Species diversity and genome evolution of the pathogenic protozoan parasite, *Neospora caninum*

Larissa Calarco[*a] and John Ellis[a]

[a] School of Life Sciences, University of Technology Sydney,

PO Box 123, Broadway,

NSW 2007, Australia

*__Author for correspondence__: Larissa Calarco

School of Life Sciences, University of Technology Sydney,

Sydney, Australia,

E: Larissa.M.Calarco@student.uts.edu.au

P: +61 2 9514 4161

ABSTRACT

Neospora caninum is a cyst-forming coccidian parasite of veterinary and economical significance, affecting dairy and beef cattle industries on a global scale. Comparative studies suggest that *N. caninum* consists of a globally dispersed, diverse population of lineages, distinguished by their geographical origin, broad host range, and phenotypic features. This viewpoint is however changing. While intraspecies diversity, and more specifically pathogenic variability, has been experimentally demonstrated in a myriad of studies, the underlying contributors and sources responsible for such diversity have remained nebulous. However, recent large-scale sequence and bioinformatics studies have aided in revealing intrinsic genetic differences distinguishing isolates of this species, that await further characterisation as causative links to virulence and pathogenicity. Furthermore, progress on *N. caninum* research as a non-model organism is hindered by a lack of robust, annotated genomic, transcriptomic, and proteomic data for the species, especially compared to other thoroughly studied Apicomplexa such as *Toxoplasma gondii* and *Plasmodium* species.

This review explores the current body of knowledge on intra-species diversity within *N. caninum*. This includes the contribution of sequence variants in both coding and non-coding regions, the presence of genome polymorphic hotspots, and the identification of non-synonymous mutations. The implications of such diversity on important parasite phenotypes such as pathogenicity and population structure are also discussed. Lastly, the identification of potential virulence factors from both *in-silico* and next generation sequencing studies is examined, offering new insights into potential avenues for future research on neosporosis.

Keywords: Apicomplexa; Genomics; Transcriptomics; Population genetics; Hypothetical protein; Virulence

1. Introduction

Neospora caninum is a ubiquitous cyst-forming Apicomplexan parasite that presents a risk to bovine industries worldwide. Since its initial discovery, neosporosis has been identified as the leading cause of bovine reproduction failure worldwide, as well as a serious neurological disease of canids such as dogs (Dubey et al., 1988a; Dubey and Lindsay, 1996; Reichel and Ellis, 2002; Reichel et al., 2007). It is estimated that neosporosis collectively costs the Australian and New Zealand dairy and beef cattle industries over $110 million per year in losses (Miller et al., 2002; Reichel and Ellis, 2002), and approaching US $546 million in the USA (Reichel et al., 2013). *Neospora caninum* infects a wide host range, the lifecycle of which contains several infectious stages including intracellular tachyzoites and bradyzoites (tissue cysts) found in intermediate hosts, and environmentally impervious oocysts shed by definitive hosts such as dogs (Dubey et al., 2006; Dubey et al., 2007).

While neosporosis as a disease has been recognised for over 30 years, treatment and control options remain inadequate and limited. Of the current strategies available, the balance between effectiveness, economic feasibility, and on-farm applicability is of paramount importance. Reichel and Ellis (2006) compared available options for cost and effectiveness, and concluded that for within-herd prevalences <18% for *N. caninum* infections, inaction was the optimal economic decision over a one year period. Contrastingly for within-herd infection prevalences $\geq$18%, vaccination was considered the most economic strategy. While vaccination with a killed tachyzoite has been described as highly successful in rodent models (Liddell et al., 1999), it is estimated to be only 50% efficacious in cattle, as well as costly (Romero et al., 2004). However, more successful vaccines, such as those utilising live attenuated tachyzoites, could eliminate the risk of an entire herd aborting, reducing the cost of controlling *N. caninum* to vaccination alone, and hence offering a more viable option to passivity (Guy et al., 2005; Miller et al., 2005; Williams et al., 2007).

With all things considered, there is clearly a need for and immense value in investigating and elucidating potential vaccine candidates and therapeutic targets against neosporosis, to which molecular biology, next generation sequencing (NGS) data, and bioinformatics has much to offer. However, genetic diversity within a pathogen has the potential to influence vaccination strategies in a myriad of ways, including vaccine design. This paper therefore considers the current body of knowledge that supports our understanding of *N. caninum* virulence, intraspecies diversity, and population genetics, as well as the challenges and limitations associated with this research. A broad literature search was conducted over the time period from the early 1990s to late 2019. The knowledge presented originates from multiple countries including those in Europe, North America, South America, and Asia, as well as Australia, and New Zealand, to ensure that an extensive, all-inclusive analysis of the *N. caninum* literature is presented and contextualised on a global scale.

The aim of this review was to first provide a history of *N. caninum* genome and diversity studies since the initial discovery of this species (Dubey et al., 1988a). In addition, a comprehensive summary of the current ideas on the population structure and genetics of *N. caninum* is presented, that are based on the new bioinformatics and genome sequencing techniques now at the disposal of researchers.

2. Biological and genetic diversity of *N. caninum*

Significant levels of variation within *N. caninum* populations exist, as genotypic and phenotypic features aren't rigorously conserved within the species (Al-Qassab et al., 2010b). Such diversity is expected due to the broad intermediate host range, worldwide distribution, and the capacity for sexual reproduction. While only minor differences may exist in ultrastructure, *N. caninum* isolates appear to vary in their biological and genetic characteristics, of which many studies have reported differences (Atkinson et al., 1999; Barber et al., 1995;

Barr et al., 1991; Lindsay and Dubey, 1989; McInnes et al., 2006; Rojo-Montejo et al., 2009b). Vertical transmission, and more specifically endogenous transplacental transmission, in cattle and other ruminants during pregnancy is considered not only the predominant route of transmission, but also essential for parasite maintenance within in a herd (Gonzalez-Warleta et al., 2018; Trees and Williams, 2005; Williams et al., 2009). Furthermore, the importance and frequency of sexual reproduction in definitive hosts to parasite biology, epidemiology, and evolution remain unclear, although the presence of dogs on cattle farms are consistently reported as a risk factor for bovine neosporosis (Dijkstra et al., 2001; Ribeiro et al., 2019; Robbe et al., 2016). Therefore, what still requires attention is a better understanding of the major, underlying contributors responsible for reported genetic diversity, and their relevance to important parasite phenotypes such as virulence and population genetics.

2.1. Distinguishing isolates of *N. caninum* by their phenotypes

One of the earliest observations suggesting the existence of heterogeneity within *N. caninum*, was based on *in vitro* growth rate experiments. Schock *et al.* (2001) assessed relative growth rates of six *N. caninum* isolates by measuring ^{3}[H]uracil uptake, and reported significant differences. NC-Liverpool demonstrated significantly faster growth rates in comparison to the other isolates investigated, being twice as fast as NC1, whereas the Swedish bovine isolate, NC-SweB1, was the slowest. A subsequent study demonstrated a lower tachyzoite yield and viability rate of a Spanish isolate of low virulence, NC-Spain 1H, compared to NC1, which also more extensively destroyed the cell monolayer (Rojo-Montejo et al., 2009b).

Initial observations and comparisons in mice and cattle also indicated that isolates of this species exhibit biologically distinct behaviours. NC-Liverpool for example, has been thoroughly studied in murine models, where infection leads to characteristic, clinical neosporosis of weight loss, lethargy and eventual death (Atkinson et al., 1999). Atkinson *et al.*

(1999) reported the presence of moderate to extreme necrotic brain lesions in almost all experimental mice infected with NC-Liverpool, as well as hindlimb paralysis and disco-ordinated movement. This was in contrast to mice infected with NC-SweB1 that presented with a milder meningoencephalitis and less severe symptoms. Subsequently, Miller *et al.* (2002) showed that the NC-Nowra isolate was low in virulence due to the absence of clinical signs of infection in the calf from which it was isolated. Successively, live NC-Nowra tachyzoites were used to vaccinate mice prior to pregnancy, the result of which reduced transplacental transmission of a challenge strain by 80-90% (Miller et al., 2005). Subsequent inoculation of the tachyzoites into cattle prior to pregnancy conferred very high levels of protection against foetal loss, following a challenge by the NC-Liverpool isolate (Weber et al., 2013; Williams et al., 2007).

Additional studies have investigated fetal mortality rates in cattle infected with *N. caninum*, where for example the NC1 isolate was shown to induce fetal death in heifers (Dubey et al., 1988b; Dubey et al., 1992; Innes et al., 2001; Maley et al., 2003), whereas the low virulence NC-Spain 1H isolate does not (Rojo-Montejo et al., 2009a). However, virulent NC-Spain 7 also results in higher fetal mortality rates and IgG response in infected cattle, compared to cattle inoculated with NC1 (Caspe et al., 2012). Furthermore, a recent study examined early infection dynamics of heifers infected at mid-gestation, with low virulence NC-Spain 1H and high virulence NC-Spain 7 (Jimenez-Pelayo et al., 2019). This study demonstrated that NC-Spain 7 tachyzoites reached the placenta earlier, and caused lesional development and fetal mortality following transmission, whereas in comparison, NC-Spain 1H tachyzoites were delayed in reaching the placenta, and lesion development and transmission was not observed during early infection. Based on the knowledge afforded by such comparative studies, elucidating the existence of underlying genetic variation between *N. caninum* isolates with

marked phenotypic differences, presents an invaluable avenue of research warranting investigation.

2.2. Characterising the underlying genetic diversity within the species

A plethora of molecular methods over the last three decades have lent themselves to elucidating potential sources of biological diversity in *N. caninum* at the genetic level. These include randomly amplified polymorphic DNA PCR (RAPD-PCR) (Atkinson et al., 1999; Davison et al., 1999; Schock et al., 2001; Spencer et al., 2000), multilocus mini- and microsatellite analysis (Basso et al., 2009; Regidor-Cerrillo et al., 2013; Regidor-Cerrillo et al., 2006), and multiplex PCR (Al-Qassab et al., 2010a). Initially, various molecular techniques such as these were employed to study common molecular markers including internal transcribed spacer 1 sequence (ITS1), 18S ribosomal DNA, and the Nc5 repeat. However, little to no nucleotide variation was reported for many of these molecular targets studied using a range of isolates, varying in host species, geographical origin, and pathogenicity.

For example, no significant sequence variation was reported in various studies comparing 18S rDNA from NC-Liverpool, NC1, NC-SweB1, and four other bovine isolates (BPA1, BPA2, BPA3, and BPA4) (Barber et al., 1995; Holmdahl et al., 1997; Marsh et al., 1995; Stenlund et al., 1997). Such results show that this marker is conserved within *N. caninum*, since its evolutionary divergence from its closest ancestors such as *T. gondii* (Marsh et al., 1995). Furthermore, while the repetitive Nc5 sequence and ITS1 region are the most commonly used markers for detection of *N. caninum* through PCR (Dubey and Schares, 2006), only some isolates exhibit minor sequence variation in these regions, and hence are not sufficient enough for differentiation purposes (Al-Qassab et al., 2010b; Gondim et al., 2004; Slapeta et al., 2002). Sequence analysis of the ITS1 region however was useful in establishing that equine isolates

of *Neospora hughesi*, the only other known species of the *Neospora* genus, were distinct from both bovine and canine *N. caninum* isolates (Dubey et al., 2001).

Initially, RAPD-PCR was used in various studies to distinguish *N. caninum* from related coccidia, including *T. gondii*, *Sarcocystis* species (Guo and Johnson, 1995; Schock et al., 2001), *Hammondia heydorni* (Sreekumar et al., 2003), *Cryptosporidium parvum* (Schock et al., 2001), and *N. hughesi* (Spencer et al., 2000). This technique was then exploited to investigate intraspecies genetic diversity within the species, where Atkinson *et al.* (1999) was first able to generate distinct profiles for NC-Liverpool and NC-SweB1. After obtaining the first isolate of *N. caninum* from cattle in the UK, Davison *et al.* (1999) reported differences between this designated NC-LivB1 isolate, and NC-Liverpool, NC1, and NC-SweB using RAPD-PCR techniques. Shortly after, Schock *et al.* (2001) analysed DNA from six *N. caninum* isolates from both bovine and canine hosts, as well as three strains of *T. gondii*, *Sarcocystis* sp., and *C. parvum* using RAPD-PCR techniques. A total of 222 of 434 markers were found to be conserved within the *N. caninum* isolates, however distinct from the other Apicomplexa studied. Furthermore, 54 unique *N. caninum* markers exhibited sufficient variability to distinguish each isolate. The RAPD-PCR results also identified *T. gondii* as *N. caninum's* closest neighbour but failed to cluster the individual isolates based on either host or geography.

The markers and techniques discussed that were initially employed to explore genetic diversity within *N. caninum*, while of value at the time, in most cases failed to provide a significant level of genetic heterogeneity between isolates (Al-Qassab et al., 2010b; Regidor-Cerrillo et al., 2006). Furthermore, results produced by RAPD-PCR experiments in particular, were to be interpreted cautiously, due to considerations such as DNA purity, primer selection, and extraneous DNA (Al-Qassab et al., 2010b; Gondim et al., 2004; Spencer et al., 2000). The lack of discriminatory power offered by these methods, also meant that epidemiological and intraspecies population structure analyses remained uncharted and unresolved for *N. caninum*.

Subsequently, the gold standard for assessing genetic diversity in *N. caninum,* became the analysis of polymorphic mini- and microsatellites (Al-Qassab et al., 2010b; Regidor-Cerrillo et al., 2006). Such repetitive elements offered a higher degree of discriminatory power compared to other routinely used phylogenetic markers, where distinctive patterns exhibited by various isolates proved useful in distinguishing and characterising genetic diversity within the species. Mini- and microsatellite techniques have also been extensively used for genotyping other Apicomplexan species. For *T. gondii,* Ajzenberg *et al.* (2010) developed a simple genotyping method that incorporated 15 microsatellite markers across 11 chromosomes in one multiplex PCR assay. These markers were validated for 26 *T. gondii* reference isolates and were able to differentiate isolates both at the typing level (i.e. the three main clonal lineages compared to atypical isolates), and the fingerprinting level for distinguishing related isolates within lineages.

Regidor-Cerrillo *et al.* (2006) first identified and analysed 12 microsatellites for nine *N. caninum* isolates of different host and geographical origin. Multi-locus microsatellite analysis revealed unique profiles for each isolate, where 11 of the 12 markers were subsequently applied to clinical samples resulting in the detection of new alleles (Pedraza-Diaz et al., 2009). A larger subsequent study investigated 25 cultured *N. caninum* isolates and reported variation within repetitive sequences of eight of the 27 loci studied (Al-Qassab et al., 2009). A multiplex PCR assay was consequently developed incorporating three minisatellite and three microsatellite markers, presenting a simple, efficient, and informative method for genotyping and distinguishing new and existing *N. caninum* isolates (Al-Qassab et al., 2010a).

These genotyping techniques have also contributed to the study of epidemiology and population genetics within the species (Basso et al., 2009; Basso et al., 2010; Regidor-Cerrillo et al., 2013; Regidor-Cerrillo et al., 2006). A large scale genotyping study evaluated nine microsatellite markers for a total of 108 reference and clinical *N. caninum* isolates, collected

over a 10 year period from four countries and two continents (Regidor-Cerrillo et al., 2013). A total of 96 microsatellite multilocus genotypes across seven loci revealed extensive levels of genetic diversity across the samples. Data analysis suggested that these microsatellite markers were partially correlated with geographical origin, with sub-structuring present for each country population of samples. Genetically distinct clustering of microsatellite genotypes was also found to be associated with abortion, and the results of the study suggested a clonal propagation of microsatellite markers for Spanish *N. caninum* isolates in cattle.

Basso *et al*. (2010) characterised DNA extracted from the brains of 18 foetuses infected with *N. caninum* from epidemic abortion storms, using ten microsatellite markers. The same microsatellite pattern was present in all foetuses from each abortion outbreak, and unique to each herd. These results, in conjunction with avidity analysis, provided evidence that infection had originated from a common source, and therefore that horizontal transmission was a determinant of epidemic abortions. Recently, Cabrera *et al.* (2019) reported that four isolated Uruguayan strains of *N. caninum*, represented three distinct genetic lineages, as determined by microsatellite typing. Phylogenetic analysis further revealed that three of these four strains clustered closely with strains from regional Argentina and Brazil. As the remaining Uruguayan strain was found to group with an unrelated cluster, this suggested that the local strains were from multiple origins.

3. New perspectives: Elucidating the degree of sequence variation between *N. caninum* isolates

The popular typing methods used in genetic diversity studies, as discussed in Section 2.2, commonly exploit repetitive elements such as mini- and microsatellites. While valuable, such techniques generally target non-coding sequences, and their influence on parasite biology and phenotype remains largely unknown. What still remains to be addressed is whether a

relationship exists between genetic markers, and heterogeneity in clinical manifestations in infected hosts (Al-Qassab et al., 2010b; Goodswen et al., 2013). Alternatively, investigating variation present within protein-coding genes of a species can expand our understanding of whether such mutations alter gene function, and consequently an organism's phenotype.

Until recently, the presence of polymorphisms, including single nucleotide polymorphisms (SNPs) and insertions and deletions (indels), within coding sequences of *N. caninum* isolates remained unexplored, representing a gap in knowledge that has been extensively studied in related apicomplexan parasites. For example based on genome-wide SNP studies, the within-lineage variation distinguishing three of the four major lineages of *T. gondii* in the Northern Hemisphere, is <0.01%, compared to an estimated 1-3% variation present between lineages (Boyle et al., 2006). Furthermore, extensive research has revealed that only a limited number of sexual recombination events within the feline definitive host of *T. gondii,* are sufficient to influence population structure and the evolution of pathogenic variability between linages (Boyle et al., 2006; Grigg et al., 2001; Khan et al., 2011a; Khan et al., 2009). In the malaria causing *Plasmodium falciparum*, mutations in various molecular markers was identified as part contributors to the growing global threat of drug resistance. For instance, mutations in the propeller domain of the Kelch 13 gene were associated with artemisinin resistance, hence classifying this domain as a useful marker to monitor the emergence of resistance across South East Asia (Ariey et al., 2014).

3.1. The identification and annotation of hypervariable gene hotspots in the *N. caninum* genome

To address the important, and until now, neglected question of the impact of sequence variation in protein-coding genes on *N. caninum* diversity, Calarco *et al.* (2018) generated NGS data from the tachyzoite lifecycle stage to subsequently perform *in-silico* analyses. RNA-seq data

produced for NC-Liverpool and NC-Nowra isolates was used in a variant analysis pipeline (Figure 1), which resulted in the identification of more than 3,000 differentiating SNPs. The allocation of this SNP callset within the *N. caninum* genome revealed their concentration within specific genomic windows on chromosomes VI, XI, and XII, comprising a total of 19 tachyzoite-associated SNP "hotspots" within transcriptionally active coding regions.

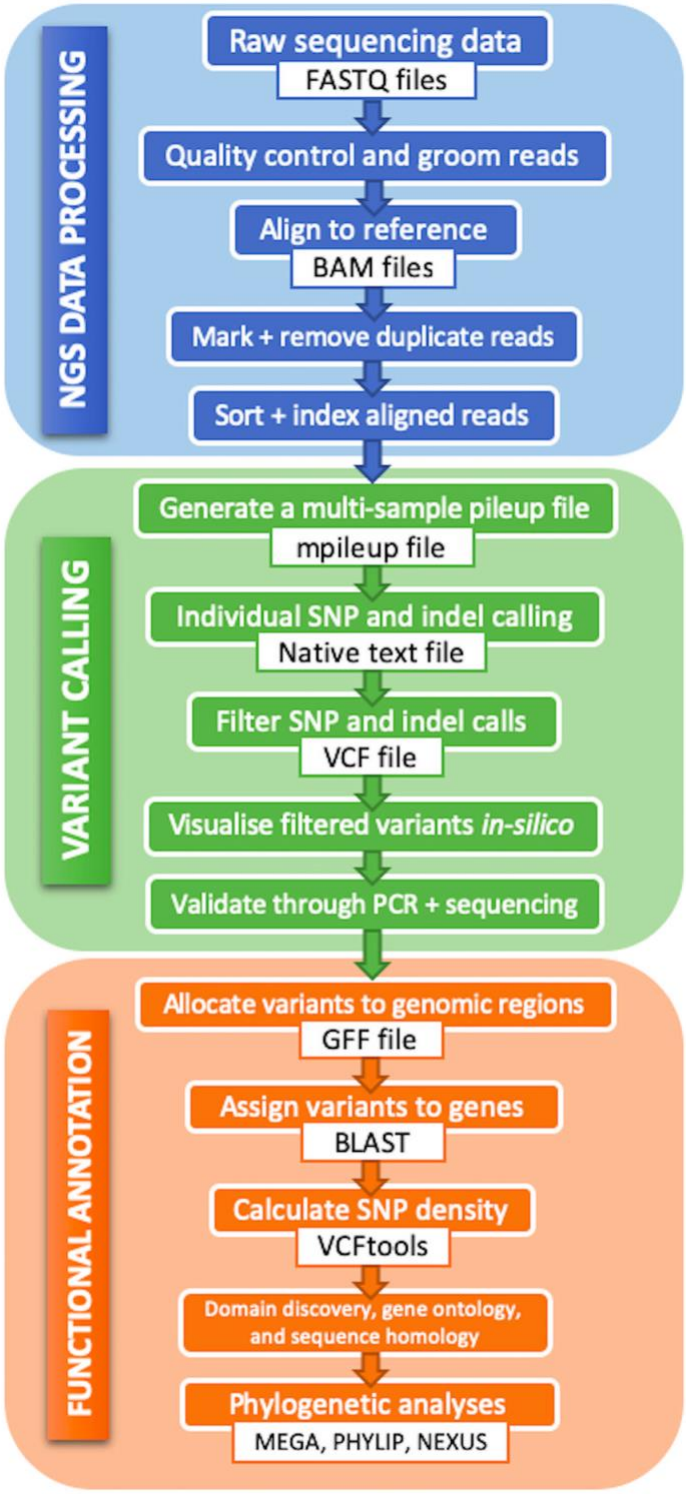

Figure 1. Variant analysis pipeline for detecting SNPs and indels, from raw high-throughput sequencing data generation through to functional analysis of identified variants. Sequence reads produced by next generation sequencing (NGS) platforms, require initial quality control. This can involve trimming the ends of reads where base quality tends to

decrease, discarding short sequences, and removing low quality base calls. Groomed reads can then be aligned to a reference genome or transcriptome, where optimising default alignment parameters can help improve overall mapping quality, especially for non-model organisms such as *N. caninum*. Aligned reads can then be further processed to remove duplicate reads produced through PCR, followed by sorting and indexing BAM files, which are additional steps required by many variant callers. For multi-sample variant calling, an 'mpileup' file can subsequently generated from multiple aligned BAM files, which can then be fed into a variant calling algorithm to detect SNPs and indels, in comparison to a reference FASTA file. Filtered variants should then be visualised to assess their quality, and a subset of variants should be selected for confirmation through PCR and sequencing analysis. Annotating high-quality variant calls can consequently involve allocating them to coding and non-coding genomic regions, performing BLAST analyses to assign polymorphic loci to genes, and also identifying protein sequence domains and repeats, and gene ontologies to assign functional significance to variable loci.

Annotation of these hypervariable genes, many of which contained non-synonymous mutations, revealed their implication in protein-protein interactions, transcription and translation processes, protein binding, ribosomal subunit formation, kinase activity, and protein phosphorylation. Moreover, sequence annotation of the hotspot genes revealed noteworthy protein superfamilies and domains such as WD40 repeat containing (IPR036322), ARM-like helical (IPR011989), P-loop containing nucleoside triphosphate hydrolase (IPR027417), ABC transporter (IPR036640), EF-hand calcium binding domain (IPR018247), AAA+ ATPase domain (IPR003593), and tetratricopeptide repeat (TPR; IPR019734).

The SNP-dense protein-coding genes identified in this study, many of which were characterised as non-synonymous mutations, could represent transcriptionally active, novel determinants of tachyzoite virulence, that distinguish *N. caninum* isolates with reported differences in pathogenicity. Upon further sequence analysis, various hotspot genes also contained transmembrane (TM) domains and/or a signal peptide (SP), indicating the presence of structural sequence features often associated with proteins that are located in the membrane or excreted/secreted that may be potential vaccine candidates.

Interestingly, also reported in this study was almost a complete absence of sequence polymorphisms located within routinely used genotyping markers for *T. gondii*, including surface antigens (SAG1 and SRS2), dense granule proteins (GRA; GRA6), and rhoptry proteins (ROP; ROP30 and ROP37). These results supported previous studies investigating such markers, where Marsh *et al.* (1999) reported complete sequence conservation in genes encoding SAG1 and SRS2, between five *N. caninum* isolates from different continents and host species. Similarly, Walsh *et al.* (2001) compared *GRA6* and *GRA7* gene sequences between four bovine and canine isolates, and found no nucleotide differences.

3.2. The contribution of non-coding SNPs to *N. caninum* intraspecies genetic diversity

The metabolically expensive presence and hence potential functional importance of non-coding DNA has been a topic of growing debate and contention. While initially deemed "junk DNA", non-coding sequences are being elevated in status as potential contributors to the evolutionary diversity and protein repertoire of various species (Fedorova and Fedorov, 2003; Gilbert, 1985; Irimia and Roy, 2008). Furthermore, it was suggested that sequence polymorphisms within non-coding DNA have the potential to influence genotype-phenotype relationships and gene expression (Cooper, 2010).

Calarco and Ellis (2020) exploited both genomic and transcriptomic data to investigate SNPs located within non-coding and intron regions between isolates. *In-silico* variant analysis was performed using NC-Nowra genomic reads, aligned against the NC-Liverpool reference genome from ToxoDB. The results suggested that the *N. caninum* genome was largely monomorphic, however hypervariable regions were identified on chromosomes VI and XI. Within the variable hotspots, more than half of the identified SNPs on chromosome VI were located within introns, and approximately 70% of SNPs on chromosome XI were in either introns or non-coding genomic regions. The annotation of these hypervariable genes containing SNP-dense introns, revealed gene ontologies (GOs) related to transcription and translation processes, protein kinase activity, ATP binding and ATPase activity, transmembrane transport, and ion binding. There was also a distinct overlap of and correlation between genes containing SNP-dense introns and coding regions. Prioritisation based on SNP density within their respective introns, resulted in the identification of a set of functionally significant genes located on chromosomes VI and XI, including an ABC transporter involved in transmembrane transport and ATP binding (*NCLIV_015830*), a calcium-activated ion channel protein with eight TM domains and a SP (*NCLIV_015820*), and a kinesin-like protein involved in microtubule movement (*NCLIV_056770*).

Further to the topical discussion on non-coding DNA, the role of introns in evolutionary diversity through mechanisms of alternative splicing (AS) is gaining momentum and attracting debate. This includes the impact of phenomena such as exon shuffling, intron retention, mRNA surveillance, transcriptional regulation of gene expression, and recombination (Duret, 2001; Fedorova and Fedorov, 2003; Gilbert, 1985; Lynch and Richardson, 2002). With respect to protozoan parasites, approximately 22% of protein-coding genes in *T. gondii* (Yeoh et al., 2015), and 16% in *Plasmodium* species (Iriko et al., 2009) are estimated to undergo AS, where studies suggest that AS mechanisms can generate distinct, functionally significant transcripts in Apicomplexan parasites, where such gene products are also considered as prospective vaccine candidates (Agarwal et al., 2011; Delbac et al., 2001; Gabriel et al., 2015; Kern et al., 2014).

By implementing *in-silico* structural variant identification pipelines, Calarco and Ellis (2020) reported 315 and 503 deletions by the Pindel and BreakDancer algorithms respectively, spanning more than 50bp between RNA-seq reads generated from NC-Liverpool and NC-Nowra tachyzoites. Many of these predicted deletions were subsequently identified as introns, present in transcripts from NC-Liverpool, however absent from, and likely removed through splicing in, NC-Nowra transcripts. The corresponding genes containing these introns, presented GOs for RNA processing, protein binding, proteolysis, methyltransferase activity, and oxidation-reduction process, where many also contained TM domains and/or a SP.

In addition, many predicted *in-silico* deletions were instead in regions either just before or after a gene sequence (i.e. non-coding or intergenic sequences). Annotations for such genes including those putatively encoding microneme proteins (MIC3, NCLIV_010600; MIC8, NCLIV_062770), a protein kinase (NCLIV_050650), and a translation initiation factor subunit protein (NCLIV_011760). It was suggested that these deletions may belong to regulatory elements that influence gene expression, or alternatively that reference gene sequence

annotations are inaccurate. With respect to functionally important micronemes, following the discovery of the *N. caninum* MIC8 protein, which shows a high similarity to that of its homologue in *T. gondii*, this protein was reported to play a critical role in host-cell invasion, and also to form a complex with NcMIC3, most likely during transportation (Wang et al., 2017). Subsequently, Zhang *et al.* (2019) demonstrated that immunising mice with recombinant NcMIC3 and NcMIC8, provided at least part protection against neosporosis following a challenge with *N. caninum* tachyzoites, with mice also exhibiting a lower parasite burden in brain tissue. Such studies demonstrate the importance of microneme proteins in not only host cell invasion and interaction, but also as vaccine candidates against neosporosis.

Overall, these studies revealed and characterised the contribution of variable, non-coding and intron DNA, to the intraspecies diversity reported for *N. caninum*. Such research improves upon our understanding of the underlying genetic forces occurring in this species. It also suggests that mechanisms of alternative splicing, specifically intron retention in transcripts, distinguish isolates of this species, and may play a larger role in the context of parasite biology, than previously thought. By collating these new studies, we start to gain a clearer picture of the underlying, intrinsic genetic differences occurring in isolates of *N. caninum*.

It is evident that regions on chromosomes VI and XI are highly polymorphic between isolates, in both transcriptionally active tachyzoite-associated genes, as well as intergenic and intron sequences. These data hence expand upon our understanding of the driving forces responsible for important parasite phenotypes and biological mechanisms, conducive to the species' success. In a historical context for example, do these genetic diversity studies indicate that sexual recombination in definitive hosts plays a more important role that previously thought? Do the concentrated polymorphic regions identified point towards sex being of greater significance and influence in parasite evolution, epidemiology, and intraspecies diversity?

4. The evolution and heterogeneity of coccidian genomes

How can we contextualise variation at the genetic and transcriptomic levels to establish its relationship with the population structure, epidemiology, and phylogeny of a species? Revealing the existence of a species' underlying population structure can offer insight into the spread and determination of virulence factors, pathogenicity, and genotyping markers (Khan et al., 2011b). Recent efforts towards identifying genome-wide SNP markers for a range of parasite species, is providing a new framework based on population genetics, with implications in molecular evolution, species divergence, and association studies.

For example, initial genotyping studies for *T. gondii,* and their interpretation in terms of a clonal population structure for the species, were limited by the small number of isolates studied and their host and geographical origins (Darde et al., 1988; Howe and Sibley, 1995; Sibley and Boothroyd, 1992). However, subsequent studies incorporating multilocus techniques for a wider range of isolates, revealed the existence of a more genetically diverse species with a complex population structure (Ajzenberg et al., 2004). Large-scale sequencing efforts for *T. gondii* resulted in over 250 SNPs within random fragment length polymorphism (RFLP) markers, being observed between the three main clonal lineages (Khan et al., 2005). Based on genome-wide SNP studies, it was estimated that the polymorphism rate between the lineages is 0.65% (Boyle et al., 2006), which equates to approximately one variant per 100bp throughout almost all chromosomes (Khan et al., 2005). The three predominate lineages are also classified as either type I, II, or III SNPs (Boyle et al., 2006), with each exhibiting extensive bi-allelism (Grigg et al., 2001; Khan et al., 2005). The distribution of these SNP types also suggests that types I and III represent first and second generation offspring, resulting from a cross between a type II strain and an ancestral strain (Boyle et al., 2006).

It is well understood that diversification, duplication, and expansion of loci is ubiquitous and prevalent in apicomplexan genomes, where this is especially true of protein-coding genes found in a parasite's secretory pathway, and/or expressed on their surface (Blank and Boyle, 2018). The SRS superfamily of surface antigens for example, has been identified as one of the most divergent and rapidly evolving protein families within this phylum (Jung et al., 2004; Manger et al., 1998; Reid et al., 2012; Wasmuth et al., 2009; Wasmuth et al., 2012). As a result, gene duplication and expansion events have been used to study the phenotypic differences and phylogenetic relationships between Apicomplexan species, and more specifically, within the *Toxoplasmatinae* subfamily, with implications in species divergence and evolution (Adomako-Ankomah et al., 2014; Lorenzi et al., 2016; Reid et al., 2012). The genome and transcriptome sequencing of related Coccidia such as *T. gondii*, *N. caninum*, and *H. hammondi* has revealed a high degree of conservation and genomic synteny between these species, with respect to size, protein-coding genes, GC content, and gene catalogue (Reid et al., 2012; Walzer et al., 2013). However, the study of large structural variations at expanded loci have revealed sources of evolutionary divergence within this subfamily.

For example, more than double the number of *SRS* genes were reported for *N. caninum* by Reid *et al.* (2012) compared to *T. gondii*, where there was also a divergence in secreted virulence factors between the species, namely rhoptry kinases. These results suggested that a small set of genes implicated in host-parasite interactions, have influenced the ecological niches and pathogenic capabilities of these species. By comparing the genomes of 62 globally dispersed *T. gondii* isolates to *N. caninum*, *H. hammondi*, and *S. neurona*, Lorenzi *et al.* (2016) reported that these closely related, but phenotypically diverse, parasites could be distinguished based on the tandem amplification and diversification of secreted, pathogenic determinants. Duplicated genes and copy number variation (CNV) events were specifically in *T. gondii* genes encoding ROPs, MICs, and SAG/SRS, many of which also showed evidence of positive

selection based on non-synonymous mutation rates. Comparing the orthologous *GRA*, *ROP*, and *SRS* genes between related Coccidia, suggested that divergence within these genes may be responsible for phenotypic differences observed between these species, whereas MIC genes were highly conserved. Furthermore, Marsh *et al.* (1999) indicated that while the SAG1 and SRS2 protein sequences were conserved across six *N. caninum* isolates originating from different hosts and geographical regions, there was a 6% and 9% difference in these amino acid sequences respectively, when compared to *N. hughesi*. The variation within these protein-coding genes therefore represented additional molecular markers for the distinction of these two *Neospora* species.

4.1. Current ideas on *N. caninum* population structure

Through the *in-silico* identification and laboratory confirmation of a set of SNPs using NGS data, Calarco *et al.* (2018) developed a multilocus sequencing approach for nine *N. caninum* isolates, differentiated by geographical origin, host, and/or reported pathogenicity. This data elucidated a population structure consisting of two major clades, one of which included the virulent NC-Liverpool isolate, and the other containing the low virulence NC-Nowra isolate. Interestingly, the clustering of the isolates based on polymorphic loci did not appear to conform to a pattern based on their geographical segregation or host. Furthermore, the inclusion of *N. hughesi* in the multilocus sequence typing, bolstered support for the two-clade structure for *N. caninum*.

However, there are a limited number of studies documenting and comparing the pathogenicity of the *N. caninum* isolates investigated in this study, in both murine and bovine host models. While this limits the assumptions that can be drawn from such results, it does highlight the value of SNP data in conjunction with population structure analyses, in enriching our understanding of intraspecies genetic diversity and important parasite phenotypes. These

results emphasised that the two representative populations of *N. caninum* presented not only differ in their pathogenicity as previously demonstrated but are also genetically distinct. Such knowledge can expedite the identification of novel virulence markers, and hence provide candidates for vaccine and drug development against neosporosis.

Furthermore, few studies have been conducted to address the crucial question of whether the genetic diversity associated with *N. caninum*, and hence population genetics, effects the pathogenicity of the species. The paucity of knowledge available to address this question can be attributed to the fact that available studies only focus on a limited number of *N. caninum* isolates, which lack variability by geographic distribution and/or host range (Khan et al., 2019). Additionally, such analyses are generally of low resolution, where they are based on minimal genetic markers, or fail to employ a suite of appropriate population genetic tools. A recent study published by Khan *et al.* (2019) however, aimed to address these shortcomings, and elucidate and improve upon the current understanding of *N. caninum's* population genomics, by exploiting NGS data. In this study, 47 isolates from a wide range of hosts and geographic locations were genotyped, along with three *N. hughesi* isolates, across a total of 19 markers.

Existing molecular epidemiological studies, as discussed earlier in this review, have collectively suggested that extensive genetic diversity exists between *N. caninum* isolates (Al-Qassab et al., 2010a; Basso et al., 2009; Pedraza-Diaz et al., 2009; Regidor-Cerrillo et al., 2013; Regidor-Cerrillo et al., 2006). However, the phylogenetic and population structure analyses conducted by Khan *et al.* (2019) supports a model whereby a single, highly inbred *N. caninum* genome has recently evolved from a common ancestor, and experienced a global selective sweep. Phylogenetic analyses based on nine sequenced markers resulted in all *N. caninum* isolates being tightly clustered, with *N. hughesi* representing a second population. Pairwise F_{ST} calculations classified the geographically dispersed *N. caninum* isolates as closely genetically related, as well as being highly divergent from *N. hughesi* isolates, indicating a non-subdivided

population amongst *N. caninum* isolates. Additionally, this study used whole-genome sequencing (WGS) from six *N. caninum* and one *N. hughesi* isolate, to perform genome-wide SNP analysis against the NC-Liv reference genome. The SNP distribution plots generated demonstrated a monomorphic *N. caninum* genome, with the exception of six distinct sequence haploblocks, each of which presented elevated SNP densities. These results were consistent with previous SNP data from small *N. caninum* genomic windows in both coding and non-coding regions, specifically on chromosomes VI, XI, and XII (Calarco et al., 2018; Calarco and Ellis, 2020). By also integrating seroprevalence data from African cattle, this study presented the impact of migration of European domesticated cattle on the genetic sweep of the *Neospora* genome, resulting in the recent evolution and global expansion of a highly inbred, single lineage of *N. caninum* through unisexual mating and vertical transmission.

4.2. Exploring non-nuclear DNA in apicomplexan parasites

Another potential source of genetic variation recently investigated and reported for *N. caninum* was non-nuclear DNA, where Calarco and Ellis (2020) assembled and annotated the apicoplast genome for NC-Liverpool, based on NGS data. The apicoplast is a secondary plastid organelle unique to most species within the apicomplexan phylum, and while it lacks photosynthetic ability, it is essential for parasite survival (Gardner et al., 1991; Howe, 1992; Williamson et al., 1994; Wilson et al., 1996). The published apicoplast genomes of *P. falciparum* (Wilson et al., 1996), *T. gondii* (EMBL accession number U87145), and *Eimeria tenella* (Cai et al., 2003) have been shown to be highly similar with respect to gene catalogue and structure. This includes an inverted repeat region containing duplicated small and large subunit ribosomal RNAs (*SSU rRNA* and *LSU rRNA*), and tRNA genes. The conservation of this organellar genome between genera also extends to the transcription of genes, where half of the apicoplast

circle is transcribed in a clockwise direction, and the other half is transcribed counter-clockwise.

While most orthologous genes from these three coccidian species align to one another, with the exception of hypothetical open reading frames (ORFs), their sequences are divergent. Contrastingly, the apicoplast genomes of piroplasmids such as *Theileria parva* (Gardner et al., 2005) and *Babesia bovus* (Brayton et al., 2007) do not contain an inverted repeat region, and most of their genes are single-copy and transcribed uni-directionally. The sequencing of apicoplast genomes from various apicomplexan species is conducive to elucidating the evolution, phylogeny, population structure, and biological mechanisms of these parasites, which can subsequently contribute to our understanding of the epidemiology and pathogenicity within and between species.

For example, to address limitations associated with nuclear SNP barcodes, including a lack of geographic specificity, Preston *et al.* (2014) exploited the extra-nuclear mitochondrial and apicoplast organelles for *P. falciparum* isolates. Sequencing data from a total of 711 isolates across 14 countries was used to identify high quality SNPs within these genomes, and subsequently analyse the geographic diversity and extent of recombination within the population. A total of 151 mitochondrial and 488 apicoplast SNPs were documented, where variation within these combined non-nuclear genomes resulted in a 23-SNP barcode able to discriminate between the region of sample origin with an accuracy of 92%. High linkage disequilibrium between the catalogued SNPs supported the co-transmitted and non-recombining nature of the organelles, revealing novel haplotypes unique for different geographic regions.

Variant analysis of reads from NC-Nowra aligned to an assembled NC-Liverpool apicoplast genome, revealed that this organelle is highly conserved between these isolates with marked biological differences. Calarco and Ellis (2020) only reported three SNPs and one

insertion three base pairs long present in NC-Nowra reads across the ~35kb apicoplast genome. The SNPs were located within RNA polymerase β subunit (*rpoB*), RNA polymerase β' subunit (*rpoC2.2*), and ORF-F, and the insertion was positioned just before the start codon of the *tufA* gene. High sequence similarity was also reported between the NC-Liverpool apicoplast genes when aligned to closely related coccidia such as *T. gondii*, *H. hammondi*, *S. neurona*, and *Cryptosporidium suis*, with identities between 79-93%. The structure, gene content, and arrangement of the NC-Liverpool apicoplast genome was also highly similar between these related coccidia and *Plasmodium* species, suggesting the evolutionary conservation of this essential organelle. Figure 2 presents the NC-Liverpool *N. caninum* annotated apicoplast genome published by Calarco and Ellis (2020) (MK770339), as aligned to a reference apicoplast genome sequence from *T. gondii* (U87145), which highlights how the nucleotide sequence, and gene arrangement and catalogue of this organelle are highly conserved between these related species.

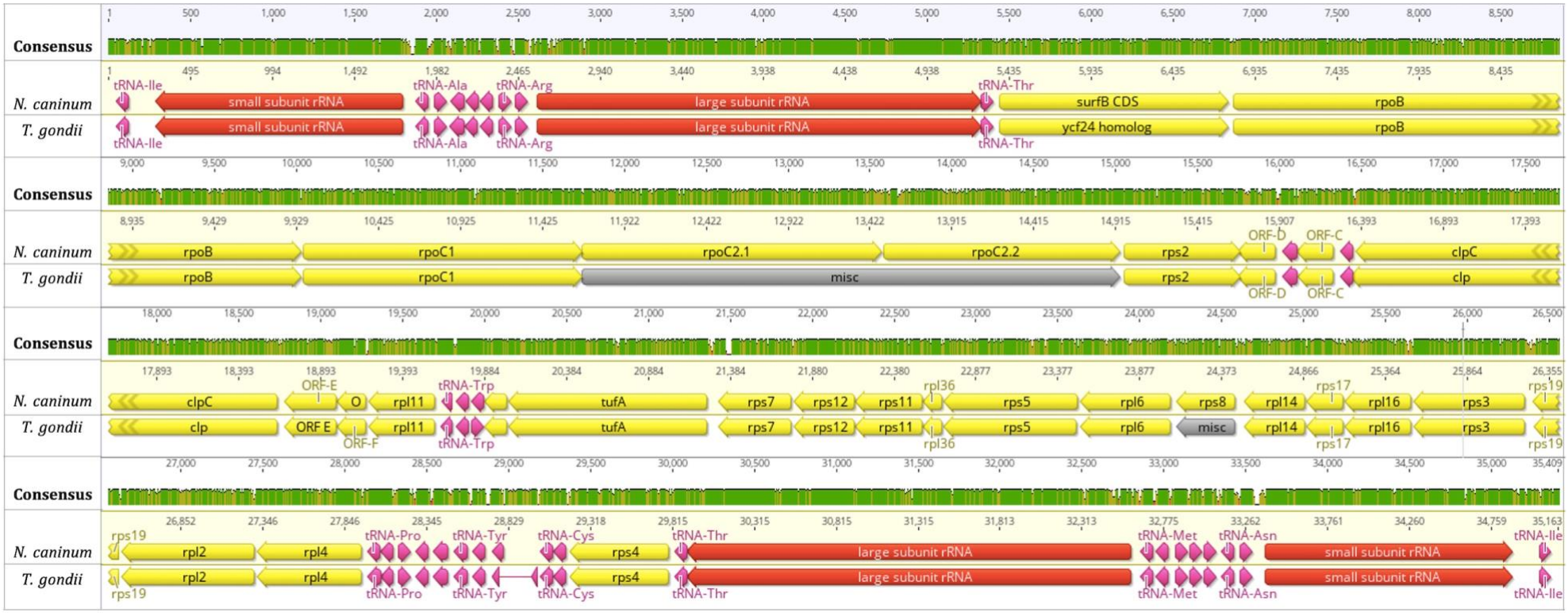

Figure 2. Comparison of the gene catalogue and structure of the *N. caninum* and *T. gondii* apicoplast genomes. The NC-Liverpool apicoplast genome (accession MK770339) was aligned to a *T. gondii* reference apicoplast genome (accession U87145) in Geneious Prime (www.geneious.com; version 2020.0.4). Genes are presented in yellow, rRNAs are in red, and tRNAs are shown in pink. The top consensus track shows the percentage identity between the two apicoplast nucleotide sequences, where the green represents 100% sequence similarity, yellow is between 30-100% sequence similarity, and red is <30% sequence similarity at each base. This figure highlights how the ~35 kb apicoplast genome is highly conserved between these two closely related species.

As mitochondria are inherited maternally and less susceptible to genetic exchange, these genomes are commonly exploited to differentiate between asexual and sexual reproduction (Hutchison et al., 1974). When attempting to improve upon the current understanding of *N. caninum* population genetics, Khan *et al.* (2019) used three mitochondrial sequenced markers out of a total of 12 to genotype 47 *N. caninum* isolates. Sequence polymorphisms present in the *COX1*, *COX3*, and *CYTB* mitochondrial markers were subjected to maximum-likelihood analysis to determine how the *N. caninum* mitochondrial genomes were genetically related, and hence indicative of sexual recombination. The results revealed that the mitochondrial genomes of NcSp7 and NcSp9, strains which were isolated from different cattle on the same Spanish farm, had been inherited from a distinct source, based on the presence of two unique SNPs. Two additional strains, Rhino1 and Rhino2 isolated from the one host, also possessed a single SNP within these sequenced markers. One possible explanation for these results is that the two sets of divergent strains evolved through distinct recombination events. Alternatively, the SNPs distinguishing the mitochondrial genomes of these strains may be the result of genetic drift. However, as the nuclear genomes of all four of these strains were shown to be identical to the remaining *N. caninum* investigated in this study, the authors (Khan et al., 2019) postulated that unisexual expansion is occurring within the *N. caninum* population. This study demonstrated the value of exploiting both nuclear and non-nuclear sequenced markers, to improve upon our understanding of species diversity and population genomics.

5. *Neospora caninum* as a non-model organism: What are we still missing?

One of the greatest challenges currently plaguing the advancement of treatment and control strategies against neosporosis, is at least partly attributable to the status of *N. caninum* as a non-model organism. Consequently, there are limited resources available for the species, and for the data that has been generated, there is the question of the robustness, accuracy, and

completeness of such records, including the reference genome. This is further compounded by the lack of gene and protein annotations presented in reference resources and databases, where putative gene descriptions commonly stem from sequence homology with closely related and thoroughly studied Apicomplexa. Currently, more than 60% of genes present in the *N. caninum* genome remain uncharacterised, and instead are described as coding for 'hypothetical' or 'unnamed' proteins (Calarco and Ellis, 2019).

The SNP hotspot regions identified by Calarco *et al.* (2018) mostly contained genes that were simply labelled "hypothetical" or "unspecified". Annotation using several *in-silico* tools, led to gene classification by homology, protein families, domains, and/or GO terms. However, one quarter of the SNP hotspot genes still remained unannotated after all methods were exhausted. This clearly highlights how such genetic diversity analyses are only as valuable as the annotation available for important loci identified, and hence overall for the organism under investigation. It also alludes to the disturbing and sizeable existence of uncharacterised, theoretically significant proteins that await description for *N. caninum*.

5.1. Virulence factors of Apicomplexa

Members of the Apicomplexa, whilst biologically specialized, share a plethora of cellular and molecular characteristics that are conducive to their success as intracellular parasites. This includes chiefly an apical complex, and molecules that aid in parasite motility, host cell adhesion, and subsequent invasion (Huynh et al., 2014). Rhoptries, micronemes, and dense granules are the three secretory vesicles constituting the apical organelle, which release effector proteins essential for host cell recognition, invasion, and modification, and are therefore instrumental in the success of parasite virulence (Blackman and Bannister, 2001).

Rhoptry proteins (ROPs) are amongst the most comprehensively studied family of proteins in Apicomplexa, where many have been identified as key virulence factors (Talevich

and Kannan, 2013). These include ROP18, which modulates parasite growth, and ROP5, which together play a role in impeding the host immune response (Behnke et al., 2012; Saeij et al., 2006; Talevich and Kannan, 2013). In a gene knockout study for example, Ma *et al.* (2017b) deleted the *ROP5* gene from the NC1 strain, and subsequently assessed the knockout strains at the gene, protein, and transcription levels. Hallmarks of ROP5-deficient parasites when compared with the parent strain included decreased host-cell invasion capacity and reduced proliferation and virulence. This study characterized *ROP5* as a crucial *N. caninum* virulence factor and contributed vital information to the underlying mechanisms of host-parasite interactions. In a similar study, the same gene-knockout technique was used to construct a ROP16-deficient *N. caninum* strain, where mice infected with this strain presented with significantly reduced cerebral parasite load and decreased parasite proliferation (Ma et al., 2017a). The results showed *ROP16* to be a key *N. caninum* virulence factor, which induced host-cell apoptosis through the continuous phosphorylation of STAT3.

The microneme (MICs) family of proteins can be classified by the presence of TM domains, a SP, and adhesive motifs. Such structural features within these secreted proteins enable the parasite to interact with host cell receptors, thereby facilitating successful attachment to and invasion of host cells. In *T. gondii,* the MIC2 complex was implicated in gliding motility and cell invasion (Jewett and Sibley, 2003), and is acknowledged as a major virulence determinant in *Toxoplasma* infection. MIC2-deficient parasites have also been exploited in the development of a live attenuated vaccine (Beghetto et al., 2005; Dautu et al., 2007; Huynh and Carruthers, 2006). Furthermore, while MIC2 was initially described by Lovett *et al.* (2000), it remains listed as a "hypothetical protein' in NCBI, UniProt, and ToxoDB reference databases

Following host cell invasion, dense granule (GRA) proteins are released and serve to modify the conditions of the parasitophorous vacuole (PV) membrane and environment,

ensuring the successful intracellular maintenance and replication of the parasite (Cesbron-Delauw, 1994; Mercier et al., 2005; Mercier et al., 2002). A total of 12 GRA proteins were identified in *T. gondii*, many of which contain a SP sequence between 25-30 amino acids, and/or TM domains, and as such are classified as excretory/secretory antigens, associated with the secretory pathway of the parasite (Ahn et al., 2005; Cesbron-Delauw, 1994; Mercier et al., 2005; Michelin et al., 2009; Rome et al., 2008). Nishikawa *et al.* (2018) for example reported that GRA7 plays a regulatory role in neosporosis pathogenesis, through its ability to modulate the host immune response. Using the CRISPR/Cas9 system, GRA7-deficient parasites were generated and demonstrated reduced virulence in mice, and a decreased level of parasite burden and necrosis in brain tissue, thereby identifying GRA7 as an important virulence factor in *N. caninum* infection. Comparably, Yang *et al.* (2018) first identified a novel GRA protein, GRA17, and subsequently employed the CRISPR/Cas9 gene editing system to generate an *N. caninum* GRA17 knockout strain. This strain resulted in reduced parasite virulence and proliferation in mice and the development of PVs with abnormal morphology, again identifying GRA17 as in important regulator in *N. caninum* pathogenicity and PV formation.

Many key players contributing to crucial parasite mechanisms of invasion, motility, and adhesion, contain distinct sequence features that can be identified *in-silico*. For example, proteins containing TM domains are implicated in a range of crucial biological processes such as cell attachment, invasion, and molecule transport (Reynolds et al., 2008), whereas proteins transported to secretory organelles commonly require an N-terminal SP (Chen et al., 2008). Such aforementioned effector molecules implicated in parasite invasion, host-cell signaling modulation, and adhesion, generally contain important structural features such as these, and have hence been identified as potential vaccine candidates for apicomplexan parasites (Kim and Weiss, 2004). In terms of gene expansion and diversification in *T. gondii*, and its relevance to phenotypic diversity within and between species, Adomako-Ankomah *et al.* (2014) reported

that a total of 42 of 53 tandemly expanded loci identified were predicted to be part of the parasite's secretory pathway, and 29 of these contained an N-terminal SP. Such studies emphasize the on-going importance of bioinformatics and the need for novel algorithms to facilitate the detection of crucial virulence factors buried within the *N. caninum* proteome.

5.2. The value of bioinformatics in the quest to resolve the 'hypothetical'

The advent of second and third generation sequencing technologies has seen the mass generation and availability of data for a wide range of organisms, increasingly so for *N. caninum*. To accommodate for this burgeoning field of technology, *in-silico* tools and pipelines are constantly being developed and optimised, to ensure researchers are able to process, analyse, and make such data publicly available. Many well studied and non-model organisms have benefitted from *in-silico* analysis of hypothetical proteins. For example, Oladele *et al.* (2011) identified several sequences that could represent biomarkers of malaria, through bioinformatics analysis of hypothetical *P. falciparum* proteins. Syn *et al.* (2018) presented an *in-silico* pipeline to identify *T. gondii* proteins influencing host cell epigenetic regulation. The tools used aimed to identify proteins secreted via classical and non-classical pathways, predict proteins localised to the nucleus, and prioritise proteins associated with epigenetic regulation. From a total of 8313 proteins in the *T. gondii* proteome, 57 proteins were predicted to be involved in both secretory pathways, and epigenetic functions.

By combining current knowledge of conserved apicomplexan virulence factors and their sequence features, with suitable bioinformatics tools to facilitate their identification, Calarco *et al.* (2019) curated a list of proteins described as "hypothetical" or "unnamed" from reference resources. A total of 4,264 uncharacterised proteins were then subjected to a range of tools for classification by topology, sequence homology, adhesin-like properties, GOs, and protein domains. From this initial list, 125 uncharacterised *N. caninum* proteins were predicted

to contain TM domains and a SP sequence. After these protein sequences were submitted to a malarial adhesins predictor, a final set of 32 proteins were classified as adhesin-like TM proteins with a SP sequence. Following functional annotation of prioritised proteins, enriched GOs included cell adhesion, proteolysis, antigen binding, protein serine/threonine phosphatase complex, locomotion, and ATP binding. Various hypothetical proteins annotated through this workflow contained structural features known to be implicated in important parasite mechanisms, such as those described as MICs, GRAs, proteases, peptidases, and surface antigens.

The development of a bioinformatics workflow in this study, resulted in the prioritisation of all predicted, uncharacterised proteins within the *N. caninum* proteome, based on sequence structure and features of known, conserved apicomplexan virulence factors. This resulted in the identification of a novel and focused subset of proteins, that are potentially involved in biological processes pertaining to parasite motility, adhesion, invasion, signalling, and interaction with host cells. Consequently, these proteins could be targeted in new avenues of investigation in the near future, for the development of treatment and control options against neosporosis. This includes comparing these proteins between *N. caninum* isolates with reported phenotypic differences, at both a molecular and biological level. This includes nucleotide and amino acid sequence diversity through variant detection, gene expression studies, gene knockout experiments, and immunoassays.

6. Summary

Extensive efforts have recently been made to reveal the key contributors to and major sources of, genetic diversity characterising isolates of the pathogenic protozoa, *N. caninum*. All avenues investigated thus far have provided evidence that *N. caninum* exists as a diverse, globally distributed population of isolates, that exhibit genetic and biological heterogeneity.

Initial efforts dedicated to elucidating the genetic diversity between *N. caninum* isolates, consisted of analysing repetitive elements and popular phylogenetic molecular markers. However, these studies generally focused on non-coding genomic regions and hence failed to explain the relationship of such markers to important parasite phenotypes. As a result, recent new efforts have been dedicated to the identification of sequence polymorphisms within protein-coding genes, non-nuclear DNA, and introns, to reveal their contribution to and impact on the biological diversity reported between isolates.

Bioinformatics and *in-silico* analysis of sequencing data both original and publicly accessible, has the potential to reveal and contribute to our understanding of apicomplexan virulence factors and possible vaccine and drug targets. It is clear however, that the development of treatment and control options against neosporosis can only progress as fast as the improvement of gene and protein descriptions. Current analyses of the *N. caninum* genome, proteome, and transcriptome are only as valuable as the annotations available for such data, and hence efforts in the near future should be dedicated and prioritised as such.

Acknowledgements

This study was completed by LC in partial fulfillment of the Ph.D. degree at the University of Technology Sydney.

Declaration of interest

None

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Adomako-Ankomah, Y., Wier, G.M., Borges, A.L., Wand, H.E., Boyle, J.P., 2014. Differential locus expansion distinguishes Toxoplasmatinae species and closely related strains of Toxoplasma gondii. MBio 5, e01003-01013.
Agarwal, S., Kern, S., Halbert, J., Przyborski, J.M., Baumeister, S., Dandekar, T., Doerig, C., Pradel, G., 2011. Two nucleus-localized CDK-like kinases with crucial roles for malaria parasite erythrocytic replication are involved in phosphorylation of splicing factor. J Cell Biochem 112, 1295-1310.
Ahn, H.J., Kim, S., Nam, H.W., 2005. Host cell binding of GRA10, a novel, constitutively secreted dense granular protein from Toxoplasma gondii. Biochem Biophys Res Commun 331, 614-620.
Ajzenberg, D., Banuls, A.L., Su, C., Dumetre, A., Demar, M., Carme, B., Darde, M.L., 2004. Genetic diversity, clonality and sexuality in Toxoplasma gondii. International journal for parasitology 34, 1185-1196.
Ajzenberg, D., Collinet, F., Mercier, A., Vignoles, P., Darde, M.L., 2010. Genotyping of Toxoplasma gondii isolates with 15 microsatellite markers in a single multiplex PCR assay. J Clin Microbiol 48, 4641-4645.
Al-Qassab, S., Reichel, M.P., Ellis, J., 2010a. A second generation multiplex PCR for typing strains of Neospora caninum using six DNA targets. Mol Cell Probes 24, 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 23, 132-139.
Al-Qassab, S.E., Reichel, M.P., Ellis, J., 2010b. On the biological and genetic diversity in Neospora caninum. diversity 2, 27.
Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C.M., Bout, D.M., Menard, S., Rogers, W.O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.C., Fairhurst, R.M., Benoit-Vical, F., Mercereau-Puijalon, O., Menard, D., 2014. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature 505, 50-55.
Atkinson, R., Harper, P.A., Ryce, C., Morrison, D.A., Ellis, J.T., 1999. Comparison of the biological characteristics of two isolates of Neospora caninum. Parasitology 118 (Pt 4), 363-370.
Barber, J.S., Holmdahl, O.J., Owen, M.R., Guy, F., Uggla, A., Trees, A.J., 1995. Characterization of the first European isolate of Neospora caninum (Dubey, Carpenter, Speer, Topper and Uggla). Parasitology 111 (Pt 5), 563-568.
Barr, B.C., Anderson, M.L., Dubey, J.P., Conrad, P.A., 1991. Neospora-like protozoal infections associated with bovine abortions. Veterinary pathology 28, 110-116.
Basso, W., Schares, S., Barwald, A., Herrmann, D.C., Conraths, F.J., Pantchev, N., Vrhovec, M.G., Schares, G., 2009. Molecular comparison of Neospora caninum oocyst isolates from naturally infected dogs with cell culture-derived tachyzoites of the same isolates using nested polymerase chain reaction to amplify microsatellite markers. Veterinary parasitology 160, 43-50.
Basso, W., Schares, S., Minke, L., Barwald, A., Maksimov, A., Peters, M., Schulze, C., Muller, M., Conraths, F.J., Schares, G., 2010. Microsatellite typing and avidity analysis suggest a common source of infection in herds with epidemic Neospora caninum-associated bovine abortion. Veterinary parasitology 173, 24-31.
Beghetto, E., Nielsen, H.V., Del Porto, P., Buffolano, W., Guglietta, S., Felici, F., Petersen, E., Gargano, N., 2005. A combination of antigenic regions of Toxoplasma gondii microneme

proteins induces protective immunity against oral infection with parasite cysts. J Infect Dis 191, 637-645.
Behnke, M.S., Fentress, S.J., Mashayekhi, M., Li, L.X., Taylor, G.A., Sibley, L.D., 2012. The polymorphic pseudokinase ROP5 controls virulence in Toxoplasma gondii by regulating the active kinase ROP18. PLoS pathogens 8, e1002992.
Blackman, M.J., Bannister, L.H., 2001. Apical organelles of Apicomplexa: biology and isolation by subcellular fractionation. Molecular and biochemical parasitology 117, 11-25.
Blank, M.L., Boyle, J.P., 2018. Effector variation at tandem gene arrays in tissue-dwelling coccidia: who needs antigenic variation anyway? Curr Opin Microbiol 46, 86-92.
Boyle, J.P., Rajasekar, B., Saeij, J.P.J., Ajioka, J.W., Berriman, M., Paulsen, I., Roos, D.S., Sibley, L.D., White, M.W., Boothroyd, J.C., 2006. Just one cross appears capable of dramatically altering the population biology of a eukaryotic pathogen like Toxoplasma gondii. Proc Natl Acad Sci U S A 103, 10514-10519.
Brayton, K.A., Lau, A.O., Herndon, D.R., Hannick, L., Kappmeyer, L.S., Berens, S.J., Bidwell, S.L., Brown, W.C., Crabtree, J., Fadrosh, D., Feldblum, T., Forberger, H.A., Haas, B.J., Howell, J.M., Khouri, H., Koo, H., Mann, D.J., Norimine, J., Paulsen, I.T., Radune, D., Ren, Q., Smith, R.K., Jr., Suarez, C.E., White, O., Wortman, J.R., Knowles, D.P., Jr., McElwain, T.F., Nene, V.M., 2007. Genome sequence of Babesia bovis and comparative analysis of apicomplexan hemoprotozoa. PLoS Pathog 3, 1401-1413.
Cabrera, A., Fresia, P., Berna, L., Silveira, C., Macias-Rioseco, M., Arevalo, A.P., Crispo, M., Pritsch, O., Riet-Correa, F., Giannitti, F., Francia, M.E., Robello, C., 2019. Isolation and molecular characterization of four novel Neospora caninum strains. Parasitol Res 118, 3535-3542.
Cai, X., Fuller, A.L., McDougald, L.R., Zhu, G., 2003. Apicoplast genome of the coccidian Eimeria tenella. Gene 321, 39-46.
Calarco, L., Barratt, J., Ellis, J., 2018. Genome Wide Identification of Mutational Hotspots in the Apicomplexan Parasite Neospora caninum and the Implications for Virulence. Genome Biol Evol 10, 2417-2431.
Calarco, L., Ellis, J., 2019. Annotating the 'hypothetical' in hypothetical proteins: In-silico analysis of uncharacterised proteins for the Apicomplexan parasite, Neospora caninum. Veterinary parasitology 265, 29-37.
Calarco, L., Ellis, J., 2020. Contribution of introns to the species diversity associated with the apicomplexan parasite, *Neospora caninum*. Parasitol Res, 1-15.
Caspe, S.G., Moore, D.P., Leunda, M.R., Cano, D.B., Lischinsky, L., Regidor-Cerrillo, J., Alvarez-Garcia, G., Echaide, I.G., Bacigalupe, D., Ortega Mora, L.M., Odeon, A.C., Campero, C.M., 2012. The Neospora caninum-Spain 7 isolate induces placental damage, fetal death and abortion in cattle when inoculated in early gestation. Veterinary parasitology 189, 171-181.
Cesbron-Delauw, M.F., 1994. Dense-granule organelles of Toxoplasma gondii: their role in the host-parasite relationship. Parasitol Today 10, 293-296.
Chen, Z., Harb, O.S., Roos, D.S., 2008. In silico identification of specialized secretory-organelle proteins in apicomplexan parasites and in vivo validation in Toxoplasma gondii. PLoS One 3, e3611.
Cooper, D.N., 2010. Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. Hum Genomics 4, 284-288.
Darde, M.L., Bouteille, B., Pestre-Alexandre, M., 1988. Isoenzymic characterization of seven strains of Toxoplasma gondii by isoelectrofocusing in polyacrylamide gels. Am J Trop Med Hyg 39, 551-558.

Dautu, G., Munyaka, B., Carmen, G., Zhang, G., Omata, Y., Xuenan, X., Igarashi, M., 2007. Toxoplasma gondii: DNA vaccination with genes encoding antigens MIC2, M2AP, AMA1 and BAG1 and evaluation of their immunogenic potential. Exp Parasitol 116, 273-282.
Davison, H.C., Guy, F., Trees, A.J., Ryce, C., Ellis, J.T., Otter, A., Jeffrey, M., Simpson, V.R., Holt, J.J., 1999. In vitro isolation of Neospora caninum from a stillborn calf in the UK. Res Vet Sci 67, 103-105.
Delbac, F., Sanger, A., Neuhaus, E.M., Stratmann, R., Ajioka, J.W., Toursel, C., Herm-Gotz, A., Tomavo, S., Soldati, T., Soldati, D., 2001. Toxoplasma gondii myosins B/C: one gene, two tails, two localizations, and a role in parasite division. J Cell Biol 155, 613-623.
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. International journal for parasitology 31, 747-752.
Dubey, J.P., Buxton, D., Wouda, W., 2006. Pathogenesis of bovine neosporosis. J Comp Pathol 134, 267-289.
Dubey, J.P., Carpenter, J.L., Speer, C.A., Topper, M.J., Uggla, A., 1988a. Newly recognized fatal protozoan disease of dogs. Journal of the American Veterinary Medical Association 192, 1269-1285.
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. Journal of the American Veterinary Medical Association 193, 1259-1263.
Dubey, J.P., Liddell, S., Mattson, D., Speert, C.A., Howe, D.K., Jenkins, M.C., 2001. Characterization of the Oregon isolate of Neospora hughesi from a horse. J Parasitol 87, 345-353.
Dubey, J.P., Lindsay, D.S., 1996. A review of Neospora caninum and neosporosis. Veterinary parasitology 67, 1-59.
Dubey, J.P., Lindsay, D.S., Anderson, M.L., Davis, S.W., Shen, S.K., 1992. Induced transplacental transmission of Neospora caninum in cattle. Journal of the American Veterinary Medical Association 201, 709-713.
Dubey, J.P., Schares, G., 2006. Diagnosis of bovine neosporosis. Veterinary parasitology 140, 1-34.
Dubey, J.P., Schares, G., Ortega-Mora, L.M., 2007. Epidemiology and control of neosporosis and Neospora caninum. Clin Microbiol Rev 20, 323-367.
Duret, L., 2001. Why do genes have introns? Recombination might add a new piece to the puzzle. Trends Genet 17, 172-175.
Fedorova, L., Fedorov, A., 2003. Introns in gene evolution. Genetica 118, 123-131.
Gabriel, H.B., de Azevedo, M.F., Palmisano, G., Wunderlich, G., Kimura, E.A., Katzin, A.M., Alves, J.M., 2015. Single-target high-throughput transcription analyses reveal high levels of alternative splicing present in the FPPS/GGPPS from Plasmodium falciparum. Sci Rep 5, 18429.
Gardner, M.J., Bishop, R., Shah, T., de Villiers, E.P., Carlton, J.M., Hall, N., Ren, Q., Paulsen, I.T., Pain, A., Berriman, M., Wilson, R.J., Sato, S., Ralph, S.A., Mann, D.J., Xiong, Z., Shallom, S.J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoaibi, A., Domingo, A.R., Wasawo, D., Crabtree, J., Wortman, J.R., Haas, B., Angiuoli, S.V., Creasy, T.H., Lu, C., Suh, B., Silva, J.C., Utterback, T.R., Feldblyum, T.V., Pertea, M., Allen, J., Nierman, W.C., Taracha, E.L., Salzberg, S.L., White, O.R., Fitzhugh, H.A., Morzaria, S., Venter, J.C., Fraser, C.M., Nene, V., 2005. Genome sequence of Theileria parva, a bovine pathogen that transforms lymphocytes. Science 309, 134-137.

Gardner, M.J., Williamson, D.H., Wilson, R.J., 1991. A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. Mol Biochem Parasitol 44, 115-123.
Gilbert, W., 1985. Genes-in-pieces revisited. Science 228, 823-824.
Gondim, L.F., Laski, P., Gao, L., McAllister, M.M., 2004. Variation of the internal transcribed spacer 1 sequence within individual strains and among different strains of Neospora caninum. J Parasitol 90, 119-122.
Gonzalez-Warleta, M., Castro-Hermida, J.A., Calvo, C., Perez, V., Gutierrez-Exposito, D., Regidor-Cerrillo, J., Ortega-Mora, L.M., Mezo, M., 2018. Endogenous transplacental transmission of Neospora caninum during successive pregnancies across three generations of naturally infected sheep. Vet Res 49, 106.
Goodswen, S.J., Kennedy, P.J., Ellis, J.T., 2013. A review of the infection, genetics, and evolution of Neospora caninum: from the past to the present. Infect Genet Evol 13, 133-150.
Grigg, M.E., Bonnefoy, S., Hehl, A.B., Suzuki, Y., Boothroyd, J.C., 2001. Success and virulence in Toxoplasma as the result of sexual recombination between two distinct ancestries. Science 294, 161-165.
Guo, Z.G., Johnson, A.M., 1995. Genetic comparison of Neospora caninum with Toxoplasma and Sarcocystis by random amplified polymorphic DNA-polymerase chain reaction. Parasitol Res 81, 365-370.
Guy, C.S., Ellis, J., Williams, D.J.L., Smith, A.J., Trees, A.J., 2005. Vaccination against Neospora-associated abortion in cattle, Proceedings of the 20th International Conference of the World Association for the Advancement of Veteriniary Parasitology, Christchurch.
Holmdahl, J., Bjorkman, C., Stenlund, S., Uggla, A., Dubey, J.P., 1997. Bovine Neospora and Neospora caninum: One and the same. Parasitol Today 13, 40-41.
Howe, C.J., 1992. Plastid origin of an extrachromosomal DNA molecule from Plasmodium, the causative agent of malaria. J Theor Biol 158, 199-205.
Howe, D.K., Sibley, L.D., 1995. Toxoplasma gondii comprises three clonal lineages: correlation of parasite genotype with human disease. J Infect Dis 172, 1561-1566.
Hutchison, C.A., 3rd, Newbold, J.E., Potter, S.S., Edgell, M.H., 1974. Maternal inheritance of mammalian mitochondrial DNA. Nature 251, 536-538.
Huynh, M.H., Boulanger, M.J., Carruthers, V.B., 2014. A conserved apicomplexan microneme protein contributes to Toxoplasma gondii invasion and virulence. Infection and immunity 82, 4358-4368.
Huynh, M.H., Carruthers, V.B., 2006. Toxoplasma MIC2 is a major determinant of invasion and virulence. PLoS Pathog 2, e84.
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 31, 1523-1534.
Iriko, H., Jin, L., Kaneko, O., Takeo, S., Han, E.T., Tachibana, M., Otsuki, H., Torii, M., Tsuboi, T., 2009. A small-scale systematic analysis of alternative splicing in Plasmodium falciparum. Parasitol Int 58, 196-199.
Irimia, M., Roy, S.W., 2008. Spliceosomal introns as tools for genomic and evolutionary analysis. Nucleic Acids Res 36, 1703-1712.
Jewett, T.J., Sibley, L.D., 2003. Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. Mol Cell 11, 885-894.
Jimenez-Pelayo, L., Garcia-Sanchez, M., Vazquez, P., Regidor-Cerrillo, J., Horcajo, P., Collantes-Fernandez, E., Blanco-Murcia, J., Gutierrez-Exposito, D., Roman-Trufero, A., Osoro, K., Benavides, J., Ortega-Mora, L.M., 2019. Early Neospora caninum infection dynamics in cattle after inoculation at mid-gestation with high (Nc-Spain7)- or low (Nc-Spain1H)-virulence isolates. Vet Res 50, 72.

Jung, C., Lee, C.Y., Grigg, M.E., 2004. The SRS superfamily of Toxoplasma surface proteins. International journal for parasitology 34, 285-296.
Kern, S., Agarwal, S., Huber, K., Gehring, A.P., Strodke, B., Wirth, C.C., Brugl, T., Abodo, L.O., Dandekar, T., Doerig, C., Fischer, R., Tobin, A.B., Alam, M.M., Bracher, F., Pradel, G., 2014. Inhibition of the SR protein-phosphorylating CLK kinases of Plasmodium falciparum impairs blood stage replication and malaria transmission. PLoS One 9, e105732.
Khan, A., Dubey, J.P., Su, C., Ajioka, J.W., Rosenthal, B.M., Sibley, L.D., 2011a. Genetic analyses of atypical Toxoplasma gondii strains reveal a fourth clonal lineage in North America. International journal for parasitology 41, 645-655.
Khan, A., Fujita, A.W., Randle, N., Regidor-Cerrillo, J., Shaik, J.S., Shen, K., Oler, A.J., Quinones, M., Latham, S.M., Akanmori, B.D., Cleaveland, S., Innes, E.A., Ryan, U., Slapeta, J., Schares, G., Ortega-Mora, L.M., Dubey, J.P., Wastling, J.M., Grigg, M.E., 2019. Global selective sweep of a highly inbred genome of the cattle parasite Neospora caninum. Proc Natl Acad Sci U S A 116, 22764-22773.
Khan, A., Miller, N., Roos, D.S., Dubey, J.P., Ajzenberg, D., Darde, M.L., Ajioka, J.W., Rosenthal, B., Sibley, L.D., 2011b. A monomorphic haplotype of chromosome Ia is associated with widespread success in clonal and nonclonal populations of Toxoplasma gondii. MBio 2, e00228-00211.
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 5, e1000404.
Khan, A., Taylor, S., Su, C., Mackey, A.J., Boyle, J., Cole, R., Glover, D., Tang, K., Paulsen, I.T., Berriman, M., Boothroyd, J.C., Pfefferkorn, E.R., Dubey, J.P., Ajioka, J.W., Roos, D.S., Wootton, J.C., Sibley, L.D., 2005. Composite genome map and recombination parameters derived from three archetypal lineages of Toxoplasma gondii. Nucleic Acids Res 33, 2980-2992.
Kim, K., Weiss, L.M., 2004. Toxoplasma gondii: the model apicomplexan. International journal for parasitology 34, 423-432.
Liddell, S., Jenkins, M.C., Collica, C.M., Dubey, J.P., 1999. Prevention of vertical transfer of Neospora caninum in BALB/c mice by vaccination. The Journal of parasitology 85, 1072-1075.
Lindsay, D.S., Dubey, J.P., 1989. In vitro development of Neospora caninum (Protozoa: Apicomplexa) from dogs. The Journal of parasitology 75, 163-165.
Lorenzi, H., Khan, A., Behnke, M.S., Namasivayam, S., Swapna, L.S., Hadjithomas, M., Karamycheva, S., Pinney, D., Brunk, B.P., Ajioka, J.W., Ajzenberg, D., Boothroyd, J.C., Boyle, J.P., Darde, M.L., Diaz-Miranda, M.A., Dubey, J.P., Fritz, H.M., Gennari, S.M., Gregory, B.D., Kim, K., Saeij, J.P., Su, C., White, M.W., Zhu, X.Q., Howe, D.K., Rosenthal, B.M., Grigg, M.E., Parkinson, J., Liu, L., Kissinger, J.C., Roos, D.S., Sibley, L.D., 2016. Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. Nat Commun 7, 10147.
Lovett, J.L., Howe, D.K., Sibley, L.D., 2000. Molecular characterization of a thrombospondin-related anonymous protein homologue in Neospora caninum. Mol Biochem Parasitol 107, 33-43.
Lynch, M., Richardson, A.O., 2002. The evolution of spliceosomal introns. Curr Opin Genet Dev 12, 701-710.
Ma, L., Liu, G., Liu, J., Li, M., Zhang, H., Tang, D., Liu, Q., 2017a. Neospora caninum ROP16 play an important role in the pathogenicity by phosphorylating host cell STAT3. Veterinary parasitology 243, 135-147.
Ma, L., Liu, J., Li, M., Fu, Y., Zhang, X., Liu, Q., 2017b. Rhoptry protein 5 (ROP5) Is a Key Virulence Factor in Neospora caninum. Front Microbiol 8, 370.

Maley, S.W., Buxton, D., Rae, A.G., Wright, S.E., Schock, A., Bartley, P.M., Esteban-Redondo, I., Swales, C., Hamilton, C.M., Sales, J., Innes, E.A., 2003. The pathogenesis of neosporosis in pregnant cattle: inoculation at mid-gestation. J Comp Pathol 129, 186-195.
Manger, I.D., Hehl, A.B., Boothroyd, J.C., 1998. The surface of Toxoplasma tachyzoites is dominated by a family of glycosylphosphatidylinositol-anchored antigens related to SAG1. Infect Immun 66, 2237-2244.
Marsh, A.E., Barr, B.C., Sverlow, K., Ho, M., Dubey, J.P., Conrad, P.A., 1995. Sequence analysis and comparison of ribosomal DNA from bovine Neospora to similar coccidial parasites. J Parasitol 81, 530-535.
Marsh, A.E., Howe, D.K., Wang, G., Barr, B.C., Cannon, N., Conrad, P.A., 1999. Differentiation of Neospora hughesi from Neospora caninum based on their immunodominant surface antigen, SAG1 and SRS2. International journal for parasitology 29, 1575-1582.
McInnes, L.M., Irwin, P., Palmer, D.G., Ryan, U.M., 2006. In vitro isolation and characterisation of the first canine Neospora caninum isolate in Australia. Veterinary parasitology 137, 355-363.
Mercier, C., Adjogble, K.D., Daubener, W., Delauw, M.F., 2005. Dense granules: are they key organelles to help understand the parasitophorous vacuole of all apicomplexa parasites? International journal for parasitology 35, 829-849.
Mercier, C., Dubremetz, J.F., Rauscher, B., Lecordier, L., Sibley, L.D., Cesbron-Delauw, M.F., 2002. Biogenesis of nanotubular network in Toxoplasma parasitophorous vacuole induced by parasite proteins. Mol Biol Cell 13, 2397-2409.
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 39, 299-306.
Miller, C., Quinn, H., Ryce, C., Reichel, M.P., Ellis, J.T., 2005. Reduction in transplacental transmission of Neospora caninum in outbred mice by vaccination. International journal for parasitology 35, 821-828.
Miller, C.M., Quinn, H.E., Windsor, P.A., Ellis, J.T., 2002. Characterisation of the first Australian isolate of Neospora caninum from cattle. Australian veterinary journal 80, 620-625.
Nishikawa, Y., Shimoda, N., Fereig, R.M., Moritaka, T., Umeda, K., Nishimura, M., Ihara, F., Kobayashi, K., Himori, Y., Suzuki, Y., Furuoka, H., 2018. Neospora caninum Dense Granule Protein 7 Regulates the Pathogenesis of Neosporosis by Modulating Host Immune Response. Appl Environ Microbiol 84.
Oladele, T.O., Sadiku, J.S., Bewaji, C.O., 2011. *In silico* characterisation of some hypothetical proteins in the proteome of *Plasmodium falciparum*. Centrepoint J. 17, 129-139.
Pedraza-Diaz, S., Marugan-Hernandez, V., Collantes-Fernandez, E., Regidor-Cerrillo, J., Rojo-Montejo, S., Gomez-Bautista, M., Ortega-Mora, L.M., 2009. Microsatellite markers for the molecular characterization of Neospora caninum: application to clinical samples. Veterinary parasitology 166, 38-46.
Preston, M.D., Campino, S., Assefa, S.A., Echeverry, D.F., Ocholla, H., Amambua-Ngwa, A., Stewart, L.B., Conway, D.J., Borrmann, S., Michon, P., Zongo, I., Ouedraogo, J.B., Djimde, A.A., Doumbo, O.K., Nosten, F., Pain, A., Bousema, T., Drakeley, C.J., Fairhurst, R.M., Sutherland, C.J., Roper, C., Clark, T.G., 2014. A barcode of organellar genome polymorphisms identifies the geographic origin of Plasmodium falciparum strains. Nat Commun 5, 4052.
Regidor-Cerrillo, J., Diez-Fuertes, F., Garcia-Culebras, A., Moore, D.P., Gonzalez-Warleta, M., Cuevas, C., Schares, G., Katzer, F., Pedraza-Diaz, S., Mezo, M., Ortega-Mora, L.M.,

2013. Genetic diversity and geographic population structure of bovine Neospora caninum determined by microsatellite genotyping analysis. PLoS One 8, e72678.
Regidor-Cerrillo, J., Pedraza-Diaz, S., Gomez-Bautista, M., Ortega-Mora, L.M., 2006. Multilocus microsatellite analysis reveals extensive genetic diversity in Neospora caninum. J Parasitol 92, 517-524.
Reichel, M.P., Alejandra Ayanegui-Alcerreca, M., Gondim, L.F., Ellis, J.T., 2013. What is the global economic impact of Neospora caninum in cattle - the billion dollar question. International journal for parasitology 43, 133-142.
Reichel, M.P., Ellis, J., 2006. If control of Neospora caninum infection is technically feasible does it make economic sense? Veterinary parasitology 142, 23-34.
Reichel, M.P., Ellis, J.T., 2002. Control options for Neospora caninum infections in cattle--current state of knowledge. N Z Vet J 50, 86-92.
Reichel, M.P., Ellis, J.T., Dubey, J.P., 2007. Neosporosis and hammondiosis in dogs. J Small Anim Pract 48, 308-312.
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 8, e1002567.
Reynolds, S.M., Kall, L., Riffle, M.E., Bilmes, J.A., Noble, W.S., 2008. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. PLoS Comput Biol 4, e1000213.
Ribeiro, C.M., Soares, I.R., Mendes, R.G., de Santis Bastos, P.A., Katagiri, S., Zavilenski, R.B., de Abreu, H.F.P., Afreixo, V., 2019. Meta-analysis of the prevalence and risk factors associated with bovine neosporosis. Trop Anim Health Prod 51, 1783-1800.
Robbe, D., Passarelli, A., Gloria, A., Di Cesare, A., Capelli, G., Iorio, R., Traversa, D., 2016. Neospora caninum seropositivity and reproductive risk factors in dogs. Exp Parasitol 164, 31-35.
Rojo-Montejo, S., Collantes-Fernandez, E., Blanco-Murcia, J., Rodriguez-Bertos, A., Risco-Castillo, V., Ortega-Mora, L.M., 2009a. Experimental infection with a low virulence isolate of Neospora caninum at 70 days gestation in cattle did not result in foetopathy. Vet Res 40, 49.
Rojo-Montejo, S., Collantes-Fernandez, E., Regidor-Cerrillo, J., Alvarez-Garcia, G., Marugan-Hernandez, V., Pedraza-Diaz, S., Blanco-Murcia, J., Prenafeta, A., Ortega-Mora, L.M., 2009b. Isolation and characterization of a bovine isolate of Neospora caninum with low virulence. Veterinary parasitology 159, 7-16.
Rome, M.E., Beck, J.R., Turetzky, J.M., Webster, P., Bradley, P.J., 2008. Intervacuolar transport and unique topology of GRA14, a novel dense granule protein in Toxoplasma gondii. Infect Immun 76, 4865-4875.
Romero, J.J., Perez, E., Frankena, K., 2004. Effect of a killed whole Neospora caninum tachyzoite vaccine on the crude abortion rate of Costa Rican dairy cows under field conditions. Veterinary parasitology 123, 149-159.
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 314, 1780-1783.
Schock, A., Innes, E.A., Yamane, I., Latham, S.M., Wastling, J.M., 2001. Genetic and biological diversity among isolates of Neospora caninum. Parasitology 123, 13-23.
Sibley, L.D., Boothroyd, J.C., 1992. Virulent strains of Toxoplasma gondii comprise a single clonal lineage. Nature 359, 82-85.

Slapeta, J.R., Modry, D., Kyselova, I., Horejs, R., Lukes, J., Koudela, B., 2002. Dog shedding oocysts of Neospora caninum: PCR diagnosis and molecular phylogenetic approach. Veterinary parasitology 109, 157-167.
Spencer, J.A., Witherow, A.K., Blagburn, B.L., 2000. A random amplified polymorphic DNA polymerase chain reaction technique that differentiates between Neospora species. J Parasitol 86, 1366-1368.
Sreekumar, C., Hill, D.E., Fournet, V.M., Rosenthal, B.M., Lindsay, D.S., Dubey, J.P., 2003. Detection of Hammondia heydorni-like organisms and their differentiation from Neospora caninum using random-amplified polymorphic DNA-polymerase chain reaction. J Parasitol 89, 1082-1085.
Stenlund, S., Bjorkman, C., Holmdahl, O.J., Kindahl, H., Uggla, A., 1997. Characterization of a Swedish bovine isolate of Neospora caninum. Parasitol Res 83, 214-219.
Syn, G., Blackwell, J.M., Jamieson, S.E., Francis, R.W., 2018. An in silico pipeline to filter the Toxoplasma gondii proteome for proteins that could traffic to the host cell nucleus and influence host cell epigenetic regulation. Mem Inst Oswaldo Cruz 113, e170471.
Talevich, E., Kannan, N., 2013. Structural and evolutionary adaptation of rhoptry kinases and pseudokinases, a family of coccidian virulence factors. BMC evolutionary biology 13, 117.
Trees, A.J., Williams, D.J., 2005. Endogenous and exogenous transplacental infection in Neospora caninum and Toxoplasma gondii. Trends Parasitol 21, 558-561.
Walsh, C.P., Vemulapalli, R., Sriranganathan, N., Zajac, A.M., Jenkins, M.C., Lindsay, D.S., 2001. Molecular comparison of the dense granule proteins GRA6 and GRA7 of Neospora hughesi and Neospora caninum. International journal for parasitology 31, 253-258.
Walzer, K.A., Adomako-Ankomah, Y., Dam, R.A., Herrmann, D.C., Schares, G., Dubey, J.P., Boyle, J.P., 2013. Hammondia hammondi, an avirulent relative of Toxoplasma gondii, has functional orthologs of known T. gondii virulence genes. Proc Natl Acad Sci U S A 110, 7446-7451.
Wang, J., Tang, D., Li, W., Xu, J., Liu, Q., Liu, J., 2017. A new microneme protein of Neospora caninum, NcMIC8 is involved in host cell invasion. Exp Parasitol 175, 21-27.
Wasmuth, J., Daub, J., Peregrin-Alvarez, J.M., Finney, C.A., Parkinson, J., 2009. The origins of apicomplexan sequence innovation. Genome Res 19, 1202-1213.
Wasmuth, J.D., Pszenny, V., Haile, S., Jansen, E.M., Gast, A.T., Sher, A., Boyle, J.P., Boulanger, M.J., Parkinson, J., Grigg, M.E., 2012. Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of Toxoplasma virulence. MBio 3.
Weber, F.H., Jackson, J.A., Sobecki, B., Choromanski, L., Olsen, M., Meinert, T., Frank, R., Reichel, M.P., Ellis, J.T., 2013. On the efficacy and safety of vaccination with live tachyzoites of Neospora caninum for prevention of neospora-associated fetal loss in cattle. Clin Vaccine Immunol 20, 99-105.
Williams, D.J., Guy, C.S., Smith, R.F., Ellis, J., Bjorkman, C., Reichel, M.P., Trees, A.J., 2007. Immunization of cattle with live tachyzoites of Neospora caninum confers protection against fetal death. Infect Immun 75, 1343-1348.
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 136, 1895-1900.
Williamson, D.H., Gardner, M.J., Preiser, P., Moore, D.J., Rangachari, K., Wilson, R.J., 1994. The evolutionary origin of the 35 kb circular DNA of Plasmodium falciparum: new evidence supports a possible rhodophyte ancestry. Mol Gen Genet 243, 249-252.
Wilson, R.J., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W., Williamson, D.H., 1996. Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. J Mol Biol 261, 155-172.

Yang, C., Liu, J., Ma, L., Zhang, X., Zhang, X., Zhou, B., Zhu, X., Liu, Q., 2018. NcGRA17 is an important regulator of parasitophorous vacuole morphology and pathogenicity of Neospora caninum. Veterinary parasitology 264, 26-34.
Yeoh, L.M., Goodman, C.D., Hall, N.E., van Dooren, G.G., McFadden, G.I., Ralph, S.A., 2015. A serine-arginine-rich (SR) splicing factor modulates alternative splicing of over a thousand genes in Toxoplasma gondii. Nucleic Acids Res 43, 4661-4675.
Zhang, T.T., Zhang, X., Liu, Q., Zu, J.H., Liu, J., 2019. Protective efficacy of vaccination with NcMIC3 and NcMIC8 against Neospora caninum infection in mice. Frontiers of Agricultural Science and Engineering 6, 188-196.