue cysts, containing 20-100 and 50-500+ bradyzoites respectively (Speer et al. 1999). Figure 1.4 (adapted from Dubey, Buxton & Wouda (2006) and Dubey, Lindsay & Speer (1998)) shows histological images of a tachyzoite and a bradyzoite tissue cyst of *N. caninum,* the latter with a thick cyst wall; and a tissue cyst of *T. gondii* with a comparatively thinner cyst wall.

The definitive hosts of *T. gondii* are cats, capable of disseminating millions of oocysts in their faeces (Dubey & Frenkel 1972). In comparison, dogs produce relatively few oocysts of *N. caninum* (Gondim, McAllister & Gao 2005) measured a total of approximately 500 000 shed by a dog), which has led to speculation that there exists another, more efficient definitive host (Lindsay, Ritter & Brake 2001), although the dog is widely accepted as the definitive host and

none further than other than additional members of the *Canis* genus have been identified to date.

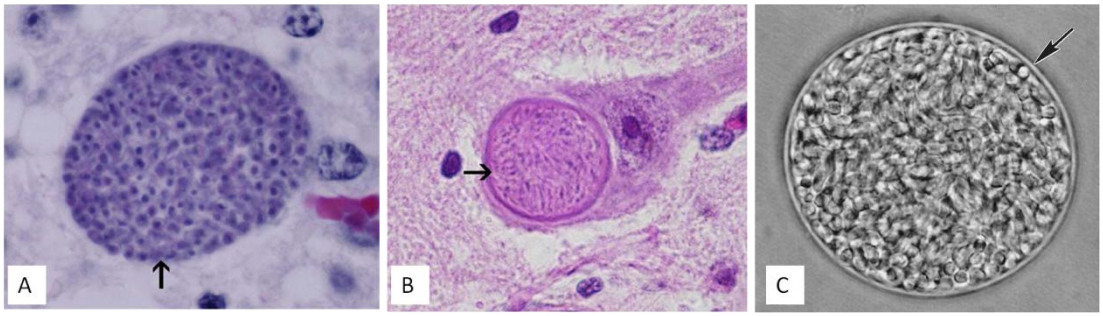



**Figure 1.4:** Tachyzoites and bradyzoite tissue cysts. A) A large group of apparently intracellular *N. caninum* tachyzoites (arrow); B) A thick walled (arrow) *N. caninum* tissue cyst within a neuron in the spinal cord of a 3-day old calf; C) *T. gondii* tissue cyst freed from mouse brain, cyst wall (arrow) enclosing hundreds of bradyzoites. Haematoxylin & Eosin x 600 (A & B), unstained impression smear (C), adapted from Dubey, Buxton & Wouda (2006) and Dubey, Lindsay and Speer (1998).

## 1.6 Host immune response and pathology

There is a catch-22 situation in the host response to a parasitic infection, in that the host must produce an immune response sufficient to control the spread of a pathogen, whilst not causing pathology so extensive that it is of detriment to itself. *N. caninum* and related coccidians generally induce a mixed immune response involving both cellular and humoral immunity (Andrianarivo *et al.* 2001; De Marez *et al.* 1999; Guy *et al.* 2001; Innes *et al.* 2002; Lunden *et al.* 1998; Williams *et al.* 2000). However, due the intracellular nature of the infection, cell-mediated immunity is expected to play the major role and proinflammatory cytokines such as interferon-gamma (IFN-γ) and interleukin-twelve (IL-12) are up regulated during an infection and have a controlling effect on proliferation of the parasite (Innes *et al.* 2002; Innes *et al.* 1995; Suzuki *et al.* 1988; Williams *et al.* 2000). Neosporosis is primarily a disease causing reproductive pathologies in cattle. During pregnancy, the maternal immune response is altered to cope with the presence of a foetus: regulatory cytokines such as interleukin-10 (IL-10) are favoured (Entrican 2002), which counteract the inflammatory cytokines such as IFN-γ, which may be dangerous to a pregnancy (Tangri & Raghupathy 1993). As yet, the triggers that allow the recrudescence of an infection, when bradyzoites reactivate to become tachyzoites and can cross the placenta to infect the foetus, are not well understood (Innes *et al.* 2007), but it is likely to be a result of the changes in the maternal Type I immune response during pregnancy (Innes *et al.* 2001). When a placenta becomes infected with *N. caninum,* there is a

significant increase in IFN-γ levels and the timing of this event is crucial to whether the foetus is aborted, born persistently infected, or born healthy (Trees & Williams 2005; Williams *et al.* 2000). Bovine neosporosis differs from ovine toxoplasmosis in that while both involve exogenous infections acquired during pregnancy, ovine toxoplasmosis does not commonly involve recrudescence of endogenous infection or have an effect on successive pregnancies (Innes *et al.* 2007).

## 1.7 Proteomics

Advances in proteomic techniques in recent years have made possible the high throughput analysis of complex mixtures of proteins. This has enabled the move to a 'systems biology' approach to research, as opposed to the study of individual proteins in isolation. To improve resolution of mixtures however, it is common practice to separate proteins out prior to enzymatic digestion, or alternatively digest the lysate prior to separation in peptide space. Peptides can be identified sensitively and accurately by mass spectrometry then assigned to proteins by bioinformatic searching of predicted protein databases.

### 1.7.1 Electrophoresis

Complex mixtures of proteins can be separated out on a gel matrix by the application of an electric field. Polyacrylamide gel electrophoresis (PAGE) of proteins denatured with the detergent sodium dodecyl sulphate (SDS) is a widely used and well established method for reliable separation according to molecular weight (Shapiro, Vinuela & Maizel 1967).

The commonly used method developed by Laemmli (1970) employs a short layer of large-pore gel to concentrate dilute/large volume samples into a thin front; which then passes on to a longer, small pore-sized resolving gel where the molecules separate out, allowing for improved resolution of the protein mixture. This discontinuous one dimensional gel method is referred to as 1-DE for the remainder of this thesis.

An additional initial step of isoelectric focusing allows protein separation on a second dimension, utilising a pH gradient, this type of gel-separation is known as two-dimensional SDS-PAGE, or 2-DE (Gorg, Weiss & Dunn 2004). 2-DE can offer improved resolution of proteins compared to 1-DE as proteins are spread horizontally across the gel in addition to vertically, and hence there are likely to be fewer protein species present in one spot than in a typical protein band on a 1-D gel.

Once separated on a gel, bands or plugs of protein can be excised and digested with proteolytic enzymes and the resulting peptides identified by mass spectrometry. Trypsin, a commonly used enzyme in proteomics (Rosenfeld et al. 1992; Shevchenko et al. 2006; Shevchenko et al. 1996), cleaves proteins after lysine and arginine residues, leaving a mixture of peptides which can be bioinformatically predicted in order to generate databases against which mass spectrometric data can be searched.

### 1.7.2 Reverse-phase high performance liquid chromatography (LC)

For complex mixtures of peptides, it is beneficial to introduce further stages of separation in order to simplify the sample and allow for better resolution of the different peptides it comprises. High performance liquid chromatography (HPLC) is a process whereby molecules contained within a liquid mobile phase are forced through a stationary phase contained within a closely-packed column, under pressure. In the stationary phase, compounds bind to the column based upon their chemical properties. For reverse phase (RP) HPLC, this means non-polar molecules are retained while polar molecules are first to elute. The mobile phase solvent buffer is altered along a gradient, by gradually increasing the acetonitrile content, in order to elute the more polar molecules by increasing hydrophobicity (Simpson 2003).

### 1.7.3 Tandem mass spectrometry (MS/MS)

A property known as the mass to charge ratio ($m/z$) is utilized by mass spectrometry technology to differentiate peptides. Proteomic mass spectrometry requires peptides to be converted to the gas phase by an ionisation source. When a mass spectrometer is coupled downstream to an LC system, this is usually an electrospray ionisation source (ESI), as found on an LTQ (linear trap quadrupole) mass spectrometer (Thermo Finnigan). Gas phase ions are identified by an initial survey spectrum (MS) then the three most abundant peptides fragmented by collision-induced dissociation (CID) before a second scan is performed to measure the $m/z$ of the resulting fragments (MS/MS).

A recent advance in LTQ technology is the release of the Orbitrap mass spectrometer; a more accurate instrument providing increased resolution. Rather than ions being trapped in a quadrupole, they are trapped by an outer barrel-like electrode to orbit around an inner electrode (Makarov 2000).

### 1.7.4 Multidimensional protein identification technology (MudPIT)

MudPIT technology utilises an additional strong cation exchange (SCX) LC column upstream of an RP column to provide enhanced separation of peptides. As such, prior gel separation is not required and proteins are digested to peptides as a complete sample. Twelve salt step gradients elute peptides from the SCX to the RP column, which is then coupled to a mass spectrometer, for example an LTQ, for analysis as described above. Bioinformatic analysis of the twelve resultant data files is complex, due to the fact that peptides to one protein could, in theory, be identified in any number of the twelve fractions. This is in contrast to techniques that adopt protein-space separation, where enzymatic digestion is applied to pre-simplified mixtures of proteins (Washburn, Wolters & Yates 2001; Wolters, Washburn & Yates 2001).

### 1.7.5 Bioinformatics – peptide/protein identification

In order to extract meaning from mass spectra, peptide identifications must be compared to an *in silico* database of all possible peptides that could be cleaved with the enzyme used (usually trypsin) from all the predicted proteins in genome of the organism(s) of interest. It is also necessary to assign a confidence score to each hit so that only confident, uniquely-identified peptide matches are believed. In general, the more abundant a protein is within a sample, the higher scoring it will be, due to the greater number of peptides that are detected from such a protein. One method of ascertaining which protein identifications to believe is false-discovery rate (FDR) scoring. This process searches peptide identifications against a nonsense database in addition to the genuine database, so that it can detect at what level the nonsense proteins begin to be identified in order to provide a cut-off point. Moreover, the different algorithms available for providing protein identifications from mass spectral data often produce slightly different results from one another. By combining more than one of these algorithms, not only can the number of protein hits be increased, but also the confidence levels assigned to them, when they are identified by more than one method. One such technique is that designed by Jones *et al.* (2009) which employs Mascot (Matrix Science), OMSSA (Geer *et al.* 2004) and X!Tandem (Craig & Beavis 2004; Fenyo & Beavis 2003) to make protein identifications and calculates an FDR score (based on hits to a decoy database) to assist determination of, and maximise, correct peptide identifications.

## 1.8 Transcriptomics

Proteomic techniques analyse proteins, the end result of transcription and translation from genetic material. The field of transcriptomics is interested in the intermediate product, RNA. The analysis of messenger RNA (mRNA) levels within a system can provide a snapshot of a rapidly changing environment to give information about cellular processes earlier than would be possible by proteomics. Microarray has been a popular technique for studying transcription, it requires synthesis of thousands of complementary DNA (cDNA) probes which are attached to a chip, before hybridisation of cDNA reverse-transcribed from mRNA in the sample. The chip is then scanned to provide an expression profile. Other ways of measuring mRNA expression include quantitative polymerase chain reaction (qPCR), which is highly accurate but more suited to smaller numbers of sequences of interest, and expressed sequence tag (EST) profiling, which sequences short regions of expressed genes and was useful for genome research before being largely superseded by the advent of RNA-Seq.

### 1.8.1 RNA-Seq

Next-generation sequencing technology has opened up opportunities in genome research by making the technique faster, cheaper and most importantly, available to the wider scientific community. One method that has been developed as a result is RNA-Seq, the sequencing of cDNA libraries created from reverse transcribing mRNA in a sample. This technique is not only highly sensitive, but provides absolute quantitation, and as such has become hugely popular in recent years.

Sanger capillary sequencing (Sanger, Nicklen & Coulson 1977) requires *in vivo* cloning of DNA in *E. coli* prior to sequencing with dye-labelled terminator dNTPs. Next generation sequencing has removed the need for *in vivo* cloning and uses either DNA synthesis (Illumina, 454) or ligation (SOLiD) instead. The 454 platform is a higher throughput alternative to Sanger sequencing that uses emulsion PCR in place of clone libraries, to perform 'pyrosequencing' (Margulies *et al.* 2005). It produces similar length reads, of approximately 500 base pairs, so is suitable for sequencing DNA or RNA for which there is no reference sequence available, as the long read lengths aid bioinformatic data processing.

During Illumina sequencing (Bennett *et al.* 2005; Bentley *et al.* 2008) a cDNA library, generated from an mRNA sample, is amplified by PCR reactions that form bridges between oligonucleotide primers on a flow cell. This generates clusters of DNA to which bases are

added which have a fluorescent dye attached, which acts as a reversible terminator. The flow cell is imaged then the dye removed (but not the base) and the next base added; and so on (Holt & Jones 2008), until the whole product has been sequenced. Because all bases are added at once, there is competition for incorporation; this increases accuracy compared to Sanger sequencing in which four separate reactions take place, however read lengths are much shorter (50-100 base pairs) so a good reference genome to align to is a requirement.

The SOLiD platform performs sequencing by rounds of ligation of dinucleotides to adapters (Shendure *et al.* 2005). This method is highly accurate due to the fact that each base is read twice (as the second nucleotide and then the first nucleotide), so genuine single nucleotide polymorphisms require two fluorescent colour changes. SOLiD also produces short reads, but sample preparation and data processing are complex compared to the Illumina platform.

A new third-generation sequencing method: 'single molecule sequencing' (Harris *et al.* 2008) has recently been introduced, which requires only tiny amounts of starting material; but it is not yet widely available.

## 1.9 Systems biology

The study of biological events is in the process of moving on from a 'reductionist' approach, where components are examined individually, to a 'systems biology' approach, due to the advances in technology described above. The advent of the genomic era and the ability to analyse whole transcriptomes and proteomes has led on from hypothesis driven experiments focusing on single genes to hypothesis-generating studies, based on the integration of all these data, to provide an overview of the system as a whole (Wastling *et al.* 2012). This is particularly valuable in parasite biology due to the importance of understanding the interaction of the parasite with its host; and the Apicomplexa are some of the most comprehensively studied to date (Wastling *et al.* 2009).

## 1.10 Current state of the *N. caninum* genome, proteome and transcriptome

During the course of this study, the *N. caninum* genome was undergoing annotation as part of a collaboration between Liverpool and the Sanger Institute (Reid *et al.* 2012). Proteomic studies published prior to this had suffered from the lack of an available predicted proteome and as such had identified only limited numbers of proteins (Belli, Walker & Flowers 2005). These included a 2-DE proteome map (Lee *et al.* 2003) and immunoblot studies (Lee *et al.* 2004), followed by a comparative analyses of *T. gondii* and *N. caninum* 2-DE proteome maps (Kang *et*

*al.* 2008b; Lee *et al.* 2005). No RNA-Seq experiments had been published for *N. caninum* and although mRNA expression had been examined by annealing control primer-based PCR (Kang *et al.* 2008a), the transcriptome data presented in Reid *et al.* 2012 and the pilot study discussed in Chapter 6.2 were the first RNA-Seq analysis performed on *N. caninum* tachyzoites. In comparison, *T. gondii, C. parvum* and *Plasmodium spp.* have been extensively studied and large proportions (25 % and more) of their predicted proteomes have been identified experimentally (Xia *et al.* (2008), Sanderson *et al.* (2008) and multiple experiments for *Plasmodium* (Belli, Walker & Flowers 2005) hosted on PlasmoDB (Aurrecoechea *et al.* 2009)).

The *N. caninum* genome is assembled into pseudochromosomes (Reid *et al.* 2012) based upon the fourteen published chromosomes of *T. gondii* (Toxo DB, Gajria *et al.* 2008). The analyses in DeBarry and Kissinger (2011) and Reid *et al.* (2012) found a high degree of synteny between the two organisms, with over 90 % of the genome comprising one-to-one orthologues and only 113 and 231 genes (excluding surface antigens) specific to *N. caninum* and *T. gondii* respectively. In addition, the G+C content was very similar (54.8 % in *N. caninum* and 52.3 % in *T. gondii*) as was the percentage of the genome that encoded proteins (29.7 % in *N. caninum* and 28.3 % in *T. gondii*, compared to 53.0 % in *P. falciparum).* Reid *et al.* (2012) also estimated that 32.1 % of the *N. caninum/T.gondii* genome consisted of eukaryotic genes, 28.8 % of Apicomplexan genes and the remaining 39.1 % of genes were coccidian specific. The majority of the divergence between *N. caninum* and *T. gondii* appeared to be attributed to surface antigen genes and to those encoding apical proteins, which will be discussed in more detail in this thesis.

A more comprehensive analysis of the *N. caninum* proteome and transcriptome was necessary to better understand the genes and proteins important to an infection; and the release of an annotated genome sequence made this a possibility.

A desire to understand the genetic basis for the differences between *N. caninum* and *T. gondii* underpinned the comparative genome analysis performed by Reid *et al.* (2012). Of particular interest were differences in apical genes, as these are the molecular effectors of host cell invasion. While the majority of apical genes were found to have a one-to-one orthologous relationship between the two organisms, and as such are hypothesized to behave in an analogous way *in vitro* and *in vivo*, there were a few exceptions. One of the omissions from the *N. caninum* genome, *ROP18*, was noted in particular, due to the gene's recent identification as a virulence determinant in *T. gondii* (Saeij *et al.* 2006; Taylor *et al.* 2006).

## 1.11 Aims and objectives

The aim of this project was to advance understanding of *N. caninum* biology by taking a systems view of the genes and proteins important to this parasite. Identifying proteins expressed by the invasive tachyzoite life stage, in particular those released by the micronemes, rhoptries and dense granules, would provide an insight into the process of host cell entry, an event that underpins Apicomplexan parasitism.

The recent sequencing of the *N. caninum* genome was a major factor in facilitating the analysis of the proteome and transcriptome. There had been only limited proteomic studies published for *N. caninum*, so the peptide data produced from a global analysis of the tachyzoite proteome was an asset to the genome annotation. The accuracy and completeness of the genome annotation has a direct effect on the quality of the proteomic identifications derived from searching against the gene models, therefore this analysis was seen as an essential prerequisite to the rest of the study.

Simplifying the proteome to contain proportionally more apical proteins in comparison to structural and housekeeping proteins allowed a more focused analysis of these proteins of interest. This was achieved by a combination of two approaches: an examination of the excreted/secreted proteome by stimulating the parasites to discharge the contents of their apical organelles; and by attempting to separate the organelles into different fractions for analysis.

Bioinformatic study comprised an invaluable part of the project, both in analysis of proteomic data and for *in silico* investigations into the predicted subproteomes. Various proteins are difficult to detect in experimental proteomics (von Hagen 2008), due to differences in expression levels, chemical and structural characteristics which affect solubility and differences in peptide flight during mass spectrometry (Bell *et al.* 2009), so by using bioinformatics it was hoped to supplement the data generated by proteomic experiments.

By combining the results of multiple proteomic analyses with the findings of collaborators annotating the *N. caninum* genome, rhoptry gene 18 emerged as a subject to examine further. The confirmation of its pseudogenisation in *N. caninum*, and its recent emergence as a virulence determinant in *T. gondii*, led to a hypothesis that exhibiting a low-virulence phenotype may have aided this parasite in its strategy for exploiting vertical transmission.

The ability to convert to a latent bradyzoite stage and later recrudesce is an important aspect of *N. caninum* transmission, and is responsible for much of the pathology associated with neosporosis in dairy herds. RNA-Seq provided a high-throughput platform from which to analyse the process by which stage conversion occurs, and to examine the differences in expression of key genes between the tachyzoite and bradyzoite stages.

Overall, it was intended that an increased knowledge of tachyzoite and bradyzoite biology would enable more informed vaccine development in the future.

# CHAPTER 2: AN ANALYSIS OF THE TACHYZOITE PROTEOME

## 2.1 INTRODUCTION

The overall aim of this thesis was to improve the understanding of host cell invasion by *N. caninum* by identifying and studying the proteins involved in the process. In order to achieve this, and due to the fact that there had been only limited proteome data published for this organism to date, a global proteome analysis for the tachyzoite stage was undertaken. It was hoped that the data would benefit the genome sequencing project by providing peptide evidence to reinforce the predicted gene models. The accuracy and completeness of the genome annotation has a direct effect on the quality of the data derived from searching against the gene models. Therefore, this analysis was seen as an essential prerequisite to the rest of the study.

Proteomic analyses by one and two dimensional electrophoresis (1-DE and 2-DE, respectively), followed by high performance liquid chromatography coupled to a tandem mass spectrometer (LC MS/MS), of the related organisms *T. gondii* and *C. parvum* had previously yielded substantial coverage of their respective predicted proteomes (Sanderson *et al.* 2008; Xia *et al.* 2008). These studies acquired peptide data for a wide range of protein types and presented a global view of protein expression for the *T. gondii* tachyzoite and *C. parvum* sporozoite life stages, providing an insight into which groups of proteins may be important during this time of rapid multiplication and invasion. These methods were therefore adopted for this equivalent study on the *N. caninum* tachyzoite, although with some slight adaptations as will be discussed in this chapter.

### 2.1.1 Global proteomic analyses on *N. caninum* and related Apicomplexa

Whilst there had been proteomic analyses carried out prior to the *N. caninum* sequence data becoming available on the genomic resource website ToxoDB (Gajria *et al.* 2007), these were able to identify only very limited numbers of proteins, due to the lack of sequence data available (Belli, Walker & Flowers 2005). One such study, utilising 2-DE, identified 31 spots corresponding to 20 different proteins, by peptide mass fingerprinting on a matrix-assisted laser desorption/ionisation-time of flight (MALDI-TOF) mass spectrometer (Lee *et al.* 2003). The same group then went on to use immunoblot analyses to identify immunogenic spots (Lee *et al.* 2004; Lee *et al.* 2003) and compare the KBA-1 and JPA2 isolates of *N. caninum* to each

other and to RH strain *T. gondii*. The combined number of proteins identified from these studies is less than 100 proteins, which, for an organism with a predicted 7082 protein coding genes (Reid *et al.* 2012), equates to < 1.5 % coverage, which is probably a result of the lack of sequence data available at this time. From the existing data for *T. gondii* and *C. parvum*, approximately 30% of the predicted proteins were detected by proteomic methods in lysates of tachyzoites or sporozoites (Sanderson *et al.* 2008; Xia *et al.* 2008). It is important to note that these analyses examine one life stage only, and that some proteins are likely to be specific to other life stages not analysed.

At the time this study commenced, there were a predicted 5587 proteins for *N. caninum*, compared to 7993 for *T. gondii* ME49. The Xia *et al.* (2008) study used a combination of 1-DE and LC MS/MS of whole cell lysate, including separate preparations of soluble and insoluble fractions, 2-DE and LC MS/MS, and multi-dimensional protein identification technology (MudPIT) to identify a total of 2252 non-redundant proteins. When these multiple experiments are broken down and looked at individually, it becomes clear that the MudPIT platform (2121 non-redundant identifications) followed by 1-DE LC MS/MS (939) are the highest yielding in terms of protein identifications. A further 40 non-redundant proteins were identified in the 2-DE experiments (547 overall). Whilst 2-DE can achieve superior protein separation to 1-DE, the aim of this study was to identify as many proteins as possible, and it was therefore decided that 2-DE was unlikely to provide much additional benefit over 1-DE and MudPIT experiments.

In comparison, *Plasmodium falciparum* is probably the most well characterised Apicomplexan, with proteomes identified over a whole range of life-stages. Total protein identifications number 3969 (multiple studies, data hosted on PlasmoDB) of the 5772 predicted genes (Aurrecoechea *et al.* 2009). This is an example of what can be achieved with a model organism for which the entire life cycle can be replicated in culture.

### 2.1.2 Peptide data as an aid to genome annotation

A bioinformatically-assembled genome annotation is a valuable resource that provides a wealth of information. However, protein-coding gene predictions are no substitute for empirical evidence that a protein exists. Not only can proteomic data be viewed alongside gene predictions to provide another level of information, but peptide identifications can be used to confirm or alter gene predictions where they are seen to agree or differ, respectively.

Over the duration of this study, the *N. caninum* genome annotation was being curated by collaborators at the Sanger Institute, so peptide data generated from these experiments was used to resolve queries about the existence of genes or their intron/exon boundaries. For instance, gene models for which the predictions had low confidence could be searched against the protein/peptide data and if present, their putative existence could be confirmed. Whilst a lack of proteomic detection is not sufficient evidence to deny the existence of a protein, the identification of even one significant uniquely matched peptide provides a reasonable level of confidence that a protein is present within a sample.

### 2.1.3 Aims and objectives

The aim of this chapter was to perform a global protein analysis of the tachyzoite stage of *N. caninum,* in order to better understand which proteins are important to this life stage of the parasite. In addition, this work was helpful in assisting the genome annotation by providing experimental evidence for predicted proteins. The quality and quantity of proteomic identifications are directly influenced by the databases which mass spectra are searched against, but these databases can in turn be improved by the availability of peptide data.

## 2.2 MATERIALS AND METHODS.

### 2.2.1 Cell culture

NcLiv-strain tachyzoites were passaged twice weekly onto a layer of African Green Monkey Kidney Fibroblasts (Vero cells) in 25 $cm^2$ tissue culture flasks, supported by IMDM (Iscove's Modified Dulbecco's Medium, Lonza). The medium was supplemented with 10 % (v/v) foetal calf serum (Labtech International), 100 U/ml penicillin and 100 µg/ml streptomycin sulphate (Sigma-Aldrich) and filter sterilised prior to use in a 0.2 µm pore-size membrane, attached to a vacuum pump. Cells and parasites were maintained at 37 °C and 5 % $CO_2$.

#### 2.2.1.1 Host cell passage

Host cell passage was performed by washing the confluent cell monolayer in 5 ml 1 M HEPES Balanced Salt Solution (Sigma-Aldrich) before incubating the cells for 5 minutes in 5 ml Trypsin-Versene (EDTA) mixture (Lonza), at 37 °C and 5 % $CO_2$. The cell suspension was centrifuged for 5 minutes at 1500 x g, and then resuspended in 5 ml IMDM. A haemocytometer was used to count a sample of the cells, before they were seeded at a ratio of 1 x $10^5$ per 5 ml IMDM/25 $cm^2$ flask.

#### 2.2.1.2 Parasite passage

Approximately 24 hours post seeding, when the vero cells had reached 10-20 % confluence, they were infected with 4 x $10^5$ tachyzoites, which were then allowed to grow for 3-4 days to coincide with egress before harvesting with a sterile cell scraper. Parasites were then either passaged into new cells, or isolated for use in experiments.

#### 2.2.1.3 Parasite isolation

After harvesting the cells/parasites, host cell debris was removed by filtration through 47 mm diameter 3 µm pore-sized Nucleopore polycarbonate membranes (Whatman). Parasites were then washed twice in phosphate-buffered saline (PBS) pH 7.4 by centrifugation at 1500 x g, for 10 minutes at 4 °C. They were washed a final time by centrifugation at 13 000 x g for 5 min, the supernatant discarded, and pellets (each containing approximately 1 x $10^8$ tachyzoites) were either used immediately for a downstream experiment, or stored at -20 °C or -80 °C until required.

### 2.2.2 Proteomic analyses

#### 2.2.2.1 Preparation of tachyzoite lysates

Samples were lysed and solubilised in sample loading buffer (50 mM Tris HCL (pH 6.8), 2 % (w/v) sodium dodecyl sulphate (SDS), 0.2 % (w/v) bromophenol blue, 10 % (w/v) glycerol, 100 mM dithiothrietol (DTT)) at concentration of 5 µg tachyzoite protein/µl and subjected to three cycles of heating to 95 °C for 3 minutes and vortexing for 3 minutes, followed by centrifugation at 13 000 x g for 5 minutes.

#### 2.2.2.2 Bradford assay

Protein concentration was estimated by carrying out a Bradford assay (Bio-Rad Protein Assay) on an equivalent sample of tachyzoites, lysed and solubilised as described above in the sample loading buffer minus SDS and bromophenol blue, due to their incompatibility with the assay reagents. This assay is based on the principle of Coomassie dye binding to protein, and comprises a comparison to bovine serum albumin standards of known concentration as detailed in Bradford (1976). The samples and standards (10 µl of each, in triplicate) were pipetted into a 96-well plate and 200 µl dye reagent added to each well and mixed thoroughly. The plate was incubated at room temperature for a minimum of 5 minutes and the absorbance measured at 595 nm. A standard curve was produced from the absorbencies of the standards and the sample concentration calculated accordingly.

#### 2.2.2.3 One dimensional electrophoresis (1-DE)

For 1-DE, 16 cm discontinuous gels were poured based on the Laemmli method (Laemmli 1970), as follows: 12 % (v/v) acrylamide resolving gel contained 40 ml acrylamide/bisacrylamide (30 %), 25 ml 1.5 M Tris/HCl pH 8.8, 1 ml 10 % (w/v) SDS, 33.5 ml $ddH_20$ (ultrapure water, Milli-Q grade, resistivity 18.2 Ω.cm), 500 µl ammonium persulphate (APS) and 100 µl tetramethyladenamine (TEMED); the 5 % (v/v) acrylamide stacking gel contained 3.4 ml acrylamide/bisacrylamide (30%), 5 ml 0.5 M Tris/HCl pH 6.8, 200 µl 10 % (w/v) SDS, 13.6 ml $ddH_20$, 200 µl APS and 20 µl TEMED.

A total of 200 µg tachyzoite lysate was loaded onto the gel alongside 10 µl Precision Plus Protein Standard (Biorad). Samples were run through the 5 % (v/v) acrylamide stacking gel in 1 x SDS running buffer (25 mM Tris base, 192 mM glycine, 0.1 % SDS in $ddH_20$) for approximately 50 minutes at 16 mA, followed by 5-6 hours at 24 mA through the resolving gel.

Gels were fixed at room temperature in fixing solution (40 % (v/v) ethanol/10 % (v/v) acetic acid) overnight, and then rinsed twice in $ddH_2O$ before staining with Colloidal Coomassie for 24 hours (20 % (v/v) methanol, 0.08 % (w/v) Coomassie Brilliant Blue G250, 0.8 % (v/v) phosphoric acid, 8 % (w/v) ammonium sulphate). After further rinsing in $ddH_2O$, bands were excised from the gel and stored individually in 1 % (v/v) acetic acid at 4 °C.

#### 2.2.2.4 In-gel tryptic digestion

Gel bands were destained twice for 15 minutes at 37 °C in a solution of 50 % (v/v) acetonitrile/50 mM ammonium bicarbonate. Reduction of cysteines was performed by a 30 minute incubation in 50 µl of 10 mM DTT/100 mM ammonium bicarbonate, followed by alkylation in 50 µl 100 mM iodoacetamide (IAA)/55 mM ammonium bicarbonate for 1 hour, in the dark; both at 37°C. The gel bands were then dehydrated for 15 minutes at 37 °C with 100 % (v/v) acetonitrile, which was then left for a further 10 minutes for the solvent to evaporate, before rehydrating with 15-25 µl (enough to ensure the gel plug was completely covered) 10 ng/µl sequencing grade trypsin (Roche Diagnostics)/25 mM ammonium bicarbonate. After 1 hour at 37 °C, the tubes were topped up with 15-25 µl of 25 mM ammonium bicarbonate and returned to incubate overnight. The reaction was stopped by freezing the following morning, and the digested bands were stored at -20°C until required for mass spectrometric analysis.

### 2.2.3 Reverse-phase high performance liquid chromatography and tandem mass spectrometry (LC MS/MS)

LC MS/MS was carried out using an LTQ ion trap mass spectrometer (Thermo Fisher Scientific) with an electrospray ionisation source, coupled downstream to an online nano pepMap100 c18 RP column (3 µm, 100 Å, 75 µm i.d. x 15 cm) on a Dionex Ultimate 3000 HPLC system (Dionex). A C18 trapping column (300 µm i.d. x 5 mm) desalted the peptides prior to their entry onto the analytical column, which was equilibrated with buffer comprising of water/2 % (v/v) acetonitrile/0.1 % (v/v) formic acid at a flow of 300 nl $min^{-1}$. Tryptic peptides were eluted using a 3 hour programme. Conditions were as follows: a linear gradient of 0-50 % (v/v) acetonitrile/0.1 % (v/v) formic acid over 140 minutes followed by 100 % (v/v) acetonitrile/0.1 % formic acid for 20 minutes and a further 20 minutes of 0 % (v/v) acetonitrile/0.1 % (v/v) formic acid. A 500 fmol $\mu l^{-1}$ solution of glufibrinopeptide ($m/z$ 785.8, $[M+2H]^{2+}$) was used to tune the LTQ. The resulting MS/MS spectra (.raw files) were converted to .dta files using TurboSequest Bioworks version 3.1 (Thermo Fisher Scientific)

using the parameters: threshold cut-off 100, group scan default 100, minimum group count 1, minimum ion count 15, peptide tolerance +/- 1.5. These were then merged into search-compatible .mgf files.

### 2.2.4 Multidimensional Protein Identification Technology (MudPIT)

Sample preparation and MudPIT was carried out at the Yates' Lab, Scripps Research Institute, La Jolla, CA, USA, by Dr. Helena Prieto. Briefly, tachyzoite pellets (cultured as described in 2.2.1 by Rebecca Norton, Liverpool) were resuspended to approximately 800 µg $ml^{-1}$ in 100 mM Tris/HCl pH 8.5, lysed by three freeze/thaw cycles and separated into soluble and insoluble fractions by centrifugation for 30 minutes at 16 000 x g. Invitrosol (Invitrogen) was added to the supernatant (soluble fraction) at 1 % (v/v), before it was heated for 5 minutes at 60 °C, vortexed for 2 minutes, denatured with 2 M urea, reduced with 5 mM Tris (2-carboxyethyl) phosphine hydrochloride and carboxyamidomethylated with 10 mM IAA. Next, 1 mM calcium chloride and trypsin was added at a ratio of 1:100 (enzyme:protein), followed by incubation at 37 °C overnight. The pellet (insoluble fraction) was resuspended in 10 % (v/v) Invitrosol (Invitrogen), heated and vortexed as before, sonicated for 1 hour, diluted to 1 % (v/v) Invitrosol with 8 M urea/100 mM Tris/HCL pH 8.5, reduced and alkylated as before, then digested with endoproteinase Lys-C for 6 hours. The solution was then further diluted to 4 M urea with 100 mM Tris/HCl pH 8.5 and finally digested with trypsin, as above (2.2.2.4).

MudPIT analysis was performed on four soluble and three insoluble replicates, with modifications to the method of Link *et al.* (1999). Samples of approximately 100 µg protein were loaded onto separate microcolumns and resolved by a fully automated 12 step chromatography process; consisting of strong cation exchange LC upstream of reverse phase LC (Washburn, Wolters & Yates 2001) on a quaternary Agilent 1100 series HPLC system (Agilent Technologies). Tandem mass spectrometry was performed with a nano-LC electrospray ionisation source on an LTQ ion trap mass spectrometer (Thermo Fisher Scientific Inc). Protein identifications were made using the SEQUEST algorithm (Eng, McCormack & Yates 1994), employing the parameter of no enzyme specificity and searched against the ToxoDB version 5.0 *N. caninum* gene models. In addition, raw data files were sent to Liverpool for processing using Bioworks as described in 2.2.3 and subsequent protein identification.

### 2.2.5 Analysis of tachyzoite lysate on an Orbitrap Velos mass spectrometer

Two different sample preparation approaches were taken as follows: a tachyzoite lysate was prepared and digested 'in-solution' prior to analysis in a single run on an Orbitrap Velos mass spectrometer, and 50 µg of tachyzoite lysate was separated on a 10 cm 12 % acrylamide gel prior cutting into 10 bands for in-gel digestion and Orbitrap analysis.

#### 2.2.5.1 In-solution digestion and analysis

A pellet of 1.43 x $10^8$ tachyzoites was solubilised in 200 µl 50 mM ammonium bicarbonate and then sonicated in a sonicating waterbath at 30 % amplitude for three cycles of 10 seconds on, 50 seconds off. An additional 100 µl was added to the sample and the sonication repeated. The sample was centrifuged at 13 000 x g for 10 minutes at 4 °C and the supernatant collected. A 50 µg aliquot of the sample (as determined by Bradford assay (2.2.2.2)) was diluted in 25 mM ammonium bicarbonate to a volume of 160 µl, prior to the addition of 10 µl 1 % Rapigest (Waters) (0.05 % (w/v) final concentration). The sample was heated to 80 °C for 10 minutes, with a brief vortex at the half way point; 10 µl of DTT (3 mM final concentration) was added and the sample incubated for 10 minutes at 60 °C, before being cooled to room temperature. IAA was added to a final concentration of 9 mM and the sample incubated in the dark for 30 minutes at room temperature. Trypsin was added at a 50:1 trpsin:protein ratio and the digest incubated overnight at 37 °C. Trifluoroacetic acid (TFA) was added to a final concentration of 0.5 % (v/v) and the digest incubated for 45 minutes at 37 °C. Finally, the sample was centrifuged for 15 minutes at 13 000 x g at room temperature and the supernatant collected for subsequent mass spectrometric analysis.

#### 2.2.5.2 1-DE and in-gel digestion

A tachyzoite lysate was prepared as in 2.2.2.1 and 50 µg was run on a 10 cm 12 % acrylamide gel (2.2.2.3) with limiting factors of 200 V, 70 mA and 10 W for 50 minutes. The stained gel was sectioned into 10 contiguous bands, which were then further halved, to enable tryptic digestion, which was performed as in 2.2.2.4. Tryptic digests were re-pooled to return to 10 samples, which were purified using StageTips (Thermo Scientific) and analysed by mass spectrometry on an Orbitrap Velos.

#### 2.2.5.3 Orbitrap Velos mass spectrometry

Mass spectrometry was performed by Stuart Armstrong (Liverpool) as follows: peptide mixtures (1 µg) were analyzed by on-line nanoflow liquid chromatography using the nanoACQUITY-nLC system (Waters) coupled to an LTQ-Orbitrap Velos (ThermoFisher Scientific) mass spectrometer equipped with the manufacturer's nanospray ion source. The analytical column (nanoACQUITY UPLC™ BEH130 C18 15 cm x 75 µm, 1.7 µm capillary column) was maintained at 35 °C and a flow-rate of 300 nl/min. The gradient consisted of 3-40 % acetonitrile in 0.1 % formic acid for 90 minutes then a ramp of 40-85 % acetonitrile in 0.1 % formic acid for 3 minutes. Full scan MS spectra (m/z range 300-2000) were acquired by the Orbitrap at a resolution of 30,000. Analysis was performed in data dependant mode. The top 20 most intense ions from MS1 scan (full MS) were selected for tandem MS by collision induced dissociation (CID) and all product spectra were acquired in the LTQ ion trap. Ion trap and orbitrap maximal injection times were set to 50 ms and 500 ms, respectively.

### 2.2.6 Protein identification using multiple search engines

The .mgf files resulting from mass spectrometric analysis were either submitted to Mascot (Matrix Science) or to a multiple search engine algorithm (Jones et al. 2009) which employs Mascot, OMSSA (Geer et al. 2004) and X!Tandem (Craig & Beavis 2004; Fenyo & Beavis 2003). To make protein identifications, data was searched against a locally-mounted database comprising the *N.caninum* gene predictions (various releases (Reid *et al.* 2012), hosted on ToxoDB (Gajria et al. 2007)). Unless otherwise stated, search parameters were as follows: fixed carbamidomethyl modification of cysteine, variable oxidation of methionine, one missed trypsin cleavage, peptide tolerance ±1.5 Da, fragment ion tolerance ±0.8 Da and peptide charge state of +1, +2 and +3. Target protein identifications at a false discovery rate (FDR) of 1 % were accepted as correct. For data searched using Mascot alone, a cut-off score of 50 was applied.

### 2.2.7 Assignment of proteins to MIPS Functional Catalogue categories

Protein identifications were manually assigned to MIPS Functional Catalogue categories (Ruepp *et al.* 2004) to determine the range of protein types identified. For each protein, the 'biological process' Gene Ontology (GO) term (Ashburner *et al.* 2000) was downloaded from ToxoDB (Gajria *et al.* 2008), and used as a search term in the MIPS Functional Catalogue Database. If multiple GO biological processes were assigned, all were searched in the MIPS

functional catalogue and a consensus category assigned. If there was no 'biological process' GO term assigned, the 'molecular function', followed by 'cellular component' GO term was used instead. If there was no GO term, but the gene description on ToxoDB contained information alluding to the function of the protein, then literature searching was employed to qualitatively determine the most appropriate MIPS category. Any protein that was known to be invasion-related was assigned to category 32: cell rescue, defence and virulence. If there was no GO term or other indicator of function on ToxoDB, or in the literature, a Basic Local Alignment Search Tool (BLAST) search (NCBI) was performed to identify possible orthologues for which a function was known. Any proteins for which this yielded no useful result, were assigned 'unclassified'. Table 2.1 shows the MIPS Functional Categories used for this analysis.

**Table 2.1:** MIPS Functional Categories

| **Category** | **Description** |
|---|---|
| 01 | Metabolism |
| 02 | Energy |
| 04 | Storage protein |
| 10 | Cell cycle and DNA processing |
| 11 | Transcription |
| 12 | Protein synthesis |
| 14 | Protein fate (folding, modification, destination) |
| 16 | Protein with binding function or cofactor requirement |
| 18 | Regulation of metabolism and protein function |
| 20 | Cellular transport, transport facilities and transport routes |
| 30 | Cellular communication/signal transduction mechanism |
| 32 | Cell rescue, defence and virulence |
| 34 | Interaction with the environment |
| 40 | Cell fate |
| 42 | Biogenesis of cellular components |
| 70 | Subcellular localisation |

## 2.3 RESULTS

### 2.3.1 One dimensional gel electrophoresis (1-DE) and mass spectrometry (LC MS/MS) analysis of whole tachyzoite lysate

The 16 cm gel resulting from 1-DE (Figure 2.1) showed good resolution of protein bands and was loaded with sufficient protein to enable downstream detection of peptides. A total of 129 contiguous bands were excised manually and taken forward for tryptic digestion.

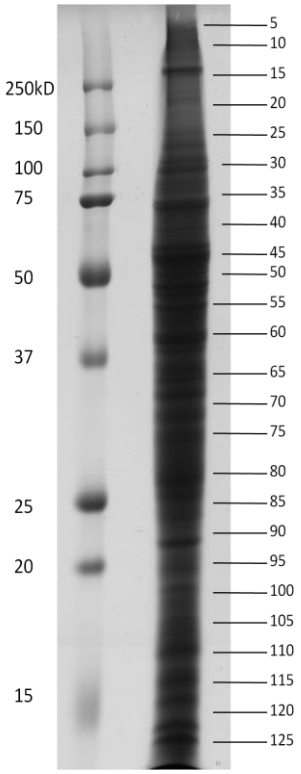



**Figure 2.1:** 1-DE gel of *N. caninum* whole tachyzoite lysate. Lane one: protein standards, lane 2: 200µg tachyzoite lysate. A total of 129 contiguous bands were excised for tryptic digestion and LC MS/MS as indicated at 5-band intervals in the figure.

From the LC MS/MS analysis of the 129 bands on an LTQ mass spectrometer 5563 *N. caninum* proteins (Appendix I) were identified when employing a false discovery rate (FDR) of 1 % and using the ToxoDB version 5.0 genome release. When redundancy between gel bands was taken into account, this left 965 individual protein identifications. With the resubmission of the data to version 6.0 of the genome annotation, the number of identifications increased to 975 non redundant protein hits, equating to 13.8 % genome coverage. These numbers

correlate very closely to those found by Xia *et al.* (2008) for their 1-DE analysis of *T. gondii* (939 non redundant hits, 11.7 % coverage).

As the focus of this project was on the apical proteins, it was useful to examine whether any known apical proteins had been identified by this whole-tachyzoite lysate approach. Searching the genome annotation for rhoptry, microneme and dense granule proteins will be discussed in more detail in Chapter 3 and correlated with proteomic results in Chapter 4, however a brief initial analysis was made at this point.

A text search of the ToxoDB version 5.0 genome annotation for the word 'microneme' yielded 15 proteins, of which only two were identified by this proteomic experiment: the Nc MIC11 precursor protein NC_LIV_081370 and a 'microneme associated protein' NC_LIV_113750.

However, more success was found with dense granule proteins, of which four out of five were identified: two annotated as DG32 (NC_LIV_030870 and NC_LIV_031880), a DG1 precursor (NC_LIV_082090) and a DG4 precursor (NC_LIV_131370). With the rhoptry proteins, peptides for all eight were identified; they comprised a rhoptry surface protein (NC_LIV_001550), two annotated as 'rhoptry protein related' (NC_LIV_040440 and NC_LIV_130570), RON4 (NC_LIV_100030), RON1 (NC_LIV_130800), ROP13 (NC_LIV_132210) and two putative rhoptry antigens (NC_LIV_140500 and NC_LIV_140510).

In order to assess the range of proteins identified, all version 6.0 hits were assigned functional annotation as described in 2.2.8. The results, seen in Figure 2.2, show that proteins over the whole range of categories were identified using this method, but due to a lack of available information in the genome annotation, the most heavily populated category was 'unclassified'.

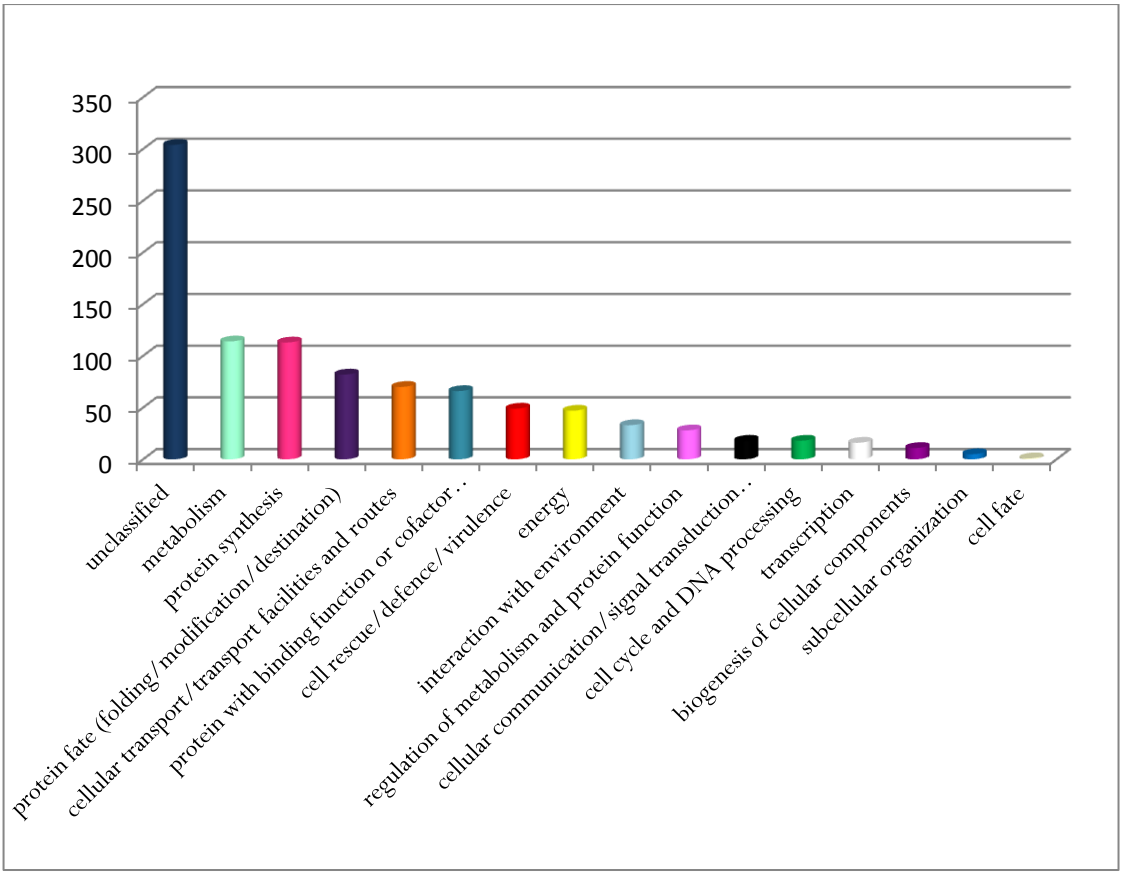



**Figure 2.2:** Functional classification of proteins identified in whole tachyzoite lysate analysed by 1-DE and LC MS/MS. Proteins (n = 975) were identified from the *N. caninum* genome annotation hosted on ToxoDB version 6.0. Proteins were assigned to functional categories according to the MIPS Functional Catalogue.

### 2.3.2 Multidimensional proteomic identification technology (MudPIT) analysis of whole tachyzoite lysate

MudPIT data comprising soluble and insoluble fractions were collected at the Scripps Research Institute and processed as SEQUEST (Thermo Finnigan) search results. A total of 447 and 436 non-redundant proteins were identified from the insoluble and soluble fractions, respectively (Appendix I). When combined and further redundancy was removed this resulted in 635 non-redundant protein identifications from the MudPIT platform. This figure was far lower than was expected (in previous experiments (Xia *et al.* 2008) MudPIT had outperformed 1-DE two-fold in terms of non-redundant protein hits). Thus, the original .RAW files generated from the MS/MS were converted to .mgf files and subjected to the triple-search algorithm (Jones *et al.* 2009) used with the 1-DE data, with the aim of improving the yield of protein hits.

From the triple-search algorithm, against ToxoDB version 5.0 of the *N. caninum* genome annotation, a total of 1002 non redundant hits were identified (645 insoluble and 658 soluble with a 1 % FDR) of which 487 were newly identified and had not been present in the SEQUEST search results. Therefore, while it proved propitious to re-search the data with the Jones *et al.* (2009) algorithm against Mascot (Matrix Science), OMMSA (Geer *et al.* 2004) and X!Tandem (Fenyo & Beavis 2003) combined, the result was similar in yield to that obtained from 1-DE, and hence half of that anticipated (Figure 2.3).

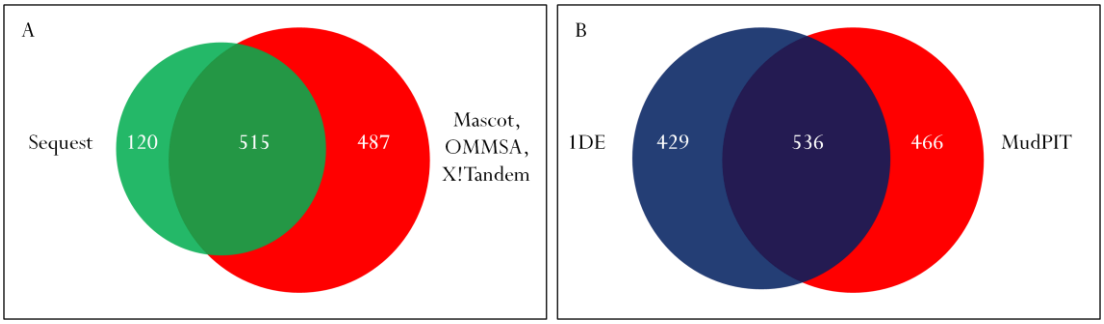



**Figure 2.3:** Venn diagrams showing MudPIT non-redundant protein identifications from whole tachyzoite lysates. A) MudPIT non-redundant identifications (n = 1122) using SEQUEST compared to Mascot/OMMSA/X!Tandem searching algorithm; B) Comparison of non-redundant identifications (n = 1431) made from MudPIT and 1-DE LC MS/MS proteomic techniques.

When the MudPIT data was combined with the ToxoDB version 5.0 1-DE data, there were a total of 1431 non redundant protein identifications, equating to 25.6 % coverage of the *N. caninum* predicted genome at that time (5587 gene models). Unfortunately, due to the lower than expected protein-hit yield of the MudPIT data and the excessive amount of server time that the database search required, it was decided not to resubmit the data to the ToxoDB version 6.0 release of the genome annotation, especially as this approach had not resulted in a significant increase in hits for the 1-DE data. Instead, Mascot searches alone were performed and submitted to the false discovery rate rescoring algorithm (Jones *et al.* 2009) to yield 699 non-redundant protein identifications (1 % FDR) in the ToxoDB version 6.0 annotation, so that when combined with the 975 identifications from the 1-DE analysis, a total of 1267 proteins overall form the tachyzoite stage had been detected, equivalent to 17.9 % of the version 6.0 total predicted proteome.

### 2.3.3 Analysis of tachyzoite lysate using an Orbitrap Velos mass spectrometer

#### 2.3.3.1 In-solution analysis

From the in-solution digestion of tachyzoite lysates a total of 323 *N. caninum* protein identifications achieving a Mascot score ≥ 50 were made from a single one-hour analysis of this preparation on the Orbitrap Velos. These identifications are presented in Appendix I. Novel identifications (n = 28), those not previously identified by 1-DE LC MS/MS or MudPIT analyses, are displayed in Table 2.2. Of these, fifteen were designated 'hypothetical' and had no functional annotation. Over half of these proteins (15/28) achieved a Mascot score of ≤ 72 and as such were in the lowest scoring third of the total identifications, of which the top-hitting third of proteins scored between 1193 and 117.

**Table 2.2:** Novel protein identifications from the Orbitrap Velos tachyzoite analysis (n = 28). Proteins were identified by searching the ToxoDB version 6.0 *N. caninum* genome annotation, and a Mascot cut-off score of ≥ 50 was employed.

| Protein I.D. | Description | Mascot Score |
|---|---|---|
| NCLIV_006470 | hypothetical protein, conserved | 170 |
| NCLIV_000860 | SPATR, related | 137 |
| NCLIV_048550 | 26S proteasome non-ATPase regulatory subunit 4, putative | 103 |
| NCLIV_062890 | hypothetical protein | 98 |
| NCLIV_067050 | hypothetical protein, conserved | 85 |
| NCLIV_004860 | hypothetical protein | 83 |
| NCLIV_013150 | hypothetical protein, conserved | 82 |
| NCLIV_043300 | nucleolar phosphoprotein nucleolin, putative | 82 |
| NCLIV_006570 | serine/threonine protein phosphatase, putative | 77 |
| NCLIV_008890 | tim10/DDP zinc finger domain-containing protein, putative | 77 |
| NCLIV_031320 | hypothetical protein, conserved | 74 |
| NCLIV_013700 | CHCH domain-containing protein, putative | 73 |
| NCLIV_013180 | GM04207p, related | 72 |
| NCLIV_043870 | hypothetical protein, conserved | 71 |
| NCLIV_017440 | hypothetical protein | 69 |
| NCLIV_031310 | clathrin coat assembly protein AP50, putative | 69 |
| NCLIV_064230 | pre-mRNA-splicing factor, putative | 68 |
| NCLIV_063760 | hypothetical protein | 66 |
| NCLIV_049140 | hypothetical protein | 63 |
| NCLIV_003100 | serine proteinase inhibitor TgPI-2, putative | 58 |

| | | |
|---|---|---|
| NCLIV_069400 | hypothetical protein, conserved | 57 |
| NCLIV_017430 | hypothetical protein | 55 |
| NCLIV_011800 | hypothetical protein, conserved | 54 |
| NCLIV_029240 | YbaK / prolyl-tRNA synthetases associated domain containing protein, putative | 53 |
| NCLIV_068960 | hypothetical protein | 53 |
| NCLIV_044610 | hypothetical protein, conserved | 50 |
| NCLIV_054110 | YHL017Wp-like protein, related | 50 |

#### 2.3.3.2 1-DE separation prior to LC MS/MS

A further experiment was designed with the intention of combining the successful 1-DE separation achieved in 2.3.1, with the increased resolution of the Orbitrap Velos indicated by the results in 2.3.3.1, where just a single run yielded 28 protein identifications not previously made by the extensive analyses on the LTQ. A tachyzoite lysate equivalent to that analysed in 2.3.1 was prepared and separated on a 10 cm gel, before being excised into 10 sections for trypsin digestion (Figure 2.4). Resulting digests were analysed by LC MS/MS on the Orbitrap Velos and a total of 765 non-redundant protein identifications were made (Mascot score ≥ 50). The identifications, per section, are presented in Appendix I. There was redundancy between sections, nevertheless, overall this analysis yielded 110 novel protein identifications compared to the previous analyses, 59 of which were annotated as 'hypothetical'. Hypothetical proteins are potentially Apicomplexan-specific, as they are often lacking in functional annotation due to a lack of homology to genes of known function. Proteins are often present in more than one area on a gel due to high abundance, post-translational modifications or multiple isoforms migrating at different speeds. Similar occurences were observed in the 1-DE gel described in 2.3.1.

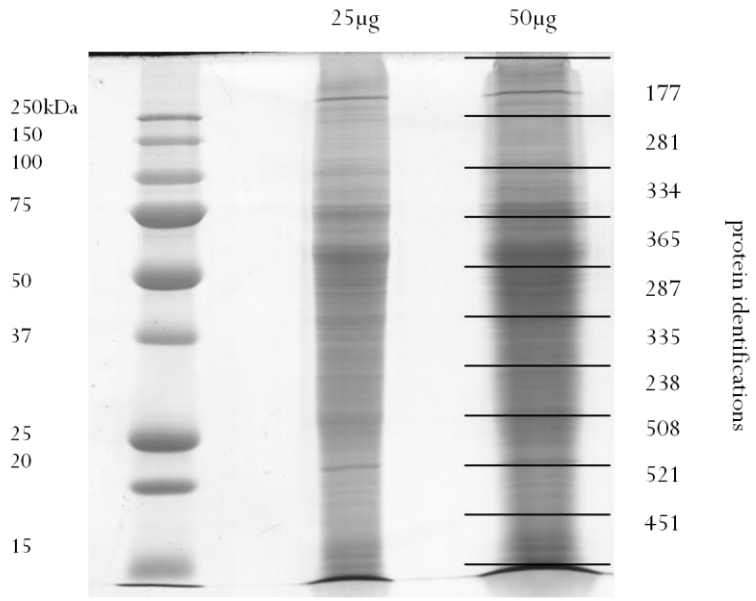



**Figure 2.4:** 1-DE gel of tachyzoite lysates for Orbitrap. Samples were analysed on a 12 % (w/v) acrylamide gel under denaturing conditions, visualised by Colloidal Coomassie staining. Lanes are as follows, lane 1: protein standards, lane 2: 25 µg tachyzoite lysate, lane 3: 50 µg tachyzoite lysate. Horizontal lines indicate excision of 10 bands from the 50 µg sample and the numbers to the right indicate the number of protein identifications made for each with a Mascot score ≥ 50.

### 2.3.4 OFFGEL separation

Appendix III contains an additional in-solution analysis of a tachyzoite lysate by OFFGEL separation, not analysed by LC MS/MS due to protein losses during the clean up process.

## 2.4 DISCUSSION

The analyses presented in this chapter resulted in a combined yield of 1404 protein identifications made using the ToxoDB version 6.0 release of the *N. caninum* genome annotation (Reid *et al.* 2012). The 1-DE gel separation, analysed by LC MS/MS on an LTQ ion trap, produced 975 non-redundant identifications alone, which make up the bulk of the data. These proteins were found to be representative of a wide range of functions and localisations within the cell (Figure 2.2), including the known apical proteins MIC11, DG32, GRA1, GRA7, RON1, RON4 and ROP13. Due to a lack of genes annotated at the time as apical, it was not possible to detect further ROP, GRA and MIC proteins from within those identified. A bioinformatic analysis of the genome to identify further apical genes is presented in Chapter 3, and Table 4.4 summarizes all the apical protein identifications made in the 1-DE LC MS/MS analysis in addition to those made in Chapter 4: Proteomic analysis of the apical organelles. Further discussion on the biological significance of the proteomic identifications is made in Chapter 4.

The number of protein identifications generated by this approach (975) compared favourably to that of Xia *et al.* (2008) who achieved a very similar number (939). Furthermore, the range of protein types identified and the spread of the proteins within the various categories, as seen in Figure 2.2, were comparable in the two studies (Figure 5b in Xia *et al.* 2008) with the categories 'unclassified', 'metabolism', 'protein synthesis', 'protein fate' and cellular transport' being the most densely populated. However, there are two important differences to note that affect this comparison: firstly, approximately 1000 additional genes have been identified in the *T. gondii* genome compared to that of *N. caninum*, depending upon which strain is in question; and secondly, the analysis here comprised three hour mass spectrometry analysis runs as opposed to the one hour runs employed by Xia *et al.* (2008). It became apparent during the data analysis that the three hour chromatography resulted in a larger number of peptide identifications compared to the one hour gradient, but not significantly more protein identifications. Hence, confidence scores assigned to protein hits were higher than they might have been using a one hour gradient, but when this was balanced with a three-fold increase in instrument time, the consensus for future experiments (Chapter 4) was to use one hour chromatography gradient.

Furthermore, the *T. gondii* data were a result of one gel analogous to that analysed here for *N. caninum*, in addition to Tris-soluble and Tris-insoluble samples analysed in the same way.

These additional gels were not a component of the experimental design for this chapter due to the fact that together they identified only a further 82 proteins (Xia *et al.* 2008) which is not a good return on the amount of instrument time the analysis required. This is similar to the reasoning that excluded a 2-DE experiment from the *N. caninum* study. However, there is a further reason why 1-DE was selected in preference to 2-DE for the purpose of maximising protein identifications: a 1-DE gel can be cut into contiguous bands so that every part of the sample can be digested and analysed, whereas it is possible to lose material in a 2-DE gel that has not been sufficiently stained, as spots only are excised, not the whole gel. 2-DE has superior resolving power compared to 1-DE, as a result of the additional dimension, but for this analysis the main concern was to identify as much of the proteome as possible, regardless of how the proteins separated.

The identification of peptides for 'hypothetical' genes confirms the existence of the proteins they encode. The mass spectrometry data generated in this analysis was used by researchers working on the *N. caninum* genome annotation to assess the accuracy of the annotation and solve queries. For example, a novel microneme protein (NCLIV_033690) was identified that appeared to be a duplication of MIC2, which was not present in the *T. gondii* genome. By searching the peptide data for the 1-DE LC MS/MS experiment, peptides uniquely mapping to both genes were identified, thereby confirming experimentally that the novel gene, MIC2B, did exist.

When the data were resubmitted to the version 6.0 annotation, the expectation was that the number of protein identifications would increase, as the gene number had increased ~1.27 fold from 5587 to 7082. However, this was not the case: the 965 identifications for the 1-DE LC MS/MS analysis increased by ten proteins to 975, which equates to a 1.01 fold increase in identifications. The reason for this is not known, but one speculation is that some property of peptides/proteins that makes them readily identified by mass spectrometry also makes the genes they are encoded by easier to detect by gene finding algorithms, so that most of the peptides identified in the analysis were attributed to genes already present in version 5.0.

With regard to the MudPIT data, it was not possible to perform a resubmission to the new genome annotation, due to the disproportionate amount of server time the original .mgf files had taken when searched with the triple-search algorithm (Jones *et al.* 2009). On closer inspection, the spectra were found to contain non-tryptic peptides, an indication that the digestion had not been entirely successful. This adversely affects database searching because,

although it is possible to search with the parameter 'no enzyme specificity', the resulting increase in the database size dramatically reduces confidence levels assigned to peptide hits, as there are so many more peptides possible that the probability of hitting them uniquely is much lower. This is likely to be the main factor in the disappointing number of protein identifications made by this platform (635 non-redundant SEQUEST identifications) when compared to data for *T. gondii* (2121, (Xia *et al.* 2008)) and *C. parvum* (1672, (Sanderson *et al.* 2008), which is 2.6 fold greater despite the smaller genome size of *C. parvum*). Nevertheless, these data did provide novel protein identifications not previously identified by 1-DE, which combined, equated to 25.6 % proteome coverage of the version 5.0 annotation, solely in the tachyzoite stage. It should be noted, that when making comparisons between *N. caninum* and *T. gondii* experimental data, both refer to the tachyzoite stage, while the *C. parvum* data was generated by analysis of sporozoites. The proteins identified by this analysis must therefore be assumed to be either ubiquitously expressed genes over multiple *N. caninum* life stages, or tachyzoite-specific, as any other life-stage specific genes will elude detection based on this experimental design. If culture methods advance or *in vivo* production of sexual stages becomes possible in the future, it would be interesting to analyse additional life stages to gain more perspective of the proteins important to this parasite.

The additional mass spectrometry analyses utilizing an Orbitrap Velos mass spectrometer achieved an additional 138 protein identifications overall, taking the combined non redundant total to 1404, or 19.8 % coverage of the ToxoDB version 6.0 genome annotation. These analyses primarily identified the same proteins as the previous 1-DE LC MS/MS experiment on an LTQ: the novel identifications were lower scoring (though still confident) protein hits, meaning that the most abundant proteins were common to all the different analyses. One advantage that the Orbitrap Velos analysis had was the amount of instrument time it required: the whole analysis of 10 sections of a 10 cm gel took 10 hours of mass spectrometry, compared to 129 bands from a 16cm gel each being analysed for 3 hours on the LTQ, so while the 1-DE LC MS/MS experiment was most successful in identifying the largest proportion of the proteome, it required considerable resources and input to do so. The data generated however, have been useful in aiding genome annotation, confirming the presence of genes previously considered hypothetical, and identifying almost 20 % of the predicted proteome that is now known to be expressed in the tachyzoite stage. The Orbitrap Velos is a more sensitive platform than the LTQ used previously (Makarov 2000), which may explain why it was able to detect

some less-abundant proteins not identified in the previous analyses. Alternatively, their identification may be a result of the different sample preparation procedure employed.

Among the proteins identified, were 22 of the 33 that have been linked to the electron transport chain (Liverpool Library of Apicomplexan Metabolic Pathways (LAMP) database). This finding, and the 21 of 31 glycolysis-related proteins and 17 of 22 tricarboxylic acid cycle proteins (LAMP database) detected, appears to be an indication of a high demand for energy in the tachyzoite, which fits with its rapidly invading and reproducing phenotype. In addition, there were a number of proteins identified which are involved with phospholipid metabolism, which could be related to the production of phospholipid bilayers for daughter cell membranes. Appendix II shows all the proteins involved in *N. caninum* metabolic pathways (downloaded from the LAMP database) and which of these were identified proteomically in the tachyzoite stage.

This analysis has utilized the powerful technique of gel separation followed by liquid chromatography and mass spectrometry to confidently identify almost 20 % of the *N. caninum* genome. These results have been corroborated by using more than one proteomic platform, as the vast majority of identifications were made by multiple analyses. Among the proteins identified were representatives from a range of functional categories, including cell rescue, defence and virulence, into which the proteins involved in the host cell invasion process would fit. In order to find out more about these proteins, it will be necessary to investigate the predicted genome in more depth to identify proteins that may be apical but are not yet annotated as such, and then to experiment with other methods of sample preparation to analyse the proteins of interest. These analyses of whole tachyzoite lysates were a useful precursor to the work in subsequent chapters and were a timely experiment in assisting the *N. caninum* genome annotation project.

# CHAPTER 3: BIOINFORMATIC MINING OF THE GENOME FOR PREDICTED APICAL PROTEINS

## 3.1 INTRODUCTION

At the beginning of this study the *N. caninum* genome was undergoing annotation and, as discussed in Chapter 2. Section 3.2, very few apical proteins were annotated in Release 5.0 of ToxoDB (Gajria *et al.* 2008; Reid *et al.* 2012). This chapter describes an effort to identify further genes encoding apical proteins, based on their similarity to known *N. caninum* apical genes and to those of *T. gondii*, whose genome is largely syntenic to that of *N. caninum* (DeBarry & Kissinger 2011; Reid *et al.* 2012).

### 3.1.1 Definitions

The following definitions describe the use of these words in the context of this thesis.

*Synteny*: collinear conservation of gene order along the chromosomes of different species (Blood, Studdert & Gay 2007).

*Homology*: the structural similarity of two genes/proteins that is a result of common descent (Koonin 2005).

*Orthologue*: a gene/protein homologous to another, but existing within the genome of another species that would have arisen from a common ancestor (Koonin 2005).

*Paralogue*: a homologue of another gene/protein found within the same genome, i.e. a gene that has been duplicated at some point (Koonin 2005).

### 3.1.2 ToxoDB

ToxoDB (Gajria *et al.* 2008) is an open source database that has been developed as a component of EuPathDB (Aurrecoechea *et al.* 2007); a collection of websites hosting a wealth of genomic information on a range of eukaryotic pathogens. ToxoDB is not only the home of the various *T. gondii* genomes that have been sequenced to date (Lis Caler at the J. Craig Venter Institute, hosted on ToxoDB) but also the *N. caninum* genome (Reid *et al.* 2012), as there is no 'NeosporaDB'. Figure 3.1 shows a representative screen shot of a ToxoDB gene page and the orthologous *T. gondii* genes can be seen shaded with the *N. caninum* gene

highlighted in yellow. ToxoDB provides links to an orthologue database named OrthoMCL (Chen *et al.* 2006; Li, Stoeckert & Roos 2003) where more information is available on the orthologue group a gene has been assigned to.

ToxoDB also hosts data on gene expression that has been submitted by researchers around the world, and a number of genes are linked to comments and PubMed citations of publications in which they feature. Information about the gene such as coding sequence, predictions of signal peptides, number of orthologues and gene ontology classifications can be downloaded and as such this database serves as an extremely useful tool for anyone working on these organisms.

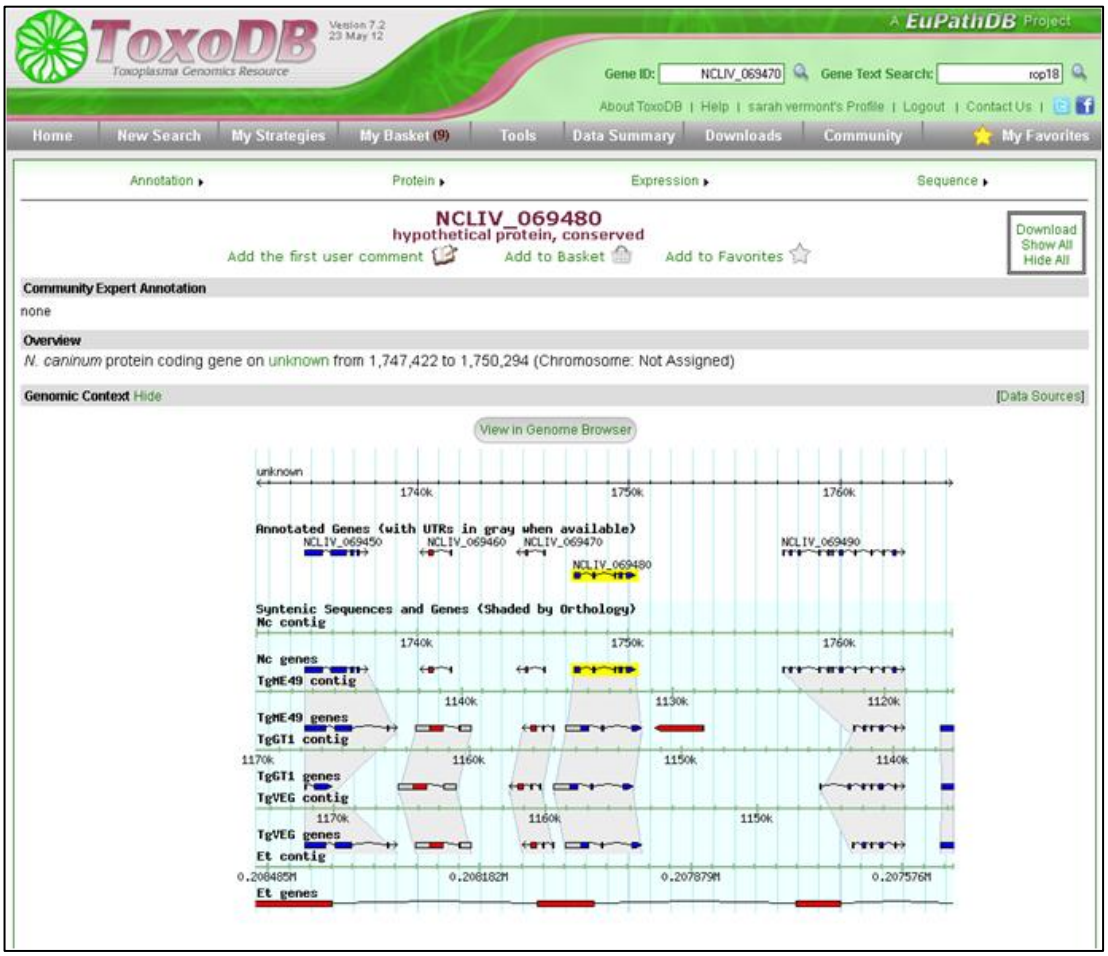



**Figure 3.1**: Example gene page from ToxoDB for *N. caninum* gene NCLIV_069480 (highlighted in yellow). *T. gondii* genes are aligned below with orthology shown by shading between the genes, for this example there is a syntenic orthologue present in each *T. gondii* strain.

### 3.1.3 The Basic Local Alignment Search Tool (BLAST)

BLAST (NCBI) is a bioinformatic resource that allows comparisons of nucleotide or protein sequences. A BLAST alignment provides sequence identity score (%) in addition to an E-value

(expect value: the number of hits that are expected to be seen with that score by chance). The blastp algorithm performs protein-protein BLAST searches whilst the blastn algorithm uses nucleotide sequences. A search can be set up to analyse a database of all non-redundant proteins (blastp) which includes sequences from GenBank and SwissProt for all organisms (an example is shown in Figure 3.2). Alternatively, specific databases can be selected, or there is a BLAST search facility incorporated into ToxoDB for searching against *T. gondii* and *N. caninum* databases. BLAST is very useful for inferring orthologous relationships and for searching for possible functions when a gene is annotated simply as 'hypothetical'.

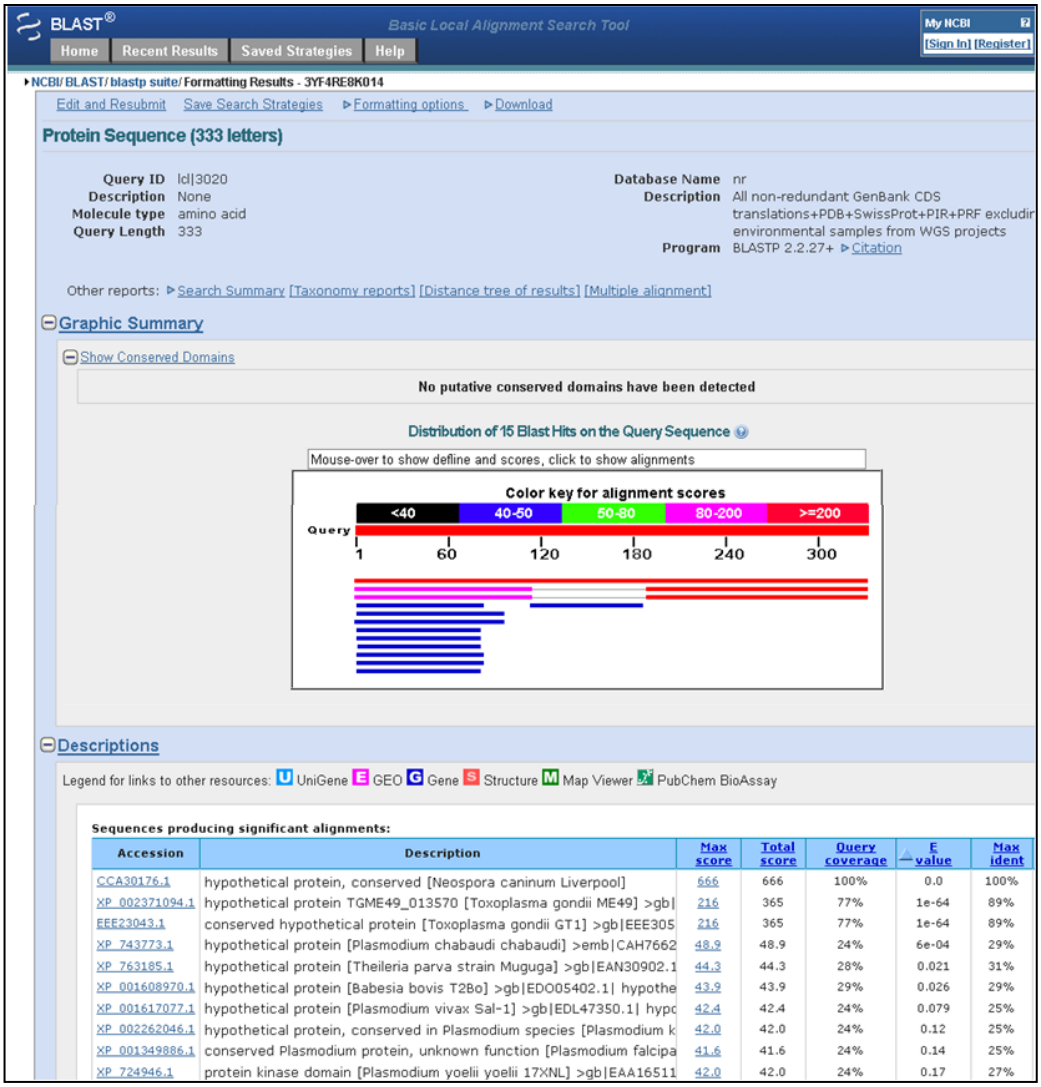



**Figure 3.2:** Example of a protein-protein BLAST alignment of NCLIV_069480 against the 'all non redundant protein sequences' database on NCBI. The top hit is to the query gene itself and has 100 % identity, as would be expected. *T. gondii* ME49 and GT1 strains have orthologues with 89 % identity and highly significant E-values (1 x $10^{-64}$). The *P. chabaudi* protein is also significant to a cut off of 0.001 (E-value 0.0006) despite only having an identity score of 29 %. Other lower scoring hits are also to apicomplexan proteins.

### 3.1.4 Aims and objectives

After the global analysis of the tachyzoite proteome (Chapter 2), the next step was to simplify the sample to an enriched preparation of apical proteins (Chapter 4). Any proteomic study is highly reliant on the database that the peptides are searched against, as a list of hits to proteins which relate to poorly annotated genes does not provide much useful information. It was therefore important to carry out bioinformatic and proteomic analyses side by side, so that ideally, they would complement one another. This was especially important because the proteomic techniques were unable to produce pure samples of apical proteins, so that contamination from other organelles and structural and host proteins were always present. Having a comprehensive list of predicted apical proteins was therefore necessary to enable the identification of the apical proteins present within the data sets generated by proteomic experiments (in Chapters 2 and 4).

At the time this analysis began, when the experiments in Chapters 2 and 4 were underway, the available version of the *N. caninum* genome was ToxoDB Release 5.0, for which annotation was still in progress. As the genome annotation continued, and more information was made available, the data presented in this chapter were updated and expanded. As such, it represents an amalgamation of information taken from ToxoDB versions 5.0, 6.0 and 7.0, in addition to published literature. Moreover, the *T. gondii* genome (ME49 strain) was annotated to a far greater extent than the *N. caninum* genome, so there was an opportunity to exploit this information and seek orthologous relationships that had not already been identified in the first annotation (ToxoDB Release 5.0). The aim was that this work would contribute to the genome annotation (being carried out by the Sanger Institute) as part of a collaborative effort. Clearly, genome annotation is ever-evolving and there will doubtless be additional apical genes identified in the future, and corrections made to existing annotations.

## 3.2 MATERIALS AND METHODS

As a component of the *N. caninum* genome project (Reid *et al.* 2012), a bioinformatic search was undertaken to identify genes encoding proteins that were hypothesized as apical in origin. The author and Adam Reid each performed bioinformatic and literature searches and the results of each were taken into account for the analysis in Reid *et al.* (2012).

While some genes have experimental data regarding their localisation and can be confidently assigned to an apical organelle, others have only sequence similarity to known apical genes (which may include conserved domains) to suggest that they may be a member of the same family. The work presented in this chapter is highly qualitative and somewhat subjective, and has undergone constant revision as new versions of the genome annotation were released.

### 3.2.1 Identification of potential apical genes

Initially, a range of text searches were performed in ToxoDB for the phrases 'rhoptry', 'rop' 'ron' 'dense granule' 'gra' 'microneme' 'mic' 'ama'. The results of these were used as a starting point to build upon, by performing protein-protein BLAST searches (in ToxoDB) to *T. gondii* coding sequences and identifying potential orthologues in *N. caninum*. Some orthologous relationships were already identified by ToxoDB. Criteria for assigning an orthologue based on BLAST sequence similarity were as follows: BLAST identity scores $\geq$ 30 %, with an E-value $\leq$ 0.001 and being the top hit within that organism (i.e. the *N. caninum* orthologue of a *T. gondii* gene had to be the top hitting *N. caninum* gene to that sequence).

In cases where a *T. gondii* gene with an apical annotation had an *N. caninum* syntenic orthologue (identified by ToxoDB) that was only described as 'hypothetical', the *N. caninum* gene was considered likely to be of similar function and designated the same. Where there was an apparent expansion at a locus in either organism and as a result one organism had an extra copy (or copies) of that gene, the extra copy was labelled 'B' (for example MIC2, and MIC2B). When gene comments on ToxoDB contradicted each other, the related literature was examined to attempt to discern which was correct. Published literature on apical proteins was also consulted, but in many cases authors did not specify gene accession numbers (I.D.'s) so it was not always possible to determine the exact gene that they referred to. SignalP 3.0 predictions for signal peptides (Bendtsen *et al.* 2004) were downloaded from ToxoDB version 6.0.

## 3.3 RESULTS AND DISCUSSION

The genes in Tables 3.1, 3.2 and 3.3 have been presented with their ToxoDB version 6.0 annotations. It should be noted that this analysis has evolved over time and when it initially began there were only fifteen, eight and five genes identified by a text search of 'microneme', 'rhoptry' and 'dense granule', respectively, on ToxoDB Release 5.0. This annotation has now been superseded by Release 6.0 (there have been more recent releases of ToxoDB but Release 6.0 was the last to contain a new *N. caninum* annotation) and when searches were repeated in version 6.0 the yield of *N. caninum* genes was higher (twenty seven, forty-two, and eight for 'microneme', 'rhoptry' and 'dense granule', respectively). This is an example of how better databases facilitate identification of genes. The tables also include SignalP 3.0 predictions for signal peptides: a number of apical proteins are predicted to have signal peptides, there are also many apical proteins that do not, so whether this is due to their secretion being via a non-classical pathway (and hence not recognised by the SignalP algorithm) or whether they are non-secretory proteins is unknown.

### 3.3.1 Microneme Genes

The list of putative microneme genes identified is presented in Table 3.1. There were 42 identified, including five which did not have orthologues in *T. gondii* (MCP 5, MCP6 and MCP7, MIC2B and MIC19). There were no *T. gondii*-specific microneme genes identified.

MIC2B (NCLIV_033690) BLASTs to MIC2 with an E-value 2.3 x $10^{-153}$ and 45% identity and has been dubbed a duplication event (Reid *et al.* 2012); the two genes can be viewed in Figure 3.3. MIC2B has been known by a number of aliases; it was initially named MIC2B in recognition of its paralogous relationship with MIC2, but then became MIC14 in the publication Reid *et al.* (2012). It is currently annotated as MIC26 in comment 105330 on ToxoDB, but this is possibly due to human error as '2b' resembles '26'. MIC2B has no further orthologues in *T. gondii* other than those syntenic to MIC2. Peptides for both genes were identified in the tachyzoite experiment in Chapter 2 (Mass Spectrometry data in Appendix I). The function of MIC2B is currently unknown but would make an interesting target for functional analyses due to the fact that it is specific to *N. caninum*. MIC2 is known to be involved in gliding and motility during invasion (Carruthers & Sibley 1997) and the trafficking of MIC2-associated protein (M2AP) (Jewett & Sibley 2004); and a reduction in expression results in reduced virulence in *T. gondii* (Huynh & Carruthers 2006).

**Table 3.1:** Putative list of microneme genes (n = 42). Signal peptide predictions were downloaded from ToxoDB version 6.0 and had been made using SignalP 3.0. The comments column contains information regarding genes which did not have a straightforward syntenic-orthologous relationship with a known *T. gondii* microneme gene; '-' denotes no gene identified/no comment.

| Protein Description | I.D. | ToxoDB Annotation | Signal peptide? | Comments | *T. gondii* ME49 orthologue(s) |
|---|---|---|---|---|---|
| MIC1 | NCLIV_043270 | microneme protein MIC1 | Yes | Contains a MAR domain so is the equivalent of MCP1 | TGME49_091890 |
| MIC2 | NCLIV_022970 | microneme protein 2 | Yes | Paralogue of NCLIV_033690 | TGME49_001780 |
| MIC2 paralogue | NCLIV_033690 | hypothetical protein | Yes | Putatively MIC2B - duplication of MIC2 (NCLIV_022970), also known as 'MIC26' and 'MIC14' | - |
| M2AP | NCLIV_051970 | MIC2-associated protein M2AP | Yes | - | TGME49_014940 |
| MIC3 | NCLIV_010600 | microneme protein MIC3 | Yes | - | TGME49_119560 |
| MIC4 | NCLIV_002940 | microneme protein MIC4 | Yes | - | TGME49_008030 |
| MIC5 | NCLIV_068520 | microneme TgMIC5 protein | Yes | - | TGME49_077080 |
| MIC5 or MIC17 | NCLIV_038120 | Microneme protein 5 (Precursor), related | Yes | Paralogues: NCLIV_038100, NCLIV_068830 and NCLIV_038110 looks like incorrect annotation as MIC5, also known as 'MIC17A' by Sohn *et al.* (2011) | TGME49_000230 TGME49_000250 TGME49_000270 TGME49_000240 |
| MIC17 paralogue | NCLIV_038100 | hypothetical protein | Yes | Known as 'MIC17C' by Sohn *et al.* (2011) | |
| MIC17 paralogue | NCLIV_038110 | hypothetical protein | Yes | Known as 'MIC17B'; confirmed by immunolocalisation (Sohn *et al.* 2011) | |
| MIC17 paralogue | NCLIV_068830 | PAN domain-containing protein , related | Yes | Paralogue of NCLIV_038100, NCLIV_038110 and NCLIV_038120 | |
| MIC6 | NCLIV_061760 | microneme protein MIC6, putative | No | - | TGME49_018520 |
| MIC7 | NCLIV_025710 | microneme protein 7, putative | No | - | TGME49_061780 |
| MIC8 | NCLIV_062770 | microneme protein 8, putative | No | MIC8 and MIC9 are tandemly arranged | TGME49_045490 |
| MIC8B | NCLIV_062750 | EGF-like domain-containing protein | No | | TGME49_086740 |
| MIC9 | NCLIV_062760 | hypothetical protein | No | | TGME49_045490 |
| MIC10 | NCLIV_066250 | microneme protein 10, putative | Yes | - | TGME49_050710 |
| MIC11 | NCLIV_020720 | microneme protein MIC11, putative | Yes | - | TGME49_004530 |

| | | | | | |
|---|---|---|---|---|---|
| MIC12 | NCLIV_069310 | microneme protein, putative | No | Paralogue of NCLIV_022530 | TGME49_002400<br>TGME49_067680 |
| MCP2 | NCLIV_026810 | NcMCP2, putative | Yes | Also known as MIC13 | TGME49_060190 |
| MCP3 | NCLIV_003260 | NcMCP3, putative | Yes | - | TGME49_008740 |
| MCP4 | NCLIV_003250 | NcMCP4, putative | Yes | - | TGME49_008730 |
| MCP5 | NCLIV_066750 | NcMCP5, putative | No | No *N. caninum* or *T. gondii* homologues | - |
| MCP6 | NCLIV_054450 | NcMCP6, putative | Yes | No *T. gondii* orthologues, paralogue to NcMCP7 | - |
| MCP7 | NCLIV_054425 | NcMCP7, putative | Yes | No *T. gondii* orthologues, paralogue to NcMCP6 | - |
| - | NCLIV_007140 | microneme protein, putative | No | MIC22 (Reid *et al.* 2012) | TGME49_075790 |
| - | NCLIV_008720 | microneme protein, putative | Yes | MIC24 (Reid *et al.* 2012) | TGME49_054430 |
| - | NCLIV_013920 | microneme protein, putative | No | - | TGME49_086740 |
| - | NCLIV_015580 | microneme protein, putative | No | MIC20 (Reid *et al.* 2012) | TGME49_038210 |
| - | NCLIV_015590 | microneme protein, putative | No | MIC21 (Reid *et al.* 2012) | TGME49_038220 |
| - | NCLIV_018780 | microneme protein, putative | No | MIC15 (Reid *et al.* 2012) | TGME49_044180 |
| - | NCLIV_058210 | microneme protein, putative | No | MIC16 (Reid *et al.* 2012) | TGME49_115520 |
| - | NCLIV_058240 | microneme protein, putative | No | MIC25 (Reid *et al.* 2012) | TGME49_115550 |
| AMA1 | NCLIV_028680 | apical membrane antigen 1, putative | No | - | TGME49_055260 |
| AMA2 | NCLIV_058410 | apical membrane antigen, putative | No | - | TGME49_115730 |
| AMA3 | NCLIV_064590 | apical membrane antigen, putative | Yes | - | TGME49_100130 |
| - | NCLIV_038380 | hypothetical protein | No | Similarity to a *Plasmodium* microneme protein (geneDB), but no *T. gondii* ME49 orthologue (MIC19, Reid *et al*. 2012) | - |
| - | NCLIV_058230 | hypothetical protein | No | Adjacent and BLASTs to putative microneme protein NCLIV_058240 with identity of 49 % and E-value of 5.2 x10 $^{-72}$ | TGME49_115540 |
| - | NCLIV_036660 | hypothetical protein, conserved | Yes | BLASTs to putative microneme protein NCLIV_069310 with 33 % identity and E-value of 1.1 x $10^{-12}$ | TGME49_069930 |
| - | NCLIV_022530 | fibrillin-2 precursor, putative | No | Paralogue of putative microneme protein NCLIV_069310 | TGME49_002400<br>TGME49_067680 |
| - | NCLIV_029340 | EGF-like domain-containing protein, putative | Yes | Syntenic orthologue to *T. gondii* VEG microneme protein TGVEG_078790 | TGME49_055460 |
| SUB1 | NCLIV_021050 | hypothetical protein | Yes | Orthologue of microneme protein TgSUB1 (Miller *et al.* 2001) | TGME49_004050 |

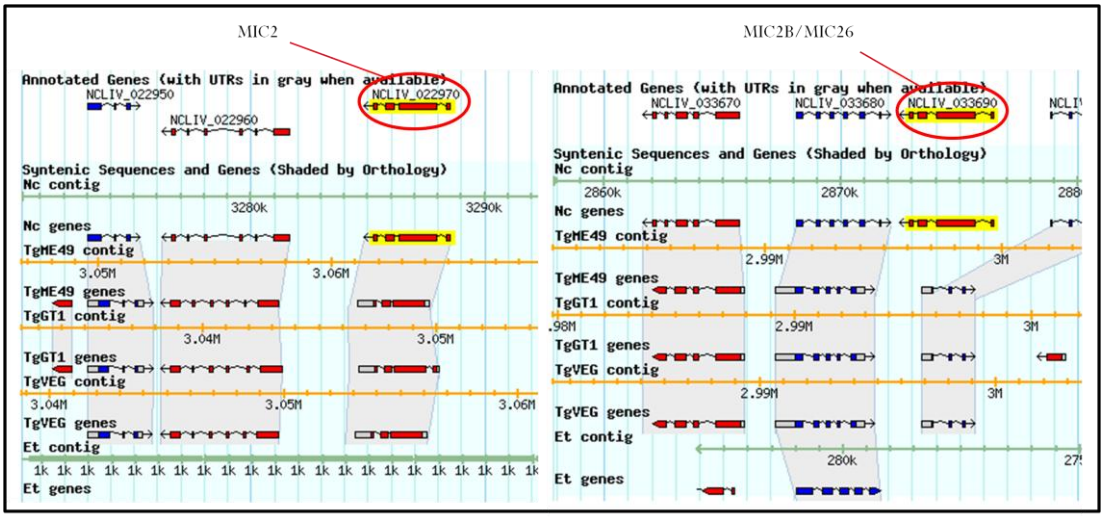



**Figure 3.3**: MIC2 gene duplication. ToxoDB screen shots of MIC2 (NCLIV_022970) and its paralogue NCLIV_033690.

MIC5 and MIC17 provide another area of confusion in the genome annotation. Both NCLIV_068520 and NCLIV_038120 have been labelled as MIC5 ('microneme TgMIC5 protein, putative' and 'Microneme protein 5 (Precursor), related', respectively) but NCLIV_068520 is likely to be the true MIC5 as it is a syntenic orthologue of *T. gondii* MIC5 (TGME49_077080). Sohn *et al.* (2011) proposed the name MIC17B for a novel microneme protein (NCLIV_038110) identified in their monoclonal antibody localisation experiment. MIC17B is flanked by its paralogues NCLIV_038120 and NCLIV_038100, which were designated MIC17A and MIC17C respectively: this locus appears to have undergone gene duplication at some point, while the *T. gondii* genes appear to have behaved in a similar way (Figure 3.4). These three microneme genes are highly homologous and are 89 and 64 % identical (BLAST) to another paralogue, NCLIV_068830, (not identified by Sohn *et al.* (2011)) which is located on the 'unknown' chromosome of *N. caninum*.

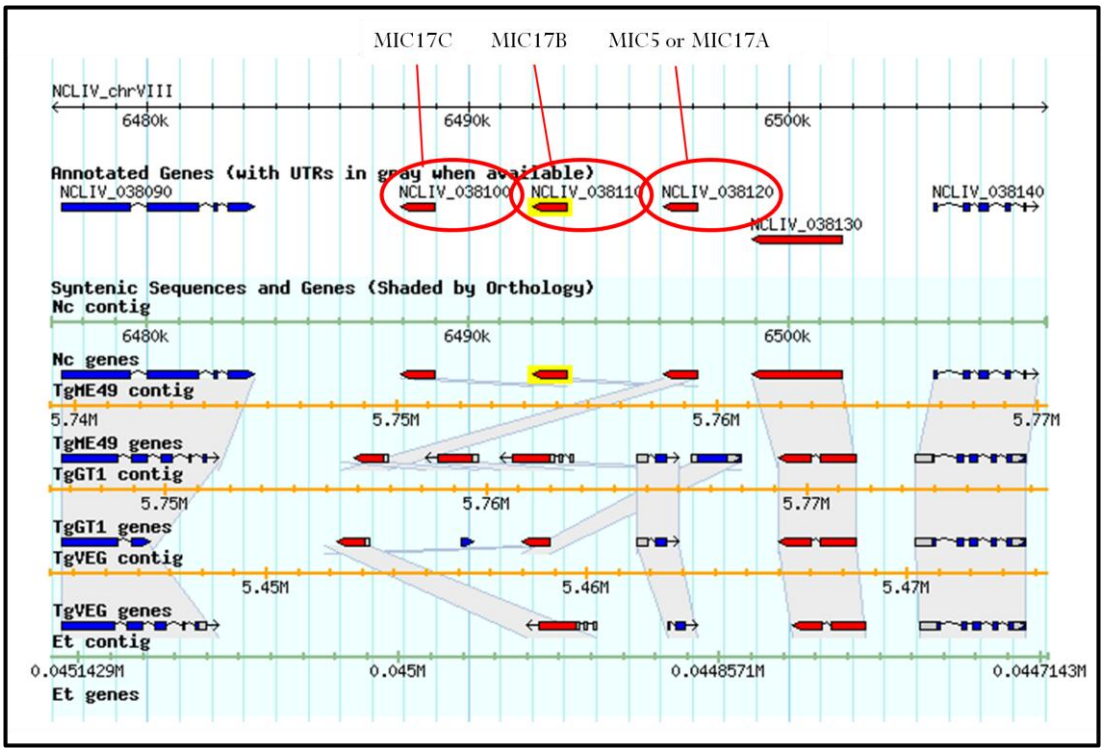



**Figure 3.4:** Apparent expansion at the MIC17 locus. MIC17A, B and C are all paralogues of one another.

The 'microneme adhesive repeat' (MAR) domain containing proteins (MCPs) are coccidian-specific proteins thought to be involved in the ability of apicomplexan parasites to exhibit different tissue tropisms (Friedrich *et al.* 2010), by binding to sialic acid and other ligands on host cells. The fact that *N. caninum* possesses three more MCPs than *T. gondii* is very interesting as *N. caninum* is generally considered to exhibit a narrower host range than *T. gondii*, for example, *N. caninum* is not zoonotic, despite the fact that *in vitro* it is capable of infecting human cells. There may be other roles for MCPs that have not yet been identified, as suggested by Friedrich *at al*. (2010); due to the fact that not all MCPs identified would have the ability to bind host cells. Hence, the three *N. caninum* MCPs not shared by *T. gondii* would make interesting subjects for further investigation, as a better understanding of the differences in host cell recognition between the two genera could help to inform vaccine development for both neosporosis and toxoplasmosis.

### 3.3.2 Rhoptry Genes

The rhoptries are the apical organelles that exhibit the greatest difference in gene/protein repertoire between *N. caninum* and *T. gondii*. Rhoptry proteins have been the subject of a large number of studies and reviews due to their importance in invasion and virulence. As such, many of the genes have been known by a range of different names and the different publications are sometimes contradictory. This analysis attempted to combine the naming strategies used by different researchers, but due to the changeable nature of the genome annotation it was at times difficult to keep up to date with current information for every rhoptry gene. The review by Boothroyd and Dubremetz (2008) listed *T. gondii* rhoptry proteins, but was published using ToxoDB version 4.2 gene models, as opposed to the version 6.0 gene models in this analysis. The naming conventions used were suggested by Bradley *et al*. (2005), who also identified two distinct groups of rhoptry proteins localised to the rhoptry neck (RONs) and the rhoptry bulb (ROPs), respectively. Reid *et al.* (2012) contained an analysis of the rhoptry genes of *N. caninum* and *T. gondii,* and included updated annotations not yet available on ToxoDB, identified by gene-finding algorithms employed by A. Reid. The putative rhoptry genes presented here are those which were identified using the approaches outlined in the methods (Section 3.2) and any additional information on name allocations (from Reid *et al.* (2012)) has been inserted in the comments column of Table 3.2.

**Table 3.2:** Putative list of rhoptry genes (n = 58). Signal peptide predictions were downloaded from ToxoDB version 6.0 and had been made using SignalP 3.0. The comments column contains information regarding genes which did not have a straightforward syntenic-orthologous relationship with a known *T. gondii* rhoptry gene; '-' denotes no gene identified/no comment.

| Protein description | I.D. | ToxoDB annotation | Signal peptide? | Comments | *T. gondii* ME49 orthologue |
|---|---|---|---|---|---|
| ROP1 | NCLIV_053840 | hypothetical protein | Yes | - | TGME49_109590 |
| ROP2A | - | - | - | ROP2 not present in *N. caninum* | TGME49_015780 |
| ROP2B | - | - | - | | TGME49_075300 |
| ROP4 | NCLIV_001970 | hypothetical protein | Yes | Orthologue of ROP7 (NCLIV_001950) and *T. gondii* ROP4 (TGME49_095110) and ROP8 (TGME49_015770) | TGME49_095110 |
| ROP5 | NCLIV_060730 | hypothetical protein | Yes | Paralogue of NCLIV_060740 | TGME49_108080 |
| ROP5B | NCLIV_060740 | ROP 2, related | Yes | Paralogue of ROP5 (NCLIV_060730), incorrectly annotated as ROP2 | - |
| ROP6 | NCLIV_027850 | rhoptry protein 6, putative | Yes | NCLIV_027840 | TGME49_058660 |
| ROP7 | NCLIV_001950 | Rhoptry protein ROP7 | Yes | NCLIV_001970, TGME49_015770=ROP8, Toxo ROP2A and 2B | TGME49_095110 |
| ~~ROP8~~ | NCLIV_004220 | Rhoptry antigen ROP8 (EC 2.7.11.26), related | Yes | Incorrect annotation, more similar to *T. gondii* ROP42, ROP43, ROP 44 (all syntenic orthologues). No ROP8 exists in *N. caninum* | TGME49_015770 |
| ROP9 | NCLIV_018420 | p36 protein, putative | Yes | - | TGME49_043730 |
| ROP10 | NCLIV_058180 | Rhoptry protein 10, related | No | - | TGME49_115490 |
| ROP11 | NCLIV_045580 | hypothetical protein, conserved | No | - | TGME49_027810 |
| ROP12 | NCLIV_021100 | hypothetical protein, conserved | Yes | - | TGME49_003990 |
| ROP13 | NCLIV_055850 | hypothetical protein, conserved | Yes | - | TGME49_112270 |
| ROP14 | NCLIV_057960 | hypothetical protein | No | - | TGME49_115220 |
| ROP15 | NCLIV_011690 | ROP15 protein, related | Yes | - | TGME49_011290 |
| ROP16 | NCLIV_025120 | Rhoptry kinase family protein ROP16, putative | No | - | TGME49_062730 |
| ROP17 | NCLIV_027930 | Rhoptry kinase family protein ROP17, putative | No | - | TGME49_058580 |
| ROP18 | - | - | - | Pseudogene in *N. caninum* - stop codons, see Chapter 5 | TGME49_005250 |

| | | | | | |
|---|---|---|---|---|---|
| ROP19 | NCLIV_017440 | hypothetical protein | Yes | Orthologue of NCLIV_017410 (ROP38), NCLIV_ 017420 (Zgc), NCLIV_017430 (ROP29) | TGME49_042240 |
| ROP20 | NCLIV_028170 | Rhoptry kinase family protein ROP20, putative | No | - | TGME49_058230 |
| ROP21 | NCLIV_024700 | Rhoptry kinase family protein ROP21, putative | Yes | - | TGME49_063220 |
| ROP22 | NCLIV_002650 | Rhoptry kinase family protein ROP22 (incomplete catalytic triad), putative | No | - | TGME49_007700 |
| ROP23 | NCLIV_016220 | Rhoptry kinase family protein ROP23 (incomplete catalytic triad), putative | Yes | - | TGME49_039600 |
| ROP24 | NCLIV_068850 | hypothetical protein | Yes | - | TGME49_052360 |
| ROP25 | NCLIV_022130 | Rhoptry kinase family protein ROP25, putative | No | - | TGME49_002780 |
| ROP26 | NCLIV_011730 | Rhoptry kinase family protein ROP26 (incomplete catalytic triad), putative | No | - | TGME49_011260 |
| ROP27 | NCLIV_056620 | Rhoptry kinase family protein ROP27, putative | Yes | - | TGME49_113330 |
| ROP28 | NCLIV_028130 | Rhoptry kinase family protein ROP28, putative | No | - | TGME49_058370 |
| ROP29 | NCLIV_017430 | hypothetical protein | Yes | Orthologue of NCLIV_017410 (ROP38), NCLIV_ 017420 (Zgc), NCLIV_017440 (ROP19) | TGME49_042230 |
| ROP30 | NCLIV_046000 | Rhoptry kinase family protein ROP30, putative | No | - | TGME49_027010 |
| ROP31 | NCLIV_027710 | Rhoptry kinase family protein ROP31, putative | No | - | TGME49_058800 |
| ROP32 | NCLIV_035860 | Rhoptry kinase family protein ROP32, putative | No | - | TGME49_070920 |
| ROP33 | NCLIV_023260 | Rhoptry kinase family protein ROP33, putative | Yes | - | TGME49_001130 |
| ROP34 | NCLIV_000650 | Rhoptry kinase family protein ROP34, putative | Yes | - | TGME49_040090 |
| ROP35 | NCLIV_044410 | Rhoptry kinase family protein ROP35, putative | Yes | - | TGME49_104740 |
| ROP36 | NCLIV_002580 | Rhoptry kinase family protein ROP36 (incomplete catalytic triad), putative | No | - | TGME49_007610 |
| ROP37 | NCLIV_001460 | Rhoptry kinase family protein ROP37 (incomplete catalytic triad), putative | No | - | TGME49_094560 |
| ROP38 | NCLIV_017410 | hypothetical protein | Yes | Orthologue of NCLIV_017430 (ROP29), NCLIV_ 017420 (Zgc), NCLIV_017440 (ROP19) | TGME49_042110 |
| ROP40 | NCLIV_012920 | Rhoptry kinase family protein ROP40 (incomplete catalytic triad), putative | No | - | TGME49_091960 |
| ROP41 | NCLIV_048060 | Rhoptry kinase family protein ROP41, putative | No | - | TGME49_066100 |

| | | | | | |
|---|---|---|---|---|---|
| ROP42<br>ROP43<br>ROP44 | NCLIV_004230<br><br>NCLIV_004220 | Rhoptry antigen ROP8, related | Yes<br><br>Yes | ROP8 incorrectly annotated in *N. caninum*, there is no orthologue to *T. gondii* ROP8. NCLIV_004220 and NCLIV_004230 are homologous to *T. gondii* ROP42, 43 and 44 | TGME49_009980<br>TGME49_010090<br>TGME49_010110 |
| ROP45 | NCLIV_023580 | Rhoptry kinase family protein ROP45 (incomplete catalytic triad), putative | Yes | - | TGME49_081670<br>TGME49_081790 |
| ROP46 | NCLIV_030990 | Rhoptry kinase family protein ROP46, putative | No | - | TGME49_030470 |
| RON1 | NCLIV_054120 | sushi domain containing protein | Yes | - | TGME49_110010 |
| RON2 | NCLIV_064620 | rhoptry neck protein 2, putative | Yes | - | TGME49_100100 |
| RON2L1 | NCLIV_001400 | hypothetical protein, conserved | Yes | Identified from John Boothroyd comment 2009 (ToxoDB, comment ID 23680) | TGME49_094400 |
| RON2L2 | NCLIV_040110 | hypothetical protein, conserved | Yes | Identified from John Boothroyd comment 2007 (ToxoDB, comment ID 19792) | TGME49_065120 |
| RON3 | NCLIV_048590 | hypothetical protein, conserved | Yes | - | TGME49_023920 |
| RON3L1 | NCLIV_020340 | Hypothetical protein | No | - | TGME49_005370<br>TGME49_005360 |
| RON4 | NCLIV_030050 | hypothetical protein, conserved | Yes | - | TGME49_029010 |
| RON4L1 | NCLIV_007800 | Tg65, related | Yes | Identified as rhoptry from Boothroyd & Dubremetz (2008) | TGME49_053370 |
| RON5 | NCLIV_055360 | hypothetical protein, conserved | Yes | - | TGME49_111470 |
| - | NCLIV_065640 | rhoptry kinase family protein | Yes | - | TGME49_049470 |
| - | NCLIV_057950 | lipase maturation factor 2, related | No | Syntenic to 'rhoptry protein' in *T. gondii,* ROP14B in Reid *et al.* (2012) | TGME49_115210 |
| - | NCLIV_058560 | rhoptry protein, putative | No | ROP48 (Reid *et al.* 2012) | TGME49_115940 |
| - | NCLIV_007770 | Rhoptry kinase family protein, truncated (incomplete catalytic triad), putative | Yes | - | TGME49_053330 |
| - | NCLIV_017420 | Zgc:55863, related | Yes | Orthologue of NCLIV_017430 (ROP29), NCLIV_ 017410 (ROP38), NCLIV_017440 (ROP19), ROP47 in Reid *et al.* (2012) | TGME49_042120 |
| - | - | - | - | ROP19 in *T. gondii*- no apparent syntenic one to this ROP19 in *N. caninum* (2 different genes annotated as ROP19 in *T. gondii*) | TGME49_042250 |
| - | NCLIV_051340 | toxofilin, putative | Yes | Toxofilin (Bradley *et al.* 2005) | TGME49_014080 |
| Toxopain 1 | NCLIV_069550 | Cathepsin B-like protease (Precursor), related | Yes | Toxopain1, John Boothroyd comment 2007 (ToxoDB, comment ID 19790) | TGME49_049670 |
| - | NCLIV_068890 | hypothetical protein | Yes | ROP51 (Reid *et al.* 2012), no *T. gondii* orthologue | - |

ROP2A and ROP2B, as present in *T. gondii*, were not identified in the *N. caninum* genome. ROP2 has been implicated in invasion and has been localised to the parsitophorous vacuole membrane (Sinai & Joiner 2001) but is also part of a wider family (the 'ROP2 family') of rhoptry kinases that includes the orthologues ROP4, ROP5 and ROP7. These latter rhoptry proteins do exist in *N. caninum*, in fact there is an additional ROP5 orthologue named ROP5B, so it may be that there is a certain amount of redundancy in this family that allows the two parasites (*T. gondii* and *N. caninum*) to function effectively despite differing ROP2 family repertoires. However, ROP5 has recently been identified as a major effector of virulence in *T. gondii* and while apparently present in the genome as one gene only (TGME49_108080) it is actually present in a range of copy numbers (four to ten) dependent on the *T. gondii* strain (Reese *et al.* 2011; Reese & Boothroyd 2011). The relationship between copy number and virulence is not yet clear as the intermediate-virulence type II strain ME49 has the highest copy number (ten) compared to the virulent RH strain (six) and the avirulent VEG strain (four).

The *N. caninum* gene NCLIV_004220 appears to have been incorrectly annotated as ROP8. Figure 3.5 shows the NCLIV_004220 and NCLIV_004230 region: NCLIV_004220 is syntenically orthologous to *T. gondii* ROP42 (ME49_009980) and NCLIV_004230 BLASTs to NCLIV_004220 with 91 % identity and an E-value of $1 \times 10^{-76}$ despite not being identified by ToxoDB as an orthologue of ROP42, ROP43 or ROP44. NCLIV_004230 has therefore been included in the set of rhoptry genes as an orthologue of *N. caninum* ROP42 (putatively NCLIV_004220), which could possibly represent *N. caninum* ROP43 or ROP44. It shares 37 and 35 % identity with *T. gondii* ME49 ROP44 and ROP43 (E-values $2.8 \times 10^{-13}$ and $1.5 \times 10^{-12}$) respectively.

Reid *et al.* (2012) identified NCLIV_068890 as an *N. caninum* specific rhoptry gene and named it ROP51. The gene has not been assigned to a chromosome but the protein product was identified in the tachyzoite lysate analysed in Chapter 2 and BLASTs to a number of other *N. caninum* and *T. gondii* rhoptry proteins.

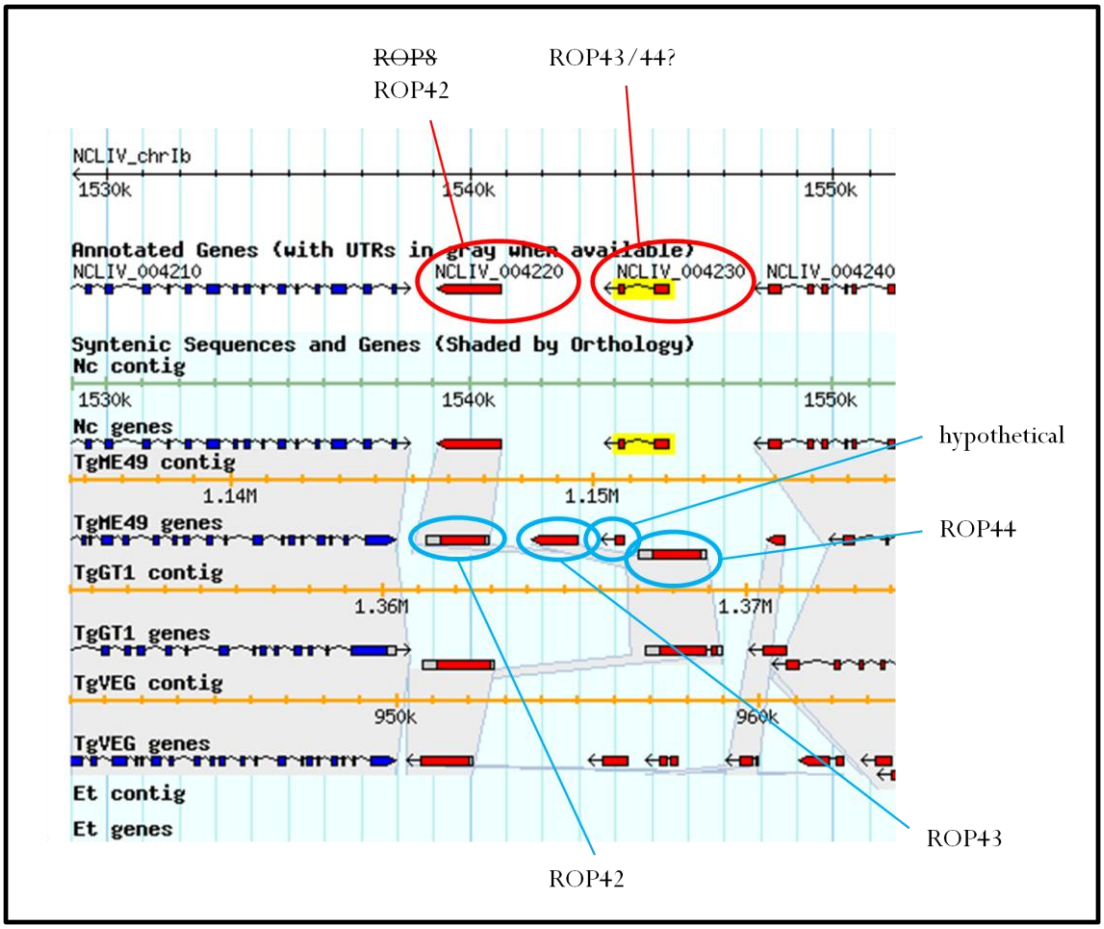



**Figure 3.5:** *N. caninum* ROP42 appears to have been incorrectly annotated as ROP8. Screen shot of ToxoDB gene page for NCLIV_004230. *N. caninum* genes are ringed in red, *T. gondii* ME49 genes are ringed in blue.

*N. caninum* does not have a ROP18 gene annotated, there are stop codons present within the coding sequence so it has been hypothesized to be a pseudogene. This is examined in depth in Chapter 5.

### 3.3.3 Dense Granule Genes

The genes identified by this analysis as putatively dense granule-associated are presented in Table 3.3. There were twenty two in total, incorporating orthologues of 15 of the 17 known *T. gondii* dense granule genes (GRAs 1-14 (excluding GRA13), NTPases I and II and TgPIs I and II), the exceptions being GRA11 and GRA12. However, there is some confusion over whether GRA12 in *T. gondii* is TGME49_088650 (GRA12 according to V. Carruthers' ToxoDB comment (comment ID 102930)) or TGME49_075850 (GRA12 according to M. Cesbron-Delauw's ToxoDB comment (comment I.D. 47270)). Both researchers refer to Michelin *et al.* (2009) but examination of said publication reveals that TGME49_088650 is GRA12, for which there is an *N. caninum* orthologue: NCLIV_041120. Therefore, NCLIV_041120 has been designated GRA12 in this analysis. A BLAST search of the TGME49_075850 protein sequence yields hits to 18 genes within ToxoDB, all of which are annotated simply as 'hypothetical protein' or 'conserved hypothetical protein'. Regarding information presented on ToxoDB; this example goes to show that, while an immensely useful resource, it is sometimes necessary to delve deeper into the information provided and verify comments and annotation, if possible, with the literature available for the gene in question.

**Table 3.3:** Putative list of dense granule genes (n = 22). Signal peptide predictions were downloaded from ToxoDB version 6.0 and had been made using SignalP 3.0. The comments column contains information regarding genes which did not have a straightforward syntenic-orthologous relationship with a known *T. gondii* dense granule gene. '-' denotes no gene identified/no comment.

| Protein description | I.D. | ToxoDB annotation | Signal peptide? | Comments | *T. gondii* ME49 orthologue |
|---|---|---|---|---|---|
| GRA1 | NCLIV_036400 | dense granule protein 1/major antigenp24, putative | Yes | - | TGME49_070250 |
| GRA2 | NCLIV_045650 | 28 kDa antigen, putative | Yes | - | TGME49_027620 |
| GRA3 | NCLIV_045870 | Putative dense granule protein 3 | Yes | - | TGME49_027280 |
| GRA4 | NCLIV_054830 | hypothetical protein | No | - | TGME49_110780 |
| GRA5 | NCLIV_014150 | hypothetical protein | Yes | Not designated as an orthologue of TGME49_086450 on ToxoDB but 36 % BLAST identity (E-value of 5.5 x $10^{-11}$) | TGME49_086450 |
| GRA6 | NCLIV_052880 | granule antigen protein GRA6, putative | Yes | - | TGME49_075440 |
| GRA7 | NCLIV_021640 | dense granule protein 7, putative | Yes | - | TGME49_003310 |
| GRA8 | NCLIV_008990 | hypothetical protein | Yes | - | TGME49_054720 |
| GRA9 | NCLIV_066630 | GRA9 protein, putative | Yes | - | TGME49_051540 |
| GRA10 | NCLIV_037450 | dense granular protein GRA10, putative | No | - | TGME49_068900 |
| GRA11 | - | - | - | No *N. caninum* orthologue for GRA11 | TGME49_012410 |
| GRA12 | NCLIV_041120 | hypothetical protein, conserved | Yes | Confusion over GRA12, no *N. caninum* orthologue exists for TGME49_075850, but NCLIV_041120 is an orthologue of TGME49_088650 | TGME49_088650 |
| GRA12? | - | - | - | | TGME49_075850 |
| GRA13 | - | - | - | No GRA13 found for either organism | |
| GRA14 | NCLIV_016360 | hypothetical protein, conserved | Yes | - | TGME49_039740 |
| DG32 | NCLIV_005560 | dense-granule antigen DG32, putative | Yes | - | TGME49_022170 |
| NTPase I | NCLIV_068460 | hypothetical protein | No | - | TGME49_077240 |
| NTPase II | NCLIV_068400 | NTPase | Yes | - | TGME49_077270 |
| NTPase | NCLIV_067130 | hypothetical protein | Yes | - | TGME49_078880 |

| | | | | | |
|---|---|---|---|---|---|
| Nc PI-1 | NCLIV_003120 | serine proteinase inhibitor, putative | Yes | - | TGME49_008450 |
| Nc PI-2 | NCLIV_003100 | serine proteinase inhibitor TgPI-2, putative | Yes | - | TGME49_008430 |
| 14-3-3 | NCLIV_024820 | 14-3-3 protein homolog | No | 14-3-3 proteins are tentative additions to this table, based upon association with dense granules observed by Assossou *et al*. (2004). | TGME49_063090 |
| 14-3-3 | NCLIV_036630 | 14-3-3 protein, putative | No | | TGME49_069960 |
| 14-3-3 | NCLIV_036930 | 14-3-3 protein, putative | No | | TGME49_069590 |
| 14-3-3 | NCLIV_045450 | 14-3-3-like protein, related | No | No *T. gondii* ME49 orthologue | |

However, a complication that became apparent in this analysis, was that publications very rarely quoted gene identifiers, so when multiple genes were referred to by the same common name, it was often impossible to determine which one specifically was the object of an experiment. The example above (Michelin *et al.* 2009) was one of only very few publications found to do so, and despite the fact that the gene identifiers were of an older format (ToxoDB version 4.0) it was still possible to track the genes they corresponded to.

GRA5 is present in the *T. gondii* ME49, GT1 and VEG genomes (TGME49_086450, TGGT1_037870 and TGVEG_033680 respectively) but there is no *N. caninum* orthologue listed for any of these genes on ToxoDB. The TGME49_086450 protein sequence (120 amino acids) has a BLAST identity score of 36 % with an E-value of $5.5 \times 10^{-11}$ to NCLIV_014150, which is syntenic to *T. gondii* GRA5 and has been designated here as *N. caninum* GRA5.

Assossou *et al.* (2004) observed an association of a secreted 14-3-3 protein with *T. gondii* dense granules, using electron microscopy and immunogold labelling. However, 14-3-3 proteins are expressed by all eukaryotic cells and are involved in various functions, including signal transduction, apoptosis and cell cycle control (reviewed in Fu, Subramanian & Masters (2000)) so their inclusion in this list of dense granule genes is only putative, and requires further experimental investigation. The 14-3-3 protein NCLIV_024820 was readily identified in the whole tachyzoite proteomic analysis in Chapter 1 in addition to the excretory/secretory and rhoptry/dense granule protein preparations in Chapter 4 (to follow) so it is likely that it is an abundant protein and the observation by Assossou *et al.* (2004) that a 14-3-3 protein seemed to be associated with the dense granules may be a result of the protein's wide abundance.

### 3.3.4 Bioinformatics resources

ToxoDB was the primary resource exploited for this genome-mining study. Whilst extremely user friendly, and hosting a wealth of data, it is not without its problems; as may be expected for a database handling this amount of information. The ease with which members of the public can upload user comments can result in conflicting information as different researchers disagree on the annotation of some genes (as discussed above with regard to GRA12, Michelin *et al.* (2009), ToxoDB comment ID 102930 and comment I.D. 47270). Nevertheless, on balance it is a useful feature for sharing and accessing information and overall the database provides a 'one-stop-shop' for gene and protein data relating to *T. gondii* and *N. caninum*.

Another resource available to the scientific community, though not utilised here, is ApiLoc (http://apiloc.biochem.unimelb.edu.au). This database is designed to host information on the published sub-cellular localisations of apicomplexan genes, but unfortunately it contains only 18 entries in total for apically-located *N. caninum* proteins to date (correct as of August 2012, all of which had already been identified in this analysis using the methods described) and as such is not as useful as it might have been, had further entries been made. However, the version currently accessible online (Version 3) was only curated until May 2011 so this is likely to be the reason behind the limited amount of information available.

### 3.3.5 Requirement for experimental validation

The lists of proteins presented in Tables 3.1 to 3.3 represent a useful resource for the analysis of proteomic and transcriptomic data that will be presented in the remainder of this thesis. Very large datasets are typical of these platforms and pose difficulties in drawing any meaningful biological conclusions from the data. The ability to narrow down results to enable the examination of sub-sets of genes or proteins, renders the data more easily handled. However, it cannot be emphasized enough that a hypothetical identification as an apical gene by *in silico* methods is no substitute for experimental determination of sub-cellular localisation, using techniques such as immunofluorescence. Sohn *et al.* (2011) and a number of other researchers have gone some way to identifying protein sub-cellular localisations in *N. caninum,* but it is likely that the identification of further apical proteins will be ongoing over many more years of research.

# CHAPTER 4: PROTEOMIC ANALYSIS OF THE APICAL ORGANELLES

## 4.1 INTRODUCTION

Following the analysis of whole tachyzoite lysate in Chapter 2, the aim was to achieve a simplified *N. caninum* proteome which was enriched for proteins involved in host-cell invasion. Hence, the objective was to prepare material from the apical organelles of the parasite and analyse this proteomically, using the 1-DE LC MS/MS platform discussed in Chapters 1 and 2. There are benefits to analysing simplified proteomes; proteins vary in abundance over a wide dynamic range and heavily abundant proteins can mask the detection of low-abundance proteins. Furthermore, large data sets can be very time consuming to analyse, so the optimal proteomic experiment would analyse samples which contained a high proportion of biologically-interesting proteins with house-keeping/structural proteins (that are not the target of this experiment) removed.

There are various approaches to studying material from the apical organelles. The most obvious, and perhaps the gold standard, is to isolate these from the cell and obtain a purified sample. An alternative is possible due to the fact that these organelles are secretory: the parasites can be incubated in media and secreted/excreted material harvested from the supernatant when the parasites are removed by centrifugation.

### 4.1.1 Excretory/secretory antigen (ESA) analysis

Excretory/secretory antigenic material (ESA) has been much studied in *T. gondii*, where ethanol (EtOH) and a calcium ionophore, A23187, have been shown to stimulate the calcium ion ($Ca^{2+}$) -triggered release of material from the micronemes (Carruthers, Moreno & Sibley 1999; Carruthers & Sibley 1999; Zhou et al. 2005).

A transient rise of free $Ca^{2+}$ in the parasite cytoplasm has been demonstrated to trigger secretion of the micronemal contents through the apical tip (Carruthers & Sibley 1999). A signal transduction cascade leads to elevated $Ca^{2+}$ in the cytoplasm which directly activates secretory vesicles; this can be observed in a wide variety of eukaryotic cells (Burgoyne & Morgan 1993). Cellular stores of calcium can be mobilized from the endoplasmic reticulum by the addition of a calcium ionophore into the culture medium *in vitro*; simulating the external stimulus that would induce this phenomenon *in vivo* (Carruthers & Sibley 1999).

Carruthers and Sibley (1997) first observed the sequential release of micronemal, rhoptry and dense granule contents during host cell invasion by *T. gondii,* using the markers MIC2 (microneme), ROP1 (rhoptry), GRA1 and NTPase (dense granule). They then went on to determine that free $Ca^{2+}$ in the parasite cytoplasm acted as the trigger for micronemal secretion and that intracellular stores of $Ca^{2+}$ were sufficient to produce this response (Carruthers & Sibley 1999). Ethanol stimulates release of $Ca^{2+}$ and hence, micronemal secretion and is more effective at inducing micronemal secretion than calcium ionophore

A23187: increased levels of MIC2 and MIC4 were measured in the EtOH-stimulated samples when compared to the A23187-stimulated samples, while GRA1 was reduced (Carruthers, Moreno & Sibley 1999). However, EtOH has been observed to be less effective than A23187 at stimulating secretion in experiments with the less closely related apicomplexan *Cryptosporidium parvum* (Nadine Randle, personal communication). A study by Chen *et al.* (2004) also utilised A23187 in the presence of extracellular calcium to restore apical discharge to *C. parvum* parasites that had had their calcium stores chelated. Parasites can also be stimulated to secrete from their apical organelles simply by incubation in culture medium at 37 °C (Ahn *et al.* 2001; Song & Nam 2003) but this method is less effective than when a stimulus is used (Carruthers, Moreno & Sibley 1999). Large scale techniques for the purification of *T. gondii* ESA were developed by Opitz *et al.* (2002) and Zhou *et al.* (2005) using incubation of tachyzoites in culture media (without bovine serum albumin) with EtOH at a concentration of 1 %, for 1 hour or 20 minutes respectively.

Zhou *et al.* (2005) used a combination of N-terminal sequencing; two-dimensional gel electrophoresis (2-DE) and matrix-assisted laser desorption-ionisation (MALDI) mass spectrometry; and two-dimensional liquid chromatography followed by tandem mass spectrometry to analyse *T. gondii* ESA material. These approaches yielded a total of seventy six protein identifications, eleven of which were from the micronemes, seven the dense granules and one from the rhoptries. Novel identifications ('hypothetical' proteins) accounted for seven of the identifications, the remainder were proteins known to localise to other cellular locations such as the cytosol and endoplasmic reticulum (Zhou *et al.* 2005).

### 4.1.2 Rhoptry/dense granule - enriched fraction analysis

Bradley *et al.* (2005) undertook a comprehensive analysis of the *T. gondii* rhoptry proteome by optimising a protocol developed by Dubremetz and co-workers for fractionation using a Percoll gradient (Leriche & Dubremetz 1991). This enabled preparation of a fraction highly enriched for both rhoptries and dense granules (R/DG), which was then further purified by sucrose floatation, to leave a rhoptry fraction (R) with dense granules and contaminating mitochondria removed.

The analysis of the R fraction identified thirty eight novel proteins, eleven of which they then went on to confirm as rhoptry, by immunolocalisation. The previously known rhoptry proteins ROP2, ROP4, ROP8 and ROP9 were identified, while Toxofilin was found to localise to the rhoptries. The rhoptry neck protein family were identified for the first time, and the name 'RON' proposed. The dense granule proteins GRA3, GRA7, NTPase I and NTPase II were also identified within the R fraction, as not all dense granule contamination was successfully removed during the sucrose floatation. This technique was otherwise deemed to have been highly successful at enriching for the rhoptries in *T. gondii* (Bradley *et al.* 2005) and vastly improved on the knowledge at the time of rhoptry contents.

### 4.1.3 Aims and objectives

The aims of this chapter were two-fold: to identify *N. caninum* proteins from those predicted to be from the apical complex (Chapter 3) and to explore methods for protein preparations enriched in proteins from the micronemes, rhoptries and dense granules, which had been developed for *T. gondii*. This would be achieved by incubating tachyzoites to harvest their excretions and secretions, and by collaborating with Peter Bradley to analyse rhoptry and dense granule preparations.

## 4.2 MATERIALS AND METHODS

### 4.2.1 Excretory/secretory antigen analysis

As there was no published ESA preparation method for *N. caninum*, the approaches discussed in section 4.1 were taken into consideration and this final protocol, adapted from Zhou *et al.* (2005) was decided upon after a number of preliminary experiments, discussed in more detail in section 4.4.

#### 4.2.1.1 Preparation of ESA material from parasite cultures

Tachyzoites were harvested as in Chapter 2.2.1, except that washes were performed in DO medium (Dulbecco's modified Eagle's medium, 2 Mm glutamine, 10 Mm HEPES) instead of PBS. Pellets of approximately 1 x $10^8$ tachyzoites were immediately resuspended in 1 ml DO medium and supplemented with EtOH to a final concentration of 1 % and complete mini protease inhibitor cocktail (Roche) to a final 1x working concentration. Incubation of tachyzoites was performed at 37 °C with gentle agitation for a period of 20 minutes. After incubation, tachyzoites were pelleted at 1500 x g for 5 minutes at 4 °C and the supernatant collected, kept on ice for 5 minutes and then stored at -20 °C. A control sample of DO media was incubated without parasites. A second control consisted of a tachyzoite lysate, as described in Chapter 2.2.2.1 (but loading only 20 µg).

#### 4.2.1.2 Tricarboxylic acid (TCA) precipitation of ESA

Prior to one dimensional gel electrophoresis (1-DE), samples were defrosted then subjected to trichloroacetic acid (TCA) precipitation as follows: 1 volume of 100 % (w/v) TCA was added to 4 volumes of sample and incubated at 4 °C for 20 minutes. Precipitates were pelleted by centrifuging for 5 minutes at 13 000 x g and then washed twice in 200 µl ice-cold acetone, before being dried in a heat block at 95 °C for 5 minutes. The precipitates were then solubilised in 20 µl of SDS-sample buffer (50 mM Tris HCl (pH 6.8), 2 % (w/v) SDS, 0.2 % (w/v) bromophenol blue, 10 % (w/v) glycerol, 100 mM dithiothrietol (DTT)) by heating to 95 °C for 5 minutes followed by centrifugation at 13 000 x g for 5 minutes.

#### 4.2.1.3 One dimensional gel electrophoresis (1-DE) of ESA

Supernatants (20 µl) were run on 10 cm 12 % acrylamide gels (Chapter 2.2.2.3) with limiting factors of 200 V, 70 mA and 10 W for approximately 50 minutes. Gels were fixed at room temperature in fixing solution (40 % (v/v) ethanol, 10 % (v/v) acetic acid) overnight, then rinsed twice in $ddH_2O$ before staining with Colloidal Coomassie (20 % (v/v) methanol, 0.08 % (w/v) Coomassie Brilliant Blue G250, 0.8 % (v/v) phosphoric acid, 8 % (w/v) ammonium sulphate) for 24 hours. After further rinsing in $ddH_2O$, bands were excised from the gel and stored individually in 1 % (v/v) acetic acid at 4 °C.

### 4.2.2 Rhoptry (R) and rhoptry/dense granule (R/DG) fractions

In collaboration with Dr. Peter Bradley, an *N. caninum* rhoptry fraction analogous to that previously produced for *T. gondii* (Bradley *et al.* 2005) was prepared for analysis by 1-DE LC

MS/MS. A summary of the Percoll centrifugation and sucrose floatation steps carried out by the Bradley group can be seen in Figure 4.1 (reproduced from Bradley *et al.* 2005).

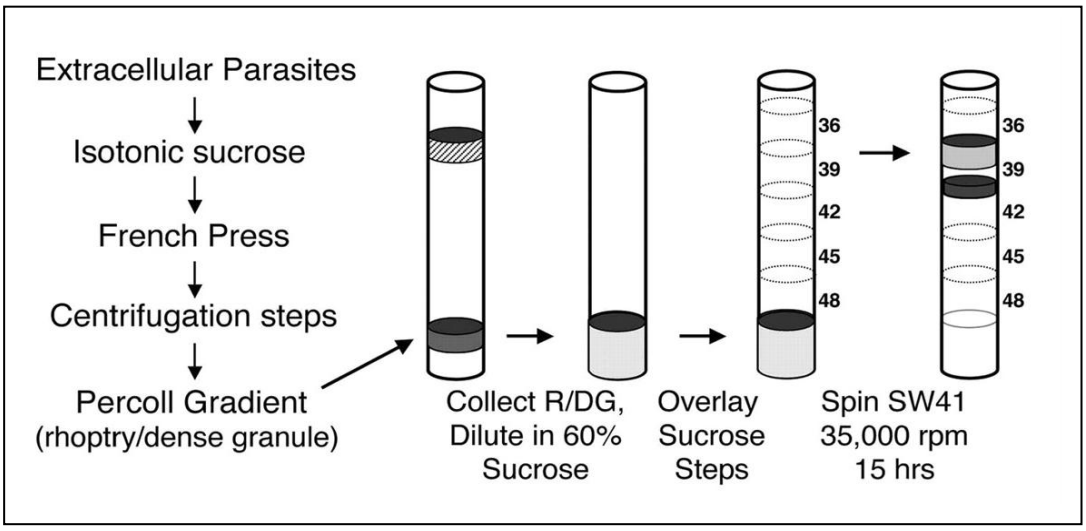



**Figure 4.1:** Schematic of the isolation of rhoptries from *T. gondii*. A rhoptry/dense granule (R/DG) preparation was isolated by Percoll gradient. The R/DG fraction was then subjected to a sucrose floatation gradient to isolate the purified rhoptry (R) fraction that corresponds to the top band in the gradient (reproduced from Bradley *et al.* 2005).

Two samples were produced, R/DG refers to the fraction collected after the Percoll gradient step and the R fraction was that resulting from sucrose floatation.

#### 4.2.2.1 Preparation of rhoptry and rhoptry/dense granule fractions

Preparations enriched for *Neospora caninum* rhoptries and dense granule organelles were prepared according to the methods detailed in Figure 4.1 and Bradley *et al* (2005). Briefly, freshly harvested tachyzoites of the *N. caninum* NC1 strain were suspended in R buffer (250 mM sucrose, 10 mM MOPS, pH 7.2, 2 mM dithiothreitol, 1 mM EDTA, 1 x complete mini protease inhibitor cocktail (Roche) and passed through a French press in order to disrupt the cells. Intact parasites and large debris were removed by centrifugation at 1 300 x g for 20 minutes, before the supernatant was subjected to further centrifugation at 25 000 x g for 25 minutes. The resulting pellet was resuspended in R buffer containing 30 % Percoll, for a 25 minute centrifugation at 61 500 x g after which the R/DG fraction (containing a mixed fraction of rhoptries, dense granules, mitochondria and apicoplasts) was collected from the brown band at the bottom of the gradient and repelleted to remove the Percoll at 100 000 x g for 90 minutes. To obtain a cleaner fraction of rhoptries, this material was resuspended in 300 µl R buffer and mixed with 2 ml 60 % (w/v) sucrose in S buffer (10 mM MOPS, 2 mM dithiothreitol, 1 mM EDTA, 1 x complete mini protease inhibitor cocktail (Roche)) before being subjected to centrifugation for 18 hours at 150 000 x g over a sucrose gradient (48, 45, 42, 39 and 36 %). The peak rhoptry fractions were pooled in R buffer and pelleted for 90

minutes at 100 000 x g. Samples were stored in R buffer at -80 °C and sent to Liverpool for subsequent 1-DE LC MS/MS and data analysis.

#### 4.2.2.2 TCA precipitation of R and R/DG fractions

A 40 µl aliquot from each of the R and R/DG samples were precipitated with TCA as follows: 10 µl TCA was added to the sample and the resulting solution stored overnight at -20 °C. These samples were then centrifuged at 13 000 x g for 25 minutes, the pellets washed in ice-cold acetone, twice, for 25 minutes at 13 000 x g and then dried in a heat block at 95 °C for 5 minutes. The pellets were resuspended in 40 µl 1 x sample buffer (Chapter 2.2.2.1), heated to 95 °C for 5 minutes, vortexed briefly, heated again to 95 °C for 5 minutes and pelleted at 13 000 x g for 5 minutes.

#### 4.2.2.3 1-DE of R and R/DG fractions

The R and R/DG supernatents (40 µl) were loaded onto a 12 % gel (Chapter 2.2.2.3) and run as described in 4.2.1.3. Contiguous bands were excised and stored at 4 °C in 1 % (v/v) acetic acid prior to trypsin digestion.

### 4.2.3 Trypsin digestion of samples for mass spectrometry

ESA, R and R/DG samples were digested with sequencing grade trypsin (Roche) as described in Chapter 2.2.2.4. Digests were stored at -20 °C prior to LC MS/MS.

### 4.2.4 Liquid chromatography and tandem mass spectrometry analysis (LC MS/MS)

LC MS/MS was performed on an LTQ ion trap mass spectrometer (Thermo Fisher Scientific) as in Chapter 2.2.3, except that a 1 hour gradient was employed as follows: a linear gradient of 0-50 % (v/v) acetonitrile/0.1 % (v/v) formic acid over 30 minutes followed by 100 % (v/v) acetonitrile/0.1 (v/v) % formic acid for 5 minutes and a further 20 minutes of 0.1 % (v/v) formic acid.

### 4.2.5 Bioinformatic Analyses

#### 4.2.5.1 Protein identification using Mascot

Raw mass spectra were processed as described in Chapter 2.2.3. Resulting .mgf files were submitted to Mascot (Matrix Science) and searched against a locally-mounted database comprising the *N. caninum* gene predictions (various releases) hosted on ToxoDB (Gajria et al. 2007). Search parameters were as follows: fixed carbamidomethyl modification of cysteine, variable oxidation of methionine, one missed trypsin cleavage, peptide tolerance ±1.5 Da, fragment ion tolerance ±0.8 Da and peptide charge state of +1, +2 and +3, with a cut-off Mascot score of ≥ 50.

#### 4.2.5.2 Assignment of proteins to MIPS Functional Catalogue categories

Protein identifications were manually assigned to MIPS Functional Catalogue categories (as described in Table 2.1) (Ruepp *et al.* 2004) to determine the range of protein types identified

and to enable comparison between the different experiments. For each protein, the 'biological process' Gene Ontology (GO) term (Ashburner *et al.* 2000) was downloaded from ToxoDB (Gajria *et al.* 2008), and used as a search term in the MIPS Functional Catalogue Database, where the top hitting suggestion for a catalogue assignment was taken. Proteins for which there were multiple biological process GO terms were searched using each one, and a consensus MIPS category used, as in most cases they yielded the same result. If there was no 'biological process' GO term assigned, the 'molecular function', followed by 'cellular component' GO term was used instead. If there was no GO term, but the gene description on ToxoDB contained information alluding to the function of the protein, then literature searching was employed to qualitatively determine the most appropriate MIPS category. Any protein that was known to be invasion-related was assigned to category 32: cell rescue, defence and virulence. If there was no GO term or other indicator of function on ToxoDB, or in the literature, a Basic Local Alignment Search Tool (BLAST) search (NCBI) was performed to identify possible orthologues for which a function was known. Any proteins for which this yielded no useful result, were assigned 'unclassified'.

#### 4.2.5.3 Signal peptide predictions

Protein sequences were submitted to SignalP version 3.0 (Bendtsen *et al*. 2004) to obtain predictions as to whether they possess a processed/cleaved signal peptide, which could indicate that the protein is secretory. SignalP uses a hidden Markov model (HMM) to calculate a probability score, for which a cut off of >0.9 was employed.

## 4.3 RESULTS

### 4.3.1 Analysis of excretory/secretory antigens from *N. caninum*

A total of 78 non-redundant protein identifications were made from 24 bands excised from two 1-DE gels (Table 4.1). An ESA lysate gel is shown in Figure 4.2 with a tachyzoite lysate for comparison. The ESA preparation was considerably less complex than the tachyzoite lysate and there was sufficient protein to enable band visualisation and subsequent proteomic detection. A control sample of DO media (culture media without bovine serum albumin, which might mask proteomic detection of parasite proteins), incubated in 1 % EtOH but without the addition of parasites, is shown for comparison.

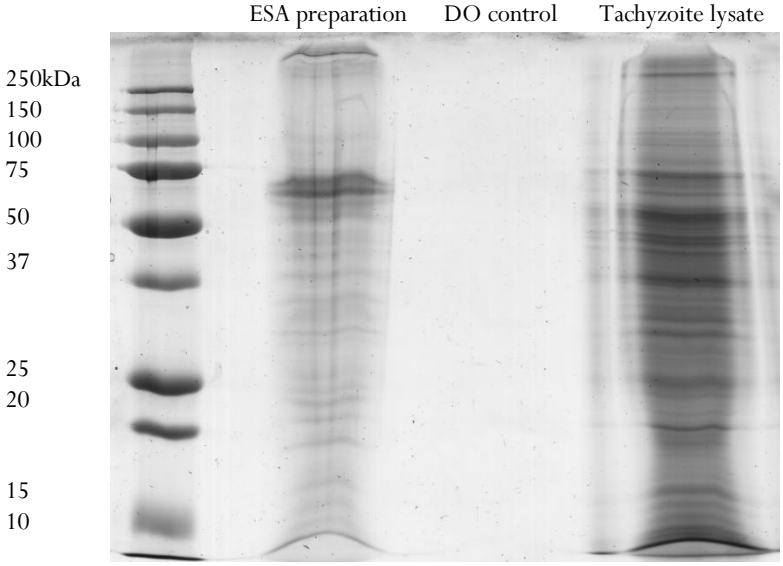



**Figure 4.2:** 1-DE gel of excretory/secretory preparation. Samples were analysed on a 12 % (w/v) acrylamide gel under denaturing conditions, visualised by Colloidal Coomassie staining. Lanes are as follows, lane 1: protein standards, lane 2: ESA supernatent (after 20 minute incubation of tachyzoites in 1% EtOH), lane 3: control (DO media incubated without parasites), lane 4: 20µg whole tachyzoite lysate.

While proteins were identified from a broad spectrum of functional classifications, such as 'protein synthesis' (n = 13), 'protein with binding function' (n = 5) and 'metabolism' (n = 6); 24 % (n = 19) were involved in 'cell rescue, defence and virulence' and hence likely to be involved in the process of the parasite invading and maintaining itself within the host cell (Figure 4.3). A further 13 % (n = 10) were unable to be designated a function as they were insufficiently annotated. A table including all the identifications made in this analysis is provided in Appendix IV. Of the top fifteen protein hits, five were predicted micronemal (hypothetical protein (NCLIV_021050), which is the *N. caninum* orthologue of *T. gondii* SUB1, MIC1, M2AP, MIC4 and AMA1) and four were predicted dense granule proteins (14-3-3 protein homologue, NTPase, hypothetical protein (NCLIV_068460), which is an NTPase orthologue; and GRA1). There were also four surface antigens (SRS proteins (Jung, Lee & Grigg 2004)), a hypothetical protein (NCLIV_050370) that is annotated as being involved with the biological process of glycolysis; and malate dehydrogenase, which is also part of the glycolysis metabolic pathway (LAMP database). Other proteins identified included MIC3,

MIC6, MIC10, GRA2, GRA3, GRA6, GRA7, GRA9 and ROP9. The hypothetical proteins NCLIV_043760, NCLIV_059430, NCLIV_031510, NCLIV_020140, NCLIV_000430, NCLIV_032270 and NCLIV_046030 were unable to be assigned to MIPs functional categories due to their lack of homology to proteins of known function, hence they are potentially novel apicomplexan specific proteins. While this analysis identified just one microneme protein (MIC3) not previously identified by the whole tachyzoite analysis (Chapter 2), it did proportionally increase the number of microneme and dense granule proteins within the preparation by almost ten-fold (22 % and 2.6 %, respectively).

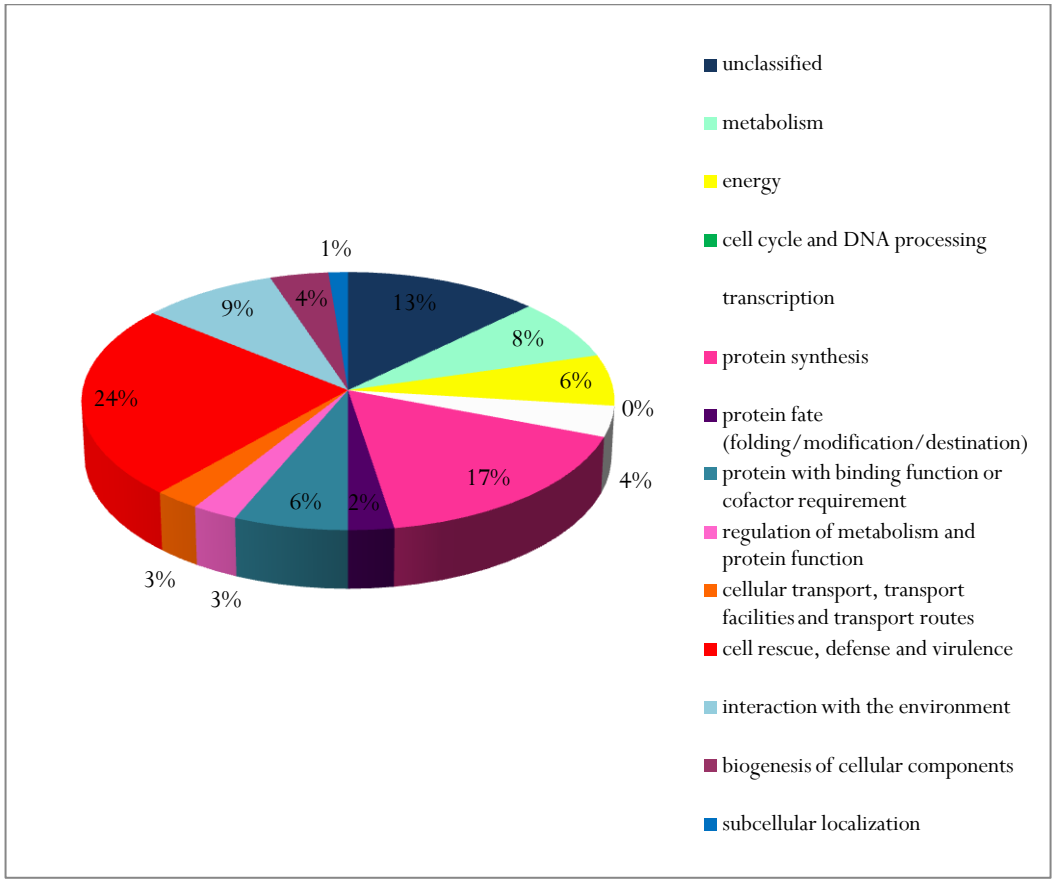



**Figure 4.3**: Functional classification of proteins identified by 1-DE analysis of the excretory/secretory preparation. Proteins (n = 78) were assigned to functional categories according to the MIPS Functional Catalogue.

Protein identifications were submitted to SignalP 3.0 and 27% (n = 21) were predicted to be secretory. This represents enrichment with respect to both the whole predicted proteome of *N. caninum* (Reid *et al*. 2012) hosted on ToxoDB (170 % more secretory proteins) and the whole tachyzoite lysate proteome (80 % more secretory proteins) (Figure 4.4).

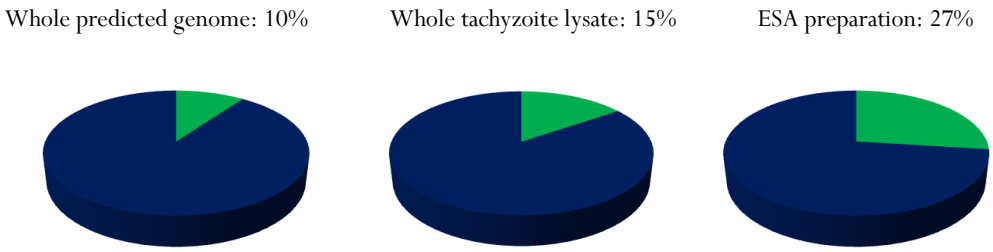



**Figure 4.4:** Enrichment for proteins with predicted signal peptides by different proteomic samples**.** Green sections denote proteins predicted by SignalP 3.0 to contain a signal peptide (HMM probability score >0.9).

### 4.3.2 Rhoptry and rhoptry/dense granule-enriched fraction analyses

The rhoptry (R) and rhoptry/dense granule (R/DG) fractions were visualised on a 1-DE gel then analysed by LC MS/MS. The R fraction yielded 655 protein identifications with a Mascot score ≥ 50; 138 (21 %) of which were non-redundant (Appendix V Table V.1). The two fractions appeared to contain almost identical bands (Figure 4.5); when the protein identifications were compared, 83 proteins, including both rhoptry and dense granule proteins, were shared between the two preparations which equates to 60 % of the R fraction and 75 % of the R/DG fraction (Figure 4.6). Overall 111 non-redundant protein identifications were made for the R/DG fraction, of which 28 (25 %) were novel identifications when compared to the R fraction (Appendix V Table V.2).

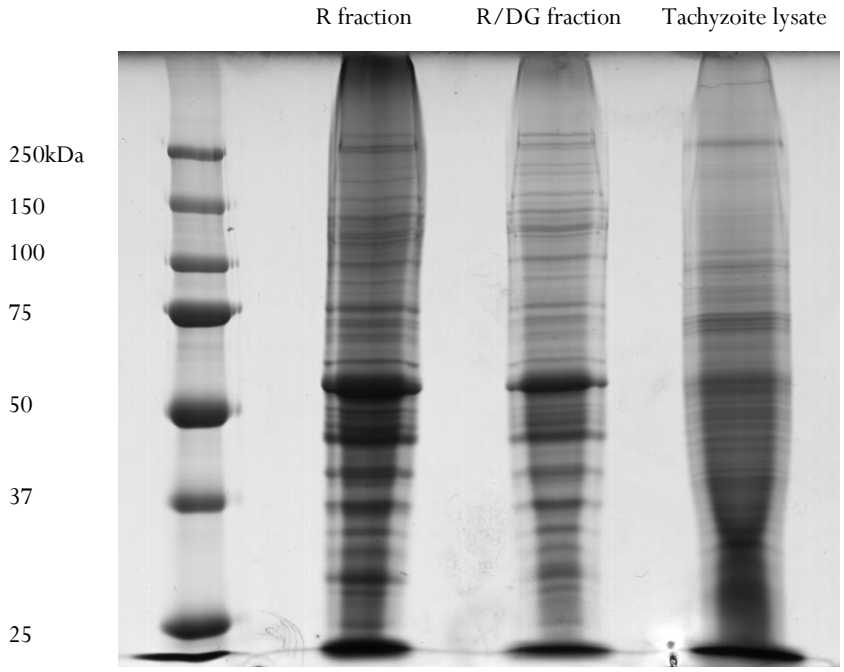



**Figure 4.5: 1-DE gel of rhoptry and rhoptry/dense granule–enriched fractions.** Samples were analysed on a 12 % (w/v) acrylamide gel under denaturing conditions, visualised by Colloidal Coomassie staining. Lanes are as follows, lane 1: protein standards, lane 2: 40 µl rhoptry (R) fraction, lane 3, 40 µl rhoptry/dense granule (R/DG) fraction, (protein concentration unknown: insufficient sample for protein assay) lane 4: 20 µg whole tachyzoite lysate.

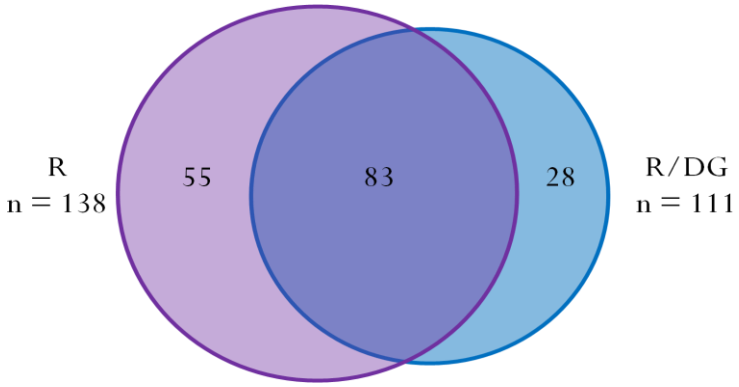



**Figure 4.6:** Venn diagram illustrating redundancy between rhoptry and rhoptry/dense granule fractions. Protein identifications were made using Mascot with a cut off score of 50. Total non redundant identifications: n = 166.

In Table V.1 (Appendix V) and Figure 4.7, six proteins have been assigned to a new category: 'possibly defence' ('?' in Table V.1). This represents proteins identified in this experiment that were considered to be candidate novel rhoptry proteins. NCLIV_031550 is annotated as 'hypothetical' on ToxoDB and was identified in the R fraction. It has no homologues predicted within the *N. caninum* or *T. gondii* genomes and when searched using BLAST (see Chapter 3.1.4) its top hits were the rhoptry proteins ROP37, ROP28 and ROP25 (*T. gondii* and *N. caninum*). NCLIV_069110 is predicted to be a homologue of ROP1 and has a signal peptide predicted by SignalP. The protein NCLIV_067490 is annotated as 'protein phosphatase 2c'. Another protein phophatase 2c is known to be of rhoptry origin (Gilbert *et al.* 2007), so this protein was selected for further attention. However, when the predicted amino acid sequence was subjected to BLAST, it was apparent that this is a common type of protein in Apicomplexa and other eukaryotes such as mammals, so there was no solid foundation for hypothesizing that this was a novel rhoptry protein. NCLIV_004270 is a 'putative protein kinase' which BLASTs to other protein kinases and hypothetical proteins, as well as to *N. caninum* ROP27 (27 % identity, E-value 0.0048). Predicted by SignalP to have a signal peptide, NCLIV_061160 is another potential ROP. It BLASTs to other apicomplexan acid and serine threonine protein phosphatases, but also to a glideosome protein in *P. falciparum* (XP_001352051, BLAST identity 44 %, E-value 4 x $10^{-108}$) and a secreted acid phosphatase in *C. parvum* (XP_628593, BLAST identity 27 %, E-value 1 x $10^{-23}$). Finally, the 'hypothetical protein' NCLIV_022270 (predicted to have a signal peptide by SignalP) is a homologue of other 'hypotheticals' in *T. gondii* and BLASTs to *T. gondii* ROP43 (32 % identity, E-value 1.3 x $10^{-22}$) and ROP18 (30 % identity, E-value 2.8 x $10^{-22}$) as well as to various *N. caninum* and *T. gondii* hypothetical proteins.

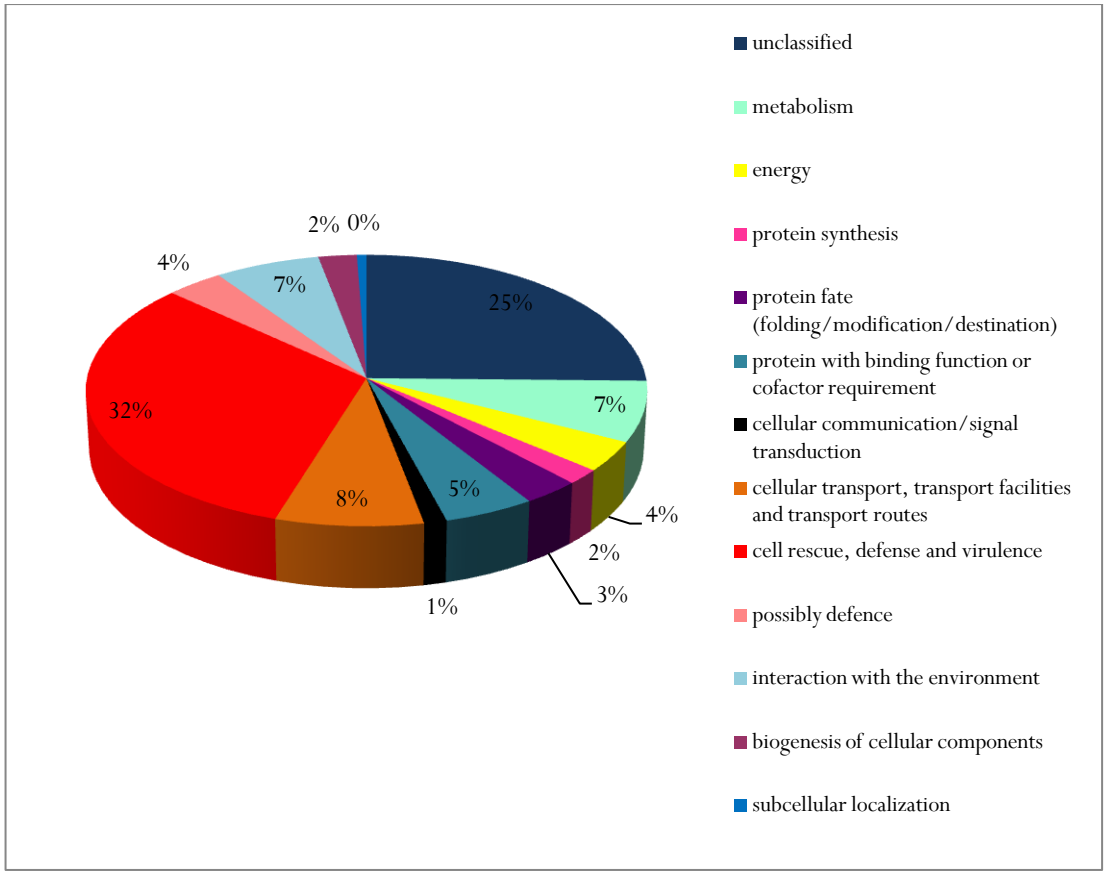



**Figure 4.7:** Functional classification of proteins identified by 1-DE analysis of the rhoptry and rhoptry/dense granule-enriched fraction. Proteins (n = 166) were assigned to functional categories according to the MIPS Functional Catalogue.

Overall, 32 % (n = 53) of the identifications within these two analyses combined (n = 166) were from the 'cell rescue, defence and virulence' category (Figure 4.7). Figure 4.8 shows a comparison between the whole tachyzoite analysis in Chapter 2, the ESA analysis and the combined R and R/DG fractions; this figure demonstrates the sequential increase in the proportion of proteins involved in defence and virulence by these approaches respectively. A total of 31 (53 %) of the 58 proteins putatively assigned to the rhoptries (Chapter 3) were identified in this analysis. The fractionation of rhoptries and dense granules did not appear to have been totally successful as six dense granule proteins were identified in the R fraction. However, as can be seen in Table 4.1, seven ROP proteins not previously identified by the tachyzoite and ESA analyses were identified, these were as follows: RON4, ROP43/44, ROP37, ROP23, ROP19, ROP12 and ROP5B in addition to the microneme protein MIC17B. This is in contrast to the ESA analysis, which only uniquely identified MIC3 when compared to the rhoptry/dense granule and tachyzoite analyses.

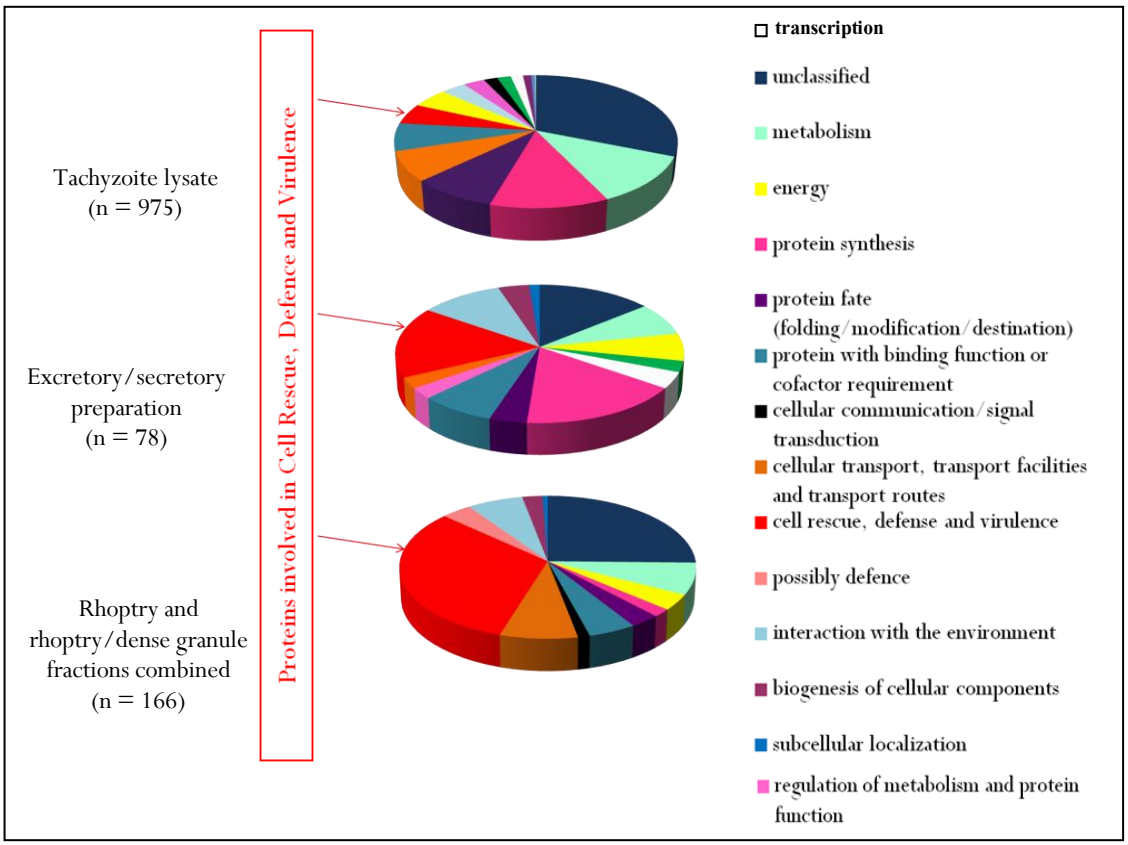



**Figure 4.8:** Enrichment of cell rescue, defence and virulence proteins. Arrows on pie charts show larger proportion of proteins in the 'cell rescue, defence and virulence' category of functional classification in the apical organelle proteome analyses (ESA, 24 %; R + R/DG, 32 %) when compared to the tachyzoite analysis (5 %). Proteins were assigned to functional categories according to the MIPS Functional Catalogue.

Overall, Table 4.1 shows that in the tachyzoite stage, 31 % of predicted microneme proteins, 68 % of predicted dense granule proteins and 60 % of predicted rhoptry proteins (Chapter 3) have been identified by the combination of whole tachyzoite and apical protein preparations being analysed by 1-DE and LC MS/MS. Of the remaining 48 % that have so far eluded detection, not all will necessarily be translated in the tachyzoite stage. Furthermore, a number of the proteins identified in these analyses which are currently annotated as hypothetical are potentially novel apical proteins, such as the six possible rhoptry proteins described above and NCLIV_007450, a putative rhoptry kinase (Talevich, Mirza & Kannan 2011) that was not identified in the analysis in Chapter 3.

**Table 4.1:** Apical proteins identified in one or more of the following: whole tachyzoite analysis, excretory/secretory antigen analysis and rhoptry and dense granule analyses. '✓' indicates a protein for which peptides were identified with a Mascot score ≥ 50, '✗' a protein for which peptides were not identified.

| Protein Description | I.D. | Identified in | | |
|---|---|---|---|---|
| | | Tachyzoite | ESA | R & R/DG |
| **Microneme proteins** | | | | |
| MIC1 | NCLIV_043270 | ✓ | ✓ | ✓ |
| MIC2 | NCLIV_022970 | ✓ | ✗ | ✗ |
| MIC2 paralogue | NCLIV_033690 | ✓ | ✗ | ✗ |
| M2AP | NCLIV_051970 | ✓ | ✓ | ✓ |
| MIC3 | NCLIV_010600 | ✗ | ✓ | ✗ |
| MIC4 | NCLIV_002940 | ✓ | ✓ | ✓ |
| MIC17B | NCLIV_038110 | ✗ | ✗ | ✓ |
| MIC6 | NCLIV_061760 | ✓ | ✓ | ✗ |
| MIC8 | NCLIV_062770 | ✓ | ✗ | ✗ |
| MIC10 | NCLIV_066250 | ✓ | ✓ | ✗ |
| MIC11 | NCLIV_020720 | ✓ | ✗ | ✗ |
| AMA1 | NCLIV_028680 | ✓ | ✓ | ✗ |
| SUB1 | NCLIV_021050 | ✓ | ✓ | ✓ |
| **Rhoptry Proteins** | | | | |
| ROP1 | NCLIV_053840 | ✓ | ✗ | ✓ |
| ROP4 | NCLIV_001970 | ✓ | ✗ | ✓ |
| ROP5 | NCLIV_060730 | ✓ | ✗ | ✓ |
| ROP5B | NCLIV_060740 | ✗ | ✗ | ✓ |
| ROP6 | NCLIV_027850 | ✓ | ✗ | ✓ |
| ROP9 | NCLIV_018420 | ✓ | ✓ | ✓ |
| ROP10 | NCLIV_058180 | ✓ | ✗ | ✓ |
| ROP11 | NCLIV_045580 | ✓ | ✗ | ✓ |
| ROP12 | NCLIV_021100 | ✗ | ✗ | ✓ |
| ROP13 | NCLIV_055850 | ✓ | ✗ | ✓ |
| ROP15 | NCLIV_011690 | ✓ | ✗ | ✓ |
| ROP17 | NCLIV_027930 | ✓ | ✗ | ✓ |
| ROP19 | NCLIV_017440 | ✗ | ✗ | ✓ |
| ROP20 | NCLIV_028170 | ✓ | ✗ | ✓ |
| ROP23 | NCLIV_016220 | ✗ | ✗ | ✓ |
| ROP24 | NCLIV_068850 | ✓ | ✗ | ✓ |
| ROP26 | NCLIV_011730 | ✓ | ✗ | ✓ |
| ROP32 | NCLIV_035860 | ✓ | ✗ | ✗ |
| ROP35 | NCLIV_044410 | ✓ | ✗ | ✗ |
| ROP37 | NCLIV_001460 | ✗ | ✗ | ✓ |
| ROP38 | NCLIV_017410 | ✓ | ✗ | ✓ |

| | | | | |
|---|---|---|---|---|
| ROP40 | NCLIV_012920 | ✓ | × | ✓ |
| ROP43/44 | NCLIV_004220 | × | × | ✓ |
| RON1 | NCLIV_054120 | ✓ | × | ✓ |
| RON2 | NCLIV_064620 | ✓ | × | ✓ |
| RON3 | NCLIV_048590 | ✓ | × | ✓ |
| RON4 | NCLIV_030050 | × | × | ✓ |
| RON4L1 | NCLIV_007800 | ✓ | × | ✓ |
| RON5 | NCLIV_055360 | ✓ | × | ✓ |
| RON8 | NCLIV_070010 | ✓ | × | ✓ |
| lipase maturation factor 2, related | NCLIV_057950 | ✓ | × | × |
| Rhoptry kinase family protein | NCLIV_007770 | ✓ | × | ✓ |
| Zgc:55863, related | NCLIV_017420 | ✓ | × | ✓ |
| toxofilin, putative | NCLIV_051340 | ✓ | × | ✓ |
| Toxopain 1 | NCLIV_069550 | ✓ | × | × |
| hypothetical protein | NCLIV_068890 | ✓ | × | ✓ |
| **Dense Granule Proteins** | | | | |
| GRA1 | NCLIV_036400 | ✓ | ✓ | ✓ |
| GRA2 | NCLIV_045650 | ✓ | ✓ | ✓ |
| GRA3 | NCLIV_045870 | ✓ | ✓ | ✓ |
| GRA4 | NCLIV_054830 | ✓ | × | × |
| GRA5 | NCLIV_014150 | ✓ | × | × |
| GRA6 | NCLIV_052880 | ✓ | ✓ | ✓ |
| GRA7 | NCLIV_021640 | ✓ | ✓ | ✓ |
| GRA8 | NCLIV_008990 | ✓ | × | ✓ |
| GRA9 | NCLIV_066630 | ✓ | ✓ | × |
| GRA12 | NCLIV_041120 | ✓ | × | ✓ |
| DG32 | NCLIV_005560 | ✓ | × | × |
| NTPase I | NCLIV_068460 | ✓ | ✓ | × |
| NTPase II | NCLIV_068400 | ✓ | ✓ | ✓ |
| Nc PI-1 | NCLIV_003120 | ✓ | × | × |
| 14-3-3 protein homolog | NCLIV_024820 | ✓ | ✓ | ✓ |

## 4.4 DISCUSSION

### 4.4.1 Optimization of an ESA protocol for *N. caninum*

During the preliminary experiments with *N. caninum* both calcium ionophore- and EtOH-stimulated secretion were adopted and various steps adjusted until the optimized protocol was decided upon. As with *T. gondii*, EtOH was observed to stimulate secretion most effectively and resulted in the highest enrichment of micronemal proteins in the analysis. Incidentally, tachyzoites incubated at 37 °C in their normal culture medium with no additives were also able to secrete micronemal proteins although more were seen in the EtOH stimulated samples.

Limitations to the practicality of culturing large numbers of parasites to incubate for ESA preparations freshly after egress meant that the amount of material harvested was low and required concentration. Previous studies on *T. gondii* utilised 4 x $10^9$ tachyzoites per preparation (Zhou *et al.* 2005) but here the method was successfully scaled down to use just 1 x $10^8$ tachyzoites per preparation which is a more achievable number.

One step in the protocol that required care was centrifugation of tachyzoites to pellet them when obtaining the supernatant containing the ESA. Rapid centrifugation lysed tachyzoites and resulted in a preparation that resembled a whole tachyzoite lysate, with structural proteins identified in subsequent LC MS/MS. Eventually, the ESA protocol from Zhou *et al.* (2005), with a few small adjustments, yielded a reproducible, simplified proteome compared to the whole tachyzoite lysate (Figure 4.2). Incubation in DO media (0 % EtOH) also stimulated micronemal secretion, but this has previously been shown to be less effective than using EtOH to a concentration of 1 % (v/v) (Carruthers, Moreno & Sibley 1999).

### 4.4.2 Enrichment for apical proteins

The proteins identified in these two preparations showed enrichment for proteins from the apical organelles when compared to a lysate of tachyzoites. However, there were a number of other proteins (for example actin, tubulin, mitochondrial proteins and surface antigens) present which would be considered structural or cytosolic and therefore were unlikely to have originated from an apical organelle. Also, while the proportion of invasion-related genes was increased, relative to the rest of the sample, the total number of identifications was considerably less so this did not represent a major increase in apical protein identifications overall. The rhoptry and dense granule fraction analysis was more successful in this respect than the ESA analysis, identifying seven rhoptry and one microneme proteins not previously identified by the tachyzoite analysis, compared to just one additional microneme protein (MIC3) identified by the ESA analysis.

Figure 4.4 shows the increased number of protein identifications for which SignalP predicted a signal peptide in the ESA analysis. This can be an indication that a protein is secretory, but the ESA lysate also contained proteins for which a signal peptide was not predicted. This could be because they are truely non-secretory, or alternatively that they are secreted by non-classical pathways and hence unable to be identified by the SignalP algorithm (Bendtsen *et al.* 2004).

Due to the identification of cytosolic and structural proteins in the apical preparations, these techniques cannot be considered to have yielded pure preparations of microneme, rhoptry and dense granule proteins. The benefit of being able to produce a totally pure fraction would be that any novel protein identified would be known to be apical and therefore warrant further study. However, a method to produce a totally pure preparation has not yet been developed; and *in vivo,* apical proteins are likely to be functioning amidst a mixture of proteins originating from other cellular locations. Furthermore, if confirmation of apical location for a protein of interest is required there are additional, downstream methods available, such as antibody labelling and immunofluorescence that can aid in determining cellular localisation, but are costly and time consuming hence better suited to smaller subsets of proteins.

### 4.4.3 Apical proteins identified

The proteins AMA1, RON2, RON4, RON5 and RON8 were identified in this analysis. These proteins form the moving junction complex in *T. gondii* host cell invasion (Besteiro *et al.* 2009). Besteiro, Dubremetz & Lebrun (2011) proposed a model where AMA1 in the parasite membrane binds to a complex of RON2 and the other associated RONs inserted into the host cell membrane. AMA1 likely has a further role in host cell invasion as it has been identified all over the parasite surface, not just localised to the moving junction (Howell *et al.* 2005).

GRA proteins, most of which (10 of 13 predicted in the *N. caninum* genome, Chapter 3) have been identified by one or more of the proteomic analyses, have been shown in *T. gondii* to interact with host cell proteins such including enzymes and structural proteins (Ahn *et al.* 2006). GRA 3, GRA5 and GRA6 have been hypothesized as implicated in preventing host cell apoptosis by interacting with a calcium modulating ligand that is present on the host endoplasmic reticulum (ER), and forms part of the attachment of the parasitophorous vacuole (PV) to the host ER (Ahn *et al.* 2006). GRA4 has also been linked to anti-host cell apoptosis mechanisms (Ahn *et al.* 2006).

The detection of both ROP5 and ROP5B corroborates their identification in the genome in Chapter 3. ROP5B is a tandem copy of ROP5 in *N. caninum*, this gene exists in different copy numbers in *T. gondii* strains (Reese *et al*. 2011; Reese and Boothroyd 2011). Conversely, MIC2B, a novel MIC2 paralogue in *N. caninum* that is not present in *T. gondii* (Reid *et al.* 2012) was proteomically identified only in the tachyzoite lysate and not in either of the apical preparations (Table 4.1). Nevertheless, regardless of which sample the identification was made from, these data still confirm the protein's existence.

It is possible that a large proportion of the proteins described as 'hypothetical' are invasion-related or at least apicomplexan-specific, which could account for why they are unannotated, as they may bear little homology to the sequences of well known proteins from better studied organisms such as humans and mice. Many of the eukaryote-wide, house-keeping genes are well annotated in the apicomplexan genomes as they are well conserved with other organisms. Of the R fraction identifications, six were noted as potential candidate rhoptry proteins. This was based on their similarity in name or function to known rhoptry proteins, and then further investigated by a combination of BLAST/homologue searching (ToxoDB and NCBI), and signal peptide prediction (SignalP 3.0). Of these six proteins, presented in Section 4.3.2,

NCLIV_031550, NCLIV_069110, NCLIV_004270 and NCLIV_022270 are the strongest candidates, based upon their lack of homologues in non-apicomplexans and their similarity to other ROP proteins. The remaining two, NCLIV_067490 and NCLIV_061160 share sequence similarity to a number of other eukaryotic proteins, so whilst this does not preclude them from being of rhoptry origin, the former are more promising candidates as rhoptry proteins. Whilst not within the scope of this thesis, it would be interesting to raise antibodies to these proteins to enable elucidation of their cellular localisation. A further 'hypothetical' identification, NCLIV_007450, was recently identified by Talevich *et al.* (2011) as a potential rhoptry kinase, it has rhoptry homologues in *N. caninum* and *T. gondii* and despite this, was not detected by the bioinformatics search carried out in Chapter 3: this goes to show the benefit of taking experimental approaches in addition to *in silico* ones, to discover the maximum number of proteins.

The following dense granule proteins were identified within the R fraction: GRA1, GRA2, GRA6, GRA7, GRA12 and a 14-3-3 protein (NCLIV_024820); hence the sucrose floatation did not appear to have been successful at removing all of the dense granule contamination. The protein identifications made from the R and R/DG fractions contained a large proportion of redundancy between the two samples (83 proteins, Figure 4.6), which, taken together with the identification of GRA proteins, implicates dense granule contamination in the rhoptry fraction. The almost identical appearance of the gel for the two samples also supports this theory (Figure 4.5); the lighter staining of the R/DG fraction compared to the R fraction was simply a result of sample loss that occurred during loading of the gel. There was insufficient sample available to perform a protein assay prior to gel loading, but both samples were revealed to contain sufficient protein to enable visualisation by Colloidal Coomassie staining and subsequent detection of tryptic peptides by LC MS/MS.

### 4.4.3 Coverage of the *N. caninum* proteome

Overall, 13 of 42 predicted microneme proteins (31 %) were identified by one or more of the experiments (including the whole tachyzoite 1-DE and MudPIT analyses in Chapter 2). Of the 22 putative dense granule proteins, only seven have eluded proteomic detection (68 % identified) and 35 (60 %) of 58 rhoptry proteins have been identified, in addition to the six candidate rhoptry proteins which need further experimental investigation to determine whether they are in fact of rhoptry origin. A recent sucrose fractionation analysis of the *N. caninum* rhoptries, using the ToxoDB version 5.2 annotation, identified up to thirteen possible rhoptry proteins (Marugan-Hernandez *et al.* 2011), one of which, ROP30, was not identified in this analysis.

The ESA experiment yielded one additional microneme protein not identified by the tachyzoite or rhoptry/dense granule experiments. However, approaches taken by other researchers have required vast numbers of parasites ($4 \times 10^9$ in Zhou *et al.* (2005)) and a pooling of results from multiple proteomic platforms, to identify eleven of the known microneme proteins. Zhou *et al.* (2005) identified the *T. gondii* microneme proteins MIC1, MIC2, MIC4, MIC5, MIC6, MIC8, MIC10, MIC11, M2AP, AMA1 and SUB1. Of these, MIC1, MIC4, MIC6, MIC10, M2AP, AMA1 and SUB1 were also identified in *N. caninum* in

this experiment. MIC2 (along with the *N. caninum*-specific MIC2B) and MIC11 were identified by the tachyzoite experiment and MIC3, which was not identified by Zhou *et al*. (2005), was identified by the ESA experiment. This study was successful in simplifying the preparation when compared to a tachyzoite lysate, and there was a proportional increase in apical proteins compared to the whole tachyzoite analysis (22 % and 5 %, respectively). Furthermore, the instrument time was significantly less for the ESA experiment: approximately 30 hours compared to 400 hours for the whole tachyzoite analysis.

Proteomic detection techniques are limited in that it is not possible to extract *all* proteins from any cell type (Gygi *et al.* 2000; Leimgruber *et al.* 2002), and some protein types (for example, hydrophobic membrane proteins that do not solubilise well (Santoni *et al.* 1999)) are not compatible with gel electrophoresis. Furthermore, during mass spectrometry some peptides are more proteotypic than others and hence are more readily detectable (Craig, Cortens & Beavis 2005) but on the whole LC MS/MS is highly sensitive and an extremely valuable technology for protein identification which allows analysis of complex mixtures for a 'systems biology' approach to research.

Overall, the data generated by these analyses open up avenues for further exploration (such as the potential novel rhoptry proteins discussed previously), and confirm the protein predictions for a number of genes currently only annotated as 'hypothetical'. Invasion proteins are successively released from the micronemes, rhoptries and dense granules over the duration of invasion (reviewed in Dubremetz *et al.* (1998)), but the speed at which this process occurs (entry usually takes just 5-10 seconds (Dubremetz *et al.* 1998)) and the non-synchronicity of the parasite cultures used in these experiments (i.e. tachyzoites within a sample will be at different points in invasion/replication) should mean that this design enabled identification of as many of the proteins involved as possible. These experiments represent the tachyzoite stage, which, while likely to share some proteins with the other invasive stages (sporozoites, bradyzoites) may not express all the possible invasion-related proteins, as there are likely to be some specific to each life cycle stage. Hence, to have proteomically identified 52 % (n = 63) of those predicted in the genome to be from the micronemes, rhoptries and dense granules (Chapter 3, n = 122) in the tachyzoite stage alone is quite considerable, and when considered in combination with the whole tachyzoite proteomic identifications in Chapter 2, represents a comprehensive analysis of the *N. caninum* proteome.

# CHAPTER 5: *ROP18* FUNCTIONAL ANALYSIS

## 5.1 INTRODUCTION

During annotation of the *N. caninum* genome, a number of pseudogenes were identified (Reid *et al.* 2012). One of these, encoding the rhoptry protein ROP18, had recently been identified by a number of studies as being implicated in *T. gondii* virulence. Therefore, the fact that it was apparently non-functional in the *N. caninum* genome was of interest and warranted further investigation.

A pseudogene is described by Vanin (1985) as a sequence containing "genetic lesions that preclude translation of a transcript from [the sequence] into a functional polypeptide equivalent to the functional gene product". An example of one of these lesions would be a stop codon occurring prematurely in the sequence, perhaps as a result of a nucleotide substitution, or alternatively there could be a deletion resulting in frame-shift.

Adam Reid (Sanger Institute) detected a number of stop codons within the region of the *N. caninum* genome syntenic to the *ROP18* gene in *T. gondii*. Various gene finding algorithms used in the genome assembly and annotation process were unable to identify an alternative location for the *ROP18* gene in *N. caninum* and it was therefore hypothesized that it was a pseudogene, hence it would not be able to produce a functional ROP18 protein. In addition, the RNA-Seq experiment carried out as part of the genome annotation project (Reid *et al*. 2012) detected very low mRNA expression levels over the *ROP18* region.

### 5.1.1 Rhoptry gene *ROP18*

*ROP18,* which was also previously known as *ROP2L2* (Bradley *et al*. 2005) is a well-characterized virulence gene in *T. gondii*. In their 2007 review, Sinai *et al*. summarized ROP18 as a stand-out member of the ROP2 family of rhoptry kinases, as it, along with ROP16, is one of only two ROP proteins to have retained their kinase function (Sinai 2007). This has since been thrown into contention as many other rhoptry kinases, once thought to be pseudokinases, have now been identified as active (Peixoto *et al.* 2010). ROP18 is secreted to the parasitophorous vacuole membrane (PVM) at the time of invasion (El Hajj *et al.* 2007). More recently the Howard group from the University of Cologne have shown ROP18 to phosphorylate host-derived GTPases on the PVM (Steinfeldt *et al.* 2010) which is described in section 5.1.5. Whether ROP18 is involved in additional roles remains to be seen, but this is

likely to be the case at it has been shown to phosphorylate a parasitic substrate of approximately 70 kDa, that could possibly be ROP4 (El Hajj *et al.* 2007).

*ROP18* shares 28% homology with *ROP5* (El Hajj *et al.* 2007) which is another *T. gondii* virulence determinant (Reese *et al.* 2011; Reese & Boothroyd 2011) which, although present in the *N. caninum* genome, has fewer tandem repeats (two) than *T. gondii* which have four or more, depending on the strain (Behnke *et al.* 2011; Reese *et al.* 2011). The rhoptry family contains a number of paralogues (*ROP2/8, ROP4/7* and *ROP18/5*) which may be suggestive of the fact that the rhoptry family is under significant selection pressure and is undergoing evolutional change (Qiu *et al.* 2009), in order to cope with the demands of parasitizing a host who will also be evolving. Interestingly, ROP kinases are not orthologues of *Plasmodium spp.* FIKKs, although they are both families of secreted kinases involved in the invasion process and containing a serine/threonine domain (Qiu *et al.* 2009; Ward *et al.* 2004).

*ROP18* expression levels vary from strain to strain in *T. gondii*; expression is 10 000 times lower in type III than type II (Saeij *et al.* 2006). Sequencing of the entire region in type I, II and III genes showed a sequence insertion in the 5' untranslated region (UTR) promoter of the type III *ROP18* allele, this is likely to be the cause of the difference in expression (Saeij *et al.* 2006). *ROP18* is variable between all three *T. gondii* types: the type II *ROP18* allele has 11 single nucleotide polymorphisms (SNPs) relative to type I and III, but the type III allele has unexpectedly large number of SNPs (85, 64 of which are non-synonymous and result in amino acid changes) relative to types I and II (Saeij *et al.* 2006); it is probable that *ROP18* is the genetic basis for the virulence QTL (quantitative trait loci) identified on chromosome VIIa (Taylor et al. 2006).

Despite this substantial genetic diversity among lineages, there are three principle alleles of *T. gondii ROP18* which account for much of the variation in acute mouse virulence among natural isolates (Khan *et al.* 2009). Behnke *et al.*(2011) state that *ROP18* cannot account for virulence differences between types I and II, but Khan *et al.*(2009) describe three decreasing levels of virulence for types I, II and III respectively.

The *T. gondii ROP18* type III allele is expressed at very low levels, it is not a pseudogene and it contains all of the essential residues in the S/T domain necessary for kinase activity (Khan *et al.* 2009). Like *T. gondii ROP18* type III, the *ROP18* gene in *N. caninum* includes an insertion in the 5' UTR promoter upstream of the gene that decreases expression levels, this is not present in the *ROP18* gene of *T. gondii* types I and II (Khan *et al.* 2009; Saeij *et al.* 2006). Alignments

using Clustal revealed that *N. caninum* was 54% identical to type III *ROP18* in this upstream region and 68% identical in the coding region (Khan *et al.* 2009). Therefore, it is interesting that *N. caninum* has gone further and has a pseudogenised *ROP18*, as that could indicate a potentially lower level of virulence than type III *T. gondii* and perhaps an adaptation to a different life-strategy.

### 5.1.2 Sequencing of the *ROP18* region to confirm presence of stop codons

Firstly, it was necessary to determine whether the sequence data was correct and the gene contained stop codons within the coding region (Figure 5.1). Furthermore, the Liverpool strain of *N. caninum,* used for the genome study, had been maintained in cell culture for a number of passages and there was the possibility that the stop codons could have occurred as a post-isolation mutation. It was important therefore to examine the sequence of a number of *N. caninum* isolates by polymerase chain reaction (PCR) to determine whether the pseudogenisation of *ROP18* was a feature of the species or of the isolate.

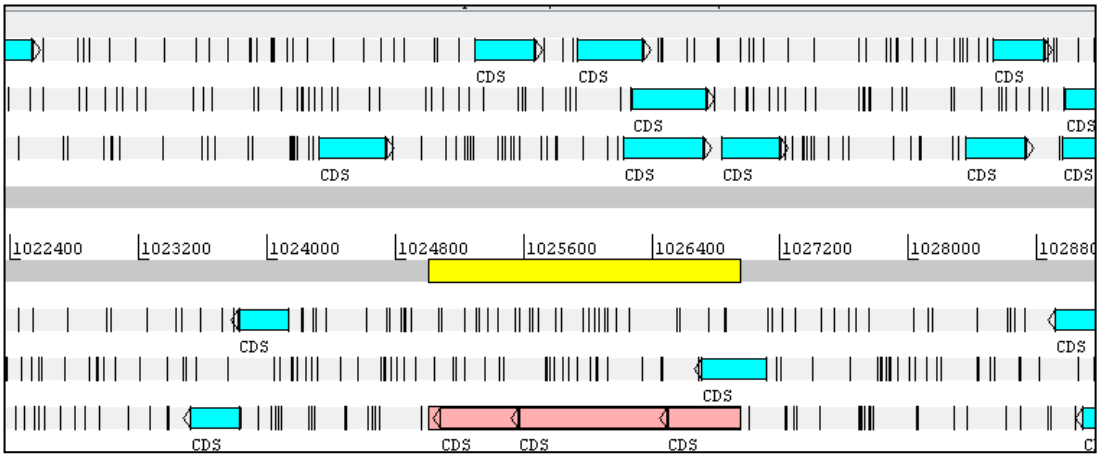



**Figure 5.1:** Artemis genome viewer (Sanger) showing a 6 frame translation of the *N. caninum ROP18* region (yellow indicates where the gene would be in *T. gondii*). The pseudogene is highlighted in pink, on the reverse strand. The vertical bars are the three stop codons within the would-be exon.

### 5.1.3 Analysis of transcriptomic and proteomic data for evidence of *ROP18* products

The *N. caninum* genome study involved a comparative RNA-Seq analysis of the transcriptome of *N. caninum* Liverpool and *T. gondii* VEG (Reid *et al.* 2012).While a pseudogene is unable to be translated into protein, transcription of the gene into mRNA is still a possibility. However, unless there is some, so far undiscovered, use for mRNA other than translation, it would not seem efficient for a cell to waste resources transcribing non-coding sequences into mRNA.

Therefore, it would make sense for there to be little or no transcription detected for the *ROP18* region in *N. caninum*. Adam Reid (personal communication) confirmed this to be the case, *ROP18* transcripts were present, but at a very low level similar to that of the avirulent type III *T. gondii* VEG isolate, shown by (Khan *et al.* 2009) to exhibit suppressed expression of this gene compared to virulent strains. Proteomically however, the hypothesis was that there would be no peptides synthesized for this region and it was therefore important to search the proteomic datasets presented in Chapters 2 and 4 to ensure that this was the case. The identification of ROP18 peptides would undermine the genome sequence data and therefore result in a revision of the proposed pseudogene hypothesis.

### 5.1.4 The interferon-γ response to coccidian infection and the role of ROP18

Interferon-gamma (IFN-γ) is known to control apicomplexan proliferation in mammalian host cells as a component of the cell-mediated immune response (Boysen *et al.* 2006; Innes *et al.* 2002; Innes *et al.* 1995; Nishikawa *et al.* 2001a; Nishikawa *et al.* 2001b; Suzuki *et al.* 1988; Yamane *et al.* 2000). As part of the host cell IFN-γ response in *T. gondii* infections, immunity-related GTPases (IRGs) are loaded onto the parasitophorus vacuole (PV), leading to its disruption and eventual parasite death (Hunn *et al.* 2008; Khaminets *et al.* 2010; Martens *et al.* 2005; Zhao *et al.* 2009a; Zhao *et al.* 2009b; Zhao *et al.* 2009c). In retaliation, virulent strains of *T. gondii* can employ rhoptry kinase protein ROP18 to phosphorylate IRGs and render them inactive, whilst preventing their accumulation on the PV (Steinfeldt *et al.* 2010).

IRG proteins are resistance factors that are induced by IFN-γ (Bekpen *et al.* 2005; Konen-Waisman & Howard 2007; Taylor 2007). They are present in a range of chordates (Li *et al.* 2009) and in their review on mammalian IRGs, Hunn *et al.* (2011) state that "the number, type and diversity of genes present differs greatly even between closely related species, probably reflecting intimate host pathogen co-evolution driven by an arms race between the IRG resistance proteins and pathogen virulence factors". Most mammals, including cattle, the *N. caninum* preferred intermediate host, contain a set of IRG proteins, although the precise repertoire tends to vary between species and has been most extensively studied in mice (Hunn *et al.* 2011) and J Hunn, personal communication).

Of the murine IRGs, some have been characterized more completely than others; IRG is the only one for which structural, biochemical and functional data are available (Ghosh *et al.* 2004; Martens *et al.* 2005; Papic *et al.* 2008; Steinfeldt *et al.* 2010; Uthaiah *et al.* 2003). IRG proteins

load onto the PV in a hierarchical manner and can be present in a variety of combinations with each other (Khaminets *et al.* 2010). IRG a6 and IRG b6 both load onto the majority of vacuoles during an avirulent *T. gondii* infection into mouse embryonic fibroblasts (MEFs); interestingly, a small proportion of vacuoles remain unloaded throughout an infection, even with avirulent parasite strains (Khaminets *et al.* 2010). Infections with virulent strains of *T. gondii* exhibit markedly reduced IRG PV loading (Zhao *et al.* 2009a) which is now known to be due to the ability of ROP18 to phosphorylate either of two threonines, T102 or T108, on IRG a6 (Steinfeldt *et al.* 2010). These threonines are highly conserved throughout the mouse IRG proteins, although phosphorylation by ROP18 has so far only been examined on IRG a6 (Steinfeldt *et al.* 2010), where it was determined by studying the infection of *T. gondii* strains into MEFs expressing various mutated forms of T108 and T102. In normal infection, IRG 6 accumulates on the majority of PVs for all strains of *T. gondii*, but the intensity of antibody staining (and hence amount of IRG a6 accumulated) is much lower for virulent strains (Holmdahl *et al.* 1997; Khaminets *et al.* 2010). Steinfeldt *et al.* (2010) confirmed their findings by conferring the ability of type I virulent strains (to phosphorylate IRG a6) to type III avirulent strains by transfection of type I *ROP18*. They suggest that destabilisation of interactions between IRG molecules and/or associations between IRGs with the vacuole, is how the phosphorylation of T102 or T108 by ROP18 affects IRG PV loading. An additional study has also indicated that IRG b6 could be another target for ROP18 (Fentress *et al.* 2010). Type II *T. gondii*, generally considered to be of an intermediate level of virulence, are only able to exhibit a very low frequency of IRG a6 phosphorylation. With regard to IRG phosphorylation, type II *T. gondii* (as well as type III) exhibit what would be considered an avirulent phenotype (Khaminets *et al.* 2010; Steinfeldt *et al.* 2010).

## 5.2 Aims and objectives

Following on from the hypothesis that *N. caninum ROP18* is pseudogenised, an experiment was undertaken to determine whether i) IRG proteins are loaded onto the PV in *N. caninum*-infected MEFs and ii) whether we would observe IRG-phosphorylation in the (hypothesized) absence of *ROP18*.

## 5.2 MATERIALS AND METHODS

### 5.2.1 *ROP18* sequencing by PCR

The *ROP18* region was amplified from five *N. caninum* isolates (Table 5.1). For the Nc Liverpool strain, parasites were cultured in vero cells as described in Chapter 2.2 and DNA was extracted using a Qiagen DNeasy Kit, according to the manufacturer's protocol. DNA for strains other than Nc Liverpool were previously extracted by Dr. Sophia Latham.

**Table 5.1:** *N. caninum* strains analysed by PCR

| Isolate | Country of origin | Source | Reference |
|---|---|---|---|
| Nc Liverpool | UK | Canine | Barber *et al.* (1993) |
| Nc 1 | USA | Canine | Dubey *et al.* (1988) |
| Nc Liverpool B1 | UK | Bovine | Davison *et al.* (1997) |
| Nc BPA | USA | Bovine | Conrad *et al.* (1993) |
| Nc JPA | Japan | Bovine | Yamane *et al.* (1996) |

Primers were designed to amplify the *ROP18* region of the *N. caninum* genome sequence using Primer3 (Rozen & Skaletsky 2000); all primers were supplied by Eurofins MWG Operon. As the region to be amplified was almost 2000 base pairs in length, pairs of overlapping primers were designed in order to amplify the region in parts: ROP18 F1+ ROP18 R3 and ROP18 F3 + ROP18 R1 (Table 5.2).

**Table 5.2:** Primers used for *ROP18* sequencing

| Primer name | Sequence (5'-3') | Expected product size (bp) |
|---|---|---|
| ROP18_F1 | GAGTGCCACGGTCCTCTAAG | 1268 |
| ROP18_R3 | ATTTGTCCGACGCAAAATTC | |
| ROP18_F3 | GGCTTCTGCTCCAGTATTCG | 890 |
| ROP18_R1 | GCCTTATAAACCACCCGTCA | |

ROP18 F1-R3 PCR reactions were set up in a final volume of 50 µl containing the following: 1.5 mM 10x PCR buffer containing 15mM $MgCl_2$, an additional 2.5 mM $MgCl_2$, 10 µl 5x Q solution, 0.2 mM of each dNTP (Sigma), 0.4 µM F1 primer, 0.4 µM R3 primer, 2.5 U Taq DNA polymerase and 2 µl DNA.

PCR reaction conditions for F1 and R3 primers were as follows: 94 °C for 3 minutes; 30 cycles of 94 °C for 30 seconds, 53 °C for 30 seconds, 72 °C for 1 minute; 72 °C for 10 minutes.

ROP18 F3-R1 PCR reactions were set up in a final volume of 50 µl containing the following: 1.5 mM 10 x PCR buffer containing 15 mM $MgCl_2$, an additional 2.5 mM $MgCl_2$, 10 µl 5x Q solution, 0.2 mM of each dNTP (Sigma), 0.4 µM F3 primer, 0.4 µM R1 primer, 2.5 U Taq DNA polymerase and 2 µl DNA.

Conditions for F3 and R1 primers (BPA and Nc1 strains) were: 94 °C for 3 minutes; 30 cycles of 94 °C for 30 seconds, 58 °C for 30 seconds, 72 °C for 1 minute; 72 °C for 10 minutes. F3 and R1 amplification of NcLiv, Nc Liv B1 and JPA were as for BPA and Nc1 but with an annealing temperature of 62 °C. PCRs were carried out using a DYAD Peltier Thermal Cycler DNA Engine (MJ Research).

Samples were visualized on a 1.5 % agarose gel and cleaned up for sequencing using a QIAQuick PCR purification kit (Qiagen) according to the manufacturer's protocol. PCR products were sequenced by Mandy Sanders at the Sanger Institute using BigDye® Terminator v3.1 Cycle Sequencing Chemistry (Applied Biosystems) and electrophoresis on a 3730xl DNA Analyser (Applied Biosystems).

### 5.2.2 Minisatellite analysis of the five strains

Primers and methods for amplification of the Contig16 minisatellite are described in Al-Quassab *et al*. (2010). The minisatellite region was as follows:

Minisatellite region Contig16: GAAGAGGAAGGGGAAGAAGAAGAGGAA
Forward primer: CTGCTGAGGCTCTGAGGAAC
Reverse primer: CCTCCCTTCCTCTCTGACC

The PCR mixture (50 µl) contained 1× DNA polymerase reaction buffer, 0.2 mM dNTPs, 2.0 mM $MgCl_2$, 1.5 U *Taq* polymerase, 0.2 µM of each primer and 1µl template DNA. A negative control containing no template DNA was also included in the experimental design.

PCR was performed in DYAD Peltier Thermal Cycler DNA Engine (MJ Research) under the following conditions: 95 °C for 5 minutes; 25 cycles of 94 °C for 1 minute, 61 °C for 1 minute, 72 °C for 1 minute; followed by 1 cycle of 72 °C for 5 minutes. Samples were visualized on a 1.5 % agarose gel.

### 5.2.3 Examination of proteomic data for ROP18 peptides

All LC MS/MS files generated in the whole tachyzoite 1DE analysis, the ESA analysis and the R/DG analysis, were resubmitted to Mascot to be searched against a new database containing an additional 'gene' comprising the *ROP18* region, using parameters previously described in Chapter 2.2.

### 5.2.4 Immunity-Related GTPase (IRG) experiments

This work was carried out while visiting the laboratory of Professor Jonathan Howard, and with the assistance of Dr. Steffi Weissman.

#### 5.2.4.1 Cell and parasite culture

C57BL/6 Mouse Embryonic Fibroblasts (MEFs), HS27 cells and Vero cells were cultured in IMDM supplemented with 10 % FCS, 100 U $ml^{-1}$ penicillin and 100 mg $ml^{-1}$ streptomycin. For IFN-γ induction, 1 day cultures of MEFs, grown on cover slips in 6-well plates, had their culture media replaced with media containing a final concentration of 200 U/ml IFN-γ (Peprotech) (otherwise same recipe as above) while non-induced MEFs simply had their media refreshed. This took place 24 hours before infection with parasites. *N.caninum* Liverpool isolate tachyzoites were passaged in Vero cells as described in Chapter 2.2 and *T.gondii* RH and ME49 isolates were kindly provided by S. Könen-Waisman from HS27 cell passage as described in Martens *et al.* (2005). They were then used for infection of untreated and IFN-γ -induced MEFs at a multiplicity of infection (m.o.i.) from 5 to 10, before being replaced in the incubator at 37 °C. At 2 hours post-infection, cells were washed three times with 2 ml PBS (phosphate-buffered saline) then fixed in PBS/3 % paraformaldehyde for 20 minutes at room temperature. After a further three washes in 2 ml PBS, cells were covered with parafilm and stored at 4 °C ready for permeabilisation and immunostaining.

#### 5.2.4.2 Immunological reagents and immunofluorescence analysis

Cells fixed as above (5.2.4.1) were permeabilised in 2 ml 0.1 % saponin/PBS for 10 minutes at room temperature, then blocked with 2 ml 3 % bovine serum albumin (BSA)/0.1 % saponin/PBS for 1 hour at room temperature or 4°C overnight, except for slides stained with T102-555 which were permeabilised in ice cold methanol for 20 minutes and blocked in 1 % BSA/PBS for 30 minutes.

**Table 5.3:** Primary antibodies used and their dilutions. Secondary antibodies were all diluted 1:1000. Antibodies were all kindly provided by J. Howard with the exception of 20B5D5 which was provided by P. Bradley.

| Primary Antibody | Dilution | Secondary Antibody | Reference if applicable |
|---|---|---|---|
| rabbit anti- b1-b2 C15A | 1:5000 | Alexa 488 labelled donkey anti- rabbit sera | |
| rabbit anti-Irga6 antiserum 165 | 1:8000 | Alexa 488 labelled donkey anti- rabbit sera | (Martens *et al.* 2005) |
| mouse anti-Irga6 mAb10E7 | 1:500 | Alexa 488-labelled donkey anti-mouse sera | (Papic et al. 2008) |
| anti Irga6 phosphopeptide Ab T102-555 | 1:5000 | Alexa 555-labelled donkey anti- rabbit sera | (Steinfeldt *et al.* 2010) |
| goat anti-Irgb6 antiserum A20 | 1:100 | Alexa 488-labelled donkey anti-goat sera | |
| rabbit anti-Irgb6 antiserum 141/1 | 1:4000 | Alexa 488 labelled donkey anti- rabbit sera | N. Pawlowski, unpublished data |
| rabbit anti-Irgm2 antiserum H53/3 | 1:500 | Alexa 488 labelled donkey anti- rabbit sera | (Martens *et al.* 2004; Martens *et al.* 2005) |
| rabbit anti-Irgb10 antiserum | 1:2000 | Alexa 488 labelled donkey anti- rabbit sera | (Coers et al. 2008) |
| mouse anti-ROP2 family mAb 20B5D5 | 1:2000 | Alexa 555-labelled donkey anti-mouse sera | (Sohn *et al.* 2011) |
| 4′,6-Diamidine-2′-phenylindole dihydrochloride (DAPI) | 1:1000 | | |

Antibodies (Table 5.3) were diluted in 3 % BSA/0.1 % saponin/PBS and cover slips were incubated in 100 µl of the appropriate antibody solution for 1 hour at room temperature or 4 °C overnight. After three washes in 0.1 % saponin/PBS, cover slips were incubated for 30 minutes at room temperature, in the dark, in 100µl of the secondary antibody solution. A further three washes in 0.1 % saponin/PBS were carried out before the cover slips were mounted onto slides with 18 µl Prolong Gold Antifade Reagent (Invitrogen) and left to harden in the dark for a minimum of 2 hours. Slides were stored ready for microscopy at 4 °C in the dark.

Images were taken with an Axioplan II fluorescence microscope and AxioCam MRm camera and processed by Axiovision 4.7 (all Zeiss). Intracellular parasites were identified by observing the vacuolar localization of the *N.caninum* ROP2 family member protein or *T. gondii* protein GRA7 and by distinct pathogen appearance in phase contrast.

## 5.3 RESULTS

### 5.3.1 Examination of proteomic data for ROP18 peptides

No peptides corresponding to the *ROP18* region of the *N. caninum* genome, nor *T. gondii ROP18*; were identified in tachyzoite, excretory/secretory or rhoptry/dense granule *N. caninum* lysates. This is supportive of the pseudogene hypothesis but is not conclusive.

### 5.3.2 PCR sequencing of the *ROP18* region

All five strains of *N. caninum* analysed exhibited the three stop codons identified in *N. caninum* Nc Liverpool within their *ROP18* regions. The conservation of these stop codons can be seen in Figure 5.2, these are TAA, the ochre stop codon, twice; and TGA, the opal or umbre stop codon. As the presence of just one stop codon is sufficient to render a gene pseudogenised, it is safe to conclude that the pseudogenisation of this gene is conserved across the isolates and ROP18 protein would not be produced for any of these strains. Furthermore, the strains were selected based upon their dissimilarity to each other in geographical location and had been isolated from both cattle and dogs, so should be representative of the *N. caninum* population as a whole, rather than a subset of closely related strains.

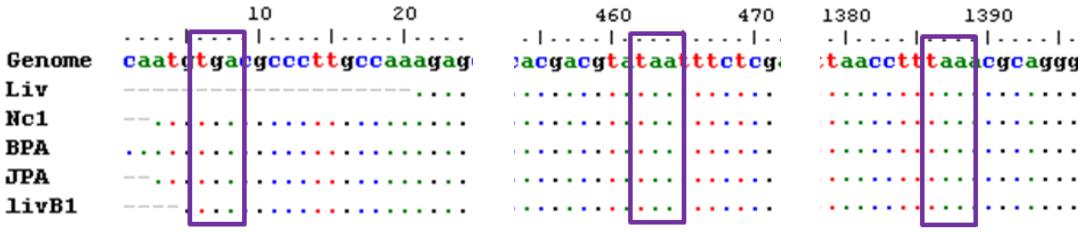



**Figure 5.2:** Truncated *ROP18* region of five strains of *N. caninum* amplified by PCR and aligned to the reference genome using BioEdit version 7.0.5.3. Stop codons are identified by purple boxes. This region is displayed in the 5' to 3' orientation, transcribed from the antisense strand of chromosome VIIIa. No SNPs were identified in any of the isolates analysed. The whole *ROP18* region is shown in Appendix VI.

Because the region analysed showed no single nucleotide polymorphisms (SNPs) between strains, it was decided to carry out a minisatellite PCR analysis to ensure that the DNA being amplified was indeed different for the five strains and not simply a contamination with Nc Liverpool. The contig16 minisatellite (Al-Qassab *et al.* 2009) enabled distinction between the strains and the resulting PCR gel can be seen in Figure 5.3, where the different sizes of the products from different *N. caninum* strains are apparent. While BPA did not amplify, it is clear

that the remaining strains exhibit a range of product sizes within the expected range of 854-965 base pairs and hence these results confirm that the DNA used for the *ROP18* PCRs did show variation. This result confirms that the lack of SNPs seen in the different *ROP18* PCR products was not attributable to contamination. This is also supported by a number of reports reviewed by (Al-Qassab, Reichel & Ellis 2010) all of which found no nucleotide differences in various protein-coding genes between *N. caninum* isolates.

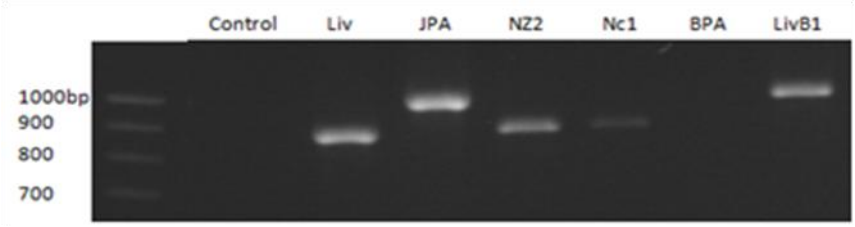



**Figure 5.3:** Minisatellite analysis of the five *N. caninum* strains to confirm amplification of DNA from different strains. Lane 1 = 1000bp ladder, lane 2 = negative control, lanes 3-8 = *N. caninum* isolates.

### 5.3.3 Immunofluorescence analyses

Immunity-related GTPase (IRG) protein rings (staining the parasitophorus vacuole (PV)) were observed in *N. caninum* infected IFN-γ stimulated MEFs, stained with antibodies against IRG a6, IRG b1-b2, IRG b6, IRG m2 and IRG b10 (Figure 5.4). PVs were approximately 60% and 45% positive for IRG a6 and IRG b1-b2, respectively (two replicates of 100 PVs (Figures 5.5 - 5.6), identified by phase contrast by systematic movement over the slide). Some cross reactivity of the antibody with the parasite was observed for the other IRG proteins (b6, m2 and b10) so that it was not possible to determine the frequency of IRG-positive PVs. However, vacuoles were sufficiently visible by phase contrast that it was clear that IRG loading on the PV was occurring in the IFN-γ stimulated cells and not in the non-stimulated negative controls, where just the parasite, and not the vacuole, exhibited staining (data not shown).

#### 5.3.3.1 IRG a6

Rings were observed in *N. caninum* infected, IFN-γ stimulated cells at a frequency of ~60 % of PVs, this indicated that IRG a6 was loaded onto the vacuoles during *N. caninum* infections. Non-stimulated controls were negative for IRG a6 staining, as were extracellular parasites. Figures 5.4 and 5.5 illustrate the very visible a6 staining on the PV of two intracellular

tachyzoites, in contrast to a non-IRG-loaded PV of another parasite. The vacuole is just visible around this parasite and hence confirms that it is intracellular.

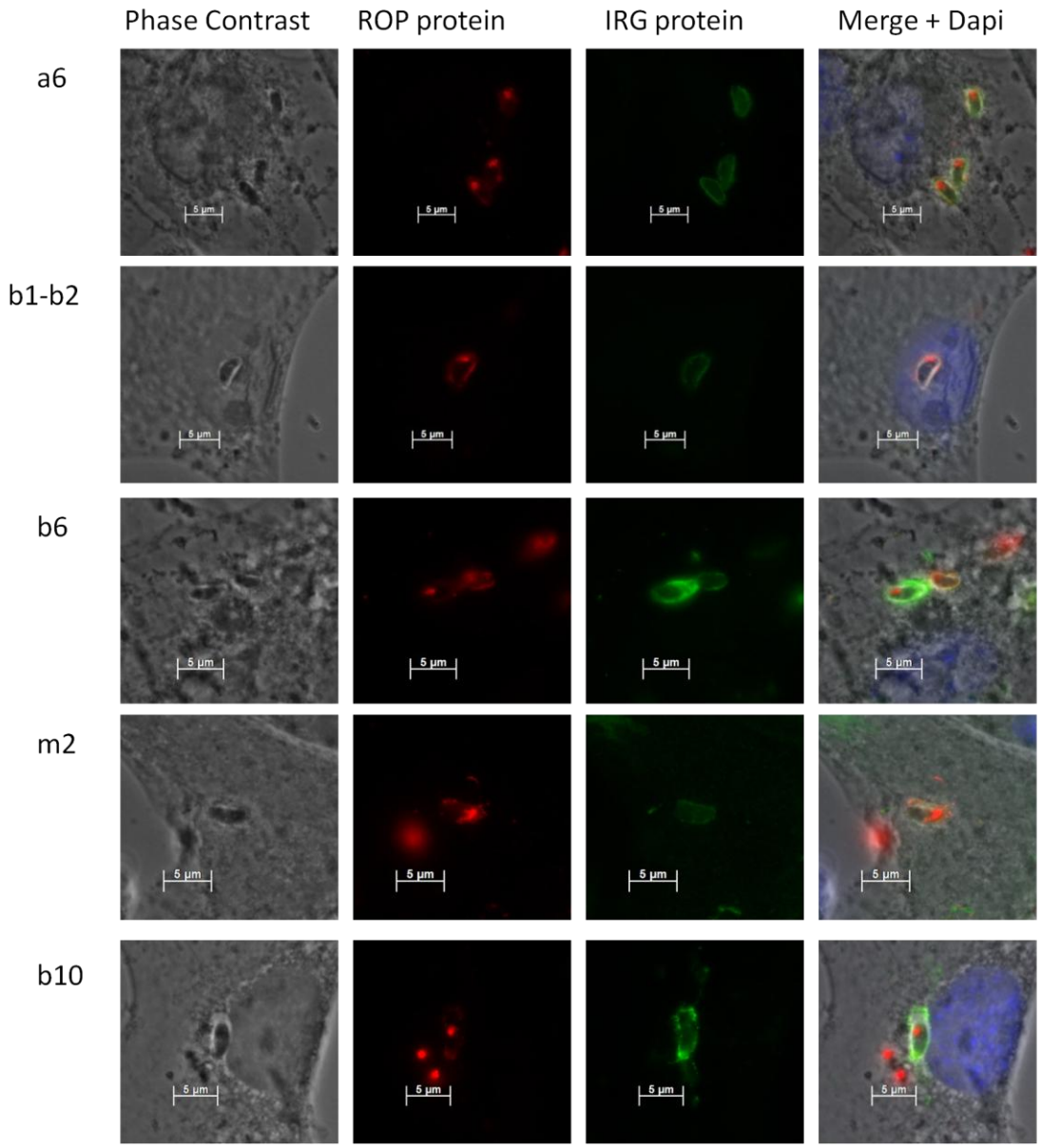



**Figure 5.4:** *N. caninum* parasitophorus vacuole IRG loading. Examples for each antibody tested (left hand labels) showing from left to right: phase contrast, localisation of tachyzoites (rhoptries) labelled with the anti-ROP2 family member antibody 20B5D5, PVs stained with the anti-IRG antibody and finally a merged picture including DAPI staining for nuclei.

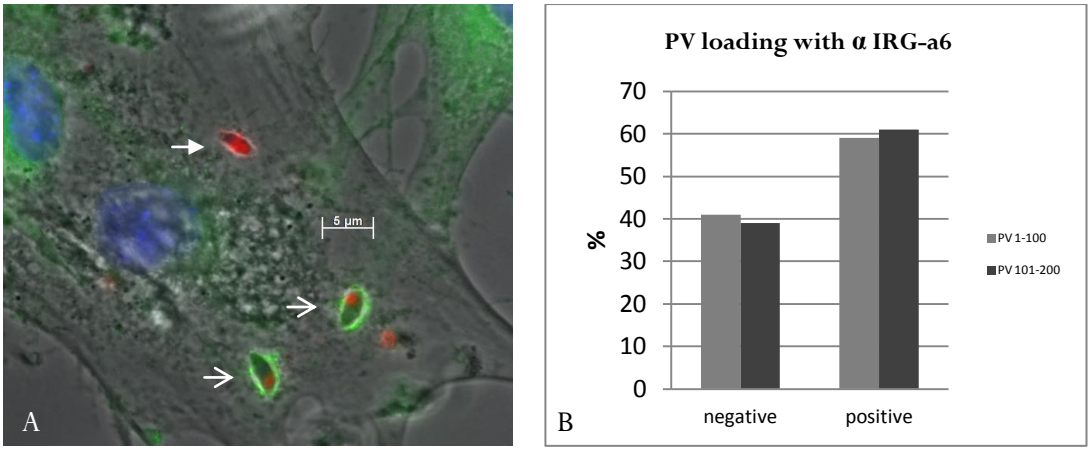



**Figure 5.5:** *N. caninum* parasitophorus vacuole IRG a6 loading. A) Localisation of tachyzoites (rhoptries) labelled with the anti-ROP2 family member antibody 20B5D5 (red), PVs stained with anti-IRG a6 (green) and nuclei with DAPI. Open arrows indicate PVs positive for IRG a6, closed arrow negative. B) Percentage of *N. caninum* parasitophorous vacuoles loaded with IRG a6, 200 PVs identified by phase contrast in two separate counts of 100 PVs.

### 5.3.2.2 IRG b1-b2

Rings were observed in *N. caninum* infected, IFN-γ stimulated cells at a frequency of ~55 % of PVs, indicating that IRG b1-b2 was loaded onto the PVM during *N. caninum* infections. As for IRG a6, non-stimulated controls and extracellular parasites were all negative for b1-b2 staining (Figures 5.4 and 5.6).

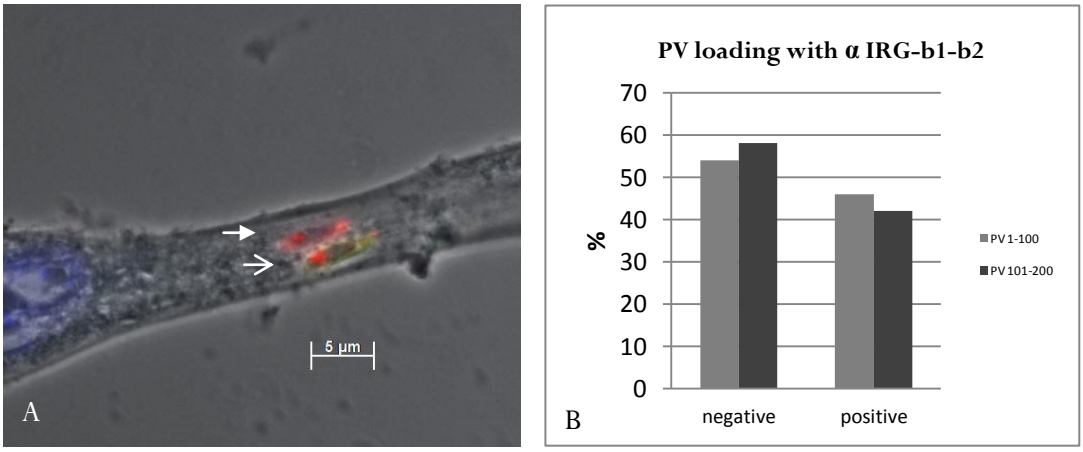



**Figure 5.6:** *N. caninum* parasitophorus vacuole IRG b1-b2 loading. A) Localisation of tachyzoites (rhoptries) labelled with the anti-ROP2 family member antibody 20B5D5 (red), PVs stained with anti-IRG b1-b2 (green) and nuclei with DAPI. Open arrow indicates PV positive for IRG a6, closed arrow negative. B) Percentage of *N. caninum* parasitophorous vacuoles loaded with IRG b1-b2, 200 PVs identified by phase contrast in two separate counts of 100 PVs

#### 5.3.3.3 IRG b6, IRG m2 and IRG b10

Cross reactivity of the anti-IRG antibodies to the parasite meant that quantification of IRG-of loading was not possible. However, qualitatively, it was possible to see the characteristic rings IRG-loaded vacuoles in a number of cases; examples of these IRG-positive PVs (rows 3, 4 and 5 showing IRG b6, m2 and b10 respectively) are presented along with further a6 and b1-b2 examples (rows 1 and 2; IRG a6 and b1-b2) in Figure 5.4.

#### 5.3.3.4 IRG a6 phosphorylation

Phosphorylation of IRG a6 was visualized by staining of the T102 a6 residue (Steinfeldt *et al.* 2010); slides were co-stained with anti-a6 (10E7 antibody) to identify a6-positive vacuoles (Figure 5.7). Whilst > 90 % of a6-positive vacuoles were phosphorylated in *T. gondii* RH infected MEFs (concurrent with findings in Steinfeldt *et al.* (2010)), no anti-T102 phosphorylated PVs were observed in *N. caninum* infected MEFs. Frequency of phosphorylation was estimated by two replicates of 100 PVs identified in the red (a6 10E7) channel by systematic movement over the slide (Figure 5.8).

There was some cross reactivity of the anti-IRG a6 antibody with the *T.gondii* RH parasites, but it was clearly distinguishable as to whether the staining pertained to the vacuole or just to the parasite (Figure 5.9). The third row in Figure 5.9 shows an a6-ringed RH parasite that was negative for phosphorylation – there is green staining but this is only on the parasite itself, not the vacuole – the red does not co-localise with the green. This is in contrast to the first row, where the IRG a6 staining (green) is staining the vacuole as can be seen clearly in the merged image. This cross reactivity was not observed with *N. caninum* or *T. gondii* type II ME49. A small number of phosphorylated IRG a6 rings were observed in ME49- infected MEFs, but staining was weak.

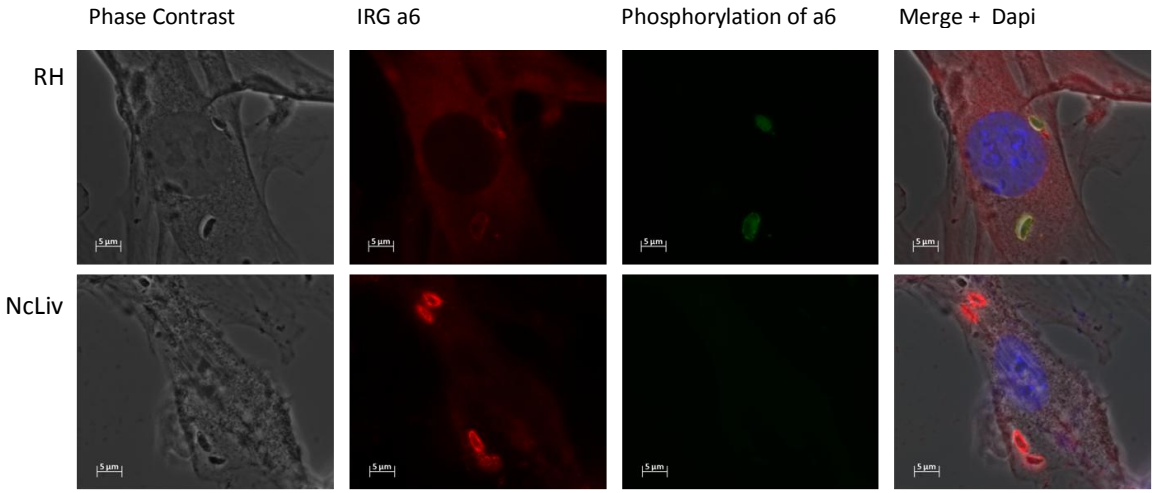



**Figure 5.7:** Phosphorylation of IRG a6- T102 was observed in *T. gondii* infected MEFs but not in *N. caninum*-infected MEFs, despite loading of IRG a6 onto the PV. From left to right: phase contrast, anti-a6 staining by 10E7, phosphorylated T102 residues of IRG-a6 stained by the T102-555 antibody and a merged image with DAPI. The first row shows MEFs infected with *T. gondii* RH and the second with *N. caninum* Liverpool.

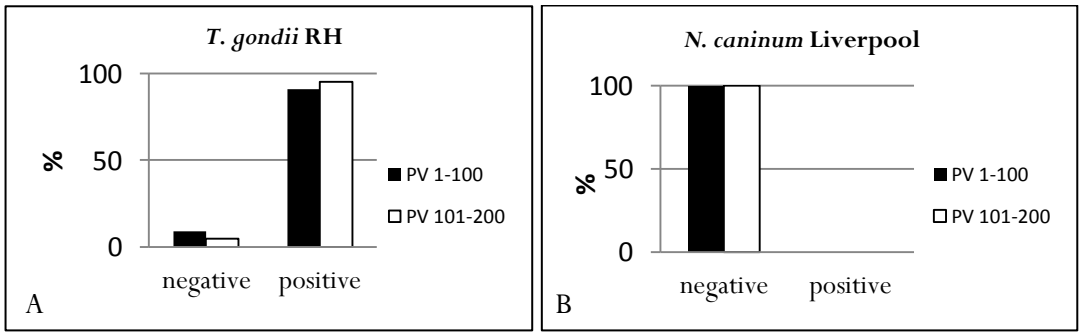



**Figure 5.8**: IRG a6 phosphorylation as a percentage of IRG a6 positive parasitophorous vacuoles for A) *T. gondii* and B) *N. caninum*. Counts were performed as two replicates of 100 a6-positive PVs.

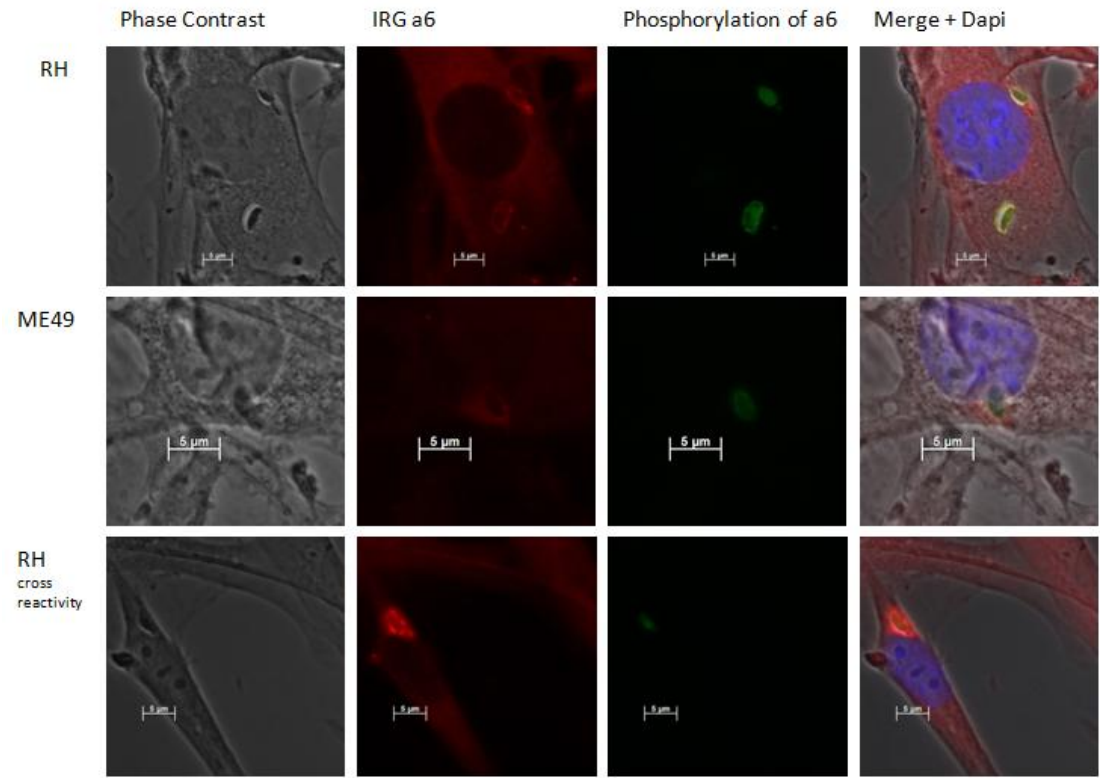



**Figure 5.9:** Phosphorylation of IRG-a6 in *T. gondii* type I (RH) and type II (ME49) isolates. From left to right: phase contrast, anti-a6 staining by 10E7, phosphorylated T102 residues of IRG-a6 stained by the T102-555 antibody and a merged image with DAPI. The first row shows positive (phosphorylated) a6-ringed vacuoles. The second row is the same phenomenon in type II ME49. The third row shows an a6-ringed RH parasite that was negative for phosphorylation. This is an example of the cross-reactivity observed on the RH slides, which was not seen for ME49 or *N. caninum* Liverpool.

## 5.4 DISCUSSION

It is not a simple procedure to search for *ROP18*, or other pseudogenes, in the *N. caninum* genome database (Gajria *et al.* 2008; Reid *et al.* 2012) due to the fact that they have no gene identifier. Proteomic and transcriptomic data tend to be presented as a list of gene identifiers followed by expression levels and/or peptides identified. However, in order to determine whether any ROP18 peptides had been identified in the proteomics experiments in chapters 2 and 4, an alternative protein database was created containing an additional FASTA sequence of the amino acids that would result if the *ROP18* region was translated. All LC MS/MS files generated in the whole tachyzoite 1DE analysis, the ESA analysis and the Rhoptry/Dense Granule analysis, were resubmitted to Mascot (Matrix Science) to be searched against the new database and no *ROP18* peptides were identified. Whilst a lack of proteomic evidence can never be sufficient to suggest a protein does not exist, when this information is taken into consideration alongside the sequence data, it supports the *ROP18* pseudogene hypothesis. Moreover, ROP18 peptides are readily detectable in equivalent *T. gondii* proteomic experiments (Gajria *et al.* 2008; Xia *et al.* 2008), hence it is unlikely that the lack of detection in the *N. caninum* proteomic experiments is due to characteristics of the protein making it unsuitable for proteomic detection.

The above results have shown loading of IRG-proteins by IFN-γ stimulated MEF host cells onto the PV of intracellular *N. caninum*. Due to *N. caninum*'s close relationship to *T. gondii* and their parallel phenotypes in host cell-invasion and parasitophorous vacuole formation, the hypothesis was that MEFs would respond to *N. caninum* infection in a similar way to *T. gondii* infection, by loading IRG proteins onto the PV, as seems to be the case.

Different frequencies of IRG-loading were observed for different strains of *T. gondii*, with virulent type I strains exhibiting fewer and/or less intensely loaded vacuoles than type II or III avirulent strains (Khaminets *et al.* 2010). The observation that approximately 60 % of *N. caninum* PVs load with IRG-a6 is consistent with previously reported findings for *T. gondii* Type II ME49 (Khaminets *et al.* 2010). Although the present study did not measure the intensities of staining on the vacuoles, it was apparent that heterogeneity existed. Whether this is related to non-synchronicity of invasion, or individual differences between parasites, is unclear; however, previous studies have observed this phenomenon even in synchronised *T. gondii* infections (Khaminets *et al.* 2010).

Of those parasitophorous vacuoles loaded with IRG-a6, 0 % and > 90 % exhibited phosphorylation in *N. caninum* and *T. gondii* infections respectively. Phosphorylation of the T102 residue of IRG-a6 has been shown by Steinfeldt *et al.* (2010) to be carried out by type I *T. gondii* ROP18 and confirmed by creating type III *T. gondii* CTG parasites transgenic for type I *ROP18*, which exhibit a type I phosphorylation phenotype. Thus, it was hypothesized that, if the presence of additional stop codons in *N. caninum ROP18* prevented translation of the protein, *N. caninum* would be unlikely to phosphorylate IRG-a6 and would be susceptible to the IFN-γ response. This is consistent with numerous previous publications that used a range of approaches to show control of *N. caninum* by IFN-γ (Boysen *et al.* 2006; Nishikawa *et al.* 2001a; Nishikawa *et al.* 2001b; Suzuki *et al.* 1988; Yamane *et al.* 2000). *In vivo*, the type I immune response is upregulated in infected tissue, (such as a bovine placenta) producing inflammatory cytokines including IFN-γ. This can be both beneficial to the host, by helping to control proliferation of the parasite, and detrimental due to the pathology it initiates (Innes 2007).

Taken together with the genome sequencing (stop codons), RNA-Seq (very low transcription level) and PCR (confirmation of stop codons across strains) analyses, in addition to a complete lack of peptide evidence for this gene from multiple proteomics experiments, these results give weight to *ROP18's* pseudogenisation in *N. caninum*.

The loss of function of *ROP18* is not the only reduced-virulence phenotype seen in *N. caninum*. The rhoptries alone account for a number of other factors that suggest avirulence. *ROP16* exhibits decreased mRNA expression in *N. caninum* (Reid *et al.* 2012) and has been associated with virulence differences in *T. gondii* strain types (Saeij *et al.* 2007; Taylor *et al.* 2006). This gene is involved with modulation of the proinflammatory host cytokine IL-12 (Butcher *et al.* 2011; Ong, Reese & Boothroyd 2010) by phosphorylation of STAT6 and has been shown to be highly expressed in all virulence-types of *T. gondii* (Bahl *et al.* 2003) despite also causing virulence differences. The *N. caninum* genome also contains a lower number of tandem repeats of *ROP5* (two compared to between four and ten in *T. gondii,* (Reid *et al.* 2012*). ROP5* has been found to be accountable for inherited variation in virulence between *T. gondii* types, albeit with an as yet unclear relationship between copy number and virulence (Behnke *et al.* 2011; Reese & Boothroyd 2011). It is possible that the loss of these virulence traits enabled *N. caninum* to specialize in the vertical transmission route in cattle. *T. gondii* is thought to involve horizontal transmission from mouse-cat very heavily in its population strategy (Boothroyd 2009; Dabritz *et al.* 2007;

Davison, Otter & Trees 1999) whereas it has been suggested for *N. caninum* that horizontal transmission may be of lesser importance to the population and that vertical transmission is the major route (Davison, Otter & Trees 1999; Pare, Thurmond & Hietala 1997). A less virulent parasite would be more likely to escape the host immune response sufficiently long to enable the host to reproduce, when the parasite can then take advantage of the host's pregnancy-induced immunosuppression to cross the placenta and infect the foetus. This allows progression onto a new host, either through the dam giving birth to a live, persistently infected offspring or through ingestion of the aborted foetus by some other potential intermediate or even definitive host. This is in agreement with *in vivo* studies by (Wiengcharoen *et al.* 2011) who found that *T. gondii* infection in a pregnant cow was more likely to result in abortion than infection by *N. caninum*. Experimental studies in viruses have shown that the evolution of vertical transmission can be associated with reduced virulence phenotypes, as demonstrated for Barley Stripe Mosaic Virus (Stewart, Logsdon & Kelley 2005) and similar strategies are apparent in sexually transmitted diseases which also benefit from extended periods of latency (Ewald 1993). Further work would be required to elucidate the precise mechanisms at play and whether these proposed hypotheses were true. IRG proteins exist in the cow but the repertoire is slightly different from those of the mouse, so this in itself could be a factor. The mouse is clearly a host of major importance to *T. gondii,* while it appears it may be less so to *N. caninum*. It may be that functional *ROP18* is not required for successful parasitism of *N. caninum*'s preferred intermediate hosts and hence its loss of function was not disadvantageous. It would have been interesting to carry out a similar experiment in a bovine cell line, but unfortunately, antibodies to bovine IRGs were not available and their development were outside the scope of this thesis.

The fact that *T. gondii ROP18* has been shown to bind a parasite derived substrate (El Hajj *et al.* 2007) and can cause virulence differences that distinguish between all three types of *T. gondii* (Khan *et al.* 2009), rather than just those that can and cannot phosphorylate IRGs indicates that it functions in additional ways not yet identified. It is not simply a matter of expression-level differences as there are a number of SNPs present between *T. gondii* strains too (Saeij *et al.* 2006). This is in contrast to the PCR results of this analysis of *N. caninum* strains, where no SNPs were identified throughout the whole *ROP18* region. However, the *T. gondii* SNPs were identified between strains known to exhibit different virulence types (I, II and III) which are not currently though to be a feature of *N. caninum* population biology (Al-Quassab, Reichel & Ellis 2010). While virulence differences are exhibited by different *N. caninum* isolates; when

compared to *T. gondii* it is likely that they would all resemble type II or III strains in terms of virulence. This potentially less divergent population stratification in *N. caninum* may be one explanation for the lack of SNPs, or there may be other roles for transcripts as yet unidentified that have led to the conservation of this region, despite the stop codons. The fact that *N. caninum* has an expression-suppressing 5'UTR promoter shared with type III *T. gondii*, in addition to a pseudogenised gene, implies that there are other stakes at play and the mutation must have been advantageous to the *N. caninum* in some way for it to have been maintained in the population. The phosphorylation of IRG proteins is probably only one of a number of roles and the elucidation of further targets will improve our understanding of the differences in host cell infection capabilities and preferences between *N. caninum* and *T. gondii*.

# CHAPTER 6: DIFFERENTIAL ANALYSIS OF GENE EXPRESSION DURING TACHYZOITE TO BRADYZOITE CONVERSION

## 6.1 INTRODUCTION

The *N. caninum* life cycle comprises a number of stages (Chapter 1, Figure 1.1) in addition to the tachyzoite that has been analysed so far in this thesis. Bradyzoites are a morphologically similar form of the parasite to the tachyzoite (Dubey, Lindsay & Speer 1998), they are able to infect host cells (Dubey, Lindsay & Speer 1998; Tunev *et al.* 2002) but they exhibit slower growth and different cyst formation, often in response to some form of stress. Safe within their thick-walled cysts, bradyzoites are able to sit quiescent, avoiding the immune system of the host, until possible recrudescence under more favourable circumstances (Buxton, McAllister & Dubey 2002; Hemphill *et al.* 2004; Innes *et al.* 2002; Tunev *et al.* 2002). This recrudescence is the key to endogenous vertical transmission.

### 6.1.1 Bradyzoites and recrudescence of infection

This reactivation of an infection by mobilisation of bradyzoites from their cysts to the bloodstream, where they can cross the placenta in a pregnant animal, is thought to be the major route of transmission of neosporosis in cattle (Anderson *et al.* 1997; Bjorkman *et al.* 1996; Guy *et al.* 2001; Innes *et al.* 2005; Schares *et al.* 1998; Trees & Williams 2005; Williams *et al.* 2009), with reported efficiencies of 78-95% (Davison, Otter & Trees 1999; Pare, Thurmond & Hietala 1997). Exogenous transplacental transmission, where the cow first acquires an infection during pregnancy, also adds to the overall spread of the parasite to successive generations (Innes 2007; Trees & Williams 2005; Williams *et al.* 2009).

After recrudescence of neosporosis, the circulating parasites are thought to be tachyzoites; having reverted back to this more proliferative life stage (Trees & Williams 2005; Williams *et al.* 2009). Tachyzoites are easily maintained in cell culture and as such are readily available for study. This, coupled with their major role in dissemination through the host during an infection, makes them an ideal life-stage for experimentation. Whilst sporozoites would make interesting subjects for study, they are produced by the definitive host in such low numbers (Gondim, Gao & McAllister 2002; McAllister *et al.* 1998) that they are unobtainable for most researchers and tend to be used mainly for *in vivo* studies with cattle, where relatively few

sporozoites (typically hundreds) are required to initiate an infection (Gondim, Gao & McAllister 2002), compared to the many millions that would be required for proteomic and transcriptomic analysis (Methods (Chapters 2.2 and 6.2)). Bradyzoites offer an opportunity to analyze an additional life stage that, whilst not easy to grow in large numbers or at 100% conversion rate, can be maintained in cell culture. The Hemphill group in Bern have worked on optimizing their production and have developed a method using sodium nitroprusside induction to trigger stage conversion ((Vonlaufen *et al.* 2004; Vonlaufen *et al.* 2002) and personal communication with A. Hemphill and T. Monney). By elucidating the differences and similarities in gene expression between bradyzoites and tachyzoites it may eventually be possible to identify crucial genes involved in stage conversion. This in turn would allow drug and vaccine developers to target either tachyzoites preparing to encyst as bradyzoites, or bradyzoites recrudescing and reverting to tachyzoites, and potentially enable arrest of these processes so imperative to the epidemiology of the disease.

### 6.1.2 RNA-Seq

High-throughput 'next generation' sequencing has developed rapidly over recent years and as such, platforms are now available from a number of manufacturers and open to commercial use by sequencing centres around the world. Early studies used this technology to analyse human, yeast and *Arabidopsis* mRNA (messengerRNA) (Kim *et al.* 2007; Lister *et al.* 2008; Nagalakshmi *et al.* 2008; Sultan *et al.* 2008; Wilhelm *et al.* 2008) but since then studies in a wide array of other organisms, including Apicomplexa, have been published ((Mader *et al.* 2011; Otto *et al.* 2010; Sorber, Dimon & DeRisi 2011; Wang, Gerstein & Snyder 2009) and many more).

Compared to microarray or complementary DNA (cDNA)/expressed sequence tag (EST) sequencing, RNA-Seq is cheaper, requires less mRNA and has a high dynamic range for effective quantitation of transcripts, with little background noise (Wang, Gerstein & Snyder 2009). Illumina sequencing, which produces 50-100 base pair reads, is ideal for situations where there is a good quality reference genome available to align the reads to. Firstly, a cDNA library is generated from the mRNA. A flow cell is then used for 'bridge' PCR reactions (between oligos on the flow cell) to generate clusters of DNA; bases are added with fluorescent dye (which is a reversible terminator) to distinguish them from one another and the flow cell is imaged. The dye is then removed (but not the base) and the next base added; and so on (Holt & Jones 2008). The competition for incorporation of bases means high

accuracy as all bases are added at once. Illumina flow cells have now been adapted to enable paired-end sequencing, which aids the bioinformatic challenge of aligning reads to the correct position on the genome (Holt & Jones 2008; Roach *et al.* 1995). Further detail on the different sequencing methods for RNA-Seq is in Chapter 1.8.1.

To date, there has not been extensive research published of *N. caninum* mRNA expression. One study, performed before the advent of RNA-Seq technology or the publication of the *N. caninum* genome sequence, identified eighty-five differentially expressed amplicons (by annealing control primer-based PCR) between tachyzoite and bradyzoite stages (Kang *et al.* 2008a). However, due to the lack of availability of gene and protein function information for *N. caninum*, they were only able to identify putative biological functions for six of these genes based on homology to known genes: GRA2, ribosomal protein S8, a possible apicomplexan specific gene, a putative sporulation protein R gene and cytochromes c and b. Of these, GRA2 and the apicomplexan specific gene were up regulated in tachyzoites with respect to bradyzoites, while for the sporulation gene and cytochromes c and b, the reverse was true. Confusingly, while included in a table of differentially expressed genes, the transcription level for the apicomplexan specific gene was described as "not changed during conversion from tachyzoites to bradyzoites" (Kang *et al.* 2008). More recently, the Wastling group, in collaboration with the Sanger Institute, performed an RNA-Seq analysis of *N. caninum* discussed in more detail in Section 6.2.

## 6.2 PILOT STUDY OF THE *N. CANINUM* TRANSCRIPTOME

The *Neospora caninum* genome sequencing and analysis project (Reid *et al.* 2012) involved an RNA-Seq analysis of the transcriptomes of *N. caninum* parasites from cultures of 2, 3, 4 and 6 days post infection. This pilot study informed the design of the experiment presented in this chapter. Day 3 or 4 tachyzoites in vero cell culture are usually at the stage where they egress from the host cells and as such are at an optimal stage for harvesting for subsequent experimentation. Consequently, Day 6 parasites had been subjected to deteriorating cell culture conditions where a lack of fresh host cells to invade meant that the parasites were under stress, and it was hypothesized that they may be in the early stages of conversion to bradyzoites. Sequencing was performed by the Sanger Institute, Cambridge on parasites cultured by Rebecca Norton, Wastling group, Liverpool, as described in Reid *et al.* (2012). The analysis presented here was performed by Sarah Vermont on data provided by Adam Reid.

### 6.2.1 Analysis of the pilot data

An analysis of the mRNA expression data revealed interesting trends in expression, especially relating to apical genes. At the time this analysis was carried out (2009), there were very few published RNA-Seq studies and little was known about the nature of the relationship between proteome and transcriptome. Whilst quantitative proteomic data for *N. caninum* was not available, the spectrum-count data for the MudPIT experiment in Chapter 2 was used to enable a cursory examination of this relationship and revealed some interesting phenomena. Firstly, both the proteome and the transcriptome exhibited a characteristic curve on a graph of genes ordered by expression level on the x axis plotted against expression level on the y axis; the vast majority of genes were expressed at a baseline level with only a small few responsible for the bulk of the transcripts or peptides (Figure 6.1). This may help to explain why < 30 % of the proteome has so far been identified experimentally (Chapter 2): if the majority of genes are expressed at very low levels only. Many of the highly expressed genes were those with apical-secretory products, such as GRA1, GRA2, MIC2 and MIC10.

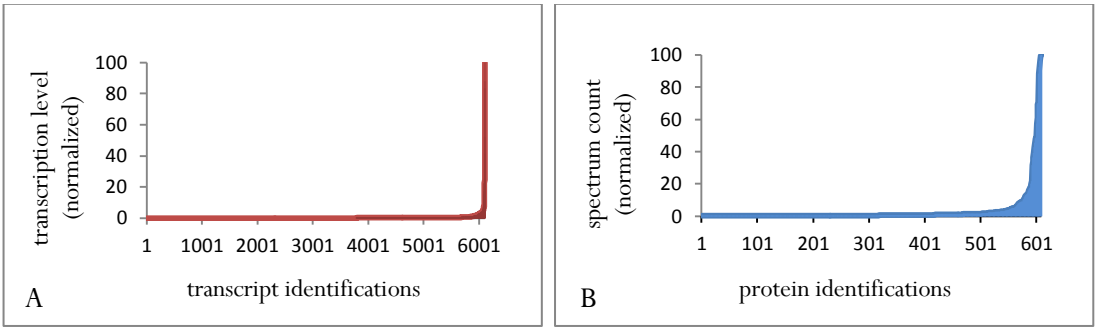



**Figure 6.1:** Normalised expression levels for all genes (x axis) for which some expression was detected, transcriptomically (A) by RNA-Seq or proteomically (B) by MudPIT analysis (Chapter 2). It is clear that a few genes exhibit extremely high expression relative to the rest of the genome or proteome (transcriptome data, Liverpool/Sanger; proteome data, Liverpool/UCLA; analysis, S. Vermont).

Secondly, a number of proteins were identified in the LC MS/MS experiments that were not detected by RNA-Seq (Figure 6.2). This was also seen when proteomic data for *T. gondii* (Xia *et al.* 2008) was compared with the *T. gondii* transcripts, with three proteins common to both proteomes but absent from the transcriptomes (NC_LIV_040580/52.m01559 alpha glucosidase II, putative; NC_LIV_041100/52.m01619 Delta-aminolevulinic acid dehydrate, isoform 3; NC_LIV_105020/59.m03479 Insulysin, putative).

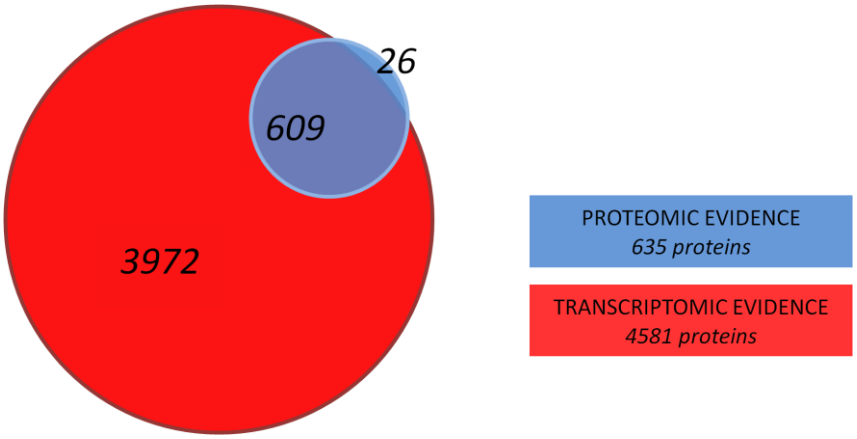



**Figure 6.2:** Venn diagram showing gene expression detected in the transcriptome and the proteome. Peptides were identified for 26 genes that were not detected by RNA-Seq (transcriptome data, Liverpool/Sanger; proteome data, Liverpool/UCLA; analysis, S. Vermont).

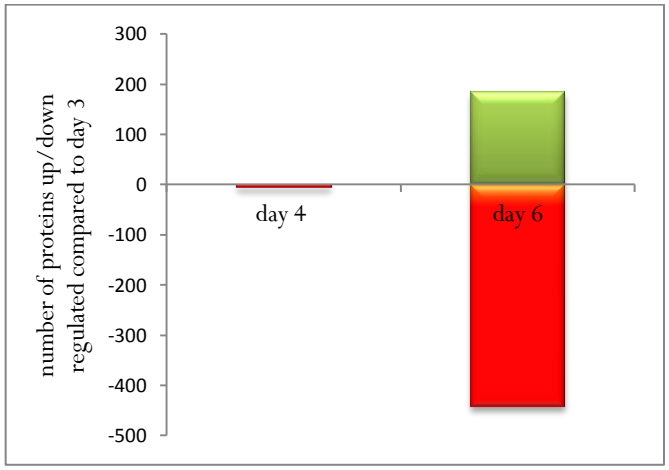



**Figure 6.3:** Comparison of Day 3 cultures to Day 4 and Day 6. Very little difference was observed between Day 3 and 4 tachyzoites, as would be expected, but by Day 6 a number of genes were up or down regulated as identified by DESeq analysis (DESeq analysis, A. Reid; data analysis, S. Vermont).

Figure 6.3 shows that there was minimal difference in expression (n=7) between Day 3 and Day 4 parasites (tachyzoites at optimal time for egress) but by Day 6 (stressed tachyzoites, possibly beginning stage conversion) many more genes were differentially expressed (n=626). The changes in expression of rhoptry, dense granule and micronemal genes, in Day 3 cultures compared to Day 6, can be seen in Figure 6.4. Most striking is the reduction in expression levels of rhoptry genes, suggesting that rhoptries may play a lesser role in bradyzoites than tachyzoites. This would support the notion that bradyzoites are less primed for invasion than tachyzoites and that their specialism lies in different areas. However, as the Day 6 cultures cannot be considered true bradyzoites, this is purely speculation, and RNA-Seq analysis of

improved bradyzoite cultures would enable further investigation of these hypotheses (Section 6.2.2). As might be expected, surface antigen expression levels indicated that a different repertoire were expressed by the Day 6 parasites than the Day 3/4 and among those identified as being up-regulated by Day 6 was the bradyzoite antigen *SAG4* (Odberg-Ferragut *et al.* 1996; Weiss & Kim 2000).

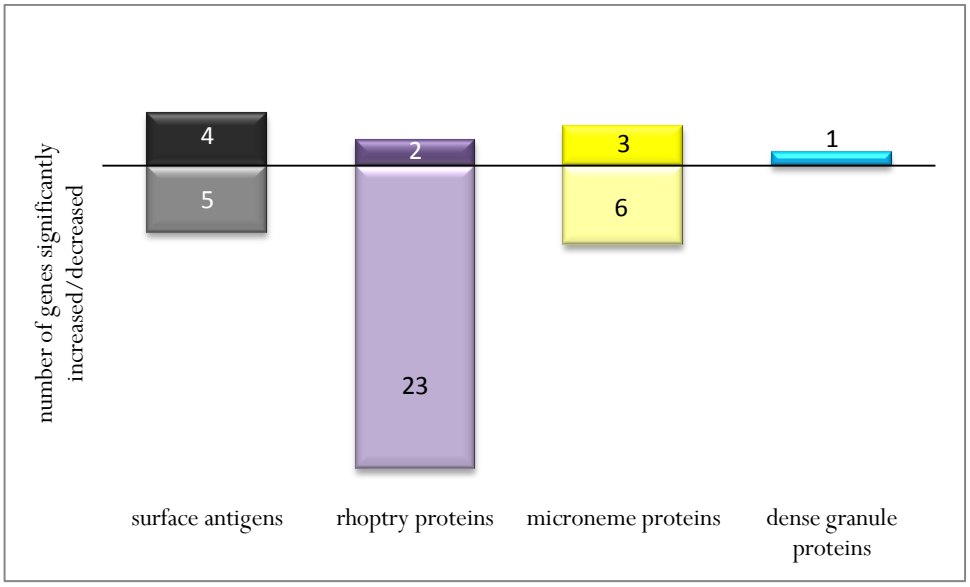



**Figure 6.4:** Changes in mRNA expression of key genes when comparing Day 6 *N. caninum* cultures to Day 3. Most notably, 23 rhoptry genes were down regulated in the putatively bradyzoite-like Day 6 cultures. Differential expression analysis was performed using DESeq (Transcriptome data Liverpool/Sanger; DESeq analysis A. Reid; data analysis S. Vermont).

### 6.2.2 Objectives for the bradyzoite/tachyzoite RNA-Seq analysis

The aim of this experiment was to understand the genes involved in tachyzoite to bradyzoite stage conversion. It was hypothesized that the two life stages would have different protein (and hence mRNA) requirements due to their different phenotypes. A particular area of interest was the apical genes known to be involved in the invasion process, as the hypothesis was that the bradyzoite stage parasites would be less specialized for host cell invasion than the tachyzoite stage. The data from the pilot RNA-Seq experiment suggested that this was the case, and so led to the design of a comprehensive RNA-Seq experiment analysing sodium nitroprusside-induced bradyzoites in addition to tachyzoites. This was achieved by an investigation of mRNA expression-level changes between tachyzoite and bradyzoite cultures, using Illumina Hi-Seq2000 next-generation sequencing technology, in collaboration with the Hemphill group, Bern.

## 6.3 MATERIALS AND METHODS

### 6.3.1 Sample collection

Cell culture was carried out with Nc Liverpool tachyzoites in vero cells by the Hemphill group, University of Bern (Vonlaufen *et al.* 2004). Stage conversion was induced by the addition of sodium nitroprusside to the cultures at a final concentration of 50 µM on Day 0. Samples were then collected as detailed in Table 6.1: two biological repeats of each sample were prepared in order to enable statistical analysis of expression level.

**Table 6.1:** Tachyzoite/bradyzoite RNA-Seq samples. Day 0 refers the time at which the tachyzoites were exposed to sodium nitroprusside.

| Sample | Description | Biological Repeats |
|---|---|---|
| **Day 0 (a and b)** | Tachyzoites, no sodium nitroprusside induction | 2 |
| **Day 1 (a and b)** | Tachyzoites after 1 day sodium nitroprusside (may be preparing to stage convert) | 2 |
| **Day 3 (a and b)** | Tachyzoites after 3 days sodium nitroprusside (may be preparing to stage convert) | 2 |
| **Day 6 (a and b)** | Tachyzoites/bradyzoites after 6 days sodium nitroprusside | 2 |

Parasites were purified on PD-10 columns (Sigma-Aldrich) on ice and counted using a Neubauer chamber. Trizol reagent (Invitrogen) was immediately added to the parasite suspensions to 0.5-1.0 x $10^7$ parasites/ml and the lysates were frozen at -80°C and shipped to Liverpool for subsequent procedures. Extraction of total RNA, isolation of mRNA, and quantitation by Qubit Assay (Invitrogen) were carried out by the author as were the eventual transcriptomic data analyses.

### 6.3.2 Total RNA extraction by hybrid chloroform/RNeasy method

Total RNA was isolated from tachyzoite/bradyzoite Trizol suspensions by a hybrid protocol which comprised incubation of the parasite/Trizol homogenate with 0.2 ml chloroform per 1 ml for 2-3 minutes (vigorously shaken for the first 15 seconds) followed by centrifugation at 12 000 x g for 15 minutes at 4 °C. The aqueous phase was then carefully transferred to a new microcentrifuge tube and an equal volume of 70 % (v/v) EtOH added (this was in the region of 500-700 µl per 1 ml starting material). Subsequent steps were taken from the Qiagen RNeasy protocol as follows: the samples were loaded onto an RNeasy column and centrifuged at 8000 x g for 30 seconds. The flow-through was discarded, and the column was washed in

buffers RW1 (700 µl, 30 second centrifugation at 8 000 x g) and RPE (500µl, 30 second centrifugation at 8000 x g) before being subjected to a 2 minute centrifugation at 8000 x g in a new collection tube to remove any remaining buffer. Finally, the column was transferred to another collection tube and the RNA eluted in 30 µl RNase-free water, by incubation at room temperature for 1 minute followed by centrifugation at 8000 x g for 1 minute. Samples were then stored at -80 °C for subsequent isolation of Poly(A) RNA and quality control (Section 6.3.4)

### 6.3.3 Isolation of mRNA by Poly(A) selection

Total RNA samples were diluted to 250 µl with RNase-free water for mRNA selection by a microPoly(A) purist kit (Invitrogen). All subsequent steps are taken from the kit protocol: Samples were mixed with an equal volume of 2x binding solution, bound to oligo(dT) cellulose, heated to 70 °C and then incubated (with rocking) for 1 hour at room temperature. The oligo(dT) cellulose was pelleted at 4 000 x g for 3 minutes at room temperature, washed twice with each of wash solutions 1 and 2, then the mRNA eluted in 200 µl RNA storage solution preheated to 70 °C, by centrifugation at 5 000 x g for 2 minutes. The eluted poly(A) RNA was precipitated overnight at -20°C by addition of 20 µl 5 M ammonium acetate, 1 µl glycogen and 550 µl 100 % (v/v) EtOH. It was recovered by 30 minutes of centrifugation at 12 000 x g at 4 °C and resuspended in RNA storage solution (included in the kit) before being quantitated on a Qubit Fluorometer (Invitrogen) and stored at -80 °C.

### 6.3.4 Quality Control

Samples were submitted to the Centre for Genomic Research (CGR), Liverpool for assessment on a Bioanalyzer chip (Agilent). Total RNA traces were examined to determine the intactness of RNA and Poly(A) RNA was evaluated to ensure that sufficient depletion of ribosomal RNA had occurred so as not to hinder analysis of mRNA. An example of a satisfactory total RNA (A) and Poly(A)-selected RNA (B) sample can be seen in Figure 6.5.

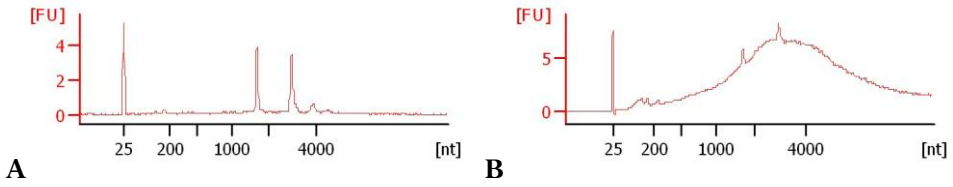



**Figure 6.5:** Bioanalyzer traces of representative samples which passed quality control. **A** shows total RNA, with ribosomal peaks clearly visible (between 1 000 and 4 000 nt); while **B** shows RNA after Poly(A) selection, ribosomal peaks can still be detected but they have been sufficiently depleted so as not to mask mRNA. Nucleotides are on the x axis with fluorescence units on the y axis. Assessment of traces for quality control was based purely upon the appearance of the curve, quantitation of RNA was measured separately by Qubit Assay (Invitrogen).

### 6.3.5 Sequencing

Library construction, amplification and sequencing took place at the Liverpool Centre for Genomic Research. Strand-specific RNA Seq libraries were prepared from eight enriched PolyA-RNA samples, then were sequenced on the Illumina HiSeq2000 platform in one lane aiming to generate in excess of 20 million paired-end reads per sample. Mapping was performed to keep only the uniquely mapped reads, this is so that a read aligning to more than one location, due to similarity of the sequences would be discarded. Therefore, RNA belonging to gene families would potentially have been excluded from the analysis, but expression spikes resulting from such a situation would have been avoided. The mapping was performed using TopHat 1.3.2 which also gave FPKM (fragments per kilobase of exon per million fragments mapped) values as an output using Cufflinks; and the raw counts were generated using htseq-count script (Python).

### 6.3.6 Data analyses

In order to assess the similarity of biological repeats, a Spearman's Rank Correlation was carried out. This test is suitable for non-parametric data and is the equivalent of the Pearson Correlation Coefficient, but carried out on ranked data (Myers & Well 2003). A Correlation Coefficient of 1 indicates perfect positive correlation while a coefficient of -1 indicates perfect negative correlation (Dytham 2003).

Raw count values were analysed in a pair wise fashion using DESeq in R. DESeq allows the identification of differentially expressed genes using a method based upon the negative binomial distribution (Anders & Huber 2010). It generates 'padj' p-values which have been subjected to the Benjamini Hochberg-correction for multiple testing.

The software package GProX (Rigbolt, Vanselow & Blagoev 2011) was used to generate cluster analyses, using a fuzzification value of 2 for 100 iterations. Apical genes were identified as detailed in Chapter 3, bradyzoite-associated genes were identified by literature searching, and by text-mining ToxoDB (Gajria *et al.* 2008).

Genes involved in *N. caninum* metabolic pathways were downloaded from the Liverpool Library of Apicomplexan Metabolic Pathways (LAMP) (Achuchuthan Shanmugasundram *et al.* unpublished) and searched against differential expression data generated by the DESeq analysis.

## 6.4 RESULTS

### 6.4.1 Sample preparation and sequencing

All samples submitted for sequencing achieved the quality control checks made by CGR and yielded reads as can be seen in Table 6.2. The percentage of reads mapped to the *N. caninum* genome is in the range 27-54 %: the remaining reads are likely to be contamination from vero cells. Nevertheless, over 90 % of all *N. caninum* genes exhibited a detectable level of mRNA expression across all eight samples, so there appeared to have been a sufficient depth of coverage of the genome. The ability to map reads specifically to a reference genome meant that host cell contamination was not a problem, as vero cell reads did not map and were therefore not included in the expression analysis. Conversion to bradyzoites in the cultures used to generate the RNA was progressive over the time course from 0 to approximately 50 %, as estimated by expression of the bradyzoite antigen *BAG1* (Hemphill group, personal communication).

**Table 6.2:** Summary of reads and mapping to the *Neospora caninum* genome. Day 0 (D0) refers to samples collected prior to sodium nitroprusside induction (tachyzoites), Day 1 and D3 (D1, D3) to parasites after one/three day(s) of sodium nitroprusside induction, respectively, and Day 6 (D6) to bradyzoite-like parasites after six days of sodium nitroprusside induction.

| Sample | Number of *N. canin*um genes | Number of transcripts mapped to genome | % of transcripts mapped to genome |
|---|---|---|---|
| **D0a** | 7227 | 7098 | 98.22 |
| **D0b** | 7227 | 7076 | 97.91 |
| **D1a** | 7227 | 6615 | 91.53 |
| **D1b** | 7227 | 6527 | 90.31 |
| **D3a** | 7227 | 6536 | 90.44 |
| **D3b** | 7227 | 6621 | 91.61 |
| **D6a** | 7227 | 6608 | 91.43 |
| **D6b** | 7227 | 6701 | 92.72 |

### 6.4.2 Spearman's Rank correlation of biological replicates

All samples were highly positively correlated with their pair for that time point, so for further analyses (such as clustering), a mean of the each two biological repeats was used to represent

the time point, as the biological repeats appear to be sufficiently similar to one another to allow this. The coefficients and p-values are displayed in Table 6.3.

**Table 6.3:** Spearman's Rank correlation of biological repeats

| Samples | Spearman's Rank Correlation Coefficient | p-value |
|---|---|---|
| Day 0a and Day 0b | 0.968 | < 0.000 |
| Day 1a and Day 1b | 0.922 | < 0.000 |
| Day 3a and Day 3b | 0.847 | < 0.000 |
| Day 6a and Day 6b | 0.908 | < 0.000 |

### 6.4.3 Differential expression

The data were analyzed using DESeq (Anders & Huber 2010) to identify significantly differently expressed genes, in a pair-wise fashion between the time points. A summary of the results, paying particular attention to apical and bradyzoite genes, can be seen in Table 6.4. Both up and down regulation of genes were identified between most samples.

As would be expected, the largest difference was apparent between Day 0 and Day 6 (99 genes differentially expressed, which corresponded to 1.4 % of the genome). The time points between which most of the differences occur coincided with when sodium nitroprusside was added to the culture (to induce stress and as such stage-conversion to bradyzoites), between Day 0 and Day 1. Pair wise comparisons which did not include the non-sodium nitroprusside-induced time point, Day 0, have noticeably fewer differentially expressed genes: there were 72 (1 % of the genome) significantly differentially expressed genes between day 0 and day 1 samples, compared to only 5 (0.7 %), 1 (0.01 %) and 11 (0.15 %) between day 1/day3, day 3/day6 and day 1/day 6 respectively.

Both of the differentially expressed genes which have been linked to the bradyzoite life cycle stage (bradyzoite surface antigen (NCLIV_019580) and Hypothetical/DnaK-TPR (NCLIV_022840)), were up regulated in the later time points when compared to earlier time points (within the pair wise comparisons). No dense granule genes were significantly differentially expressed, while microneme genes were both up and down regulated over the time course.

The trend for rhoptry gene expression to decrease as the parasites prepare to stage convert agreed with data from the previous Reid *et al.* (2012) transcriptomics experiment; ROP40

(NCLIV_012920), ROP14 (NCLIV_057960) and Toxofilin (NCLIV_051340) were down regulated in both experiments. Furthermore, the bradyzoite surface antigen NCLIV_019580 was identified in both experiments as down regulated in later time points also.

**Table 6.4:** Summary of differential expression from DESeq analysis. P-value ≤ 0.01 (padJ corrected P-value). Samples were analysed in a pair wise manner (Day x vs. Day y) where up regulation (green text) refers to higher expression in Day y when compared to Day x and down regulation the reverse (red text).

| | Day 0 vs. Day 1 | Day 0 vs. Day 3 | Day 0 vs. Day 6 | Day 1 vs. Day 3 | Day 1 vs. Day 6 | Day 3 vs. Day 6 |
|---|---|---|---|---|---|---|
| **No. genes differentially expressed** | 72 | 88 | 99 | 5 | 11 | 1 |
| **No. up regulated** [1] | 31 | 43 | 67 | 5 | 11 | 1 |
| **No. down regulated** [2] | 41 | 45 | 32 | 0 | 0 | 0 |
| **D.E. (differentially expressed) rhoptry genes** | ROP40 (NCLIV_012920); ROP37 (NCLIV_001460); Lipase maturation factor2 (NCLIV_057950); Hypothetical/ROP14 (NCLIV_057960); Toxofilin (NCLIV_051340) | Zgc:55863 (NCLIV_017420); Toxofilin (NCLIV_051340); ROP40 (NCLIV_012920) | ROP40 (NCLIV_012920) | 0 | 0 | 0 |
| **D.E. dense granule genes** | 0 | 0 | 0 | 0 | 0 | 0 |
| **D.E. microneme genes** | NcMCP7 (NCLIV_054425) | MIC3 (NCLIV_010600); Pan domain containing protein (NCLIV_068830) | MIC3 (NCLIV_010600); MIC4 (NCLIV_002940); Pan domain containing protein (NCLIV_068830); NcMCP3 (NCLIV_003260); NcMCP4 (NCLIV_003250) | NcMCP3 (NCLIV_003260); | NcMCP3 (NCLIV_003260); NcMCP4 (NCLIV_003250) | 0 |
| **D.E. bradyzoite genes** | bradyzoite surface (NCLIV_019580) | bradyzoite surface (NCLIV_019580) | bradyzoite surface (NCLIV_019580); Hypothetical/DnaK-TPR (NCLIV_022840) | Hypothetical/DnaK-TPR (NCLIV_022840) | 0 | 0 |

[1] positive log2 fold change; [2] negative log2 fold change

### 6.4.4 Cluster Analyses

Cluster analyses were performed on fragments per kilobase of exon per million fragments mapped (FPKMs) values using GProX (Rigbolt, Vanselow & Blagoev 2011). A fragment refers to a pair of reads, as generated by paired-end sequencing. Clustering RNA-Seq data is a complicated area that needs further development within the bioinformatic community. So, while the following figures can be helpful to view the general trends in the data, they should be used with caution as the algorithms have not yet been designed with analysing time-course RNA-Seq data in mind. As such, only sub sets of genes from the apical organelles (n = 22, 58 and 42 for dense granules, rhoptries and micronemes respectively) and those relating to bradyzoites (n = 28) have been clustered, simply to enable a visualisation of the way in which expression levels are changing over the time course, as opposed to the pair-wise manner of the DESeq analysis.

It is important to note that changes in expression apparent in the cluster analysis may not have been identified as significant differences in expression in the DESeq analysis, but rather *all* genes over the range of expression levels have been included.

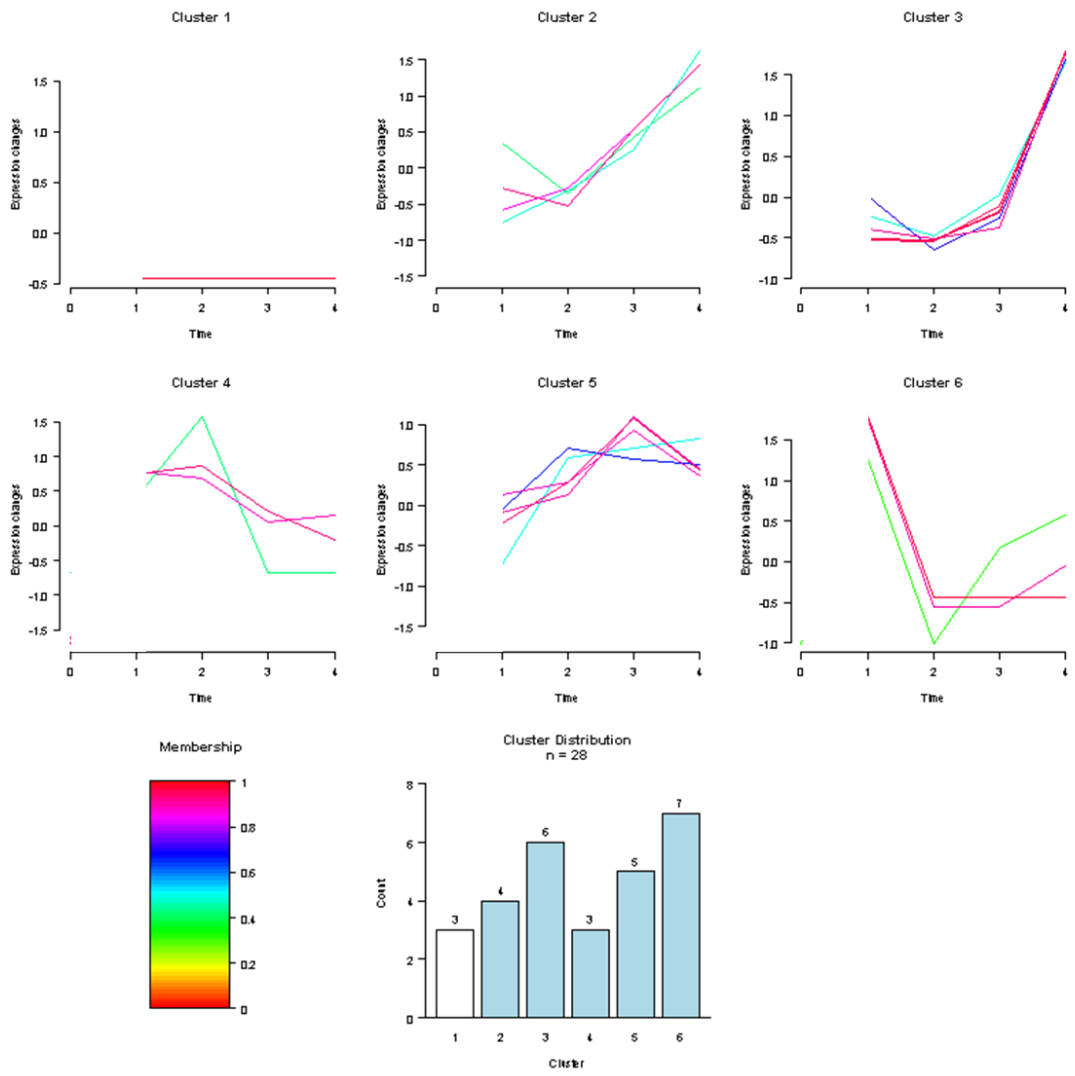



**Figure 6.6:** Bradyzoite-associated genes clustered over the time course. Biological repeats were averaged to produce a mean FPKM for each time point. N = 28, time points were as follows: 1=Day 0, 2=Day 1, 3=Day 3, 4=Day 6.

Figure 6.6 and Table 6.5 show that the majority of bradyzoite-associated genes identified in the analysis showed an upward trend in expression over the time course (clusters 2, 3 and 5), which supports the conversion procedure carried out at Bern. There were genes in clusters 4 and 6 which appear to decrease in expression, but these were not identified as significantly differently expressed in the DESeq analysis. An FPKM expression level which constitutes a biologically relevant cut-off point for 'real' expression of genes is unknown and as such it is difficult to discern how much variation in expression would be required to impact on the biology of the organism. Cluster 1 shows three genes for which no transcripts were detected at any time point.

**Table 6.5:** Membership of bradyzoite-associated gene clusters, with mean FPKM values for the two biological replicates. Green text indicates an upward overall trend in expression for that cluster over the time course.

| I.D. | Description | FPKM Day 0 | FPKM Day 1 | FPKM Day 3 | FPKM Day 6 | Cluster |
|---|---|---|---|---|---|---|
| NCLIV_067920 | SRS domain-containing protein | 0.0 | 0.0 | 0.0 | 0.0 | 1 |
| NCLIV_067930 | SRS domain-containing protein | 0.0 | 0.0 | 0.0 | 0.0 | 1 |
| NCLIV_067940 | SRS domain-containing protein | 0.0 | 0.0 | 0.0 | 0.0 | 1 |
| NCLIV_007770 | Rhoptry kinase family protein, truncated (incomplete catalytic triad), putative | 22.3 | 16.1 | 42.8 | 65.6 | 2 |
| NCLIV_019580 | SRS domain-containing protein,bradyzoite surface antigen, putative | 0.5 | 5.3 | 11.6 | 26.2 | 2 |
| NCLIV_036410 | cyst matrix protein, putative | 75.4 | 121.1 | 237.1 | 364.6 | 2 |
| NCLIV_056110 | small heat shock protein 21, putative | 40.0 | 25.1 | 41.3 | 56.2 | 2 |
| NCLIV_010810 | deoxyribose-phosphate aldolase, putative | 0.1 | 0.0 | 0.1 | 1.9 | 3 |
| NCLIV_022240 | ATPase, related | 2.1 | 1.4 | 2.8 | 7.1 | 3 |
| NCLIV_022840 | hypothetical protein, conserved | 0.1 | 0.1 | 0.7 | 3.6 | 3 |
| NCLIV_027470 | bradyzoite antigen, putative | 0.0 | 0.0 | 4.9 | 34.0 | 3 |
| NCLIV_037490 | Enolase (EC 4.2.1.11), related | 6.3 | 1.2 | 4.2 | 19.2 | 3 |
| NCLIV_042910 | Malate dehydrogenase (EC 1.1.1.37), related | 0.3 | 0.2 | 1.9 | 11.0 | 3 |
| NCLIV_012510 | hypothetical protein | 150.3 | 144.8 | 106.0 | 112.7 | 4 |
| NCLIV_032780 | small heat shock protein 20, putative | 465.1 | 490.2 | 361.1 | 277.2 | 4 |
| NCLIV_067960 | SRS domain-containing protein | 0.1 | 0.2 | 0.0 | 0.0 | 4 |
| NCLIV_011120 | Malate dehydrogenase (EC 1.1.1.37), related | 38.8 | 56.7 | 53.6 | 52.0 | 5 |
| NCLIV_019110 | HSP90-like protein, related | 105.2 | 120.0 | 187.4 | 142.5 | 5 |
| NCLIV_032330 | Malate dehydrogenase (NAD) (EC 1.1.1.37) (Precursor), related | 61.0 | 66.2 | 87.2 | 68.6 | 5 |
| NCLIV_037500 | Enolase (EC 4.2.1.11), related | 5.5 | 16.5 | 17.5 | 18.5 | 5 |
| NCLIV_040880 | HSP90, related | 178.1 | 243.0 | 348.7 | 264.1 | 5 |
| NCLIV_009990 | SRS domain-containing protein | 0.0 | 0.0 | 0.0 | 0.0 | 6 |
| NCLIV_010030 | SRS domain-containing protein,Bradyzoite surface protein BSR4 | 0.4 | 0.0 | 0.2 | 0.3 | 6 |
| NCLIV_010060 | SRS domain-containing protein | 0.1 | 0.0 | 0.0 | 0.0 | 6 |
| NCLIV_067890 | SRS domain-containing protein | 0.2 | 0.0 | 0.0 | 0.0 | 6 |
| NCLIV_067900 | SRS domain-containing protein | 0.0 | 0.0 | 0.0 | 0.0 | 6 |
| NCLIV_067910 | SRS domain-containing protein | 0.2 | 0.0 | 0.0 | 0.0 | 6 |
| NCLIV_067950 | SRS domain-containing protein | 0.3 | 0.0 | 0.0 | 0.1 | 6 |

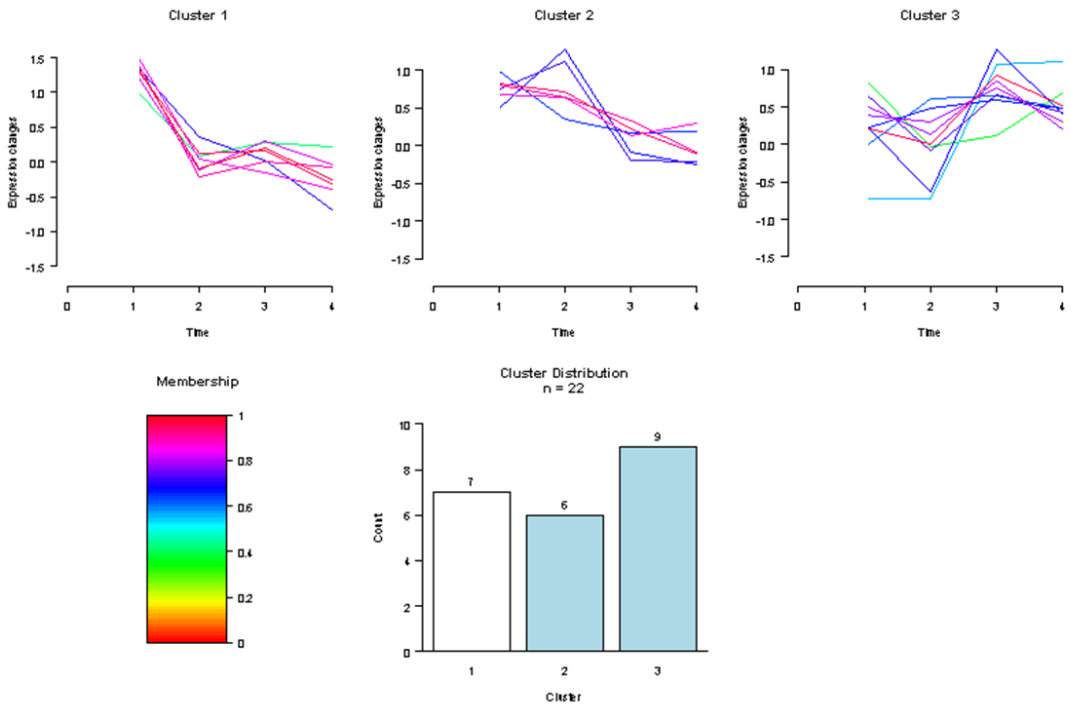



**Figure 6.7**: Dense granule-associated genes clustered over the time course. Biological repeats were averaged to produce a mean FPKM for each time point. N = 22, time points were as follows: 1=Day 0, 2=Day 1, 3=Day 3, 4=Day 6.

In Figure 6.7 and Table 6.6, dense granule genes in clusters 1 and 2 appear to decrease in expression over time, while cluster 3 contains some genes which increase as well as those which do not appear to change their expression levels. No dense granule genes were identified as having significantly different expression in the DESeq analysis, but the slight downward trend in expression seen in clusters 1 and 2 could support the hypothesis that invasion-related organelles play less of a role in bradyzoites than tachyzoites.

**Table 6.6:** Membership of dense granule-associated gene clusters, with mean FPKM values for the two biological replicates. Red text indicates an overall downward trend in expression for that cluster over the time course.

| I.D. | Description | FPKM Day 0 | FPKM Day 1 | FPKM Day 3 | FPKM Day 6 | Cluster |
|---|---|---|---|---|---|---|
| NCLIV_068460 | hypothetical protein | 270.5 | 171.9 | 193.1 | 187.3 | 1 |
| NCLIV_021640 | dense granule protein 7, putative | 996.6 | 415.5 | 344.5 | 260.1 | 1 |
| NCLIV_036400 | dense granule protein 1 / major antigenp24, putative | 8284.5 | 4002.4 | 5286.7 | 4279.1 | 1 |
| NCLIV_045650 | 28 kDa antigen, putative | 5418.8 | 2397.9 | 2968.7 | 2088.2 | 1 |
| NCLIV_045870 | Putative dense granule protein 3 | 367.4 | 197.5 | 204.3 | 141.2 | 1 |
| NCLIV_052880 | granule antigen protein GRA6, putative | 1035.4 | 384.2 | 468.7 | 437.3 | 1 |
| NCLIV_068400 | NTPase | 2.1 | 1.2 | 0.9 | 0.3 | 1 |
| NCLIV_003100 | serine proteinase inhibitor TgPI-2, putative | 81.6 | 80.1 | 63.3 | 68.9 | 2 |
| NCLIV_003120 | serine proteinase inhibitor, putative | 118.0 | 137.2 | 68.0 | 66.9 | 2 |
| NCLIV_005560 | dense-granule antigen DG32, putative | 78.8 | 60.0 | 54.3 | 55.0 | 2 |
| NCLIV_008990 | hypothetical protein | 75.5 | 72.0 | 57.3 | 47.4 | 2 |
| NCLIV_067130 | hypothetical protein | 2.6 | 3.6 | 1.8 | 1.6 | 2 |
| NCLIV_036630 | 14-3-3 protein, putative | 5.2 | 4.8 | 4.2 | 3.3 | 2 |
| NCLIV_014150 | hypothetical protein | 56.8 | 65.1 | 68.0 | 65.0 | 3 |
| NCLIV_016360 | hypothetical protein, conserved | 8.4 | 5.7 | 8.3 | 7.5 | 3 |
| NCLIV_024820 | 14-3-3 protein homolog | 137.0 | 189.1 | 191.6 | 178.5 | 3 |
| NCLIV_036930 | 14-3-3 protein, putative | 0.0 | 0.0 | 0.2 | 0.2 | 3 |
| NCLIV_037450 | dense granular protein GRA10, putative | 0.5 | 0.2 | 0.8 | 0.5 | 3 |
| NCLIV_041120 | hypothetical protein, conserved | 137.7 | 131.2 | 160.9 | 130.7 | 3 |
| NCLIV_045450 | 14-3-3-like protein, related | 5.9 | 3.8 | 4.2 | 5.5 | 3 |
| NCLIV_054830 | hypothetical protein | 30.8 | 27.2 | 42.3 | 35.6 | 3 |
| NCLIV_066630 | GRA9 protein, putative | 32.5 | 27.0 | 37.2 | 27.8 | 3 |

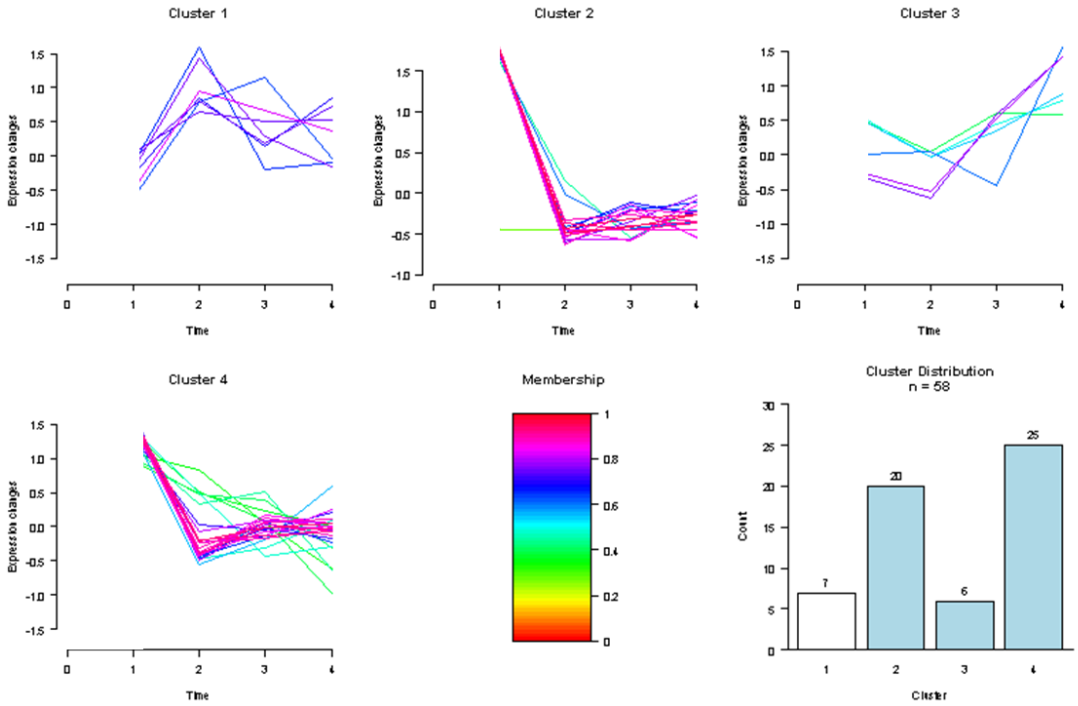



**Figure 6.8:** Rhoptry-associated genes clustered over the time course. Biological repeats were averaged to produce a mean FPKM for each time point. N = 58, time points were as follows: 1=Day 0, 2=Day 1, 3=Day 3, 4=Day 6.

In Figure 6.8 and Table 6.7, cluster 1 shows rhoptry genes which appear to increase in expression between day 0 and day 1, but then decrease. Clusters 2 and 4 show a decrease in expression over time; while cluster 3 shows six genes whose expression is highest at the most bradyzoite-like time point, day 6. The vast majority of rhoptry genes, 45 out of 58, belong to clusters 2 and 4, suggesting that in general rhoptry gene expression decreases as the parasites stage-convert to the bradyzoite form. This is in agreement with the DESeq analysis above.

**Table 6.7:** Membership of rhoptry-associated gene clusters, with mean FPKM values for the two biological replicates. Red text indicates an overall downward trend in expression for that cluster over the time course, green an increase and blue text indicates an increase in expression from Day 0 to Day 1 followed by a decrease over the remainder of the time course.

| I.D. | Description | FPKM Day 0 | FPKM Day 1 | FPKM Day 3 | FPKM Day 6 | Cluster |
|---|---|---|---|---|---|---|
| NCLIV_022130 | Rhoptry kinase family protein ROP25, putative | 0.4 | 0.7 | 0.5 | 0.7 | 1 |
| NCLIV_023260 | Rhoptry kinase family protein ROP33, putative | 2.0 | 4.9 | 2.9 | 2.1 | 1 |
| NCLIV_023580 | Rhoptry kinase family protein ROP45 (incomplete catalytic triad), putative | 0.0 | 0.1 | 0.1 | 0.0 | 1 |
| NCLIV_024700 | Rhoptry kinase family protein ROP21, putative | 1.9 | 2.5 | 2.4 | 2.4 | 1 |
| NCLIV_025120 | Rhoptry kinase family protein ROP16, putative | 0.4 | 1.1 | 0.4 | 0.4 | 1 |
| NCLIV_056620 | Rhoptry kinase family protein ROP27, putative | 3.7 | 5.6 | 4.2 | 5.4 | 1 |
| NCLIV_069550 | Cathepsin B-like protease (Precursor), related | 3.0 | 7.3 | 6.4 | 5.6 | 1 |
| NCLIV_001460 | Rhoptry kinase family protein ROP37 (incomplete catalytic triad), putative | 5.1 | 0.0 | 0.3 | 0.4 | 2 |
| NCLIV_002580 | Rhoptry kinase family protein ROP36 (incomplete catalytic triad), putative | 0.2 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_004220 | Rhoptry antigen ROP8 (EC 2.7.11.26), related | 0.3 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_004230 | hypothetical protein | 0.0 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_012920 | Rhoptry kinase family protein ROP40 (incomplete catalytic triad), putative | 20.8 | 1.2 | 1.7 | 3.2 | 2 |
| NCLIV_017410 | hypothetical protein | 0.0 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_017420 | Zgc:55863, related | 1.4 | 0.0 | 0.3 | 0.3 | 2 |
| NCLIV_017430 | hypothetical protein | 0.0 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_017440 | hypothetical protein | 1.1 | 0.1 | 0.2 | 0.3 | 2 |
| NCLIV_027710 | Rhoptry kinase family protein ROP31, putative | 0.0 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_030990 | Rhoptry kinase family protein ROP46, putative | 0.8 | 0.0 | 0.1 | 0.0 | 2 |
| NCLIV_040110 | hypothetical protein, conserved | 0.1 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_045580 | hypothetical protein, conserved | 12.9 | 4.5 | 2.4 | 2.8 | 2 |
| NCLIV_048060 | Rhoptry kinase family protein ROP41, putative | 5.5 | 2.4 | 0.9 | 1.6 | 2 |
| NCLIV_051340 | toxofilin, putative | 260.6 | 33.6 | 44.8 | 49.9 | 2 |
| NCLIV_057960 | hypothetical protein | 17.4 | 2.9 | 5.2 | 4.1 | 2 |
| NCLIV_060730 | hypothetical protein | 11.5 | 2.1 | 2.4 | 2.0 | 2 |
| NCLIV_060740 | ROP 2, related | 5.2 | 0.5 | 0.3 | 0.5 | 2 |
| NCLIV_068850 | hypothetical protein | 43.5 | 8.4 | 12.7 | 11.8 | 2 |
| NCLIV_068890 | hypothetical protein | 5.8 | 0.9 | 1.5 | 1.7 | 2 |
| NCLIV_001950 | Rhoptry protein ROP7 (EC 2.7.11.12), related | 0.6 | 0.4 | 1.4 | 2.0 | 3 |
| NCLIV_002650 | Rhoptry kinase family protein ROP22 (incomplete catalytic triad), putative | 0.4 | 0.4 | 0.3 | 1.0 | 3 |
| NCLIV_007770 | Rhoptry kinase family protein, truncated (incomplete catalytic triad), putative | 22.3 | 16.1 | 42.8 | 65.6 | 3 |
| NCLIV_011690 | ROP15 protein, related | 18.6 | 13.9 | 17.9 | 20.7 | 3 |
| NCLIV_058560 | rhoptry protein, putative | 0.8 | 0.6 | 0.8 | 1.0 | 3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| NCLIV_065640 | Rhoptry kinase family protein, truncated (incomplete catalytic triad), putative | 37.4 | 30.1 | 39.4 | 39.2 | 3 |
| NCLIV_000650 | Rhoptry kinase family protein ROP34, putative | 2.9 | 2.3 | 2.2 | 1.4 | 4 |
| NCLIV_001400 | hypothetical protein, conserved | 1.2 | 0.4 | 0.4 | 0.5 | 4 |
| NCLIV_001970 | hypothetical protein | 158.4 | 33.8 | 52.4 | 56.1 | 4 |
| NCLIV_007800 | Tg65, related | 10.1 | 4.0 | 4.7 | 4.9 | 4 |
| NCLIV_011730 | Rhoptry kinase family protein ROP26 (incomplete catalytic triad), putative | 20.3 | 5.6 | 9.8 | 8.8 | 4 |
| NCLIV_016220 | Rhoptry kinase family protein ROP23 (incomplete catalytic triad), putative | 5.8 | 3.2 | 3.0 | 3.6 | 4 |
| NCLIV_018420 | p36 protein, putative | 164.2 | 58.9 | 62.2 | 66.7 | 4 |
| NCLIV_020320 | Proteophosphoglycan ppg4, related | 0.2 | 0.1 | 0.1 | 0.0 | 4 |
| NCLIV_021100 | hypothetical protein, conserved | 29.8 | 11.6 | 12.9 | 16.9 | 4 |
| NCLIV_027850 | rhoptry protein 6, putative | 79.8 | 22.0 | 32.9 | 32.5 | 4 |
| NCLIV_027930 | Rhoptry kinase family protein ROP17, putative | 12.0 | 3.4 | 5.3 | 4.5 | 4 |
| NCLIV_028130 | Rhoptry kinase family protein ROP28, putative | 1.2 | 0.8 | 0.9 | 0.4 | 4 |
| NCLIV_028170 | Rhoptry kinase family protein ROP20, putative | 17.6 | 3.6 | 7.2 | 5.2 | 4 |
| NCLIV_030050 | hypothetical protein, conserved | 31.6 | 11.6 | 15.9 | 16.4 | 4 |
| NCLIV_035860 | Rhoptry kinase family protein ROP32, putative | 7.3 | 4.7 | 2.1 | 2.5 | 4 |
| NCLIV_044410 | Rhoptry kinase family protein ROP35, putative | 26.9 | 22.2 | 19.4 | 17.2 | 4 |
| NCLIV_046000 | Rhoptry kinase family protein ROP30, putative | 0.6 | 0.5 | 0.3 | 0.2 | 4 |
| NCLIV_048590 | hypothetical protein, conserved | 20.9 | 6.3 | 10.9 | 10.1 | 4 |
| NCLIV_053840 | hypothetical protein | 204.6 | 45.4 | 83.8 | 78.2 | 4 |
| NCLIV_054120 | sushi domain-containing protein / SCR repeat- containing protein, putative | 18.9 | 4.5 | 7.1 | 6.1 | 4 |
| NCLIV_055360 | hypothetical protein, conserved | 108.2 | 28.0 | 41.3 | 44.9 | 4 |
| NCLIV_055850 | hypothetical protein, conserved | 6.1 | 3.0 | 3.3 | 3.2 | 4 |
| NCLIV_057950 | Lipase maturation factor 2, related | 33.7 | 6.3 | 8.3 | 13.4 | 4 |
| NCLIV_058180 | Rhoptry protein 10, related | 19.2 | 4.9 | 7.6 | 13.4 | 4 |
| NCLIV_064620 | rhoptry neck protein 2, putative | 32.9 | 9.3 | 13.8 | 13.7 | 4 |

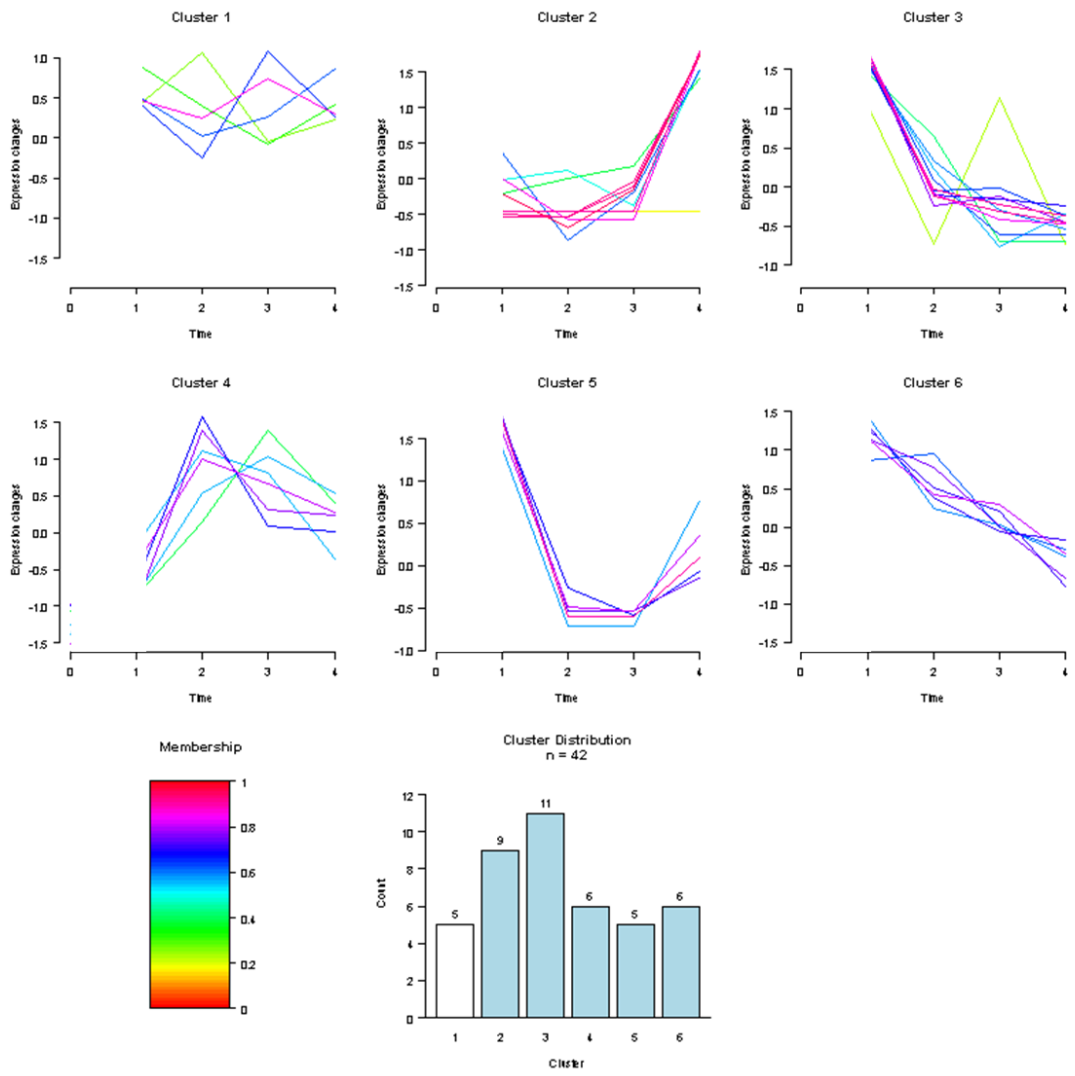



**Figure 6.9:** Microneme-associated genes clustered over the time course. Biological repeats were averaged to produce a mean FPKM for each time point. N = 42, time points were as follows: 1=Day 0, 2=Day 1, 3=Day 3, 4=Day 6.

There would appear to be different subsets of microneme genes which behave differently in their expression over time (Figure 6.9 and Table 6.8). Genes whose expression decreases over the time course belong to clusters 3 and 6. Cluster 4, similarly to rhoptry cluster 1, contains genes whose expression peaks at Day 1 after sodium nitroprusside induction; while cluster 5 genes are expressed at Day 0, show a decrease in expression at Days 1 and 3 and then increase again by Day 6. Those in cluster 1 appear to maintain expression throughout, so are perhaps important to both the tachyzoite and bradyzoite forms. Distribution of genes over the 6 clusters in Figure 6.9 is fairly uniform, suggesting that microneme genes fit into a range of different expression profiles.

**Table 6.8:** Membership of microneme-associated gene clusters, with mean FPKM values for the two biological replicates. Red text indicates an overall downward trend in expression for that cluster over the time course, green an increase and blue text indicates an increase in expression from Day 0 to Day 1 followed by a decrease over the remainder of the time course.

| I.D. | Description | FPKM Day 0 | FPKM Day 1 | FPKM Day 3 | FPKM Day 6 | Cluster |
|---|---|---|---|---|---|---|
| NCLIV_025710 | microneme protein 7, putative | 147.5 | 132.2 | 164.2 | 135.5 | 1 |
| NCLIV_036660 | hypothetical protein, conserved | 1.6 | 1.3 | 1.4 | 1.9 | 1 |
| NCLIV_050140 | subtilisin-like protease, putative | 0.2 | 0.1 | 0.3 | 0.2 | 1 |
| NCLIV_062760 | hypothetical protein | 1.1 | 1.5 | 0.9 | 1.0 | 1 |
| NCLIV_068520 | microneme TgMIC5 protein, putative | 393.2 | 313.6 | 242.0 | 316.1 | 1 |
| NCLIV_003250 | NcMCP4, putative | 2.7 | 2.3 | 9.9 | 37.0 | 2 |
| NCLIV_003260 | NcMCP3, putative,microneme protein, putative | 0.6 | 0.6 | 4.0 | 19.3 | 2 |
| NCLIV_007140 | microneme protein, putative | 0.0 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_015590 | microneme protein, putative | 0.0 | 0.0 | 0.0 | 0.0 | 2 |
| NCLIV_022530 | fibrillin-2 precursor, putative | 0.2 | 0.3 | 0.3 | 0.6 | 2 |
| NCLIV_038120 | Microneme protein 5 (Precursor), related | 0.2 | 0.0 | 0.1 | 0.3 | 2 |
| NCLIV_058240 | microneme protein, putative | 0.2 | 0.0 | 0.2 | 0.9 | 2 |
| NCLIV_058410 | apical membrane antigen, putative | 0.0 | 0.0 | 0.0 | 0.1 | 2 |
| NCLIV_068830 | PAN domain-containing protein (EC 3.4.21.27), related | 0.3 | 0.3 | 0.2 | 0.7 | 2 |
| NCLIV_002940 | microneme protein MIC4, putative | 233.0 | 66.2 | 48.4 | 35.7 | 3 |
| NCLIV_010600 | microneme protein MIC3, putative | 963.7 | 267.2 | 134.2 | 114.7 | 3 |
| NCLIV_015580 | microneme protein, putative | 0.1 | 0.0 | 0.0 | 0.0 | 3 |
| NCLIV_026810 | NcMCP2, putative,microneme protein, putative | 0.7 | 0.3 | 0.0 | 0.1 | 3 |
| NCLIV_033690 | hypothetical protein | 185.4 | 100.6 | 55.9 | 38.5 | 3 |
| NCLIV_043270 | microneme protein MIC1, putative | 427.1 | 174.4 | 178.7 | 126.0 | 3 |
| NCLIV_061760 | microneme protein MIC6, putative | 344.9 | 128.8 | 104.8 | 85.4 | 3 |
| NCLIV_062750 | EGF-like domain-containing protein (EC 3.4.21.21), related | 0.1 | 0.0 | 0.1 | 0.0 | 3 |
| NCLIV_062770 | microneme protein 8, putative | 132.3 | 49.9 | 47.4 | 42.9 | 3 |
| NCLIV_066250 | microneme protein 10, putative | 8872.3 | 2199.7 | 2608.2 | 1512.9 | 3 |
| NCLIV_066750 | NcMCP5, putative | 0.1 | 0.0 | 0.0 | 0.0 | 3 |
| NCLIV_029340 | EGF-like domain-containing protein, putative | 0.2 | 0.3 | 0.3 | 0.1 | 4 |
| NCLIV_038100 | hypothetical protein | 0.1 | 0.6 | 0.7 | 0.6 | 4 |
| NCLIV_038110 | hypothetical protein | 0.6 | 3.3 | 6.6 | 4.0 | 4 |
| NCLIV_058210 | microneme protein, putative | 0.0 | 0.3 | 0.2 | 0.2 | 4 |
| NCLIV_064590 | apical membrane antigen, putative | 0.2 | 1.9 | 0.8 | 0.8 | 4 |
| NCLIV_069310 | microneme protein, putative | 0.4 | 0.9 | 0.7 | 0.6 | 4 |
| NCLIV_008720 | microneme protein, putative | 0.3 | 0.0 | 0.0 | 0.0 | 5 |
| NCLIV_013920 | microneme protein, putative | 0.3 | 0.1 | 0.0 | 0.1 | 5 |
| NCLIV_054425 | NcMCP7, putative | 3.2 | 0.6 | 0.5 | 1.6 | 5 |
| NCLIV_054450 | NcMCP6, putative | 0.2 | 0.0 | 0.0 | 0.1 | 5 |
| NCLIV_058230 | hypothetical protein | 0.3 | 0.0 | 0.0 | 0.1 | 5 |
| NCLIV_018780 | microneme protein, putative | 0.2 | 0.1 | 0.1 | 0.0 | 6 |
| NCLIV_020720 | microneme protein MIC11, putative | 715.3 | 475.6 | 360.4 | 333.4 | 6 |
| NCLIV_022970 | microneme protein 2, putative | 162.8 | 91.5 | 79.2 | 54.7 | 6 |
| NCLIV_028680 | apical membrane antigen 1, | 349.4 | 255.5 | 238.2 | 153.0 | 6 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | putative | | | | | |
| NCLIV_038380 | hypothetical protein | 0.8 | 0.7 | 0.4 | 0.2 | 6 |
| NCLIV_051970 | MIC2-associated protein M2AP, putative | 189.1 | 197.2 | 119.4 | 96.4 | 6 |

### 6.4.5 Highly expressed genes

The data in Table 6.9 show that microneme and dense granule genes were among those most highly expressed across the whole time course. MIC10, GRA1 and GRA2 featured within the top ten most highly expressed genes at all four time points. In addition, there were large amounts of ribosomal transcripts being produced; all of the genes encoding 'hypothetical' proteins but one (NCLIV_065870) in this table are either orthologues of *T. gondii* ribosomal genes or align to other apicomplexan ribosomal genes when subjected to BLAST (Basic Local Alignment Search Tool, NCBI). The other 'hypothetical', NCLIV_065870 (in red), is more interesting as the predicted protein sequence BLASTs only to other 'hypothetical' proteins of apicomplexa. It would therefore warrant further investigation as it could prove to be a novel apicomplexan-specific gene, which, given its high mRNA expression levels, likely plays a major cellular role.

**Table 6.9:** Ten most highly expressed genes for each time point, by FPKM (fragments per kilobase of exon per million fragments mapped). Red text denotes a hypothetical gene that is potentially apicomplexan-specific.

| Day 0 | | Day 1 | | Day 3 | | Day 6 | |
|---|---|---|---|---|---|---|---|
| **Gene description** | **FPKM** | **Gene description** | **FPKM** | **Gene description** | **FPKM** | **Gene description** | **FPKM** |
| microneme protein 10, putative | 8872.3 | dense granule protein 1 / major antigenp24, putative | 4002.4 | dense granule protein 1 / major antigenp24, putative | 5286.7 | dense granule protein 1 / major antigenp24, putative | 4279.1 |
| dense granule protein 1 / major antigenp24, putative | 8284.5 | hypothetical protein NCLIV_069470 | 3821.8 | hypothetical protein NCLIV_069470 | 3694.1 | hypothetical protein NCLIV_069470 | 4256.8 |
| 28 kDa antigen (GRA2) | 5418.8 | 28 kDa antigen (GRA2) | 2397.9 | 28 kDa antigen (GRA2) | 2968.7 | 28 kDa antigen (GRA2) | 2088.2 |
| Cytochrome c oxidase subunit 1 related | 2320.2 | Ribosomal protein L37a, related | 2224.5 | microneme protein 10, putative | 2608.1 | Ribosomal protein L37a, related | 2052.2 |
| hypothetical protein | 2070.8 | microneme protein 10, putative | 2199.7 | Ribosomal protein L37a, related | 2030.7 | 40S ribosomal protein S29, putative | 1815.9 |
| Cytochrome c oxidase subunit 1 related | 1728.3 | 40S ribosomal protein S29, putative | 2037.7 | 40S ribosomal protein S29, putative | 1889.8 | microneme protein 10, putative | 1512.9 |
| hypothetical protein | 1613.7 | Cytochrome c oxidase subunit 1, related | 1604.8 | hypothetical protein NCLIV_062890 | 1528.2 | Cytochrome c oxidase subunit 1, related | 1395.7 |
| hypothetical protein, conserved | 1307.5 | hypothetical protein NCLIV_062890 | 1251.6 | Cytochrome c oxidase subunit , related | 1439.2 | hypothetical protein NCLIV_069470 | 1203.2 |
| granule antigen protein GRA6, putative | 1035.4 | hypothetical protein NCLIV_032660 | 1185.5 | hypothetical protein NCLIV_069470 | 1276.6 | hypothetical protein NCLIV_062890 | 1198.1 |
| dense granule protein 7, putative | 996.6 | hypothetical protein NCLIV_065870 | 1167.37 | hypothetical protein NCLIV_045010 | 1252.54 | hypothetical protein NCLIV_045010 | 1139.39 |

### 6.4.6 Differential expression of genes involved in metabolism

Genes involved in a number of metabolic pathways were identified as differentially expressed by DESeq analysis. These are shown in Table 6.10 with the metabolic pathway to which they belong. Both up and down regulation of metabolic genes was detected between Day 0 and Day 1; Day 0 and Day 3; and Day 0 and Day 6. One gene was up regulated from Day 1 to Day 6 and there was no significant ($P \leq 0.01$) differential expression of metabolism genes between Day 1 and Day 3 or Day 3 and Day 6. Differential expression was identified within the following metabolic pathways: purine and pyrimidine metabolism, the electron transport chain and the tricarboxylic acid (TCA) cycle, and phosphatidylcholine metabolism.

**Table 6.10:** Differential expression of genes involved in metabolic pathways. Differential expression was identified using DESeq, to a cut-off P-value (PadJ adjusted P-value) of ≤ 0.01. Red text indicates down regulation in the latter time point with respect to the earlier, green up regulation. Fragments per kilobase of exon per million fragments mapped (FPKMs) are stated for each time point, pair wise DESeq analyses were carried out on raw expression data.

| Gene I.D. | Day 0 FPKM | Day 1 FPKM | Day 3 FPKM | Day 6 FPKM | Differential Expression P-value | Metabolic pathway (according to LAMP database) |
|---|---|---|---|---|---|---|
| Genes differentially expressed from Day 0 to Day 1 | | | | | | n = 5 |
| NCLIV_000800 | 3.77 | 0.18 | 0.19 | 0.15 | 1.87E-09 | Phosphatidylcholine metabolism |
| NCLIV_032860 | 0.32 | 0.02 | 0.11 | 0.22 | 1.08E-07 | Purine metabolism |
| NCLIV_017300 | 5.59 | 0.52 | 0.59 | 1.53 | 0.000175 | Purine metabolism |
| NCLIV_052500 | 22.03 | 205.35 | 189.55 | 148.82 | 0.000572 | Electron transport chain<br>Tricarboxylic acid (TCA) cycle |
| NCLIV_019490 | 0.35 | 1.97 | 1.03 | 1.62 | 0.003911 | Lysine degradation |
| Genes differentially expressed from Day 0 to Day 3 | | | | | | n = 5 |
| NCLIV_000800 | 3.77 | 0.18 | 0.19 | 0.15 | 6.66E-07 | Phosphatidylcholine metabolism |
| NCLIV_052500 | 22.03 | 205.35 | 189.55 | 148.82 | 8.79E-05 | Electron transport chain<br>Tricarboxylic acid (TCA) cycle |
| NCLIV_017300 | 5.59 | 0.52 | 0.59 | 1.53 | 0.000297 | Purine metabolism |
| NCLIV_047410 | 39.07 | 7.05 | 8.25 | 11.85 | 0.002903 | Purine metabolism |
| NCLIV_032860 | 0.32 | 0.02 | 0.11 | 0.22 | 0.00323 | Purine metabolism |
| Genes differentially expressed from Day 0 to Day 6 | | | | | | n =5 |
| NCLIV_000800 | 3.77 | 0.18 | 0.19 | 0.15 | 4.84E-06 | Phosphatidylcholine metabolism |
| NCLIV_052500 | 22.03 | 205.35 | 189.55 | 148.82 | 0.002232 | Electron transport chain<br>Tricarboxylic acid (TCA) cycle |
| NCLIV_028420 | 0.24 | 0.86 | 1.50 | 2.03 | 0.003677 | Purine metabolism |
| NCLIV_037320 | 0.08 | 1.34 | 1.34 | 2.45 | 0.006358 | Purine metabolism<br>Pyrimidine metabolism |
| NCLIV_028410 | 0.12 | 0.67 | 1.08 | 0.68 | 0.009614 | Purine metabolism |
| Genes differentially expressed from Day 1 to Day 3 | | | | | | n = 0 |
| Genes differentially expressed from Day 1 to Day 6 | | | | | | n = 1 |
| NCLIV_032860 | 0.32 | 0.02 | 0.11 | 0.22 | 0.000116 | Purine metabolism |
| Genes differentially expressed from Day 3 to Day 6 | | | | | | n = 0 |

NCLIV_052500 is a flavoprotein subunit of succinate dehydrogenase and is a component of the electron transport chain which produces energy for the cell in the form of ATP (LAMP database). It was found to be most highly expressed at Day 1, but Days 1, 3 and 6 were all significantly higher than Day 0 (before sodium nitroprusside induction). There is another flavoprotein subunit in the *N. caninum* genome (NCLIV_068970) which was not identified as differentially expressed, but which also exhibited a peak in expression at Day 1.

Phosphoethanolamine N-methyltransferase (NCLIV_000800) is involved in the metabolism of phosphatidylcholine (LAMP database), an abundant phospholipid in *T. gondii* parasites (Gupta *et al.* 2005) that is required for synthesis of cell membranes. This gene was found to reduce expression after sodium nitroprusside induction from an FPKM of 3.77 to 0.18.

A number of genes involved in purine metabolism (NCLIV_028410, NCLIV_028420, NCLIV_032860, NCLIV_017300 and NCLIV_047410) and one involved in both purine and pyrimidine metabolism (NCLIV_037320), were also identified by this analysis. The 3',5'-cyclic-nucleotide phosphodiesterases (NCLIV_028410 and NCLIV_028420) were expressed more highly by the bradyzoites (Day 6) than the tachyzoites (Day 0), but these genes have a high amount of redundancy with twenty identified in the *N. caninum* genome (LAMP database), two more of which (NCLIV_017300 and NCLIV_032860) were expressed more highly by the Day 0 tachyzoite stage parasites. Nucleoside-diphosphate kinase (NCLIV_037320) was up regulated by Day 6 also. NCLIV_047410 (ecto-nucleoside triphosphate diphosphohydrolase) was expressed most highly by the tachyzoite stage parasites at Day 0, but, again, there are a number of other ecto-nucleoside triphosphate diphosphohydrolases present in the genome (LAMP database).

The lysine decarboxylase NCLIV_019490 increased in expression from Day 0 to Day 1. This gene encodes an enzyme involved in the degradation of lysine; there are however two additional shorter lysine decarboxylase domain-containing proteins (LAMP database) which are not differentially expressed.

## 6.5 DISCUSSION

The aim of this experiment was to determine whether there were changes in mRNA expression during stage conversion of *N. caninum* tachyzoites to bradyzoites. This was seen to be the case, and when subgroups of genes were examined in more detail, it became apparent that genes involved with invasion were some of those that exhibit differential expression. Differences were also identified within metabolic pathways, namely purine and pyramidine metabolism, the electron transport chain and the TCA cycle, and phosphatidylcholine metabolism; these differentially expressed genes (Table 6.10) could be examples of stage-specific gene expression. Phosphoethanolamine N-methyltransferase (NCLIV_000800) is an enzyme which catalyses three reaction steps in the phosphatidylcholine metabolic pathway, and produces choline phosphate (LAMP database). It exhibited reduced expression after sodium nitroprusside induction at Day 0, which could potentially be an early event signalling a reduction in the requirement for cell membrane synthesis in the less rapidly dividing bradyzoite stage. Behnke et al. (2010) analysed mRNA expression of synchronised *T. gondii* tachyzoites at different phases of the cell cycle, by microarray, and identified two distinct 'subtranscriptomes' where mRNA expression changes occurred in a cyclical manner. The parasites in this experiment, which was designed to identify differences between life stages rather than cell cycle stages, differ in that the parasite cultures were unsynchronised, so the results represent an average of the different phases of the cell cycle for each time point.

The three different types of analysis carried out in this chapter can be taken together to show that overall, the genes responsible for apical proteins are some of the most transcribed of the whole genome of over 7000 genes, for the tachyzoite and bradyzoite life cycle stages. Considering that as yet, only around 120 apical organelle-related genes have been putatively identified (Chapters 3 and 4), this is quite remarkable.

The differential expression data show that as the parasite responds to the stress of sodium nitroprusside treatment, a number of bradyzoite genes are up regulated. Vonlaufen *et al.* (2004) developed the sodium nitroprusside method of inducing stage differentiation and these results support their methods, especially as no tachyzoite markers were significantly up-regulated. Bradyzoites produced in this way are likely to be different again from bradyzoites *in vivo* as they have only converted in the previous days or hours, and as such are maybe more accurately considered to be *in the process of* stage conversion. Tissue cysts of bradyzoites removed from intermediate hosts may, in comparison, have been growing over anything from

a number of days to years (Buxton, McAllister & Dubey 2002; Tunev *et al.* 2002) and their analysis would be the gold standard. However, they are very difficult to isolate (A. Hemphill and Intervet, personal communication, (Hemphill *et al.* 2004)) and as this effectively renders them unavailable, the approach taken here serves as a promising model.

These results give an indication that rhoptry genes become less important in the bradyzoite stage than the tachyzoite stage, and this has been seen previously in the RNA-Seq data from Reid *et al*. (2012). Bradyzoites are able to invade cells (Dubey, Lindsay & Speer 1998; Tunev *et al.* 2002), but probably have less cause to do so: it is thought that a parasite which invades a cell when it is already committed to stage convert will then reproduce inside a vacuole to produce a tissue cyst of bradyzoites (Tunev *et al.* 2002). *ROP40* was down regulated between Days 0 and 6. This rhoptry kinase has not yet been well characterized but is a member of the *ROP2* family and localises to the rhoptries and PV (Peixoto *et al.* 2010). It has also been suggested that it may be under evolutionary pressure in *T. gondii* as it shows a high ratio of nonsynonymous to synonymous polymorphisms, which is a common occurrence for the rhoptry kinase (*ROPK*) family as a whole (Peixoto *et al.* 2010). Many of the ROPKs have been implicated in virulence differences between *T. gondii* strains, such as *ROP5, ROP16 and ROP18* and play roles in the host cell invasion process (see Chapter 5, (El Hajj *et al.* 2006; El Hajj *et al.* 2007; Reese *et al.* 2011; Saeij *et al.* 2006; Saeij *et al.* 2007)). Another rhoptry gene that was identified as down regulated was *Toxofilin,* which in *T. gondii* is known to be distributed throughout the tachyzoite cytoplasm during gliding and invasion (Poupel *et al.* 2000) but confined to the rhoptries in intracellular parasites (Bradley *et al.* 2005). Toxofilin binds actin to facilitate movement over and entry into the host cell (Delorme *et al.* 2003; Jan *et al.* 2007) so it's reduced expression in the more bradyzoite-like parasites could suggest a reduced requirement for gliding motility.

This reduction in rhoptry gene expression is the major finding of this experiment and it would seem intuitive that the tachyzoite, fully primed for invasion, would be producing more rhoptry products than the quiescent bradyzoite stage, as it has been well established that the major roles of rhoptry proteins are in host cell invasion, virulence and establishment of the parasite within the parasitophorous vacuole (PV) (Boothroyd & Dubremetz 2008; Peixoto *et al.* 2010).

However, when differential expression data of microneme and dense granule genes are examined, the picture is not so clear. Dense granule genes showed no significant differential expression at all (despite a slight trend for decreased expression in some clusters of genes) and

featured as some of the most highly abundant transcripts of all, over the whole time course. This could indicate an important role for dense granule proteins in the bradyzoite, which, given they are known to function generally later on in invasion than rhoptry and microneme proteins, in maintenance and establishment of the PV (Dubremetz *et al.* 1998; Nam 2009), would make sense; as bradyzoites and tachyzoites both grow in parasitophorous vacuoles (Tunev *et al.* 2002). Alternatively, it may simply be that these transcripts are remarkably robust and do not degrade as quickly as rhoptry ones, or have some other controlling element arresting their translation: this highlights that it is important to study proteomes as well as transcriptomes and genomes in order to better understand the system as a whole. *GRA1* and *GRA2* both appeared in the top ten highly expressed genes at each time point (Table 6.9), with *GRA6* and *GRA7* also in the top ten for Day 0 (tachyzoites). *T. gondii* GRA1 protein has been identified previously in both tachyzoites and bradyzoites (Cebron-Delauw et al 1989) and is secreted into the PV (Nam 2009). In this experiment *GRA1* was one of the highest-expressed transcripts across all time points and it has also been identified proteomically in excretory/secretory material and rhoptry/dense granule preparations from tachyzoites (Chapter 4). Taken together these results indicate that GRA1 is highly abundant in both the transcript and protein forms; it is likely to be an important component of the invasion process, specifically involved with the membranous tubular network (Sibley *et al.* 1995). *T. gondii* GRA2 and GRA6 have been found to interact, together with GRA4, to form a complex that may be involved with nutrient or protein transport around the vacuole (Labruyere *et al.* 1999) while *T. gondii* GRA7 has been identified in extensions protruding from the PV into the host cell cytoplasm (Bonhomme *et al.* 1998), perhaps suggesting an involvement with the acquisition of nutrients from the host cell.

There was significant differential expression of microneme genes in both an up regulatory and down regulatory sense, with respect to earlier time points; while other microneme genes remained highly expressed throughout (e.g. *MIC10*). Interestingly, MIC10 belongs to a group of microneme proteins that do not exhibit the classical adhesive domains that are associated with attachment to and invasion of the host cell (Hoff *et al.* 2001), its function remains unknown but it has been shown to exhibit varying circulation levels during infection with different *T. gondii* strains (Dautu *et al.* 2008).

Interestingly, MIC3 and MIC8, who are known to form a complex involved with invasion in *T. gondii* (Cerede *et al.* 2005) both had their transcripts designated to cluster 3 (Figure 6.9), along with *MIC1, MIC4 and MIC6*, which in *T. gondii* form the MIC1-4-6 complex (Meissner *et al.*

2002; Reiss *et al.* 2001). While *MIC3* was the only one to exhibit significantly decreased expression over the time course, all did show a downward trend in expression as the parasites became more bradyzoite like. Buchholz *et al.* (2011) also identified down regulation of *MIC3* and *MIC4* in a microarray analysis of *T. gondii* bradyzoites. Similarly to the data presented in the cluster analyses (Figures 6.7 - 6.9), Buchholz *et al.* (2011) also observed an overall decrease in invasion-related gene expression in bradyzoites relative to tachyzoites. Another finding common to both studies was the down regulation of *ROP40,* whilst *Toxofilin* was observed to be up regulated by microarray in *T.gondii* bradyzoites (Buchholz *et al.* 2011) but down regulated after sodium nitroprusside stimulation in *N. caninum* by RNA-Seq. A valuable element of the experimental design in Buchholz *et al.* (2011) was the comparison of *in vivo* and *in vitro T. gondii* tissue cysts; whilst this is not currently possible in *N. caninum* due to the lack of a suitable model intermediate host, it was interesting to note that they obtained similar and comparable results from both analyses.

The microneme adhesive repeat (MAR) domain containing proteins, or MCPs, were also among those transcripts identified as differentially expressed (Table 6.5). *MCP3* and *MCP4* were both increased in expression in the Day 6 parasites while *MCP7* decreased between Day 0 and Day 1 (immediately after sodium nitroprusside stimulation). Sialic acid is a component of glycoproteins and glycolipids found on host cell surfaces, and is often a target for host cell entry by pathogens. Friedrich *et al.* (2010) speculate that these MCPs may form a molecular basis for the coccidian parasites' ability to infect a wide range of host cells and demonstrate the MAR domain's sialic-acid dependent binding and involvement in the invasion process. They suggest that MCPs may be involved in different coccidian's tissue tropisms as they exhibit distinct binding specificities to different sialyl probes. Taken together with the results from this experiment: that MCPs 3, 4 and 7 show differential expression between tachyzoite and bradyzoite-like forms of the parasite; it could be speculated that the MCPs are involved in the different tissue tropisms of tachyzoites and bradyzoites.

Overall, these observations suggest that microneme genes may be equally important to the bradyzoite life stage as the tachyzoite, but that some microneme genes may be more stage-specific than others, with different sub groups of microneme genes playing a greater role in tachyzoite and bradyzoite biology respectively. Perhaps certain microneme genes are life-stage specific while others, such as those in Figure 6.9: cluster 1, which maintain expression throughout, are important to both the tachyzoite and bradyzoite forms. Distribution of genes

over the six clusters in figure 5 is fairly uniform, suggesting that microneme genes fit into a range of different expression profiles.

Many of the most highly expressed genes across all timepoints are apical genes, the products of which are involved in invasion (MIC10, GRA1, GRA2, GRA6 and GRA7 are all in the top 10 in Table 6.9). This gives a new angle to the earlier suggestion, from the DESeq analysis, that invasion genes show a general trend to decrease in expression level over the time course as the parasites become more bradyzoite-like: while that may be, it is also true that in many cases their expression level remains high nonetheless, when compared to the rest of the genome.

It is hoped that improvements in clustering algorithms and implementation will allow further analysis of this data in the future and enable statistical significance to be attributed to genes that cluster together so that genes statistically similar/different in expression over the whole time course can be identified. If more sophisticated clustering analysis tools were available, it would be interesting to mine the data for genes which show similar trends in expression to known bradyzoite markers as this could be a way to identify novel ones. This approach could also be used to identify further apical genes for example, but would likely generate a large proportion of false-positives, as similar expression profiles may be characteristic of wide ranging groups of genes. Nevertheless, taken together with additional bioinformatic tools such as BLAST (NCBI) and Pfam (Sanger Institute), among others, it could be a useful way to aid further annotation of the genome. As can be seen from these results and the proteomic data presented in earlier chapters, a huge proportion of the genome is still annotated as 'hypothetical' with no prediction of function or localisation available. A combination of bioinformatic and experimental approaches will be required to address this and the features of ToxoDB (Gajria *et al.* 2008) that allow the incorporation of experimental data from researchers around world will no doubt accelerate the process by enabling new functional information on genes to be shared with the community as a whole, often before it is published.

An interesting biological relationship that is as yet poorly understood is that between the proteome and transcriptome. In the pilot study, many of the highly proteomically and transcriptomically expressed tachyzoite genes and proteins were those with apical-secretory products, such as GRA1, GRA2, MIC2 and MIC10; for which it would make sense to be produced on a large scale for secretion during invasion: they are being released so not all will necessarily reach their target – it is likely that a considerable amount of redundancy and waste is accounted for. These results were echoed in the tachyzoite to bradyzoite mRNA expression

data. Despite initial hopes to perform a parallel quantitative proteomic analysis of the same samples, problems during culture and then difficulty obtaining high yields of mRNA from the parasites made this unfeasible. There is therefore scope for further study to compare the tachyzoite and bradyzoite proteomes, and compare and contrast them to the transcriptome analysed here.

# CHAPTER 7: DISCUSSION

This work presented in this thesis comprises a global analysis of gene and protein expression in *Neospora caninum*. Most of the work has focused on the tachyzoite stage, which is highly invasive and causes much of the pathology associated with neosporosis. The proteins identified by the various experiments include many known to be associated with the process of host invasion in Apicomplexa; and, when compared to published works, represent the most comprehensive analysis of the *N. caninum* proteome to date. Until 1984 (Bjerkas, Mohn & Presthus), *N. caninum* had been mistaken for *T. gondii* due to its similar histological appearance. It is interesting to compare and contrast these two organisms, due to the fact that they exhibit highly syntenic genome structure (DeBarry & Kissinger 2011; Reid *et al.* 2012), yet display important phenotypic differences in their host range and pathology.

## 7.1 Apical proteins and the potential for further study

Over 50 % of the predicted proteome putatively assigned to the apical organelles has been identified by the proteomic analyses in this thesis. While there has been much research carried out to ascertain the function of these proteins in *T. gondii*, less so has taken place with *N. caninum*, and the data produced here could be used to narrow down the choice of proteins to take forward for more in depth research. Not only are the proteins that are specific to *N. caninum* when compared to *T. gondii* interesting, such as MIC2B, MCP5, MCP6, MCP7, and NCLIV_038380 from the micronemes and ROP5B and NCLIV_068890 from the rhoptries; but all expressed invasion-related genes are worth investigating further in case their function in *N. caninum* turns out to be different from that observed previously in other Apicomplexa.

A particular method that could aid in elucidating the roles of these proteins would be to raise antibodies to proteins confirmed by proteomics, using the predicted sequences, and study their behaviour through live cell imaging. To gain an appreciation of where the apical proteins localise (for example, moving junction (MJ)/parasitophorous vacuole (PV)) and how long they are detectable for, throughout the process of invasion, would substantially improve current understanding in this field. Furthermore, this approach could help to identify any complexes formed and interactions between the different proteins. Many of these occurrences can be observed by classical techniques such as immunofluorescence microscopy of fixed slides at different points in the invasion process, and immunoprecipitation, but live cell imaging would provide an overall view of the process from start to finish. Furthermore, if this technique were

combined with bradyzoite-induction, it would be possible to confirm whether the reduction in rhoptry expression indicated by the transcriptomic study in Chapter 6, was recapitulated by rhoptry proteins *in vitro*.

Another potentially fruitful approach, that has been used successfully to study invasion in *T. gondii* and other Apicomplexa, would be to knock out genes of interest, or use interference RNA (RNAi) to block the expression of genes. Again, live cell imaging could provide a way to assess the effect of the knock out, not only during invasion but also afterwards, as the infection progressed and the tachyzoites replicated to produce cysts. It can, however, be difficult to be certain that the gene of interest is causing the effects observed, and it is usually necessary to perform a complementation experiment by knocking the gene back in to restore function.

The moving junction complex was one for which all five proteins, known to be involved in the *T. gondii* MJ, were identified proteomically. These were AMA1, RON2, RON4, RON5 and RON8: the RON protein complex gets inserted into the host cell membrane, connected to AMA1 in the parasite membrane, to enable the parasite to move into the host cell and create a parasitophorous vacuole (Besteiro, Dubremetz & Lebrun 2011). The identification of these proteins, together with the observation that they are not present in the genome with much redundancy (Boothroyd (2009) and no paralogues in ToxoDB), leads to the conclusion that the molecular structure of the *N. caninum* moving junction is likely to be analogous to that of *T. gondii*.

Another protein identified proteomically, in both the tachyzoite lysate and the rhoptry/dense granule preparation, was *N. caninum* ROP38. *T. gondii* ROP38 has recently been implicated in the down regulation of the host MAPK (mitogen-activated protein kinase) pathway which controls apoptosis and cell proliferation (Peixoto *et al.* 2010). ROP38 is typically expressed at a greater level in avirulent *T. gondii* than in virulent strains (Peixoto *et al.* 2010), and is likely to be fairly abundant in *N. caninum* as it was readily proteomically detected across multiple bands of a gel, suggesting a lower-virulence phenotype for *N. caninum*.

## 7.2 Important apical proteins not identified in the tachyzoite

Not only is there a possibility that some of the proteins *not* identified by these analyses are specific to life stages not examined, but their expression could also be affected by the host within which the parasite is infecting. All parasites studied in this thesis were cultured in Vero cells, which is a commonly used host cell for these organisms and one that the parasite grows

successfully in. However, there is now opportunity for a move toward culturing in cell systems that better replicate the situation *in vivo,* such as cell lines originating from a more likely natural host (bovine or canine, for example), or three-dimensional cell culture systems (Abbott 2003). There is also a risk that parasites which have become cell-culture adapted display a different phenotype to their wild relatives (Pelleau *et al.* 2011), but the use of parasites of the lowest possible passage number can minimise this. By studying parasites raised in (or ideally, *in vivo* isolated from-) a range of different host cells, it may be possible to establish whether slightly different repertoires of apical proteins are expressed depending upon the host cell being invaded. However, a *T. gondii* review by Boothroyd (2009) surmised that this was unlikely to be the case, as the proteins involved with the moving junction, for example, are highly conserved between strains of *T. gondii* and not encoded for by genes from large gene families, despite *T. gondii's* remarkable host range.

ROP16 was one protein known to be important in *T. gondii* virulence that was not proteomically identified; and in the transcriptome analysis, *N. caninum ROP16* mRNA was measured at basal levels only for all time points. This is an important finding, and in agreement with the transcriptome analysis in Reid *et al.* (2012), where *ROP16* was expressed highly by *T. gondii* parasites but not by *N. caninum*. ROP16 is polymorphic in *T. gondii* (Saeij *et al.* 2006) and acts as a kinase that activates STAT3 and STAT6 signalling pathways in type I and III parasites (Denkers *et al.* 2012; Saeij *et al.* 2007). This leads to reduced induction of the proinflammatory cytokine interleukin-12 (IL-12) by the host cell and illustrates modification of the host immune response by the parasite. The biology of ROP16 is complex and not yet fully understood, but enhanced growth of *T. gondii* ROP16 knockouts has been observed and linked to arginine metabolism (Butcher *et al.* 2011).

## 7.3 *ROP18* pseudogenisation and what it might mean from a population perspective

A desire to understand the genetic basis for the differences between *N. caninum* and *T. gondii* underpinned the comparative genome analysis performed by Reid *et al.* (2012). Of particular interest were differences in apical genes, as these are the molecular effectors of host cell invasion. While the majority of apical genes were found to have a one-to-one orthologous relationship between the two organisms and as such are hypothesized to behave in an analogous way *in vitro* and *in vivo* (subject to experimental validation), there were a few exceptions. One of the omissions from the *N. caninum* genome, *ROP18*, was noted in

particular, due to the recent identification of this gene as a virulence determinant in *T. gondii* (Saeij *et al.* 2006; Taylor *et al.* 2006).

As a result of its importance in *T. gondii* virulence, the *ROP18* (pseudo)gene was selected for a more in depth analysis as part of this thesis. This involved confirmation of the inclusion of stop codons in the coding sequence, not only in the reference strain of *N. caninum*, Nc Liverpool, but also in a variety of other isolates from canine and bovine hosts and a range of geographical locations. Research in *T. gondii* had established that ROP18 phosphorylated an immunity-related GTPase (IRG) called IRG a6, which was loaded onto the parasitophorous vacuole (PV) by the host cell during infection (Steinfeldt *et al.* 2010). Unphosphorylated IRG proteins result in PV disruption and parasite death (Hunn *et al.* 2008; Khaminets *et al.* 2010; Martens *et al.* 2005; Zhao *et al.* 2009a; Zhao *et al.* 2009b; Zhao *et al.* 2009c) and their loading onto the PV is an immune response by the host cell controlled by interferon-gamma (IFN-γ), experimentally determined *in vitro* in murine cells. The discovery of *ROP18*'s pseudogenisation in *N. caninum* led to the hypothesis that *N. caninum* would not be able to phosphorylate IRG a6 and as such its vacuoles would be susceptible the IFN- γ immune response and become loaded with IRG proteins. This is similar to observations in avirulent strains of *T. gondii*, which express *ROP18* only at very low levels (Khaminets *et al.* 2010; Steinfeldt *et al.* 2010). As presented in Chapter 5, for the first time IRG proteins were observed being loaded onto *N. caninum* PVs, and no phosphorylation of IRG a6 was detected, compared to 90 % of IRG a6–loaded vacuoles being phosphorylated in *T. gondii* type I controls. Overall, these data supported the theory that *ROP18* had been pseudogenised in *N. caninum* since its divergence from *T. gondii,* estimated to have occurred around 28 million years ago (Reid *et al.* 2012). It is intriguing that despite pseudogenisation, no single nucleotide polymorphisms (SNPs) were detected in this region between any of the *N. caninum* isolates analysed. In a study on sequence variation in introns of *P. falciparum*, few SNPs were identified but 71 microsatellite repeats were identified across 25 introns, which are considered along with pseudogenes to be rapidly evolving sequences (Volkman *et al.* 2001). The lack of *ROP18* sequence divergence between *N. caninum* isolates may be a result of the highly clonal population structure (Perez-Zaballos *et al.* 2005), which is in contrast to *Plasmodium*'s usual method of transmission involving sexual stages in a mosquito. Previous studies on *N. caninum* have also detected little variation between isolates (Holmdahl *et al.* 1997; Innes *et al.* 2000; Schock *et al.* 2001) and the parasite is not thought to exhibit the differences in virulence that manifest in the three lineage 'types' of *T. gondii*: I (virulent), II (intermediate) and III (avirulent) (Howe & Sibley 1995; Sibley & Boothroyd 1992).

The relative importance of vertical and horizontal transmission between *N. caninum* and *T. gondii* is another, related area of interest. Vertical transmission of *N. caninum* from cow to calf is reported to have an efficiency of between 78 and 95% (Davison, Otter & Trees 1999; Pare, Thurmond & Hietala 1997), with most seropositive cows being thought to have themselves been infected congenitally rather than exogenously (Pare, Thurmond & Hietala 1996, 1997). It was previously suggested that dogs may not be the optimum definitive host of *N. caninum* due to the low numbers of oocysts excreted (Lindsay, Ritter & Brake 2001) but Gondim *et al.* (2002) demonstrated improved production of oocysts (a mean of 160 700 total oocyst production from days 4-30 post infection) when dogs were fed infected calf tissues rather than infected mice. Gondim *et al.* (2005) then measured approximately 500 000 total oocysts shed by a dog after exposure to infected calf tissue. This number is still somewhat lower, however, than what is readily observed in cats shedding *T. gondii*, which is in the region of 20 million oocysts from days 4-13 post infection (Dubey 1995) and up to 1 billion (Dubey & Frenkel 1972). This could be an indication of a more dominant role for the sexual cycle (cat-rodent-cat transmission) in *T. gondii* population biology (which was recently found to exhibit a higher level of genetic variation than previously thought (Dubey & Su 2009)), when compared to *N. caninum* population biology, which is thought to rely heavily on vertical transmission (Dubey, Schares & Ortega-Mora 2007; Williams *et al.* 2009). The finding that *N. caninum* exhibits a low-virulence phenotype in the absence of ROP18 supports this hypothesis in that it could be an adaptation for more successful vertical transmission. Highly virulent pathogens that cause severe pathology in their host are unlikely to allow the host to survive long enough for it to reproduce, or may provoke such a strong immune response that the infection is cleared. By existing in a latent form (the bradyzoite), *N. caninum* is able to survive within the host until such time as the infection reactivates to spread to the developing foetus. Virulent *T. gondii* infecting a mouse, need only survive long enough in the mouse for it to be caught and eaten by a cat. The presence of an infection may perhaps even render the mouse less fit and make it more prone to predation. Vertical transmission does occur in *T. gondii* but ovine toxoplasmosis is thought to primarily occur in sheep infected exogenously during pregnancy, as opposed to endogenously (Innes *et al.* 2007; Rodger *et al.* 2006).

Another interesting observation regarding an additional role of ROP18 is that the *T. gondii* proteins ROP2 and ROP8, both of which have no orthologue identified in the *N. caninum* genome (Reid *et al.* 2012), were found by Qiu *et al.* (2009) to be phosphorylation substrates for ROP18. So, it appears that *N. caninum* is lacking in all three of these interacting proteins,

which perform an as yet uncharacterised function in *T. gondii*. Intriguingly, humans are the only mammals so far identified as lacking in an IRG-protein repertoire (Bekpen *et al.* 2005; Zhao *et al.* 2009). This makes it unlikely that ROP18 pseudogenisation is the cause of the difference in zoonotic capability between *N. caninum* and *T. gondii*, but this finding is still very interesting with respect to the effect it has on virulence, and requires further investigation in bovine and canine host systems.

## 7.4 Potential directions for analysis of additional life stages

An obvious next stage in proteomic and transcriptomic expression research for *N. caninum* would be to analyse other life stages, such as the oocyst. While the sexual stages are unlikely to become available by *in vivo* or *in vitro* culture methods in the next few years, the continual improvements in the sensitivity of transcriptomic and proteomic platforms should make the analysis of oocysts a possibility, as less material will be required than currently. For example, single-molecule RNA sequencing should provide opportunities to study not only oocysts, but perhaps bradyzoites isolated from *in vivo* tissue cysts, and tachyzoites that have recrudesced from a latent infection (for example, isolated from a placenta), to determine whether the findings identified by the analysis of *in vitro* cultivated parasites, as presented here, are a realistic representation of what occurs *in vivo*.

A group of surface antigens (SAGs) known as SRS (SAG-1 related sequence) proteins were found by Reid *et al.* (2012) to be extremely divergent and, in contrast to what was expected (based on the narrower host range of *N. caninum)*, approximately two-fold more abundant in the *N. caninum* than the *T. gondii* genome. The finding in Reid *et al.* (2012), that a greater number of SRS genes were expressed by the Day 6 (early bradyzoite-like) parasites than the Day 3 tachyzoites, was not repeated by the transcriptomic analysis in Chapter 6 (data not shown). This could be a reflection of the different bradyzoite preparation techniques between the two studies, the latter using the classical sodium nitroprusside-induction method of bradyzoite culture, rather than simply extending the culture period. The question of SRS expression is one that could be investigated further by analyses of gene and protein expression in additional life-stages of *N. caninum* and *T. gondii*. It would be interesting to determine whether more of the predicted *N. caninum* SRS repertoire is expressed by another stage, and whether the expression of a greater number of SRSs by *T. gondii*, despite the smaller number encoded for in the genome, is replicated by the other life stages. Another group of proteins thought to be involved in host cell recognition and preference are the MAR-domain containing

proteins (MCPs), which unexpectedly, and similarly to the SRS proteins, are present in a greater number in the *N. caninum* than the *T. gondii* genome (Friedrich *et al*. 2010; Cowper *et al*. 2012). More work on gene expression in other life stages may help unravel this paradox further.

## 7.5 Concluding remarks

This thesis has provided an overall view of the tachyzoite global proteome and in particular, identified more than half of the apical proteins predicted by the genome, including the *N. caninum* orthologues of all those involved in the *T. gondii* moving junction. The work has contributed to the annotation of the *N. caninum* genome, an important resource to the wider apicomplexan research community. In addition, the comparative analysis of the tachyzoite and bradyzoite transcriptomes identified a key difference in that rhoptry genes show a marked decrease in mRNA expression in the bradyzoite stage, likely to be a result of the decreased requirement for invasion by the bradyzoite.

*N. caninum* has long been thought to be a less virulent cousin of *T. gondii* and the confirmation of the pseudogenisation of *ROP18* has led to speculation on the role of virulence in the population biology of this parasite, and how a lower virulence phenotype may have aided the evolution of an organism specialised in vertical transmission. Additional findings supporting the low-virulence hypothesis are the low expression level of *ROP16* in the transcriptome, and the detection of ROP38 in the proteome.

Overall, these findings provide opportunities for further research into the invasion processes carried out by *N. caninum*, in the hope that it will someday lead to the discovery of vaccine candidates to prevent and control neosporosis.

# LIST OF APPENDICES

# APPENDIX III: OFFGEL SEPARATION OF A TACHYZOITE LYSATE

## Introduction

An additional tachyzoite lysate to those presented in Chapter 2 was analysed by the OFFGEL technique (Agilent). The OFFGEL fractionation system enables separation of protein or peptide mixtures over a pH gradient by isoelectric focusing, as with 2-DE, but in solution. This means that fractions are able to be digested for mass spectrometry, without the need to extract from a gel. However, the lysis buffer required contains glycerol, ampholytes, urea and thiourea, which must be precipitated out prior to mass spectrometry, therefore sufficient sample to allow for losses during precipitation must be loaded. Also, in order to focus effectively, the sample must not contain salt, so for cell preparations washed in PBS it is important to remove it prior to solubilisation.

## Methods

### Fractionation

A pellet of 1.6 x $10^8$ tachyzoites was lysed in 360 µl OFFGEL buffer (8 M urea, 2.4 M thiourea, 0.2 % (v/v) ampholytes, 10 % (v/v) glycerol) over three cycles of vortexing for 3 minutes and heating to 37 °C for 3 minutes, followed by centrifugation at 13 000 x g for 10 minutes at room temperature. The sample was stored at -20 °C overnight then added to a further 1.44 ml OFFGEL buffer. A 13 cm pH 3-10 non linear IPG strip was placed into the OFFGEL tray and an 11-well frame applied. Rehydration buffer comprising 0.56 ml OFFGEL buffer and 0.14 ml $ddH_20$ was pipetted into each well (40 µl per well). After 15 minutes, approximately 40 µg sample in 150 µl was loaded into each well and fractionation was performed for 23 hours, with parameters as follows: 20.758 kVh, 544 V, 6.5 mW and 12 µA. Each of the 11 fractions was removed and the wells washed (twice) in a further 150 µl $ddH_20$, that was then added to the appropriate fraction, so that final sample volumes were 450 µl.

### Acetone precipitation

An acetone precipitation was performed as follows: 2.25 ml ice cold acetone was added to each fraction prior to incubation at -20 °C overnight. The samples were then centrifuged at

1500 x g for 20 minutes at 4 °C, each pellet washed in acetone and resuspended in 50 µl 25 mM ammonium bicarbonate.

### 1-DE of OFFGEL fractions

Samples were visualized on a 10 cm 12 % acrylamide gel (2.2.2.2) with limiting factors of 200 V, 70 mA and 10 W for 50 minutes. The gel was fixed at room temperature in a 40 % (v/v) ethanol, 10 % (v/v) acetic acid solution overnight, then rinsed twice in $ddH_2O$ before staining with Colloidal Coomassie (20 % (v/v) methanol, 0.08 % (w/v) Coomassie Brilliant Blue G250, 0.8 % (v/v) phosphoric acid, 8 % (w/v) ammonium sulphate) for 24 hours.

### Results and Discussion

An OFFGEL fractionation was carried out for two reasons. Firstly, it enabled pre-fractionation of a tachyzoite lysate prior to trypsin digestion, so as not to separate peptides originating from the same protein (as occurs during MudPIT) to ease downstream bioinformatic searching. Secondly, it removed the requirement for gel separation, which can affect resolution of some peptides that solubilise less well, and can cause problems for more sensitive HPLC columns due to carry-over of detergent. Figure 2.4 shows the gel resulting from a tachyzoite lysate OFFGEL-fractionation in protein space.

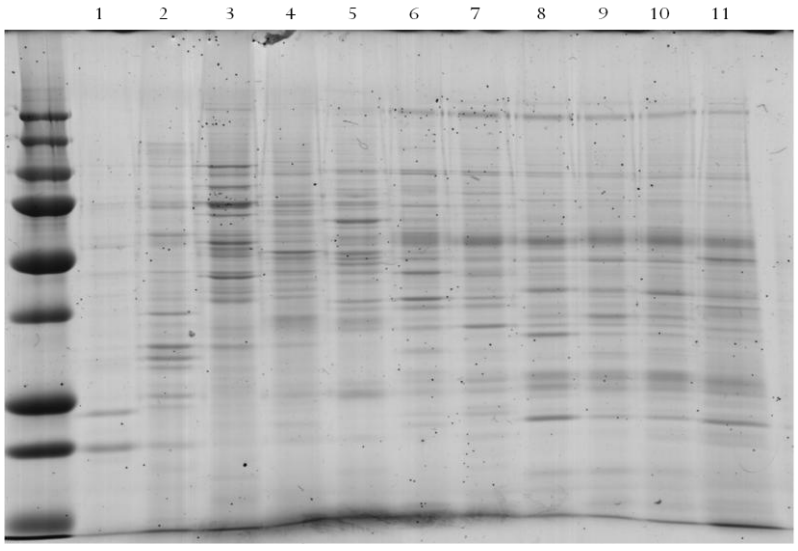



**Figure V.1:** 1-DE gel of tachyzoite OFFGEL fractions. Samples were analysed on a 12 % (w/v) acrylamide gel under denaturing conditions, visualised by Colloidal Coomassie staining. Lanes are as follows, lane 1: protein standards, lanes 2-12: 20 µl aliquots of fractions, as labelled.

Fractions 1-5 exhibited differential staining of protein bands, indicating successful fractionation of proteins, however fractions 6-11 appeared to contain many of the same bands due tosub-optimal separation. Furthermore, other researchers were reporting difficulties with the sample clean-up procedure prior to mass-spectrometry. Another complication included significant sample losses during fractionation and clean-up, which meant that the requirement for loading of initial parasite material was higher than for other methods (≥ 1 mg). As such, it was decided not to continue further with OFFGEL fractionation in this analysis.

# APPENDIX IV: EXCRETORY/SECRETORY

## PROTEOMIC IDENTIFICATIONS

**Table IV.1**: Protein identifications for excretory/secretory preparations in Chapter 4 (n = 78). Proteins were identified using Mascot and a cut off score of 50 was employed, proteins are presented in decreasing order of Mascot score. Apical proteins have been coloured according to their identification in Chapter 3: yellow indicates microneme protein; blue, dense granule and purple, rhoptry. Signal peptide predictions from SignalP 3.0 were downloaded from ToxoDB (cut off p-value ≥ 0.9). MIPs functional category assignments are abbreviated as follows: M = metabolism, E = energy, STP = storage protein, C/DNA = cell cycle and DNA processing, T = transcription, PS = protein synthesis, PF = protein fate. PB = protein with binding function or cofactor requirement, R = regulation of metabolism and protein function, CT = cellular transport, transport facilities and transport routes, CC = cellular communication/signal transduction mechanism, CRDV = cell rescue, defence and virulence, IE = interaction with the environment, CF = cell fate, BG = biogenesis of cellular components and SL = subcellular localisation.

| Protein I.D. | Description | Signal peptide predicted? | Mascot score | MIPs category assignment |
|---|---|---|---|---|
| NCLIV_021050 | hypothetical protein | Yes | 801 | CRDV |
| NCLIV_033250 | SRS domain-containing protein | No | 672 | IE |
| NCLIV_043270 | microneme protein MIC1, putative | Yes | 592 | CRDV |
| NCLIV_024820 | 14-3-3 protein homolog | No | 543 | CRDV |
| NCLIV_033230 | SRS domain-containing protein,SRS29B, putative | Yes | 513 | IE |
| NCLIV_051970 | MIC2-associated protein M2AP, putative | Yes | 433 | CRDV |
| NCLIV_032330 | Malate dehydrogenase (NAD) (Precursor), related | No | 335 | E |
| NCLIV_010720 | SRS domain-containing protein | Yes | 325 | IE |
| NCLIV_068400 | NTPase | Yes | 277 | CRDV |
| NCLIV_050370 | hypothetical protein | No | 266 | E |
| NCLIV_068460 | hypothetical protein | No | 253 | CRDV |
| NCLIV_002940 | microneme protein MIC4, putative | Yes | 242 | CRDV |
| NCLIV_028680 | apical membrane antigen 1, putative | No | 234 | CRDV |
| NCLIV_068920 | SRS domain-containing protein | Yes | 228 | IE |
| NCLIV_036400 | dense granule protein 1 / major antigenp24, putative | Yes | 204 | CRDV |
| NCLIV_010730 | SRS domain-containing protein | Yes | 199 | IE |
| NCLIV_010600 | microneme protein MIC3, putative | Yes | 195 | CRDV |
| NCLIV_047630 | 40S ribosomal protein S18, putative | No | 184 | PS |
| NCLIV_061760 | microneme protein MIC6, putative | No | 184 | CRDV |
| NCLIV_045650 | 28 kDa antigen, putative | Yes | 174 | CRDV |
| NCLIV_000610 | profilin family protein, putative | No | 170 | CT |

| | | | | |
|---|---|---|---|---|
| NCLIV_035310 | inhibitor-1 of protein phosphatase type 2A, putative | No | 168 | R |
| NCLIV_011270 | hypothetical protein | No | 156 | E |
| NCLIV_062630 | Thioredoxin-dependent peroxide reductase, mitochondrial, related | No | 155 | IE |
| NCLIV_043760 | hypothetical protein, conserved | Yes | 153 | unclassified |
| NCLIV_033950 | Heat shock protein 70, related | No | 152 | CRDV |
| NCLIV_009670 | Cold-shock protein, DNA-binding, related | No | 145 | T |
| NCLIV_059430 | hypothetical protein | No | 141 | unclassified |
| NCLIV_052880 | granule antigen protein GRA6, putative | No | 135 | CRDV |
| NCLIV_066250 | microneme protein 10, putative | Yes | 133 | CRDV |
| NCLIV_018530 | hypothetical protein, conserved | Yes | 130 | PS |
| NCLIV_066870 | hypothetical protein | No | 127 | T |
| NCLIV_029860 | hypothetical protein | No | 120 | PS |
| NCLIV_031860 | serine-threonine phosophatase 2C, putative | No | 119 | R |
| NCLIV_013780 | hypothetical protein | No | 106 | PF |
| NCLIV_031510 | hypothetical protein | No | 100 | unclassified |
| NCLIV_055710 | 60S ribosomal protein L23, putative | No | 100 | PS |
| NCLIV_000710 | hypothetical protein, conserved | Yes | 99 | SL |
| NCLIV_002390 | Nucleoside diphosphate kinase, related | No | 95 | M |
| NCLIV_021640 | dense granule protein 7, putative | Yes | 92 | CRDV |
| NCLIV_050380 | Fructose-bisphosphate aldolase, related | No | 87 | E |
| NCLIV_001300 | calmodulin, putative | No | 86 | PB |
| NCLIV_032220 | 50S ribosomal protein L30e, related | No | 85 | PS |
| NCLIV_024870 | hypothetical protein | No | 84 | PS |
| NCLIV_038780 | 60S ribosomal protein L32, related | No | 82 | PS |
| NCLIV_052380 | hypothetical protein | No | 82 | PS |
| NCLIV_066310 | DEAD-box ATP-dependent RNA helicase 34, related | No | 81 | T |
| NCLIV_066630 | GRA9 protein, putative | Yes | 81 | CRDV |
| NCLIV_012510 | hypothetical protein | No | 80 | PB |
| NCLIV_032620 | Cs1 protein, related | No | 79 | unclassified |
| NCLIV_048880 | Proteasome subunit beta type-7, related | No | 76 | PF |
| NCLIV_020140 | hypothetical protein | No | 73 | unclassified |
| NCLIV_022140 | GA11385, related | No | 72 | PB |
| NCLIV_026140 | Histone H2A, related | No | 71 | BG |
| NCLIV_037500 | Enolase, related | No | 70 | E |
| NCLIV_001670 | Elongation factor 1-alpha, related | No | 69 | PS |
| NCLIV_003440 | Actin, related | No | 69 | CT |
| NCLIV_000430 | hypothetical protein, conserved | No | 68 | unclassified |
| NCLIV_038400 | Methionine aminopeptidase, related | No | 68 | unclassified |
| NCLIV_054700 | uridine phosphorylase, putative | No | 68 | M |
| NCLIV_018420 | p36 protein, putative (ROP9) | Yes | 67 | CRDV |
| NCLIV_019770 | hypothetical protein | No | 66 | M |
| NCLIV_045870 | Putative dense granule protein 3 | Yes | 66 | CRDV |

| | | | | |
|---|---|---|---|---|
| NCLIV_025240 | Gbp1p protein, putative | No | 65 | PB |
| NCLIV_018300 | HIT family protein, related | No | 63 | unclassified |
| NCLIV_006070 | 30S ribosomal protein S10P, related | No | 61 | PS |
| NCLIV_001070 | Histone H2B, related | No | 60 | BG |
| NCLIV_030290 | transaldolase, putative | No | 60 | M |
| NCLIV_015880 | hypothetical protein | No | 59 | PS |
| NCLIV_026150 | Histone H3, related | No | 56 | BG |
| NCLIV_018290 | Ribosomal protein S26E, related | No | 55 | PS |
| NCLIV_032270 | hypothetical protein, conserved | No | 55 | unclassified |
| NCLIV_046030 | hypothetical protein | No | 55 | unclassified |
| NCLIV_061560 | hypothetical protein, conserved | No | 55 | PB |
| NCLIV_036280 | 30S ribosomal protein S15P/S13e, related | No | 54 | PS |
| NCLIV_000740 | Class I chitinase, related | Yes | 52 | M |
| NCLIV_053640 | peroxidoxin 2, putative | No | 52 | IE |
| NCLIV_041940 | Glyceraldehyde 3-phosphate dehydrogenase, related | No | 50 | M |

# APPENDIX V: RHOPTRY / DENSE GRANULE

## PROTEOMIC IDENTIFICATIONS

**Table V.1:** Protein identifications for the rhoptry (R) preparation (n = 138). Proteins were identified using Mascot and a cut off score of 50 was employed, proteins are presented in decreasing order of Mascot score. Apical proteins have been highlighted according to their identification in Chapter 3: yellow indicates microneme protein; blue, dense granule and purple, rhoptry. Signal peptides were predicted by SignalP 3.0 and downloaded from ToxoDB (cut off p-value ≥ 0.9). MIPs functional category assignments are abbreviated as follows: M = metabolism, E = energy, STP = storage protein, C/DNA = cell cycle and DNA processing, T = transcription, PS = protein synthesis, PF = protein fate. PB = protein with binding function or cofactor requirement, R = regulation of metabolism and protein function, CT = cellular transport, transport facilities and transport routes, CC = cellular communication/signal transduction mechanism, CRDV = cell rescue, defence and virulence, IE = interaction with the environment, CF = cell fate, BG = biogenesis of cellular components, SL = subcellular localisation and ? = potential ROP proteins.

| Protein I.D. | Description | Signal peptide predicted? | Mascot score | MIPs category assignment |
|---|---|---|---|---|
| NCLIV_001970 | hypothetical protein | Yes | 2346 | CRDV |
| NCLIV_060730 | hypothetical protein | Yes | 1104 | CRDV |
| NCLIV_055360 | hypothetical protein, conserved | Yes | 1002 | CRDV |
| NCLIV_053840 | hypothetical protein | Yes | 932 | CRDV |
| NCLIV_031550 | hypothetical protein | No | 920 | ? |
| NCLIV_051340 | toxofilin, putative | Yes | 761 | CRDV |
| NCLIV_018420 | p36 protein, putative (ROP9) | Yes | 568 | CRDV |
| NCLIV_036700 | M16 family peptidase, putative | Yes | 561 | PF |
| NCLIV_068850 | hypothetical protein | Yes | 552 | CRDV |
| NCLIV_018830 | hypothetical protein, conserved | No | 496 | CRDV |
| NCLIV_007770 | Rhoptry kinase family protein, truncated, putative | Yes | 487 | CRDV |
| NCLIV_011700 | hypothetical protein | Yes | 486 | unclassified |
| NCLIV_064620 | rhoptry neck protein 2, putative | Yes | 478 | CRDV |
| NCLIV_033230 | SRS domain-containing protein,SRS29B, putative | Yes | 461 | IE |
| NCLIV_006180 | duplicated carbonic anhydrase, putative | No | 455 | unclassified |
| NCLIV_069590 | hypothetical protein | Yes | 453 | CRDV |
| NCLIV_058180 | Rhoptry protein 10, related | No | 431 | CRDV |
| NCLIV_047500 | Trichohyalin, putative | Yes | 423 | unclassified |
| NCLIV_065090 | hypothetical protein, conserved | No | 411 | PF |
| NCLIV_018120 | hypothetical protein, conserved | No | 394 | unclassified |
| NCLIV_001660 | hypothetical protein, conserved | Yes | 393 | unclassified |

| | | | | |
|---|---|---|---|---|
| NCLIV_055850 | hypothetical protein, conserved | Yes | 389 | CRDV |
| NCLIV_055760 | hypothetical protein, conserved | Yes | 373 | unclassified |
| NCLIV_060740 | ROP 2, related | Yes | 361 | CRDV |
| NCLIV_068890 | hypothetical protein | Yes | 349 | CRDV |
| NCLIV_011730 | Rhoptry kinase family protein ROP26, putative | No | 344 | CRDV |
| NCLIV_056300 | hypothetical protein, conserved | Yes | 336 | unclassified |
| NCLIV_069110 | hypothetical protein | Yes | 329 | ? |
| NCLIV_027930 | Rhoptry kinase family protein ROP17, putative | No | 314 | CRDV |
| NCLIV_005500 | hypothetical protein, conserved | No | 308 | unclassified |
| NCLIV_048590 | hypothetical protein, conserved | No | 307 | CRDV |
| NCLIV_021050 | hypothetical protein | Yes | 291 | CRDV |
| NCLIV_038360 | TSP1 domain-containing protein TSP12 (Precursor), related | No | 285 | IE |
| NCLIV_070010 | hypothetical protein, conserved | No | 285 | unclassified |
| NCLIV_030050 | hypothetical protein, conserved | Yes | 278 | CRDV |
| NCLIV_027850 | rhoptry protein 6, putative | No | 268 | CRDV |
| NCLIV_017420 | Zgc:55863, related | Yes | 260 | CRDV |
| NCLIV_068230 | nucleoside-triphosphatase, putative | Yes | 260 | M |
| NCLIV_062520 | 3-ketoacyl-(Acyl-carrier-protein) reductase , related | Yes | 259 | M |
| NCLIV_054120 | sushi domain-containing protein/SCR repeat- containing protein (RON1) | Yes | 249 | CRDV |
| NCLIV_006850 | Hornerin, related | No | 243 | M |
| NCLIV_067490 | protein phosphatase 2C, putative | No | 230 | ? |
| NCLIV_047390 | hypothetical protein, conserved | Yes | 226 | unclassified |
| NCLIV_037190 | glyceraldehyde-3-phosphate dehydrogenase, putative | No | 214 | M |
| NCLIV_035910 | hypothetical protein | No | 212 | unclassified |
| NCLIV_007450 | hypothetical protein | No | 196 | CRDV |
| NCLIV_003440 | Actin, related | No | 191 | CT |
| NCLIV_025670 | ATP synthase subunit beta, related | No | 184 | E |
| NCLIV_035590 | hypothetical protein, conserved | No | 182 | unclassified |
| NCLIV_017410 | hypothetical protein | Yes | 179 | CRDV |
| NCLIV_022690 | hypothetical protein, conserved | No | 175 | unclassified |
| NCLIV_047600 | hypothetical protein | Yes | 170 | unclassified |
| NCLIV_034460 | hypothetical protein | No | 161 | PB |
| NCLIV_020340 | hypothetical protein | No | 159 | CRDV |
| NCLIV_009150 | alpha-galactosidase A, putative | Yes | 156 | M |
| NCLIV_011690 | ROP15 protein, related | Yes | 156 | CRDV |
| NCLIV_021100 | hypothetical protein, conserved | Yes | 153 | CRDV |
| NCLIV_046310 | hypothetical protein | No | 152 | unclassified |
| NCLIV_033250 | SRS domain-containing protein | No | 150 | IE |
| NCLIV_045580 | hypothetical protein, conserved | No | 146 | CRDV |
| NCLIV_066020 | hypothetical protein | No | 145 | CT |
| NCLIV_036400 | dense granule protein 1 / major antigenp24, putative | Yes | 143 | CRDV |

| | | | | |
|---|---|---|---|---|
| NCLIV_037590 | hypothetical protein, conserved | No | 143 | unclassified |
| NCLIV_043870 | hypothetical protein, conserved | No | 143 | unclassified |
| NCLIV_010730 | SRS domain-containing protein | Yes | 136 | IE |
| NCLIV_014020 | Peroxiredoxin-2E-1, related | Yes | 127 | IE |
| NCLIV_028170 | Rhoptry kinase family protein ROP20, putative | No | 125 | CRDV |
| NCLIV_032330 | Malate dehydrogenase (NAD) (Precursor), related | No | 113 | E |
| NCLIV_047530 | hypothetical protein, conserved | Yes | 113 | unclassified |
| NCLIV_063340 | hypothetical protein | Yes | 112 | PF |
| NCLIV_011410 | protein disulfide isomerase | Yes | 110 | IE |
| NCLIV_009440 | hypothetical protein | No | 108 | CT |
| NCLIV_000710 | hypothetical protein, conserved | Yes | 107 | SL |
| NCLIV_024630 | porin, putative | No | 107 | CT |
| NCLIV_004270 | protein kinase, putative | No | 106 | ? |
| NCLIV_025730 | hypothetical protein, conserved | Yes | 106 | unclassified |
| NCLIV_026070 | armadillo/beta-catenin-like repeat-containing protein, putative | No | 103 | PB |
| NCLIV_039500 | hypothetical protein | No | 98 | unclassified |
| NCLIV_066840 | hypothetical protein | No | 98 | CRDV |
| NCLIV_024820 | 14-3-3 protein homolog | No | 95 | CRDV |
| NCLIV_055490 | Heat shock protein 70 (Precursor), related | Yes | 95 | CRDV |
| NCLIV_067010 | Mitochondrial phosphate carrier protein, related | Yes | 94 | CT |
| NCLIV_046970 | hypothetical protein, conserved | No | 92 | unclassified |
| NCLIV_033860 | hypothetical protein, conserved | No | 89 | unclassified |
| NCLIV_013260 | hypothetical protein, conserved | Yes | 86 | unclassified |
| NCLIV_033270 | hypothetical protein | Yes | 85 | M |
| NCLIV_061160 | acid phosphatase, putative | Yes | 84 | ? |
| NCLIV_046830 | ATP synthase, putative | No | 83 | CT |
| NCLIV_041120 | hypothetical protein, conserved | No | 82 | CRDV |
| NCLIV_062720 | hypothetical protein | No | 82 | PS |
| NCLIV_001300 | calmodulin, putative | No | 81 | PB |
| NCLIV_025000 | hypothetical protein | Yes | 81 | PB |
| NCLIV_032390 | hypothetical protein, conserved | No | 81 | unclassified |
| NCLIV_056910 | hypothetical protein, conserved | No | 81 | unclassified |
| NCLIV_045430 | DNA-binding protein HU, putative | No | 80 | PB |
| NCLIV_031920 | ATP synthase gamma chain, putative | No | 79 | E |
| NCLIV_052880 | granule antigen protein GRA6, putative | No | 77 | CRDV |
| NCLIV_002020 | hypothetical protein | No | 76 | unclassified |
| NCLIV_003580 | hypothetical protein, conserved | No | 76 | unclassified |
| NCLIV_056480 | hypothetical protein | No | 76 | E |
| NCLIV_007800 | Tg65, related | Yes | 73 | CRDV |
| NCLIV_034990 | Transketolase, pyridine binding domain protein, related | Yes | 72 | M |
| NCLIV_047810 | hypothetical protein | No | 70 | M |

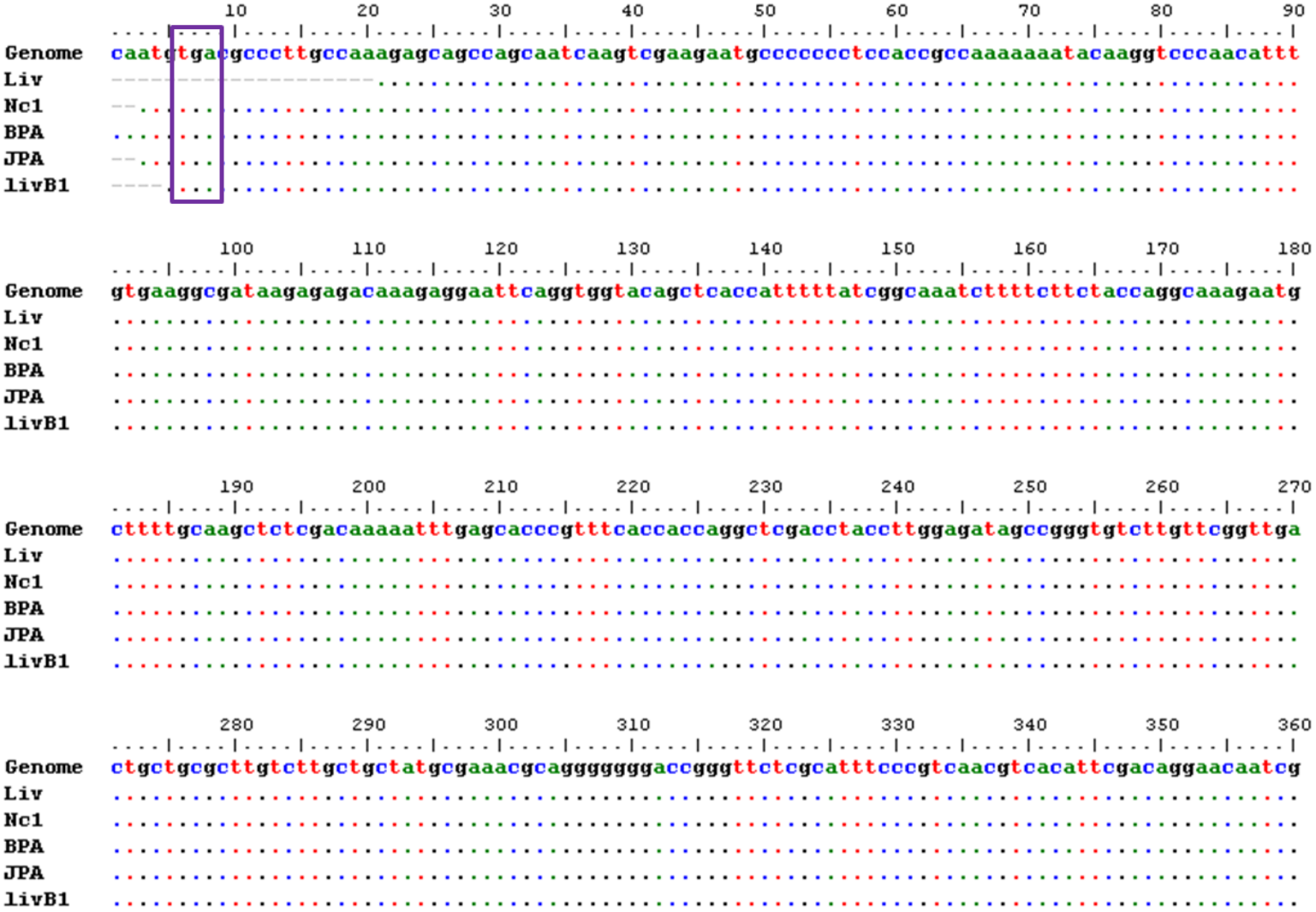
| | | | | |
|---|---|---|---|---|
| NCLIV_052190 | hypothetical protein, conserved | No | 70 | unclassified |
| NCLIV_005150 | hypothetical protein | No | 69 | CT |
| NCLIV_022270 | hypothetical protein, conserved | Yes | 69 | ? |
| NCLIV_023790 | hypothetical protein, conserved | No | 68 | unclassified |
| NCLIV_028140 | hypothetical protein, conserved | Yes | 68 | unclassified |
| NCLIV_043270 | microneme protein MIC1, putative | Yes | 67 | CRDV |
| NCLIV_043880 | hypothetical protein | No | 65 | CT |
| NCLIV_000940 | Glucose-6-phosphate dehydrogenase, putative | Yes | 64 | E |
| NCLIV_062290 | hypothetical protein, conserved | Yes | 64 | unclassified |
| NCLIV_007010 | cathepsin C2 (TgCPC2), putative | Yes | 63 | PB |
| NCLIV_018390 | hypothetical protein, conserved | No | 63 | unclassified |
| NCLIV_021640 | dense granule protein 7, putative | Yes | 63 | CRDV |
| NCLIV_051970 | MIC2-associated protein M2AP, putative | Yes | 63 | CRDV |
| NCLIV_043760 | hypothetical protein, conserved | Yes | 62 | unclassified |
| NCLIV_046260 | Iron regulatory protein-like protein, related | No | 62 | M |
| NCLIV_060700 | SRS domain-containing protein | No | 62 | IE |
| NCLIV_004220 | Rhoptry antigen ROP8, related | Yes | 61 | CRDV |
| NCLIV_036410 | cyst matrix protein, putative | Yes | 59 | IE |
| NCLIV_042510 | hypothetical protein, conserved | Yes | 59 | unclassified |
| NCLIV_002540 | GTP-binding protein, putative | No | 57 | CT |
| NCLIV_028240 | Ras family domain-containing protein, putative | No | 57 | CC |
| NCLIV_030620 | hypothetical protein, conserved | No | 57 | unclassified |
| NCLIV_033950 | Heat shock protein 70, related | No | 57 | CRDV |
| NCLIV_041930 | GE25707, related | No | 57 | CC |
| NCLIV_044690 | hypothetical protein, conserved | Yes | 57 | unclassified |
| NCLIV_060660 | SRS domain-containing protein,SRS57, putative | No | 57 | IE |
| NCLIV_001460 | Rhoptry kinase family protein ROP37 (incomplete catalytic triad), putative | No | 56 | CRDV |
| NCLIV_022950 | RNA-binding protein, putative | No | 56 | IE |
| NCLIV_033680 | Solute carrier family 25 (Mitochondrial carrier, dicarboxylate transporter), member 10, related | No | 56 | CT |
| NCLIV_057460 | Transketolase central region, related | Yes | 54 | M |
| NCLIV_026140 | Histone H2A, related | No | 52 | BG |
| NCLIV_004190 | thioredoxin, putative | Yes | 51 | IE |
| NCLIV_040550 | hypothetical protein, conserved | No | 51 | unclassified |
| NCLIV_045650 | 28 kDa antigen, putative | Yes | 51 | CRDV |
| NCLIV_034470 | hypothetical protein | No | 50 | PB |

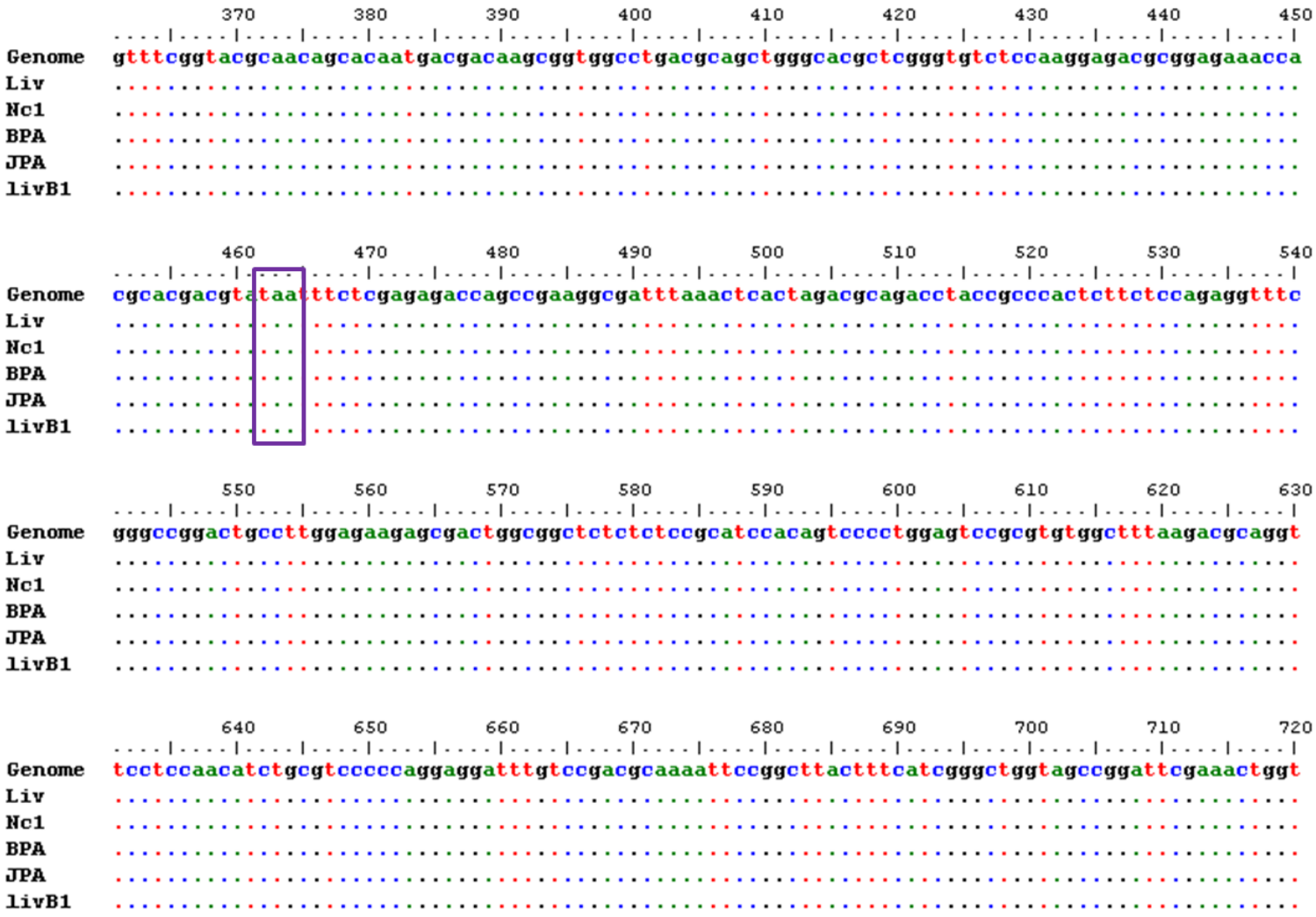
**Table V.2:** Additional protein identifications for the rhoptry/dense granule (R/DG) preparation (n = 28). Proteins were identified using Mascot and a cut off score of 50 was employed, proteins are presented in decreasing order of Mascot score. Apical proteins have been highlighted according to their identification in Chapter 3: yellow indicates microneme protein; blue, dense granule and purple, rhoptry. Signal peptides were predicted by SignalP 3.0 and downloaded from ToxoDB (cut off p-value ≥ 0.9). MIPs functional category assignments are abbreviated as follows: M = metabolism, E = energy, STP = storage protein, C/DNA = cell cycle and DNA processing, T = transcription, PS = protein synthesis, PF = protein fate. PB = protein with binding function or cofactor requirement, R = regulation of metabolism and protein function, CT = cellular transport, transport facilities and transport routes, CC = cellular communication/signal transduction mechanism, CRDV = cell rescue, defence and virulence, IE = interaction with the environment, CF = cell fate, BG = biogenesis of cellular components, SL = subcellular localisation and ? = potential ROP proteins.

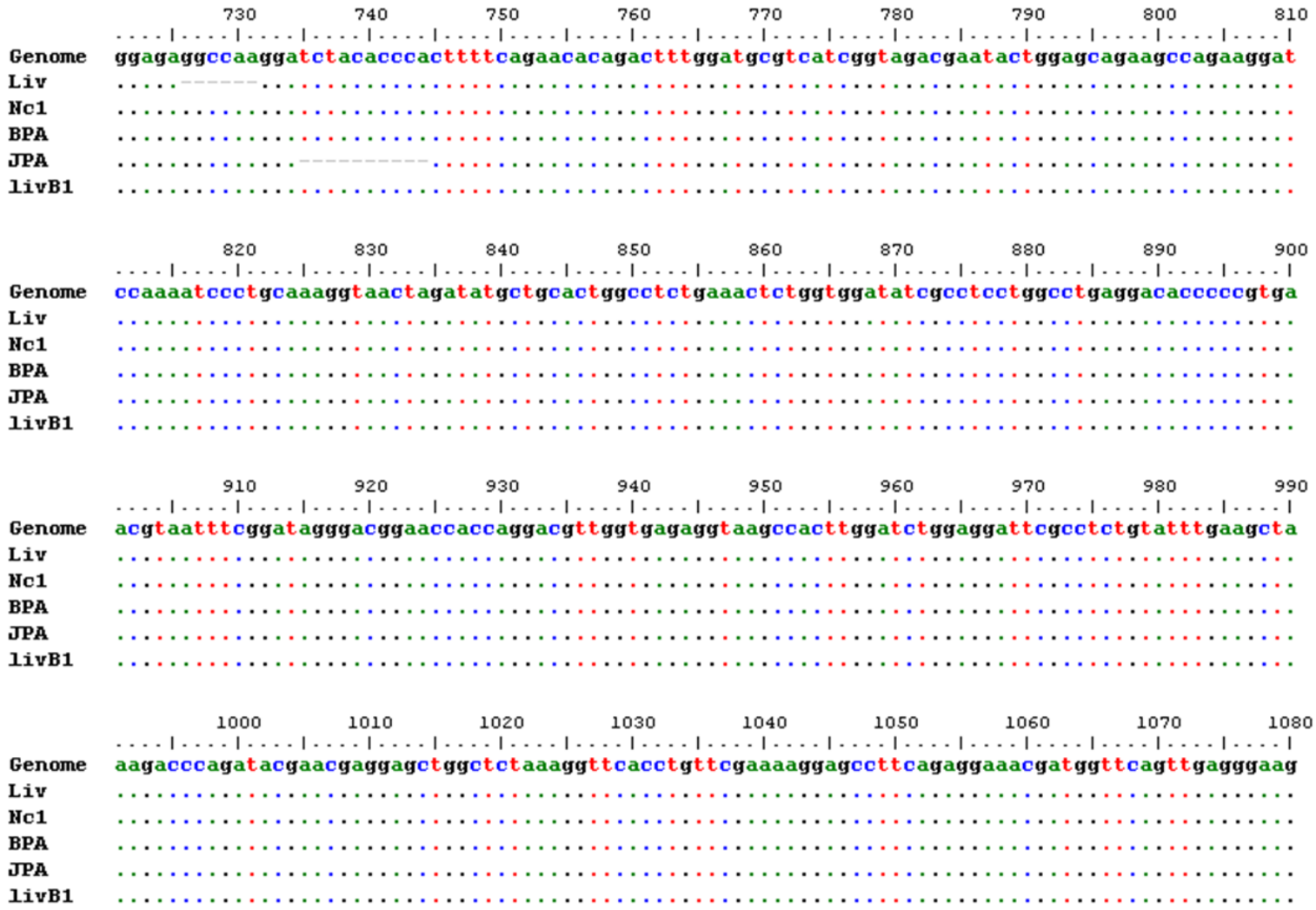
| Protein I.D. | Description | Signal peptide predicted? | Mascot score | MIPs category assignment |
|---|---|---|---|---|
| NCLIV_039100 | hypothetical protein | No | 200 | CT |
| NCLIV_012920 | Rhoptry kinase family protein ROP40, putative | No | 188 | CRDV |
| NCLIV_049900 | hypothetical protein | No | 165 | PB |
| NCLIV_068400 | NTPase | Yes | 148 | CRDV |
| NCLIV_029420 | myosin light chain TgMLC1, putative | No | 142 | CT |
| NCLIV_031770 | membrane skeletal protein IMC1, putative | No | 111 | BG |
| NCLIV_041940 | Glyceraldehyde 3-phosphate dehydrogenase, related | No | 97 | M |
| NCLIV_028540 | hypothetical protein, conserved | No | 88 | unclassified |
| NCLIV_032270 | hypothetical protein, conserved | No | 88 | unclassified |
| NCLIV_060140 | Inner membrane complex IMC3 | No | 86 | BG |
| NCLIV_002940 | microneme protein MIC4, putative | Yes | 84 | CRDV |
| NCLIV_017440 | hypothetical protein | Yes | 84 | CRDV |
| NCLIV_032780 | small heat shock protein 20, putative | No | 80 | CRDV |
| NCLIV_019770 | hypothetical protein | No | 79 | M |
| NCLIV_042000 | ubiquitin / ribosomal protein CEP52 fusion protein, putative | No | 75 | PS |
| NCLIV_050370 | hypothetical protein | No | 72 | E |
| NCLIV_016800 | TCP-1/cpn60 chaperonin family protein, putative | No | 71 | PF |
| NCLIV_052350 | hypothetical protein, conserved | No | 66 | unclassified |
| NCLIV_058890 | tubulin alpha chain | No | 64 | CT |
| NCLIV_015920 | Histone H4, related | No | 59 | BG |
| NCLIV_026340 | hypothetical protein | No | 58 | unclassified |
| NCLIV_027160 | hypothetical protein, conserved | No | 57 | unclassified |
| NCLIV_016220 | Rhoptry kinase family protein ROP23, putative | Yes | 55 | CRDV |
| NCLIV_019970 | Peptidyl-prolyl cis-trans isomerase A, related | No | 55 | PF |

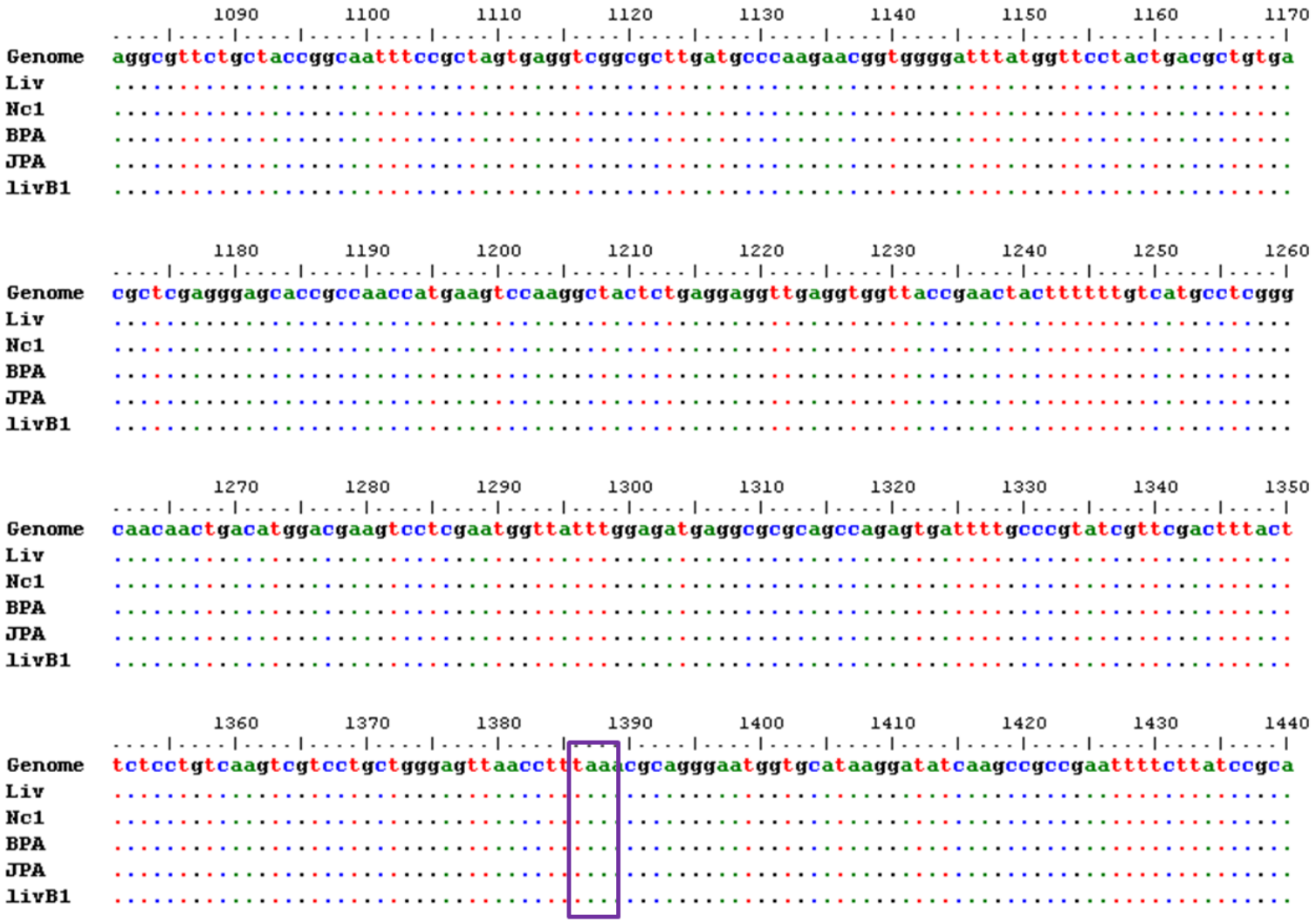
| | | | | |
|---|---|---|---|---|
| NCLIV_061830 | 60S acidic ribosomal protein P0 | No | 55 | PS |
| NCLIV_045870 | Putative dense granule protein 3 | Yes | 53 | CRDV |
| NCLIV_008990 | hypothetical protein | Yes | 50 | CRDV |
| NCLIV_038110 | hypothetical protein | Yes | 50 | CRDV |

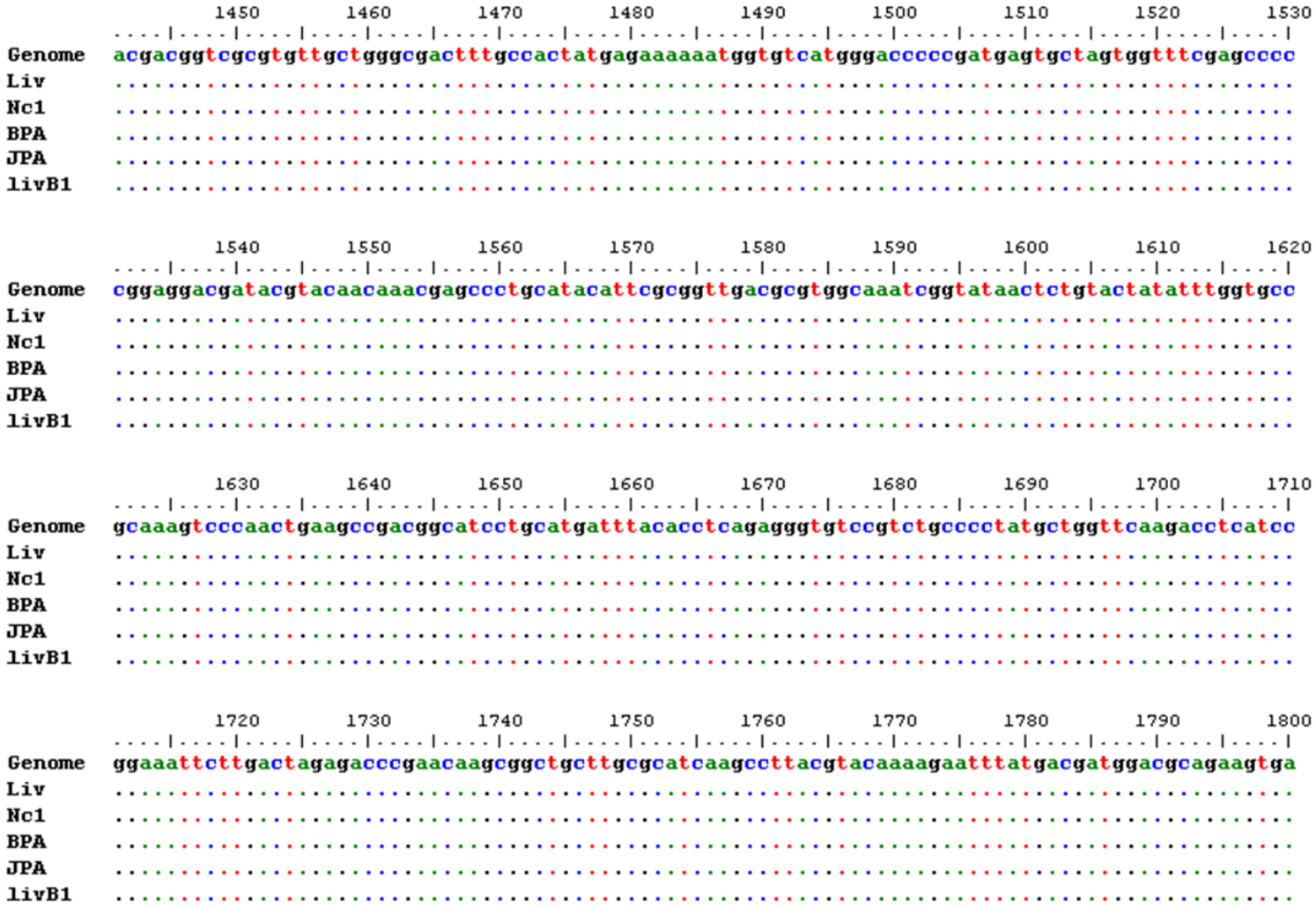
# APPENDIX VI: *ROP18* region in full, showing stop codons




370 380 390 400 410 420 430 440 450
Genome gtttcggtacgcaacagcacaatgacgacaagcggtggcctgacgcagctgggcacgctcgggtgtctccaaggagacgcggagaaacca
Liv
Nc1
BPA
JPA
livB1
460 470 480 490 500 510 520 530 540
Genome cgcacgacgtataatttctcgagagaccagccgaaggcgatttaaactcactagacgcagacctaccgcccactcttctccagaggtttc
Liv
Nc1
BPA
JPA
livB1
550 560 570 580 590 600 610 620 630
Genome gggccggactgccttggagaagagcgactggcggctctctctccgcatccacagtcccctggagtccgcgtgtggctttaagacgcaggt
Liv
Nc1
BPA
JPA
livB1
640 650 660 670 680 690 700 710 720
Genome tcctccaacatctgcgtcccccaggaggatttgtccgacgcaaaattccggcttactttcatcgggctggtagccggattcgaaactggt
Liv
Nc1
BPA
JPA
livB1


730 740 750 760 770 780 790 800 810
Genome ggagaggccaaggatctacacccacttttcagaacacagactttggatgcgtcatcggtagacgaatactggagcagaagccagaaggat
Liv .....------..........
Nc1 ..........
BPA ..........
JPA ..........
livB1 ..........
820 830 840 850 860 870 880 890 900
Genome ccaaaatccctgcaaaggtaactagatatgctgcactggcctctgaaactctggtggatatcgcctcctggcctgaggacacccccgtga
Liv ..........
Nc1 ..........
BPA ..........
JPA ..........
livB1 ..........
910 920 930 940 950 960 970 980 990
Genome acgtaatttcggatagggacggaaccaccaggacgttggtgagaggtaagccacttggatctggaggattcgcctctgtatttgaagcta
Liv ..........
Nc1 ..........
BPA ..........
JPA ..........
livB1 ..........
1000 1010 1020 1030 1040 1050 1060 1070 1080
Genome aagacccagatacgaacgaggagctggctctaaaggttcacctgttcgaaaaggagccttcagaggaaacgatggttcagttgagggaag
Liv ..........
Nc1 ..........
BPA ..........
JPA ..........
livB1 ..........


1090 1100 1110 1120 1130 1140 1150 1160 1170
Genome aggcgttctgctaccggcaatttccgctagtgaggtcggcgcttgatgcccaagaacggtggggatttatggttcctactgacgctgtga
Liv
Nc1
BPA
JPA
livB1
1180 1190 1200 1210 1220 1230 1240 1250 1260
Genome cgctcgagggagcaccgccaaccatgaagtccaaggctactctgaggaggttgaggtggttaccgaactacttttttgtcatgcctcggg
Liv
Nc1
BPA
JPA
livB1
1270 1280 1290 1300 1310 1320 1330 1340 1350
Genome caacaactgacatggacgaagtcctcgaatggttatttggagatgaggcgcgcagccagagtgattttgcccgtatcgttcgactttact
Liv
Nc1
BPA
JPA
livB1
1360 1370 1380 1390 1400 1410 1420 1430 1440
Genome tctcctgtcaagtcgtcctgctgggagttaacctttaaacgcagggaatggtgcataaggatatcaagccgccgaattttcttatccgca
Liv
Nc1
BPA
JPA
livB1


1450 1460 1470 1480 1490 1500 1510 1520 1530
Genome acgacggtcgcgtgttgctgggcgactttgccactatgagaaaaaatggtgtcatgggacccccgatgagtgctagtggtttcgagcccc
Liv
Nc1
BPA
JPA
livB1
1540 1550 1560 1570 1580 1590 1600 1610 1620
Genome cggaggacgatacgtacaacaaacgagccctgcatacattcgcggttgacgcgtggcaaatcggtataactctgtactatatttggtgcc
Liv
Nc1
BPA
JPA
livB1
1630 1640 1650 1660 1670 1680 1690 1700 1710
Genome gcaaagtcccaactgaagccgacggcatcctgcatgatttacacctcagagggtgtccgtctgcccctatgctggttcaagacctcatcc
Liv
Nc1
BPA
JPA
livB1
1720 1730 1740 1750 1760 1770 1780 1790 1800
Genome ggaaattcttgactagagacccgaacaagcggctgcttgcgcatcaagccttacgtacaaaagaatttatgacgatggacgcagaagtga
Liv
Nc1
BPA
JPA
livB1

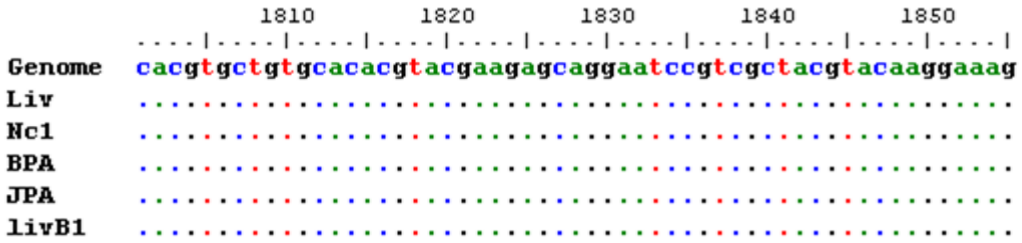



**Figure 5.2:** The *ROP18* region of five strains of *N. caninum* amplified by PCR and aligned to the reference genome using BioEdit version 7.0.5.3. Stop codons are identified by purple boxes. This region is displayed in the 5' to 3' orientation, transcribed from the antisense strand of chromosome VIIIa. No SNPs were identified in any of the isolates analysed

# REFERENCES


Abbott, A. (2003) 'Cell culture: biology's new dimension', *Nature*, vol. 424, no. 6951, pp. 870-872.

Ahn, H.J., Kim, S., Kim, H.E. & Nam, H.W. (2006) 'Interactions between secreted GRA proteins and host cell proteins across the paratitophorous vacuolar membrane in the parasitism of Toxoplasma gondii', *Korean J Parasitol*, vol. 44, no. 4, pp. 303-312.

Ahn, H.J., Song, K.J., Son, E.S., Shin, J.C. & Nam, H.W. (2001) 'Protease activity and host cell binding of the 42-kDa rhoptry protein from Toxoplasma gondii after secretion', *Biochem Biophys Res Commun*, vol. 287, no. 3, pp. 630-635.

Alexander, D.L., Mital, J., Ward, G.E., Bradley, P. & Boothroyd, J.C. (2005) 'Identification of the moving junction complex of Toxoplasma gondii: a collaboration between distinct secretory organelles', *PLoS Pathog*, vol. 1, no. 2, p. e17.

Almeria, S., Ferrer, D., Pabon, M., Castella, J. & Manas, S. (2002) 'Red foxes (Vulpes vulpes) are a natural intermediate host of Neospora caninum', *Vet Parasitol*, vol. 107, no. 4, pp. 287-294.

Al-Qassab, S., Reichel, M.P. & Ellis, J. (2010) 'A second generation multiplex PCR for typing strains of Neospora caninum using six DNA targets', *Mol Cell Probes*, vol. 24, no. 1, pp. 20-26.

Al-Qassab, S., Reichel, M.P., Ivens, A. & Ellis, J.T. (2009) 'Genetic diversity amongst isolates of Neospora caninum, and the development of a multiplex assay for the detection of distinct strains', *Mol Cell Probes*, vol. 23, no. 3-4, pp. 132-139.

Al-Quassab, S., Reichel, M.P. & Ellis, J.T. (2010) 'On the Biological and Genetic Diversity in Neospora caninum', *Diversity*, vol. 2010, no. 2, pp. 411-438.

Anders, S. & Huber, W. (2010) 'Differential expression analysis for sequence count data', *Genome Biol*, vol. 11, no. 10, p. R106.

Anderson, M.L., Reynolds, J.P., Rowe, J.D., Sverlow, K.W., Packham, A.E., Barr, B.C. & Conrad, P.A. (1997) 'Evidence of vertical transmission of Neospora sp infection in dairy cattle', *J Am Vet Med Assoc*, vol. 210, no. 8, pp. 1169-1172.

Andrianarivo, A.G., Barr, B.C., Anderson, M.L., Rowe, J.D., Packham, A.E., Sverlow, K.W. & Conrad, P.A. (2001) 'Immune responses in pregnant cattle and bovine fetuses following experimental infection with Neospora caninum', *Parasitol Res*, vol. 87, no. 10, pp. 817-825.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. & Sherlock, G. (2000) 'Gene ontology: tool for the unification of biology. The Gene Ontology Consortium', *Nat Genet*, vol. 25, no. 1, pp. 25-29.

Assossou, O., Besson, F., Rouault, J.P., Persat, F., Ferrandiz, J., Mayencon, M., Peyron, F. & Picot, S. (2004) 'Characterization of an excreted/secreted antigen form of 14-3-3 protein in Toxoplasma gondii tachyzoites', *FEMS Microbiol Lett*, vol. 234, no. 1, pp. 19-25.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J.C., Kraemer, E., Li, W., Miller, J.A., Nayak, V., Pennington, C., Pinney, D.F., Roos, D.S., Ross, C., Stoeckert, C.J., Jr., Treatman, C. & Wang, H. (2009) 'PlasmoDB: a functional genomic database for malaria parasites', *Nucleic Acids Res*, vol. 37, no. Database issue, pp. D539-543.

Aurrecoechea, C., Heiges, M., Wang, H., Wang, Z., Fischer, S., Rhodes, P., Miller, J., Kraemer, E., Stoeckert, C.J., Jr., Roos, D.S. & Kissinger, J.C. (2007) 'ApiDB: integrated resources for the apicomplexan bioinformatics resource center', *Nucleic Acids Res*, vol. 35, no. Database issue, pp. D427-430.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M.J., Gajria, B., Grant, G.R., Ginsburg, H., Gupta, D., Kissinger, J.C., Labo, P., Li, L., Mailman, M.D., Milgram, A.J., Pearson, D.S., Roos, D.S., Schug, J., Stoeckert, C.J., Jr. & Whetzel, P. (2003) 'PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data', *Nucleic Acids Res*, vol. 31, no. 1, pp. 212-215.

Bannister, L.H., Hopkins, J.M., Dluzewski, A.R., Margos, G., Williams, I.T., Blackman, M.J., Kocken, C.H., Thomas, A.W. & Mitchell, G.H. (2003) 'Plasmodium falciparum apical membrane antigen 1 (PfAMA-1) is translocated within micronemes along subpellicular microtubules during merozoite development', *J Cell Sci*, vol. 116, no. Pt 18, pp. 3825-3834.

Barber, J., Trees, A.J., Owen, M. & Tennant, B. (1993) 'Isolation of Neospora caninum from a British dog', *Vet Rec*, vol. 133, no. 21, pp. 531-532.

Barber, J.S. & Trees, A.J. (1998) 'Naturally occuring vertical transmission of *Neospora caninum* in dogs', *International Journal for Parasitology*, vol. 28, pp. 57-64.

Barling, K.S., McNeill, J.W., Thompson, J.A., Paschal, J.C., McCollum, F.T., 3rd, Craig, T.M. & Adams, L.G. (2000) 'Association of serologic status for Neospora caninum with postweaning weight gain and carcass measurements in beef calves', *J Am Vet Med Assoc*, vol. 217, no. 9, pp. 1356-1360.

Beckers, C.J., Dubremetz, J.F., Mercereau-Puijalon, O. & Joiner, K.A. (1994) 'The Toxoplasma gondii rhoptry protein ROP 2 is inserted into the parasitophorous vacuole membrane, surrounding the intracellular parasite, and is exposed to the host cell cytoplasm', *J Cell Biol*, vol. 127, no. 4, pp. 947-961.

Beckers, C.J., Wakefield, T. & Joiner, K.A. (1997) 'The expression of Toxoplasma proteins in Neospora caninum and the identification of a gene encoding a novel rhoptry protein', *Mol Biochem Parasitol*, vol. 89, no. 2, pp. 209-223.

Behnke, M.S., Khan, A., Wootton, J.C., Dubey, J.P., Tang, K. & Sibley, L.D. (2011) 'Virulence differences in Toxoplasma mediated by amplification of a family of polymorphic pseudokinases', *Proc Natl Acad Sci U S A*, vol. 108, no. 23, pp. 9631-9636.

Behnke, M.S., Wootton, J.C., Lehmann, M.M., Radke, J.B., Lucas, O., Nawas, J., Sibley, L.D. & White, M.W. (2010) 'Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of Toxoplasma gondii', *PLoS One*, vol. 5, no. 8, p. e12354.

Bekpen, C., Hunn, J.P., Rohde, C., Parvanova, I., Guethlein, L., Dunn, D.M., Glowalla, E., Leptin, M. & Howard, J.C. (2005) 'The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage', *Genome Biol*, vol. 6, no. 11, p. R92.

Bell, A.W., Deutsch, E.W., Au, C.E., Kearney, R.E., Beavis, R., Sechi, S., Nilsson, T., Bergeron, J.J.M. & Group, H.T.S.W. (2009) 'A HUPO test sample study reveals common problems in mass spectrometry-based proteomics', *Nature Methods*.

Belli, S.I., Walker, R.A. & Flowers, S.A. (2005) 'Global protein expression analysis in apicomplexan parasites: current status', *Proteomics*, vol. 5, no. 4, pp. 918-924.

Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. (2004) 'Improved prediction of signal peptides: SignalP 3.0', *J Mol Biol*, vol. 340, no. 4, pp. 783-795.

Bennett, S.T., Barnes, C., Cox, A., Davies, L. & Brown, C. (2005) 'Toward the 1,000 dollars human genome', *Pharmacogenomics*, vol. 6, no. 4, pp. 373-382.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara, E.C.M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, A.C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, J., Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R. & Smith, A.J. (2008) 'Accurate whole human genome sequencing using reversible terminator chemistry', *Nature*, vol. 456, no. 7218, pp. 53-59.

Besteiro, S., Dubremetz, J.F. & Lebrun, M. (2011) 'The moving junction of apicomplexan parasites: a key structure for invasion', *Cell Microbiol*, vol. 13, no. 6, pp. 797-805.

Besteiro, S., Michelin, A., Poncet, J., Dubremetz, J.F. & Lebrun, M. (2009) 'Export of a Toxoplasma gondii rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion', *PLoS Pathog*, vol. 5, no. 2, p. e1000309.

Bjerkas, I., Mohn, S.F. & Presthus, J. (1984) 'Unidentified cyst-forming sporozoon causing encephalomyelitis and myositis in dogs', *Z Parasitenkd*, vol. 70, no. 2, pp. 271-274.

Bjorkman, C., Johansson, O., Stenlund, S., Holmdahl, O.J. & Uggla, A. (1996) 'Neospora species infection in a herd of dairy cattle', *J Am Vet Med Assoc*, vol. 208, no. 9, pp. 1441-1444.

Blood, D.C., Studdert, V.P. & Gay, C.C. (2007) *Saunders Comprehensive Veterinary Dictionary, 3 ed.*, Elsevier, Inc.

Bonhomme, A., Maine, G.T., Beorchia, A., Burlet, H., Aubert, D., Villena, I., Hunt, J., Chovan, L., Howard, L., Brojanac, S., Sheu, M., Tyner, J., Pluot, M. & Pinon, J.M. (1998) 'Quantitative immunolocalization of a P29 protein (GRA7), a new antigen of toxoplasma gondii', *J Histochem Cytochem*, vol. 46, no. 12, pp. 1411-1422.

Boothroyd, J.C. & Dubremetz, J.F. (2008) 'Kiss and spit: the dual roles of *Toxoplasma* rhoptries', *Nature Reviews | Microbiology*, vol. 6, pp. 79-88.

Boothroyd, J.C. (2009) 'Expansion of host range as a driving force in the evolution of Toxoplasma', *Mem Inst Oswaldo Cruz*, vol. 104, no. 2, pp. 179-184.

Boysen, P., Klevar, S., Olsen, I. & Storset, A.K. (2006) 'The protozoan Neospora caninum directly triggers bovine NK cells to produce gamma interferon and to kill infected fibroblasts', *Infect Immun*, vol. 74, no. 2, pp. 953-960.

Bradford, M.M. (1976) 'A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding', *Analytical Biochemistry*, vol. 72, pp. 248-254.

Bradley, P.J., Ward, C., Cheng, S.J., Alexander, D.L., Coller, S., Coombs, G.H., Dunn, J.D., Ferguson, D.J., Sanderson, S.J., Wastling, J.M. & Boothroyd, J.C. (2005) 'Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii', *J Biol Chem*, vol. 280, no. 40, pp. 34245-34258.

Buchholz, K.R., Fritz, H.M., Chen, X., Durbin-Johnson, B., Rocke, D.M., Ferguson, D.J., Conrad, P.A. & Boothroyd, J.C. (2011) 'Identification of tissue Cyst Wall Components by Transcriptome Analysis of *In Vivo* and *In Vitro Toxoplasma gondii* Bradyzoites' *Eukaryotic Cell*, vol. 10, no.12, pp.1637-1647.

Burgoyne, R.D. & Morgan, A. (1993) 'Regulated exocytosis', *Biochem J*, vol. 293 ( Pt 2), pp. 305-316.

Butcher, B.A., Fox, B.A., Rommereim, L.M., Kim, S.G., Maurer, K.J., Yarovinsky, F., Herbert, D.R., Bzik, D.J. & Denkers, E.Y. (2011) 'Toxoplasma gondii rhoptry kinase ROP16 activates STAT3 and STAT6 resulting in cytokine inhibition and arginase-1-dependent growth control', *PLoS Pathog*, vol. 7, no. 9, p. e1002236.

Buxton, D., McAllister, M.M. & Dubey, J.P. (2002) 'The comparative pathogenesis of neosporosis', *Trends in Parasitology*, vol. 18, no. 12, pp. 546-552.

Carruthers, V.B. & Sibley, L.D. (1997) 'Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts', *Eur J Cell Biol*, vol. 73, no. 2, pp. 114-123.

Carruthers, V.B. & Sibley, L.D. (1999) 'Mobilization of intracellular calcium stimulates microneme discharge in Toxoplasma gondii', *Mol Microbiol*, vol. 31, no. 2, pp. 421-428.

Carruthers, V.B., Giddings, O.K. & Sibley, L.D. (1999) 'Secretion of micronemal proteins is associated with toxoplasma invasion of host cells', *Cell Microbiol*, vol. 1, no. 3, pp. 225-235.

Carruthers, V.B., Moreno, N.J. & Sibley, L.D. (1999) 'Ethanol and acetaldehyde elevate intracellular [$Ca^{2+}$] and stimulate microneme discharge in *Toxoplasma gondii*', *Biochemical Journal*, vol. 342, pp. 379-386.

Cerede, O., Dubremetz, J.F., Soete, M., Deslee, D., Vial, H., Bout, D. & Lebrun, M. (2005) 'Synergistic role of micronemal proteins in Toxoplasma gondii virulence', *J Exp Med*, vol. 201, no. 3, pp. 453-463.

Chen, F., Mackey, A.J., Stoeckert, C.J., Jr. & Roos, D.S. (2006) 'OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups', *Nucleic Acids Res*, vol. 34, no. Database issue, pp. D363-368.

Chen, X.M., O'Hara, S.P., Huang, B.Q., Nelson, J.B., Lin, J.J., Zhu, G., Ward, H.D. & LaRusso, N.F. (2004) 'Apical organelle discharge by Cryptosporidium parvum is temperature, cytoskeleton, and intracellular calcium dependent and required for host cell invasion', *Infect Immun*, vol. 72, no. 12, pp. 6806-6816.

Coers, J., Bernstein-Hanley, I., Grotsky, D., Parvanova, I., Howard, J.C., Taylor, G.A., Dietrich, W.F. & Starnbach, M.N. (2008) 'Chlamydia muridarum evades growth restriction by the IFN-gamma-inducible host resistance factor Irgb10', *J Immunol*, vol. 180, no. 9, pp. 6237-6245.

Conrad, P.A., Barr, B.C., Sverlow, K.W., Anderson, M., Daft, B., Kinde, H., Dubey, J.P., Munson, L. & Ardans, A. (1993) 'In vitro isolation and characterization of a Neospora sp. from aborted bovine foetuses', *Parasitology*, vol. 106 ( Pt 3), pp. 239-249.

Cowper, B., Matthews, S. & Tomley, F. (2012) 'The molecular basis for the distinct host and tissue tropisms of coccidian parasites', *Mol Biochem Parasitol*.

Craig, R. & Beavis, R.C. (2004) 'TANDEM: matching proteins with tandem mass spectra', *Bioinformatics*, vol. 20, no. 9, pp. 1466-1467.

Craig, R., Cortens, J.P. & Beavis, R.C. (2005) 'The use of proteotypic peptide libraries for protein identification', *Rapid Commun Mass Spectrom*, vol. 19, no. 13, pp. 1844-1850.

Dabritz, H.A., Miller, M.A., Atwill, E.R., Gardner, I.A., Leutenegger, C.M., Melli, A.C. & Conrad, P.A. (2007) 'Detection of Toxoplasma gondii-like oocysts in cat feces and estimates of the environmental oocyst burden', *J Am Vet Med Assoc*, vol. 231, no. 11, pp. 1676-1684.

Dautu, G., Ueno, A., Munyaka, B., Carmen, G., Makino, S., Kobayashi, Y. & Igarashi, M. (2008) 'Molecular and biochemical characterization of Toxoplasma gondii beta-hydroxyacyl-acyl carrier protein dehydratase (FABZ)', *Parasitol Res*, vol. 102, no. 6, pp. 1301-1309.

Davison, H.C., Otter, A. & Trees, A.J. (1999) 'Estimation of vertical and horizontal transmission parameters of Neospora caninum infections in dairy cattle', *Int J Parasitol*, vol. 29, no. 10, pp. 1683-1689.

Davison, H.C., Otter, A. & Trees, A.J. (1999b) 'Significance of Neospora caninum in British dairy cattle determined by estimation of seroprevalence in normally calving cattle and aborting cattle', *Int J Parasitol*, vol. 29, no. 8, pp. 1189-1194.

Davison, H.C., Trees, A.J., Guy, F., Otter, A., Holt, J.J., Simpson, V.R. & Jeffrey, M. (1997) 'Isolation of bovine Neospora in Britain', *Vet Rec*, vol. 141, no. 23, p. 607.

De Marez, T., Liddell, S., Dubey, J.P., Jenkins, M.C. & Gasbarre, L. (1999) 'Oral infection of calves with Neospora caninum oocysts from dogs: humoral and cellular immune responses', *Int J Parasitol*, vol. 29, no. 10, pp. 1647-1657.

de Melo, E.J., de Carvalho, T.U. & de Souza, W. (1992) 'Penetration of Toxoplasma gondii into host cells induces changes in the distribution of the mitochondria and the endoplasmic reticulum', *Cell Struct Funct*, vol. 17, no. 5, pp. 311-317.

DeBarry, J.D. & Kissinger, J.C. (2011) 'Jumbled genomes: missing Apicomplexan synteny', *Mol Biol Evol*, vol. 28, no. 10, pp. 2855-2871.

Del Carmen, M.G., Mondragon, M., Gonzalez, S. & Mondragon, R. (2009) 'Induction and regulation of conoid extrusion in Toxoplasma gondii', *Cell Microbiol*, vol. 11, no. 6, pp. 967-982.

Delorme, V., Cayla, X., Faure, G., Garcia, A. & Tardieux, I. (2003) 'Actin dynamics is controlled by a casein kinase II and phosphatase 2C interplay on Toxoplasma gondii Toxofilin', *Mol Biol Cell*, vol. 14, no. 5, pp. 1900-1912.

Denkers, E.Y., Bzik, D.J., Fox, B.A. & Butcher, B.A. (2012) 'An inside job: hacking into Janus kinase/signal transducer and activator of transcription signaling cascades by the intracellular protozoan Toxoplasma gondii', *Infect Immun*, vol. 80, no. 2, pp. 476-482.

Dijkstra, T., Eysker, M., Schares, G., Conraths, F.J., Wouda, W. & Barkema, H.W. (2001) 'Dogs shed Neospora caninum oocysts after ingestion of naturally infected bovine placenta but not after ingestion of colostrum spiked with Neospora caninum tachyzoites', *Int J Parasitol*, vol. 31, no. 8, pp. 747-752.

Dobrowolski, J.M. & Sibley, L.D. (1996) 'Toxoplasma invasion of mammalian cells is powered by the actin cytoskeleton of the parasite', *Cell*, vol. 84, no. 6, pp. 933-939.

Dubey, J.P. (1999a) 'Neosporosis - the first decade of research', *International Journal for Parasitology*, vol. 29, pp. 1485-1488.

Dubey, J.P. (1999b) 'Recent advances in Neospora and neosporosis', *Vet Parasitol*, vol. 84, no. 3-4, pp. 349-367.

Dubey, J.P. & Frenkel, J.K. (1972) 'Cyst-induced toxoplasmosis in cats', *J Protozool*, vol. 19, no. 1, pp. 155-177.

Dubey, J.P. & Lindsay, D.S. (1996) 'A review of Neospora caninum and neosporosis', *Vet Parasitol*, vol. 67, no. 1-2, pp. 1-59.

Dubey, J.P. & Su, C. (2009) 'Population biology of Toxoplasma gondii: what's out and where did they come from', *Mem Inst Oswaldo Cruz*, vol. 104, no. 2, pp. 190-195.

Dubey, J.P. (1995) 'Duration of immunity to shedding of Toxoplasma gondii oocysts by cats', *J Parasitol*, vol. 81, no. 3, pp. 410-415.

Dubey, J.P. (2003) 'Review of Neospora caninum and neosporosis in animals', *Korean J Parasitol*, vol. 41, no. 1, pp. 1-16.

Dubey, J.P. (2005) Neosporosis in Dogs, *30th World Congress of the WSAVA*, Mexico City, Mexico.

Dubey, J.P., Buxton, D. & Wouda, W. (2006) 'Pathogenesis of bovine neosporosis', *J Comp Pathol*, vol. 134, no. 4, pp. 267-289.

Dubey, J.P., Carpenter, J.L., Speer, C.A., Topper, M.J. & Uggla, A. (1988a) 'Newly recognized fatal protozoan disease of dogs', *J Am Vet Med Assoc*, vol. 192, no. 9, pp. 1269-1285.

Dubey, J.P., Hattel, A.L., Lindsay, D.S. & Topper, M.J. (1988) 'Neonatal Neospora caninum infection in dogs: isolation of the causative agent and experimental transmission', *J Am Vet Med Assoc*, vol. 193, no. 10, pp. 1259-1263.

Dubey, J.P., Hattel, A.L., Lindsay, D.S. & Topper, M.J. (1988b) 'Neonatal Neospora caninum infection in dogs: isolation of the causative agent and experimental transmission', *J Am Vet Med Assoc*, vol. 193, no. 10, pp. 1259-1263.

Dubey, J.P., Jenkins, M.C., Rajendran, C., Miska, K., Ferreira, L.R., Martins, J., Kwok, O.C. & Choudhary, S. (2011) 'Gray wolf (Canis lupus) is a natural definitive host for Neospora caninum', *Vet Parasitol*, vol. 181, no. 2-4, pp. 382-387.

Dubey, J.P., Koestner, A. & Piper, R.C. (1990) 'Repeated transplacental transmission of Neospora caninum in dogs', *J Am Vet Med Assoc*, vol. 197, no. 7, pp. 857-860.

Dubey, J.P., Lindsay, D.S. & Speer, C.A. (1998) 'Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts', *Clin Microbiol Rev*, vol. 11, no. 2, pp. 267-299.

Dubey, J.P., Schares, G. & Ortega-Mora, L.M. (2007) 'Epidemiology and Control of Neosporosis and *Neospora caninum*', *Clinical Microbiology Reviews*, vol. 20, no. 2, pp. 323-367.

Dubey, J.P., Sreekumar, C., Knickman, E., Miska, K.B., Vianna, M.C.B., Kwok, O.C.H., Hill, D.E., Jenkins, M.C., Lindsay, D.S. & Greene, C.E. (2004) 'Biologic, morphologic, and molecular characterisation of *Neospora caninum* isolates from littermate dogs', *International Journal for Parasitology*, vol. 34, pp. 1157-1167.

Dubey, J.P., Vianna, M.C., Kwok, O.C., Hill, D.E., Miska, K.B., Tuo, W., Velmurugan, G.V., Conors, M. & Jenkins, M.C. (2007) 'Neosporosis in Beagle dogs: clinical signs, diagnosis, treatment, isolation and genetic characterization of Neospora caninum', *Vet Parasitol*, vol. 149, no. 3-4, pp. 158-166.

Dubremetz, J.F., Garcia-Reguet, N., Conseil, V. & Fourmaux, M.N. (1998) 'Apical organelles and host-cell invasion by Apicomplexa', *Int J Parasitol*, vol. 28, no. 7, pp. 1007-1013.

Dunn, J.D., Ravindran, S., Kim, S.K. & Boothroyd, J.C. (2008) 'The Toxoplasma gondii dense granule protein GRA7 is phosphorylated upon invasion and forms an unexpected association with the rhoptry proteins ROP2 and ROP4', *Infect Immun*, vol. 76, no. 12, pp. 5853-5861.

Dytham, C. (2003) *Choosing and using statistics-A biologist's guide' second edition*, Blackwell Science

El Hajj, H., Demey, E., Poncet, J., Lebrun, M., Wu, B., Galeotti, N., Fourmaux, M.N., Mercereau-Puijalon, O., Vial, H., Labesse, G. & Dubremetz, J.F. (2006) 'The ROP2 family of Toxoplasma gondii rhoptry proteins: proteomic and genomic characterization and molecular modeling', *Proteomics*, vol. 6, no. 21, pp. 5773-5784.

El Hajj, H., Lebrun, M., Arold, S.T., Vial, H., Labesse, G. & Dubremetz, J.F. (2007) 'ROP18 is a rhoptry kinase controlling the intracellular proliferation of Toxoplasma gondii', *PLoS Pathog*, vol. 3, no. 2, p. e14.

Eng, J.K., McCormack, A.L. & Yates, J.R. (1994) 'An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database', *Journal of the American Society for Mass Spectrometry*, vol. 5, pp. 976-989.

Entrican, G. (2002) 'Immune regulation during pregnancy and host-pathogen interactions in infectious abortion', *J Comp Pathol*, vol. 126, no. 2-3, pp. 79-94.

Ewald, P.W. (1993) 'The evolution of virulence', *Sci Am*, vol. 268, no. 4, pp. 86-93.

Fentress, S.J., Behnke, M.S., Dunay, I.R., Mashayekhi, M., Rommereim, L.M., Fox, B.A., Bzik, D.J., Taylor, G.A., Turk, B.E., Lichti, C.F., Townsend, R.R., Qiu, W., Hui, R., Beatty, W.L. &

Sibley, L.D. (2010) 'Phosphorylation of immunity-related GTPases by a Toxoplasma gondii-secreted kinase promotes macrophage survival and virulence', *Cell Host Microbe*, vol. 8, no. 6, pp. 484-495.

Fenyo, D. & Beavis, R.C. (2003) 'A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes', *Analytical Chemistry*, vol. 75, no. 4, pp. 768-774.

Foth, B.J. & McFadden, G.I. (2003) 'The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites', *Int Rev Cytol*, vol. 224, pp. 57-110.

Friedrich, N., Santos, J.M., Liu, Y., Palma, A.S., Leon, E., Saouros, S., Kiso, M., Blackman, M.J., Matthews, S., Feizi, T. & Soldati-Favre, D. (2010) 'Members of a novel protein family containing microneme adhesive repeat domains act as sialic acid-binding lectins during host cell invasion by apicomplexan parasites', *J Biol Chem*, vol. 285, no. 3, pp. 2064-2076.

Fu, H., Subramanian, R.R. & Masters, S.C. (2000) '14-3-3 proteins: structure, function, and regulation', *Annu Rev Pharmacol Toxicol*, vol. 40, pp. 617-647.

Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J.C., Mackey, A.J., Pinney, D.F., Roos, D.S., Stoeckert, C.J., Jr., Wang, H. & Brunk, B.P. (2008) 'ToxoDB: an integrated Toxoplasma gondii database resource', *Nucleic Acids Res*, vol. 36, no. Database issue, pp. D553-556.

Garnett, J.A., Liu, Y., Leon, E., Allman, S.A., Friedrich, N., Saouros, S., Curry, S., Soldati-Favre, D., Davis, B.G., Feizi, T. & Matthews, S. (2009) 'Detailed insights from microarray and crystallographic studies into carbohydrate recognition by microneme protein 1 (MIC1) of Toxoplasma gondii', *Protein Sci*, vol. 18, no. 9, pp. 1935-1947.

Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W. & Bryant, S.H. (2004) 'Open mass spectrometry search algorithm', *J Proteome Res*, vol. 3, no. 5, pp. 958-964.

Ghosh, A., Uthaiah, R., Howard, J., Herrmann, C. & Wolf, E. (2004) 'Crystal structure of IIGP1: a paradigm for interferon-inducible p47 resistance GTPases', *Mol Cell*, vol. 15, no. 5, pp. 727-739.

Gibney, E.H., Kipar, A., Rosbottom, A., Guy, C.S., Smith, R.F., Hetzel, U., Trees, A.J. & Williams, D.J.L. (2008) 'The extent of parasite-associated necrosis in the placenta and foetal tissues of cattle following Neospora caninum infection in early and late gestation correlates with foetal death', *International Journal for Parasitology*, vol. 38, pp. 579-588.

Gilbert, L.A., Ravindran, S., Turetzky, J.M., Boothroyd, J.C. & Bradley, P.J. (2007) 'Toxoplasma gondii targets a protein phosphatase 2C to the nuclei of infected host cells', *Eukaryot Cell*, vol. 6, no. 1, pp. 73-83.

Gondim, L.F., Gao, L. & McAllister, M.M. (2002) 'Improved production of Neospora caninum oocysts, cyclical oral transmission between dogs and cattle, and in vitro isolation from oocysts', *J Parasitol*, vol. 88, no. 6, pp. 1159-1163.

Gondim, L.F., McAllister, M.M. & Gao, L. (2005) 'Effects of host maturity and prior exposure history on the production of Neospora caninum oocysts by dogs', *Vet Parasitol*, vol. 134, no. 1-2, pp. 33-39.

Gondim, L.F., McAllister, M.M., Pitt, W.C. & Zemlicka, D.E. (2004) 'Coyotes (Canis latrans) are definitive hosts of Neospora caninum', *Int J Parasitol*, vol. 34, no. 2, pp. 159-161.

Gorg, A., Weiss, W. & Dunn, M.J. (2004) 'Current two-dimensional electrophoresis technology for proteomics', *Proteomics*, vol. 4, no. 12, pp. 3665-3685.

Gupta, N., Zahn, M.M., Coppens, I., Joiner, K.A. & Voelker, D.R. (2005) 'Selective disruption of phosphatidylcholine metabolism of the intracellular parasite Toxoplasma gondii arrests its growth', *J Biol Chem*, vol. 280, no. 16, pp. 16345-16353.

Guy, C.S., Williams, D.J.L., Kelly, D.F., McGarry, J.W., Guy, F., Bjorkman, C., Smith, R.F. & Trees, A.J. (2001) 'Neospora caninum in persistently infected, pregnant cows: spontaneous transplacental infection is associated with an acute increase in maternal antibody', *Vet Rec*, vol. 149, no. 15, pp. 443-449.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y. & Aebersold, R. (2000) 'Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology', *Proc Natl Acad Sci U S A*, vol. 97, no. 17, pp. 9390-9395.

Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., Dimeo, J., Efcavitch, J.W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S.R., Steinmann, K., Thayer, E., Tyurina, A., Ward, R., Weiss, H. & Xie, Z. (2008) 'Single-molecule DNA sequencing of a viral genome', *Science*, vol. 320, no. 5872, pp. 106-109.

Hemphill, A. & Gottstein, B. (2000) 'A European perspective on Neospora caninum', *Int J Parasitol*, vol. 30, no. 8, pp. 877-924.

Hemphill, A., Vonlaufen, N., Naguleswaran, A., Keller, N., Riesen, M., Guetg, N., Srinivasan, S. & Alaeddine, F. (2004) 'Tissue culture and explant approaches to studying and visualizing Neospora caninum and its interactions with the host cell', *Microsc Microanal*, vol. 10, no. 5, pp. 602-620.

Hobson, J.C., Duffield, T.F., Kelton, D., Lissemore, K., Hietala, S.K., Leslie, K.E., McEwen, B., Cramer, G. & Peregrine, A.S. (2002) 'Neospora caninum serostatus and milk production of Holstein cattle', *J Am Vet Med Assoc*, vol. 221, no. 8, pp. 1160-1164.

Hoff, E.F., Cook, S.H., Sherman, G.D., Harper, J.M., Ferguson, D.J., Dubremetz, J.F. & Carruthers, V.B. (2001) 'Toxoplasma gondii: molecular cloning and characterization of a novel 18-kDa secretory antigen, TgMIC10', *Exp Parasitol*, vol. 97, no. 2, pp. 77-88.

Holmdahl, J., Bjorkman, C., Stenlund, S., Uggla, A. & Dubey, J.P. (1997) 'Bovine Neospora and Neospora caninum: One and the same', *Parasitol Today*, vol. 13, no. 1, pp. 40-41.

Holt, R.A. & Jones, S.J. (2008) 'The new paradigm of flow cell sequencing', *Genome Res*, vol. 18, no. 6, pp. 839-846.

Howe, D.K. & Sibley, L.D. (1995) 'Toxoplasma gondii comprises three clonal lineages: correlation of parasite genotype with human disease', *J Infect Dis*, vol. 172, no. 6, pp. 1561-1566.

Howell, S.A., Hackett, F., Jongco, A.M., Withers-Martinez, C., Kim, K., Carruthers, V.B. & Blackman, M.J. (2005) 'Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens', *Mol Microbiol*, vol. 57, no. 5, pp. 1342-1356.

Hunn, J.P., Feng, C.G., Sher, A. & Howard, J.C. (2011) 'The immunity-related GTPases in mammals: a fast-evolving cell-autonomous resistance system against intracellular pathogens', *Mamm Genome*, vol. 22, no. 1-2, pp. 43-54.

Hunn, J.P., Koenen-Waisman, S., Papic, N., Schroeder, N., Pawlowski, N., Lange, R., Kaiser, F., Zerrahn, J., Martens, S. & Howard, J.C. (2008) 'Regulatory interactions between IRG resistance GTPases in the cellular response to Toxoplasma gondii', *EMBO J*, vol. 27, no. 19, pp. 2495-2509.

Huynh, M.H. & Carruthers, V.B. (2006) 'Toxoplasma MIC2 is a major determinant of invasion and virulence', *PLoS Pathog*, vol. 2, no. 8, p. e84.

Innes, E.A. (2007) 'The host-parasite relationship in pregnant cattle infected with Neospora caninum', *Parasitology*, vol. 134, no. Pt 13, pp. 1903-1910.

Innes, E.A., Andrianarivo, A.G., Bjorkman, C., Williams, D.J. & Conrad, P.A. (2002) 'Immune responses to Neospora caninum and prospects for vaccination', *Trends Parasitol*, vol. 18, no. 11, pp. 497-504.

Innes, E.A., Bartley, P.M., Maley, S.W., Wright, S.E. & Buxton, D. (2007) 'Comparative host-parasite relationships in ovine toxoplasmosis and bovine neosporosis and strategies for vaccination', *Vaccine*, vol. 25, no. 30, pp. 5495-5503.

Innes, E.A., Buxton, D., Maley, S., Wright, S., Marks, J., Esteban, I., Rae, A., Schock, A. & Wastling, J. (2000) 'Neosporosis. Aspects of epidemiology and host immune response', *Ann N Y Acad Sci*, vol. 916, pp. 93-101.

Innes, E.A., Panton, W.R.M., Marks, J., Trees, A.J., Holmdahl, J. & Buxton, D. (1995) 'Interferon-Gamma Inhibits the Intracellular Multiplication of Neospora-Caninum, as Shown by Incorporation of H-3 Uracil', *Journal of Comparative Pathology*, vol. 113, no. 1, pp. 95-100.

Innes, E.A., Wright, S., Bartley, P., Maley, S., Macaldowie, C., Esteban-Redondo, I. & Buxton, D. (2005) 'The host-parasite relationship in bovine neosporosis', *Vet Immunol Immunopathol*, vol. 108, no. 1-2, pp. 29-36.

Innes, E.A., Wright, S.E., Maley, S., Rae, A., Schock, A., Kirvar, E., Bartley, P., Hamilton, C., Carey, I.M. & Buxton, D. (2001) 'Protection against vertical transmission in bovine neosporosis', *International Journal for Parasitology*, vol. 31, pp. 1523-1534.

Jan, G., Delorme, V., David, V., Revenu, C., Rebollo, A., Cayla, X. & Tardieux, I. (2007) 'The toxofilin-actin-PP2C complex of Toxoplasma: identification of interacting domains', *Biochem J*, vol. 401, no. 3, pp. 711-719.

Jewett, T.J. & Sibley, L.D. (2004) 'The toxoplasma proteins MIC2 and M2AP form a hexameric complex necessary for intracellular survival', *J Biol Chem*, vol. 279, no. 10, pp. 9362-9369.

Jones, A.R., Siepen, J.A., Hubbard, S.J. & Paton, N.W. (2009) 'Improving sensitivity in proteome studies by analysis of false discovery rates for multiple search engines', *Proteomics*, vol. 9, pp. 1220-1229.

Jung, C., Lee, C.Y. & Grigg, M.E. (2004) 'The SRS superfamily of Toxoplasma surface proteins', *Int J Parasitol*, vol. 34, no. 3, pp. 285-296.

Kang, S.W., Kweon, C.H., Lee, E.H., Choe, S.E., Jung, S.C. & Quyen, D.V. (2008) 'The differentiation of transcription between tachyzoites and bradyzoites of in vitro cultured Neospora caninum', *Parasitol Res*, vol. 103, no. 5, pp. 1011-1018.

Kang, S.W., Lee, E.H., Jean, Y.H., Choe, S.E., Van Quyen, D. & Lee, M.S. (2008b) 'The differential protein expression profiles and immunogenicity of tachyzoites and bradyzoites of in vitro cultured Neospora caninum', *Parasitol Res*, vol. 103, no. 4, pp. 905-913.

Kappe, S.H., Buscaglia, C.A., Bergman, L.W., Coppens, I. & Nussenzweig, V. (2004) 'Apicomplexan gliding motility and host cell invasion: overhauling the motor model', *Trends Parasitol*, vol. 20, no. 1, pp. 13-16.

Khaminets, A., Hunn, J.P., Konen-Waisman, S., Zhao, Y.O., Preukschat, D., Coers, J., Boyle, J.P., Ong, Y.C., Boothroyd, J.C., Reichmann, G. & Howard, J.C. (2010) 'Coordinated Loading of IRG Resistance GTPases on to the Toxoplasma gondii Parasitophorous Vacuole', *Cell Microbiol*.

Khan, A., Taylor, S., Ajioka, J.W., Rosenthal, B.M. & Sibley, L.D. (2009) 'Selection at a single locus leads to widespread expansion of Toxoplasma gondii lineages that are virulent in mice', *PLoS Genet*, vol. 5, no. 3, p. e1000404.

Kim, J.B., Porreca, G.J., Song, L., Greenway, S.C., Gorham, J.M., Church, G.M., Seidman, C.E. & Seidman, J.G. (2007) 'Polony multiplex analysis of gene expression (PMAGE) in mouse hypertrophic

cardiomyopathy', *Science*, vol. 316, no. 5830, pp. 1481-1484.

Konen-Waisman, S. & Howard, J.C. (2007) 'Cell-autonomous immunity to Toxoplasma gondii in mouse and man', *Microbes Infect*, vol. 9, no. 14-15, pp. 1652-1661.

Koonin, E.V. (2005) 'Orthologs, paralogs, and evolutionary genomics', *Annu Rev Genet*, vol. 39, pp. 309-338.

Labruyere, E., Lingnau, M., Mercier, C. & Sibley, L.D. (1999) 'Differential membrane targeting of the secretory proteins GRA4 and GRA6 within the parasitophorous vacuole formed by Toxoplasma gondii', *Mol Biochem Parasitol*, vol. 102, no. 2, pp. 311-324.

Laemmli, U.K. (1970) 'Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4', *Nature*, vol. 227, pp. 680-685.

Lagal, V., Binder, E.M., Huynh, M.H., Kafsack, B.F., Harris, P.K., Diez, R., Chen, D., Cole, R.N., Carruthers, V.B. & Kim, K. (2010) 'Toxoplasma gondii protease TgSUB1 is required for cell surface processing of micronemal adhesive complexes and efficient adhesion of tachyzoites', *Cell Microbiol*, vol. 12, no. 12, pp. 1792-1808.

Lee, E.G., Kim, J.H., Shin, Y.S., Shin, G.W., Kim, Y.H., Kim, G.S., Kim, D.Y., Jung, T.S. & Suh, M.D. (2004) 'Two-dimensional gel electrophoresis and immunoblot analysis of Neospora caninum tachyzoites', *J Vet Sci*, vol. 5, no. 2, pp. 139-145.

Lee, E.G., Kim, J.H., Shin, Y.S., Shin, G.W., Kim, Y.R., Palaksha, K.J., Kim, D.Y., Yamane, I., Kim, Y.H., Kim, G.S., Suh, M.D. & Jung, T.S. (2005) 'Application of proteomics for comparison of proteome of Neospora caninum and Toxoplasma gondii tachyzoites', *J Chromatogr B Analyt Technol Biomed Life Sci*, vol. 815, no. 1-2, pp. 305-314.

Lee, E.G., Kim, J.H., Shin, Y.S., Shin, G.W., Suh, M.D., Kim, D.Y., Kim, Y.H., Kim, G.S. & Jung, T.S. (2003) 'Establishment of a two-dimensional electrophoresis map for Neospora caninum tachyzoites by proteomics', *Proteomics*, vol. 3, no. 12, pp. 2339-2350.

Leimgruber, R.M., Malone, J.P., Radabaugh, M.R., LaPorte, M.L., Violand, B.N. & Monahan, J.B. (2002) 'Development of improved cell lysis, solubilization and imaging approaches for proteomic analyses', *Proteomics*, vol. 2, no. 2, pp. 135-144.

Leriche, M.A. & Dubremetz, J.F. (1991) 'Characterization of the Protein Contents of Rhoptries and Dense Granules of Toxoplasma-Gondii Tachyzoites by Subcellular Fractionation and Monoclonal-Antibodies', *Molecular and Biochemical Parasitology*, vol. 45, no. 2, pp. 249-260.

Li, G., Zhang, J., Sun, Y., Wang, H. & Wang, Y. (2009) 'The evolutionarily dynamic IFN-inducible GTPase proteins play conserved immune functions in vertebrates and cephalochordates', *Mol Biol Evol*, vol. 26, no. 7, pp. 1619-1630.

Li, L., Stoeckert, C.J., Jr. & Roos, D.S. (2003) 'OrthoMCL: identification of ortholog groups for eukaryotic genomes', *Genome Res*, vol. 13, no. 9, pp. 2178-2189.

Lindsay, D.S., Dubey, J.P. & Duncan, R.B. (1999) 'Confirmation that the dog is a definitive host for Neospora caninum', *Vet Parasitol*, vol. 82, no. 4, pp. 327-333.

Lindsay, D.S., Mitschler, R.R., Toivio-Kinnucan, M.A., Upton, S.J., Dubey, J.P. & Blagburn, B.L. (1993) 'Association of host cell mitochondria with developing Toxoplasma gondii tissue cysts', *Am J Vet Res*, vol. 54, no. 10, pp. 1663-1667.

Lindsay, D.S., Ritter, D.M. & Brake, D. (2001) 'Oocyst excretion in dogs fed mouse brains containing tissue cysts of a cloned line of Neospora caninum', *J Parasitol*, vol. 87, no. 4, pp. 909-911.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. & Yates, J.R. (1999) 'Direct analysis of protein complexes using mass spectrometry', *Nature Biotechnology*, vol. 17, pp. 676-682.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. & Ecker, J.R. (2008) 'Highly integrated single-base resolution maps of the epigenome in Arabidopsis', *Cell*, vol. 133, no. 3, pp. 523-536.

Lunden, A., Marks, J., Maley, S.W. & Innes, E.A. (1998) 'Cellular immune responses in cattle experimentally infected with Neospora caninum', *Parasite Immunol*, vol. 20, no. 11, pp. 519-526.

Mader, U., Nicolas, P., Richard, H., Bessieres, P. & Aymerich, S. (2011) 'Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods', *Curr Opin Biotechnol*, vol. 22, no. 1, pp. 32-41.

Makarov, A. (2000) 'Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis', *Anal Chem*, vol. 72, no. 6, pp. 1156-1162.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F. & Rothberg, J.M. (2005) 'Genome sequencing in microfabricated high-density picolitre reactors', *Nature*, vol. 437, no. 7057, pp. 376-380.

Martens, S., Parvanova, I., Zerrahn, J., Griffiths, G., Schell, G., Reichmann, G. & Howard, J.C. (2005) 'Disruption of Toxoplasma gondii parasitophorous vacuoles by the mouse p47-resistance GTPases', *PLoS Pathog*, vol. 1, no. 3, p. e24.

Marugan-Hernandez, V., Alvarez-Garcia, G., Tomley, F., Hemphill, A., Regidor-Cerrillo, J. & Ortega-Mora, L.M. (2011) 'Identification of novel rhoptry proteins in Neospora caninum by LC/MS-MS analysis of subcellular fractions', *J Proteomics*, vol. 74, no. 5, pp. 629-642.

McAllister, M.M., Dubey, J.P., Lindsay, D.S., Jolley, W.R., Wills, R.A. & McGuire, A.M. (1998) 'Dogs are definitive hosts of *Neospora caninum*', *International Journal for Parasitology*, vol. 28, pp. 1473-1478.

McCann, C.M., Vyse, A.J., Salmon, R.L., Thomas, D., Williams, D.J., McGarry, J.W., Pebody, R. & Trees, A.J. (2008) 'Lack of serologic evidence of Neospora caninum in humans, England', *Emerg Infect Dis*, vol. 14, no. 6, pp. 978-980.

Meissner, M., Reiss, M., Viebig, N., Carruthers, V.B., Toursel, C., Tomavo, S., Ajioka, J.W. & Soldati, D. (2002) 'A family of transmembrane microneme proteins of *Toxoplasma gondii* contain EGF-like domains and function as escorters', *Journal of Cell Science*, vol. 11`5, pp. 563-574.

Michelin, A., Bittame, A., Bordat, Y., Travier, L., Mercier, C., Dubremetz, J.F. & Lebrun, M. (2009) 'GRA12, a *Toxoplasma* dense granule protein associated with the intravacuolar membranous nanotubular network', *International Journal for Parasitology*, vol. 39, pp. 299-306.

Miller, S.A., Binder, E.M., Blackman, M.J., Carruthers, V.B. & Kim, K. (2001) 'A conserved subtilisin-like protein TgSUB1 in microneme organelles of Toxoplasma gondii', *J Biol Chem*, vol. 276, no. 48, pp. 45341-45348.

Mital, J. & Ward, G.E. (2008) 'Current and emerging approaches to studying invasion in apicomplexan parasites', *Subcell Biochem*, vol. 47, pp. 1-32.

Mordue, D.G., Desai, N., Dustin, M. & Sibley, L.D. (1999a) 'Invasion by Toxoplasma gondii establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring', *J Exp Med*, vol. 190, no. 12, pp. 1783-1792.

Mordue, D.G., Hakansson, S., Niesman, I. & Sibley, L.D. (1999b) 'Toxoplasma gondii resides in a vacuole that avoids fusion with host cell endocytic and exocytic vesicular trafficking pathways', *Exp Parasitol*, vol. 92, no. 2, pp. 87-99.

Mugridge, N.B., Morrison, D.A., Heckeroth, A.R., Johnson, A.M. & Tenter, A.M. (1999) 'Phylogenetic analysis based on full-length large subunit ribosomal RNA gene sequence comparison reveals that Neospora caninum is more closely related to Hammondia heydorni than to Toxoplasma gondii', *Int J Parasitol*, vol. 29, no. 10, pp. 1545-1556.

Myers, J.L. & Well, A.D. (2003) *Research Design and Statistical Analysis (2nd ed.)*, Lawrence Erlbaum.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. & Snyder, M. (2008) 'The transcriptional landscape of the yeast genome defined by RNA sequencing', *Science*, vol. 320, no. 5881, pp. 1344-1349.

Nam, H.W. (2009) 'GRA proteins of Toxoplasma gondii: maintenance of host-parasite interactions across the parasitophorous vacuolar membrane', *Korean J Parasitol*, vol. 47 Suppl, pp. S29-37.

Nichols, B.A., Chiappino, M.L. & O'Connor, G.R. (1983) 'Secretion from the rhoptries of Toxoplasma gondii during host-cell invasion', *J Ultrastruct Res*, vol. 83, no. 1, pp. 85-98.

Nishikawa, Y., Iwata, A., Nagasawa, H., Fujisaki, K., Otsuka, H. & Mikami, T. (2001a) 'Comparison of the growth inhibitory effects of canine IFN-alpha, -beta and -gamma on canine cells infected with Neospora caninum tachyzoites', *J Vet Med Sci*, vol. 63, no. 4, pp. 445-448.

Nishikawa, Y., Mishima, M., Nagasawa, H., Igarashi, I., Fujisaki, K., Otsuka, H. & Mikami, T. (2001b) 'Interferon-gamma-induced apoptosis in host cells infected with Neospora caninum', *Parasitology*, vol. 123, no. Pt 1, pp. 25-31.

Odberg-Ferragut, C., Soete, M., Engels, A., Samyn, B., Loyens, A., Van Beeumen, J., Camus, D. & Dubremetz, J.F. (1996) 'Molecular cloning of the Toxoplasma gondii sag4 gene encoding an 18 kDa bradyzoite specific surface protein', *Mol Biochem Parasitol*, vol. 82, no. 2, pp. 237-244.

Ong, Y.C., Reese, M.L. & Boothroyd, J.C. (2010) 'Toxoplasma rhoptry protein 16 (ROP16) subverts host function by direct tyrosine phosphorylation of STAT6', *J Biol Chem*, vol. 285, no. 37, pp. 28731-28740.

Opitz, C., Di Cristina, M., Reiss, M., Ruppert, T., Crisanti, A. & Soldati, D. (2002) 'Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite Toxoplasma gondii', *EMBO J*, vol. 21, no. 7, pp. 1577-1585.

Otto, T.D., Wilinski, D., Assefa, S., Keane, T.M., Sarry, L.R., Bohme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., Newbold, C. & Llinas, M. (2010) 'New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq', *Mol Microbiol*, vol. 76, no. 1, pp. 12-24.

Papic, N., Hunn, J.P., Pawlowski, N., Zerrahn, J. & Howard, J.C. (2008) 'Inactive and active states of the interferon-inducible resistance GTPase, Irga6, in vivo', *J Biol Chem*, vol. 283, no. 46, pp. 32143-32151.

Pare, J., Thurmond, M.C. & Hietala, S.K. (1996) 'Congenital Neospora caninum infection in dairy cattle and associated calfhood mortality', *Can J Vet Res*, vol. 60, no. 2, pp. 133-139.

Pare, J., Thurmond, M.C. & Hietala, S.K. (1997) 'Neospora caninum antibodies in cows during pregnancy as a predictor of congenital infection and abortion', *J Parasitol*, vol. 83, no. 1, pp. 82-87.

Peixoto, L., Chen, F., Harb, O.S., Davis, P.H., Beiting, D.P., Brownback, C.S., Ouloguem, D. & Roos, D.S. (2010) 'Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses', *Cell Host Microbe*, vol. 8, no. 2, pp. 208-218.

Pelleau, S., Bertaux, L., Briolant, S., Ferdig, M.T., Sinou, V., Pradines, B., Parzy, D. & Jambou, R. (2011) 'Differential association of Plasmodium falciparum Na+/H+ exchanger polymorphism and quinine responses in field- and culture-adapted isolates of Plasmodium falciparum', *Antimicrob Agents Chemother*, vol. 55, no. 12, pp. 5834-5841.

Perez-Zaballos, F.J., Ortega-Mora, L.M., Alvarez-Garcia, G., Collantes-Fernandez, E., Navarro-Lozano, V., Garcia-Villada, L. & Costas, E. (2005) 'Adaptation of Neospora caninum isolates to cell-culture changes: an argument in favor of its clonal population structure', *J Parasitol*, vol. 91, no. 3, pp. 507-510.

Pernas, L. & Boothroyd, J.C. (2010) 'Association of host mitochondria with the parasitophorous vacuole during Toxoplasma infection is not dependent on rhoptry proteins ROP2/8', *Int J Parasitol*, vol. 40, no. 12, pp. 1367-1371.

Poupel, O., Boleti, H., Axisa, S., Couture-Tosi, E. & Tardieux, I. (2000) 'Toxofilin, a novel actin-binding protein from Toxoplasma gondii, sequesters actin monomers and caps actin filaments', *Mol Biol Cell*, vol. 11, no. 1, pp. 355-368.

Qiu, W., Wernimont, A., Tang, K., Taylor, S., Lunin, V., Schapira, M., Fentress, S., Hui, R. & Sibley, L.D. (2009) 'Novel structural and regulatory features of rhoptry secretory kinases in Toxoplasma gondii', *EMBO J*, vol. 28, no. 7, pp. 969-979.

Ravindran, S. & Boothroyd, J.C. (2008) 'Secretion of proteins into host cells by Apicomplexan parasites', *Traffic*, vol. 9, no. 5, pp. 647-656.

Reese, M.L. & Boothroyd, J.C. (2011) 'A conserved non-canonical motif in the pseudoactive site of the

ROP5 pseudokinase domain mediates its effect on Toxoplasma virulence', *J Biol Chem*, vol. 286, no. 33, pp. 29366-29375.

Reese, Zeiner, Saeij, Boothroyd, a. & Boyle (2011) 'Polymorphic family of injected pseudokinases is paramount in Toxoplasma virulence', *Proceedings of the National Academy of Sciences of the USA*.

Reid, A.J., Vermont, S.J., Cotton, J.A., Harris, D., Hill-Cawthorne, G.A., Konen-Waisman, S., Latham, S.M., Mourier, T., Norton, R., Quail, M.A., Sanders, M., Shanmugam, D., Sohal, A., Wasmuth, J.D., Brunk, B., Grigg, M.E., Howard, J.C., Parkinson, J., Roos, D.S., Trees, A.J., Berriman, M., Pain, A. & Wastling, J.M. (2012) 'Comparative Genomics of the Apicomplexan Parasites Toxoplasma gondii and Neospora caninum: Coccidia Differing in Host Range and Transmission Strategy', *PLoS Pathog*, vol. 8, no. 3, p. e1002567.

Reiss, M., Viebig, N., Brecht, S., Fourmaux, M.N., Soete, M., Di Cristina, M., Dubremetz, J.F. & Soldati, D. (2001) 'Identification and characterization of an escorter for two secretory adhesins in Toxoplasma gondii', *J Cell Biol*, vol. 152, no. 3, pp. 563-578.

Rigbolt, K.T., Vanselow, J.T. & Blagoev, B. (2011) 'GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data', *Mol Cell Proteomics*, vol. 10, no. 8, p. O110 007450.

Roach, J.C., Boysen, C., Wang, K. & Hood, L. (1995) 'Pairwise end sequencing: a unified approach to genomic mapping and sequencing', *Genomics*, vol. 26, no. 2, pp. 345-353.

Roberts, L.S., Schmidt, G.D. & Janovy, J. (2009) *Gerald D. Schmidt & Larry S. Roberts' foundations of parasitology*, 8th edn, McGraw-Hill Higher Education, Boston.

Rodger, S.M., Maley, S.W., Wright, S.E., MacKellar, A., Wesley, F., Sales, J. & Buxton, D. (2006) 'Role of endogenous transplacental transmission in toxoplasmosis in sheep', *Veterinary Record*, vol. 159, no. 23, pp. 768-772.

Rosenfeld, J., Capdevielle, J., Guillemot, J.C. & Ferrara, P. (1992) 'In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis', *Anal Biochem*, vol. 203, no. 1, pp. 173-179.

Rozen, S. & Skaletsky, H. (2000) 'Primer3 on the WWW for general users and for biologist programmers', *Methods Mol Biol*, vol. 132, pp. 365-386.

Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M. & Mewes, H.W. (2004) 'The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes', *Nucleic Acids Res*, vol. 32, no. 18, pp. 5539-5545.

Saeij, J.P., Boyle, J.P., Coller, S., Taylor, S., Sibley, L.D., Brooke-Powell, E.T., Ajioka, J.W. & Boothroyd, J.C. (2006) 'Polymorphic secreted kinases are key virulence factors in toxoplasmosis', *Science*, vol. 314, no. 5806, pp. 1780-1783.

Saeij, J.P., Coller, S., Boyle, J.P., Jerome, M.E., White, M.W. & Boothroyd, J.C. (2007) 'Toxoplasma co-opts host gene expression by injection of a polymorphic kinase homologue', *Nature*, vol. 445, no. 7125, pp. 324-327.

Saffer, L.D., Mercereau-Puijalon, O., Dubremetz, J.F. & Schwartzman, J.D. (1992) 'Localization of a Toxoplasma gondii rhoptry protein by immunoelectron microscopy during and after host cell penetration', *J Protozool*, vol. 39, no. 4, pp. 526-530.

Sanderson, S.J., Xia, D., Prieto, H., Yates, J., Heiges, M., Kissinger, J.C., Bromley, E., Lal, K., Sinden, R.E., Tomley, F. & Wastling, J.M. (2008) 'Determining the protein repertoire of Cryptosporidium parvum sporozoites', *Proteomics*, vol. 8, no. 7, pp. 1398-1414.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proc Natl Acad Sci U S A*, vol. 74, no. 12, pp. 5463-5467.

Santoni, V., Rabilloud, T., Doumas, P., Rouquie, D., Mansion, M., Kieffer, S., Garin, J. & Rossignol, M. (1999) 'Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels', *Electrophoresis*, vol. 20, no. 4-5, pp. 705-711.

Saouros, S., Edwards-Jones, B., Reiss, M., Sawmynaden, K., Cota, E., Simpson, P., Dowse, T.J., Jakle, U., Ramboarina, S., Shivarattan, T., Matthews, S. & Soldati-Favre, D. (2005) 'A novel galectin-like domain from Toxoplasma gondii micronemal protein 1 assists the folding, assembly, and transport of a cell adhesion complex', *J Biol Chem*, vol. 280, no. 46, pp. 38583-38591.

Schares, G., Peters, M., Wurm, R., Barwald, A. & Conraths, F.J. (1998) 'The efficiency of vertical transmission of Neospora caninum in dairy cattle analysed by serological techniques', *Vet Parasitol*, vol. 80, no. 2, pp. 87-98.

Schock, A., Innes, E.A., Yamane, I., Latham, S.M. & Wastling, J.M. (2001) 'Genetic and biological diversity among isolates of Neospora caninum', *Parasitology*, vol. 123, no. Pt 1, pp. 13-23.

Shanmugasundram, A. (unpublished) LAMP database: Liverpool Library for Apicomplexan Metabolic Pathways, http://www.llamp.net/

Shapiro, A.L., Vinuela, E. & Maizel, J.V., Jr. (1967) 'Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels', *Biochem Biophys Res Commun*, vol. 28, no. 5, pp. 815-820.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D. & Church, G.M. (2005) 'Accurate multiplex polony sequencing of an evolved bacterial genome', *Science*, vol. 309, no. 5741, pp. 1728-1732.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. & Mann, M. (2006) 'In-gel digestion for mass spectrometric characterization of proteins and proteomes', *Nat Protoc*, vol. 1, no. 6, pp. 2856-2860.

Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996) 'Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels', *Anal Chem*, vol. 68, no. 5, pp. 850-858.

Sibley, L.D. & Boothroyd, J.C. (1992) 'Virulent strains of Toxoplasma gondii comprise a single clonal lineage', *Nature*, vol. 359, no. 6390, pp. 82-85.

Sibley, L.D., Niesman, I.R., Parmley, S.F. & Cesbron-Delauw, M.F. (1995) 'Regulated secretion of multi-lamellar vesicles leads to formation of a tubulo-vesicular network in host-cell vacuoles occupied by Toxoplasma gondii', *J Cell Sci*, vol. 108 ( Pt 4), pp. 1669-1677.

Simpson, R.J. (ed.) (2003) *Proteins and Proteomics*, Cold Spring Harbour Laboratory Press, New York.

Sinai, A.P. & Joiner, K.A. (2001) 'The Toxoplasma gondii protein ROP2 mediates host organelle association with the parasitophorous vacuole membrane', *J Cell Biol*, vol. 154, no. 1, pp. 95-108.

Sinai, A.P. (2007) 'The toxoplasma kinase ROP18: an active member of a degenerate family', *PLoS Pathog*, vol. 3, no. 2, p. e16.

Sohn, C.S., Cheng, T.T., Drummond, M.L., Peng, E.D., Vermont, S.J., Xia, D., Cheng, S.J., Wastling, J.M. & Bradley, P.J. (2011) 'Identification of novel proteins in Neospora caninum using an organelle purification and monoclonal antibody approach', *PLoS One*, vol. 6, no. 4, p. e18383.

Song, K.J. & Nam, H.W. (2003) 'Protease activity of 80 kDa protein secreted from the apicomplexan parasite Toxoplasma gondii', *Korean J Parasitol*, vol. 41, no. 3, pp. 165-169.

Sorber, K., Dimon, M.T. & DeRisi, J.L. (2011) 'RNA-Seq analysis of splicing in Plasmodium falciparum uncovers new splice junctions, alternative splicing and splicing of antisense transcripts', *Nucleic Acids Res*, vol. 39, no. 9, pp. 3820-3835.

Speer, C.A. & Dubey, J.P. (1989) 'Ultrastructure of tachyzoites, bradyzoites and tissue cysts of Neospora caninum', *J Protozool*, vol. 36, no. 5, pp. 458-463.

Speer, C.A., Dubey, J.P., McAllister, M.M. & Blixt, J.A. (1999) 'Comparative ultrastructure of tachyzoites, bradyzoites, and tissue cysts of Neospora caninum and Toxoplasma gondii', *Int J Parasitol*, vol. 29, no. 10, pp. 1509-1519.

Steinfeldt, T., Konen-Waisman, S., Tong, L., Pawlowski, N., Lamkemeyer, T., Sibley, L.D., Hunn, J.P. & Howard, J.C. (2010) 'Phosphorylation of mouse immunity-related GTPase (IRG) resistance proteins is an evasion strategy for virulent Toxoplasma gondii', *PLoS Biol*, vol. 8, no. 12, p. e1000576.

Stewart, A.D., Logsdon, J.M., Jr. & Kelley, S.E. (2005) 'An empirical study of the evolution of virulence under both horizontal and vertical transmission', *Evolution*, vol. 59, no. 4, pp. 730-739.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keefe, S., Haas, S., Vingron, M., Lehrach, H. & Yaspo, M.L. (2008) 'A Global View of Gene Activity and Alternative Splicing by Deep Sequencing of the Human Transcriptome', *Science*, vol. 321, pp. 956-960.

Suzuki, Y., Orellana, M.A., Schreiber, R.D. & Remington, J.S. (1988) 'Interferon-gamma: the major mediator of resistance against Toxoplasma gondii', *Science*, vol. 240, no. 4851, pp. 516-518.

Talevich, E., Mirza, A. & Kannan, N. (2011) 'Structural and evolutionary divergence of eukaryotic protein kinases in Apicomplexa', *BMC Evol Biol*, vol. 11, p. 321.

Tangri, S. & Raghupathy, R. (1993) 'Expression of cytokines in placentas of mice undergoing immunologically mediated spontaneous fetal resorptions', *Biol Reprod*, vol. 49, no. 4, pp. 850-856.

Taylor, G.A. (2007) 'IRG proteins: key mediators of interferon-regulated host resistance to intracellular pathogens', *Cell Microbiol*, vol. 9, no. 5, pp. 1099-1107.

Taylor, S., Barragan, A., Su, C., Fux, B., Fentress, S.J., Tang, K., Beatty, W.L., Hajj, H.E., Jerome, M., Behnke, M.S., White, M., Wootton, J.C. & Sibley, L.D. (2006) 'A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen Toxoplasma gondii', *Science*, vol. 314, no. 5806, pp. 1776-1780.

Tenter, A.M. & Johnson, A.M. (1997) 'Phylogeny of the tissue cyst-forming coccidia', *Adv Parasitol*, vol. 39, pp. 69-139.

Thurmond, M.C. & Hietala, S.K. (1996) 'Culling associated with Neospora caninum infection in dairy cows', *Am J Vet Res*, vol. 57, no. 11, pp. 1559-1562.

Thurmond, M.C. & Hietala, S.K. (1997) 'Effect of congenitally acquired Neospora caninum infection on risk of abortion and subsequent abortions in dairy cattle', *Am J Vet Res*, vol. 58, no. 12, pp. 1381-1385.

Thurmond, M.C., Hietala, S.K. & Blanchard, P.C. (1997) 'Herd-based diagnosis of Neospora caninum-induced endemic and epidemic abortion in cows and evidence for congenital and postnatal transmission', *J Vet Diagn Invest*, vol. 9, no. 1, pp. 44-49.

Thurmond, M.C., Hietala, S.K. & Blanchard, P.C. (1999) 'Predictive values of fetal histopathology and immunoperoxidase staining in diagnosing bovine abortion caused by Neospora caninum in a dairy herd', *J Vet Diagn Invest*, vol. 11, no. 1, pp. 90-94.

Trees, A.J. & Williams, D.J. (2005) 'Endogenous and exogenous transplacental infection in Neospora caninum and Toxoplasma gondii', *Trends Parasitol*, vol. 21, no. 12, pp. 558-561.

Trees, A.J., Davison, H.C., Innes, E.A. & Wastling, J.M. (1999) 'Towards evaluating the economic impact of bovine neosporosis', *Int J Parasitol*, vol. 29, no. 8, pp. 1195-1200.

Tunev, S.S., McAllister, M.M., Anderson-Sprecher, R.C. & Weiss, L.M. (2002) 'Neospora caninum in vitro: evidence that the destiny of a parasitophorous vacuole depends on the phenotype of the progenitor zoite', *J Parasitol*, vol. 88, no. 6, pp. 1095-1099.

Tyler, J.S. & Boothroyd, J.C. (2011) 'The C-terminus of Toxoplasma RON2 provides the crucial link between AMA1 and the host-associated invasion complex', *PLoS Pathog*, vol. 7, no. 2, p. e1001282.

Uthaiah, R.C., Praefcke, G.J., Howard, J.C. & Herrmann, C. (2003) 'IIGP1, an interferon-gamma-inducible 47-kDa GTPase of the mouse, showing cooperative enzymatic activity and GTP-dependent multimerization', *J Biol Chem*, vol. 278, no. 31, pp. 29336-29343.

Vanin, E.F. (1985) 'Processed pseudogenes: characteristics and evolution', *Annu Rev Genet*, vol. 19, pp. 253-272.

Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F. & Hartl, D.L. (2001) 'Recent origin of Plasmodium falciparum from a single progenitor', *Science*, vol. 293, no. 5529, pp. 482-484.

von Hagen, J. (ed.) (2008) *Proteomics Sample Preparation*, Wiley-VCH, Weinheim, Germany.

Vonlaufen, N., Guetg, N., Naguleswaran, A., Muller, N., Bjorkman, C., Schares, G., von Blumroeder, D., Ellis, J. & Hemphill, A. (2004) 'In vitro induction of Neospora caninum bradyzoites in vero cells reveals differential antigen expression, localization, and host-cell recognition of tachyzoites

and bradyzoites', *Infect Immun*, vol. 72, no. 1, pp. 576-583.

Vonlaufen, N., Muller, N., Keller, N., Naguleswaran, A., Bohne, W., McAllister, M.M., Bjorkman, C., Muller, E., Caldelari, R. & Hemphill, A. (2002) 'Exogenous nitric oxide triggers Neospora caninum tachyzoite-to-bradyzoite stage conversion in murine epidermal keratinocyte cell cultures', *Int J Parasitol*, vol. 32, no. 10, pp. 1253-1265.

Wang, Z., Gerstein, M. & Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature Reviews | Genetics*, vol. 10, pp. 57-63.

Ward, P., Equinet, L., Packer, J. & Doerig, C. (2004) 'Protein kinases of the human malaria parasite Plasmodium falciparum: the kinome of a divergent eukaryote', *BMC Genomics*, vol. 5, no. 1, p. 79.

Washburn, M.P., Wolters, D. & Yates, J.R., 3rd (2001) 'Large-scale analysis of the yeast proteome by multidimensional protein identification technology', *Nat Biotechnol*, vol. 19, no. 3, pp. 242-247.

Wastling, J.M., Armstrong, S.D., Krishna, R. & Xia, D. (2012) 'Parasites, proteomes and systems: has Descartes' clock run out of time?', *Parasitology*, vol. 139, no. 9, pp. 1103-1118.

Wastling, J.M., Xia, D., Sohal, A., Chaussepied, M., Pain, A. & Langsley, G. (2009) 'Proteomes and transcriptomes of the Apicomplexa--where's the message?', *Int J Parasitol*, vol. 39, no. 2, pp. 135-143.

Weiss, L.M. & Kim, K. (2000) 'The development and biology of bradyzoites of Toxoplasma gondii', *Front Biosci*, vol. 5, pp. D391-405.

Wiengcharoen, J., Thompson, R.C., Nakthong, C., Rattanakorn, P. & Sukthana, Y. (2011) 'Transplacental transmission in cattle: is Toxoplasma gondii less potent than Neospora caninum?', *Parasitol Res*, vol. 108, no. 5, pp. 1235-1241.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J. & Bahler, J. (2008) 'Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution', *Nature*, vol. 453, no. 7199, pp. 1239-1243.

Williams, D.J., Guy, C.S., McGarry, J.W., Guy, F., Tasker, L., Smith, R.F., MacEachern, K., Cripps, P.J., Kelly, D.F. & Trees, A.J. (2000) 'Neospora caninum-associated abortion in cattle: the time of experimentally-induced parasitaemia during gestation determines foetal survival', *Parasitology*, vol. 121 ( Pt 4), pp. 347-358.

Williams, D.J., Hartley, C.S., Bjorkman, C. & Trees, A.J. (2009) 'Endogenous and exogenous transplacental transmission of Neospora caninum - how the route of transmission impacts on epidemiology and control of disease', *Parasitology*, pp. 1-6.

Wolters, D.A., Washburn, M.P. & Yates, J.R., 3rd (2001) 'An automated multidimensional protein identification technology for shotgun proteomics', *Anal Chem*, vol. 73, no. 23, pp. 5683-5690.

Woodbine, K.A., Medley, G.F., Moore, S.J., Ramirez-Villaescusa, A., Mason, S. & Green, L.E. (2008) 'A four year longitudinal sero-epidemiology study of Neospora caninum in adult cattle from 114 cattle herds in south west England: associations with age, herd and dam-offspring pairs', *BMC Vet Res*, vol. 4, p. 35.

Wouda, W., Moen, A.R., Visser, I.J. & van Knapen, F. (1997) 'Bovine fetal neosporosis: a comparison of epizootic and sporadic abortion cases and different age classes with regard to lesion severity and immunohistochemical identification of organisms in brain, heart, and liver', *J Vet Diagn Invest*, vol. 9, no. 2, pp. 180-185.

Xia, D., Sanderson, S.J., Jones, A.R., Prieto, J.H., Yates, J.R., Bromley, E., Tomley, F.M., Lal, K., Sinden, R.E., Brunk, B.P., Roos, D.S. & Wastling, J.M. (2008) 'The proteome of Toxoplasma gondii: integration with the genome provides novel insights into gene expression and annotation', *Genome Biol*, vol. 9, no. 7, p. R116.

Yamane, I., Kitani, H., Kokuho, T., Shibahara, T., Haritani, M., Hamaoka, T., Shimizu, S., Koiwai, M., Shimura, K. & Yokomizo, Y. (2000) 'The inhibitory effect of interferon gamma and tumor necrosis factor alpha on intracellular multiplication of Neospora caninum in primary bovine brain cells', *J Vet Med Sci*, vol. 62, no. 3, pp. 347-351.

Yamane, I., Kokuho, T., Shimura, K., Eto, M., Haritani, M., Ouchi, Y., Sverlow, K.W. & Conrad, P.A. (1996) 'In vitro isolation of a bovine Neospora in Japan', *Vet Rec*, vol. 138, no. 26, p. 652.

Zhao, Y., Ferguson, D.J., Wilson, D.C., Howard, J.C., Sibley, L.D. & Yap, G.S. (2009a) 'Virulent Toxoplasma gondii evade immunity-related GTPase-mediated parasite vacuole disruption within primed macrophages', *J Immunol*, vol. 182, no. 6, pp. 3775-3781.

Zhao, Y.O., Khaminets, A., Hunn, J.P. & Howard, J.C. (2009b) 'Disruption of the Toxoplasma gondii parasitophorous vacuole by IFNgamma-inducible immunity-related GTPases (IRG proteins) triggers necrotic cell death', *PLoS Pathog*, vol. 5, no. 2, p. e1000288.

Zhao, Y.O., Rohde, C., Lilue, J.T., Konen-Waisman, S., Khaminets, A., Hunn, J.P. & Howard, J.C. (2009) 'Toxoplasma gondii and the Immunity-Related GTPase (IRG) resistance system in mice: a review', *Mem Inst Oswaldo Cruz*, vol. 104, no. 2, pp. 234-240.

Zhou, X.W., Kafsack, B.F., Cole, R.N., Beckett, P., Shen, R.F. & Carruthers, V.B. (2005) 'The opportunistic pathogen Toxoplasma gondii deploys a diverse legion of invasion and survival proteins', *J Biol Chem*, vol. 280, no. 40, pp. 34233-34244.

Zhu, G., Marchewka, M.J. & Keithly, J.S. (2000) 'Cryptosporidium parvum appears to lack a plastid genome', *Microbiology*, vol. 146 ( Pt 2), pp. 315-321.