COMPARATIVE ANALYSIS OF GENE DUPLICATION: IMPACT OF TANDEMLY DUPLICATED GENES ON TRAIT EVOLUTION IN *TOXOPLASMA GONDII*

by

Yaw Adomako-Ankomah

B.S. Biological Sciences, Kwame Nkrumah University of Science & Technology, 2006

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yaw Adomako-Ankomah

It was defended on

December 5, 2013

and approved by

Beth Roman, Ph.D., Assistant Professor, Biological Sciences

William Saunders, Ph.D., Associate Professor, Biological Sciences

Graham Hatfull, Ph.D., Professor, Biological Sciences

Elodie Ghedin, Ph.D., Associate Professor, Computational and Systems Biology

Dissertation Advisor: Jon Boyle, Ph.D., Assistant Professor, Biological Sciences

COMPARATIVE ANALYSIS OF GENE DUPLICATION: IMPACT OF TANDEMLY DUPLICATED GENES ON TRAIT EVOLUTION IN *TOXOPLASMA GONDII*

Yaw Adomako-Ankomah, PhD

University of Pittsburgh, 2013

Gene duplication is a well-recognized mechanism by which organism s acquire novel traits and it has been shown to play an important role in species divergence. Recent genomics studies have highlighted the significant impact that gene duplication can have in the adaptive evolution of parasite species. Traits such as drug resistance, host immune evasion and disruption, virulence and host range selection have all been shown to be affected by gene duplication events. However, relatively little is known about the importance of gene duplication in the evolution of protozoan parasites. The goal of this study is to examine the scope and impact of gene duplication on the evolution of unique traits in the human parasite *Toxoplasma gondii*.

To this end, we used comparative genomics approaches to perform a comprehensive comparative analysis of gene duplication between *T. gondii* and its close relatives *Neospora caninum* and *Hammondia hammondi*. We show that despite the high degree of synteny between these species, they differ significantly in their expanded gene profiles. Interestingly, we find that these expanded loci (EL) are under strong diversifying selection and that many exhibit copy-number variations among different strains of the same species. Additionally, we have performed functional analyses on two expanded loci: we have identified *EL4* to be responsible for host mitochondrial association, which is a trait unique to *T. gondii*, and also shown that *EL3* may be involved in regulating parasite proliferation during bradyzoite differentiation. Our results suggest that gene duplication may represent an important underlying distinction that sets *T. gondii* apart from its less pathogenic relatives such as *N. caninum* and *H. hammondi*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to acknowledge the immense support and guidance I have received from my advisor, Dr. Jon Boyle. Thank you for allowing me the space and time to explore freely and nurture my passion for science. Thank you for allowing me to take full ownership of my projects, and for encouraging me to try new things, to make my own mistakes and to learn from them. After 5 years, I realize that is all science is about after all. I know this is your first Ph.D. mentoring gig as a PI but I think you do it the right way. *You will do great!*

I am extremely thankful to my dissertation committee for being so generous with their time and readily providing guidance on matters relating to my dissertation as well as career decisions. Thank you to Drs. Beth Roman, Bill Saunders, Graham Hatful and Elodie Ghedin!

I am also grateful to all past and present members of the Boyle Lab. I have enjoyed and benefited from your help, advice, encouragement and above all your friendship. Greg, Elise, Tracy, Katie, Ananth, Adair, Heather, Rachel and Hannah, Thank you! Special thanks to my boy, Edwin Kamau, who was a one-man welcome party for me and was very instrumental in setting me up when I joined the lab. Everyone in the lab misses you!

I have had an incredible experience of learning and growing as a person and as a scientist over the past few years thanks to the amazing cordiality and supportive atmosphere in the department and I am eternally grateful to everyone who has contributed in any way to this experience. To everyone I ever interacted with in the department, the main office staff, who do

such a great job taking care of all kinds of paperwork so that graduate students can just focus on being graduate students, the fiscal office staff who handled all my purchases and payments, the teaching faculty who introduced me to the kind of exploratory science that has formed the foundation of my research career, my classmates who shared and eased the burden of the past five years, my fellow grad students who welcomed me into the program and in their own attempts to make grad school bearable for themselves made it bearable for me, and anyone who I ever bumped into in the creepy stairwell connecting Langley Hall to the LSA, thank you, thank you, thank you! .

The following Labs generously shared reagents and equipment for my work: the Chapman, Hildebrand, Roman, Campbell, Saunders, Hatfull, Schwacha and Tonsor Labs. Dr. Jennifer cott-Schwartz donated NRK-mitoRFP cells for my host mitochondrial association experiments. Dr. Carolyn Coyne helps me tremendously with my host cell invasion – HMA time-lapse microscopy. Most importantly, I have benefitted from a very fruitful collaboration with the laboratory of Dr. John Boothroyd at Stanford University on my MAF1 work and I am extremely grateful to Dr. John Boothroyd and Dr. Lena Pernas for their generous exchange of ideas, materials and data.

I dedicate this work to my parents, Dickson and Veronica Donkor for being the most loving and supportive parents anyone could wish for. You never asked for any more from me than my best effort and I have always been encouraged by the knowledge that my success is your success and that you will always be happy for me.

Above all, I am most indebted to my dear wife, Dr. Ashrifia Adomako-Ankomah, who has been a source of strength and inspiration over the years. Thank you for being there to share the highs and lows of the past 5 years with me.

1.0 INTRODUCTION

1.1 *TOXOPLASMA GONDII*: LIFE CYCLE

Toxoplasma gondii is a protozoan obligate intracellular parasite belonging to the phylum Apicomplexa. This phylum also includes other major human and animal parasites such as *Plasmodium* (malaria), *Cryptosporidium* (cryptosporidiosis), *Babesia* (babesiosis) and *Eimeria* (coccidiosis) (Robben *et al.*, 2004b). Like most of its relatives, *T. gondii* has a complex life cycle made up of two phases: the sexual phase which takes place exclusively in felines and the asexual phase which takes place in its wide variety of intermediate hosts (Boothroyd, 2009). Infection of intermediate hosts usually results from either ingestion of oocysts contaminants in food or water, or tissue cysts found in undercooked meat. Upon entering the host, the infective form initially differentiates into the fast growing form called tachyzoite. This developmental stage is highly immunogenic and elicits a strong immune response from the host. Under host immune pressure, tachyzoites differentiate into slow growing forms called bradyzoites, which eventually become encysted and remains buried in the tissues of the host for many years. Tissue cysts serve as a major transmission stage to other intermediate hosts as well as the definitive hosts. Within the definitive hosts, the parasite undergoes gametogenesis and fertilization. Eventually parasites are passed out as highly infectious and environmentally stable oocysts which may initiate infects when ingested by an intermediate host (Dubey, 2009).

1.2 *TOXOPLASMA GONDII*: TOXOPLASMOSIS

T. gondii is unique among these parasites in its ability to infect a remarkably wide range of intermediate hosts including all warm blooded-animals (Sibley *et al.*, 2008). Human infections usually result from ingestion of infective forms of the parasite in undercooked meat (tissue cysts) or food contaminated with cat feces (oocysts). Oocysts in particular are extremely environmentally stable and can survive several days outside of the host and still remain infectious (Lelu *et al.*, 2012). Vertical transmission may also occur from mother to fetus during pregnancy (Jones *et al.*, 2009).

Primary *T. gondii* infection elicits a strong, usually effective, interferon-γ-driven immune response from the host and is therefore generally asymptomatic (Gazzinelli *et al.*, 1994, Miller *et al.*, 2009). However, under host immune pressure, a primary acute infection converts to a latent chronic form which may persist for the life of the host. Severe disease outcomes are associated with reactivation of chronic *T. gondii* infections under suboptimal host immune conditions. Toxoplasmic encephalitis resulting from reactivation of latent chronic infection in the brain is among the leading causes of CNS infections in immunocompromised HIV-AIDS patients (Richards *et al.*, 1995). Reactivation in the eye may also lead to toxoplasmic chorioretinitis. Recent reports of severe acute infections in immunocompetent individuals (de Moura *et al.*, 2006) have alerted researchers to the shifting pattern of *T. gondii* infections and have highlighted the urgent need for the development of curative chemotherapeutics.

1.3 *TOXOPLASMA GONDII*: POPULATION STRUCTURE

Toxoplasma has a rather simple population structure which belies its extensive global distribution. The populations of *T. gondii* in North America and in Europe are predominated by three highly clonal lineages with only about 1% or less genetic difference between them (Howe *et al.*, 1995, Su *et al.*, 2003, Sibley *et al.*, 2008). These clonal lineages, referred to as Types I, II and III, have been shown to exhibit important and interesting phenotypic differences based on the few underlying genetic differences. It has been proposed that these lineages emerged from either one or two genetic recombination event(s) involving two divergent ancestral genetic pools followed by natural selection and stabilization of the three fittest progeny genotypes (Grigg *et al.*, 2001a, Saeij *et al.*, 2005a, Boyle *et al.*, 2006).

Two main reasons have been proposed to explain the maintenance of such a highly clonal population structure for an organism that has a clearly defined sexual cycle. Firstly, *Toxoplasma* has the ability to bypass its sexual host and cycle indefinitely among intermediate hosts. Within these intermediate hosts, the parasite undergoes asexual replication and produces infective forms, which can be transmitted up and down the food chain (Boothroyd, 2009). This life history trait significantly decreases the opportunity for recombination of genetic material from genetically distinct strains through sexual replication. In fact, the evolution of oral infectivity in *T. gondii*, which is a key adaptive feature for lateral transmission between intermediate hosts, seems to have occurred concurrently with the emergence of the clonal population structure ~10,000 years ago (Su *et al.*, 2003). Secondly, during infection of the feline sexual host, *Toxoplasma* most commonly undergoes self-fertilization from gametocytes of the same parental strain to produce millions of highly clonal infective forms (Su *et al.*, 2003). The combination of a very short window for gametogenesis and the rare occurrence of multi-strain natural infections in cats favor

self-fertilization, rather than out-crossing, as a mechanism for enhancing the transmission of recently emerged clonal recombinants and thus perpetuating the clonal population structure of *T. gondii* in Europe and North America (Boothroyd, 2009, Wendte *et al.*, 2010).

1.4 GENE EXPANSION AND TRAIT EVOLUTION

The widespread existence of multi-gene families provides a clear indication that gene expansion has been centrally involved in evolution and acquisition of new traits in all life forms (Pays *et al.*, 1981, Hurles, 2004, Sudmant *et al.*, 2010, Lu *et al.*, 2012, Fares *et al.*, 2013). Several mechanisms can drive the initial gene duplication event, the most prominent of which is unequal crossing over between repetitive elements during meiosis (Hurles, 2004). After a duplication event has occurred, there is a narrow window of time where relaxed selective pressure on duplicate genes allows for adaptive changes through accumulation of non-synonymous mutations before selective pressure is reestablished as these genes begin to drift apart (Lynch *et al.*, 2000, Bergthorsson *et al.*, 2007). While this selective constraint ensures that a significantly large proportion of gene duplicates are eliminated from the population, a precious few may evolve new functions which may be evolutionarily advantageous and may become fixed in the population. The nature of the duplication event, i.e. small-scale duplications (SSD) or whole-genome duplication (WGD) is believed to influence the fate of duplicated genes. In *Saccharomyces cerevisiae*, it has been suggested that SSD events were more likely to produce duplicate genes which would go on to evolve divergent functions (neo-functionalization) while WGD events would give birth to duplicate genes with shared ancestral functions (sub-functionalization) notwithstanding the large amount of genetic raw material produced by this

type of duplication (Fares *et al.*, 2013). Regardless of the mechanism or nature of gene duplication, it is considered as a vehicle which allows organisms to explore the evolutionary space for new and possibly advantageous traits without immediately compromising original gene function.

There are several examples of evolution by gene duplication in protozoan parasites. The human parasite *Trichomonas vaginalis* has undergone perhaps the most extensive gene expansion of any protozoan parasite, with about 59 expanded gene families which average over 600 copies per gene (Carlton *et al.*, 2007). In this organism, there appears to have been selective expansion of genes which are important for its particular parasitic lifestyle. These include components of its membrane trafficking machinery which is important for secretion of pathogenic proteins into the host cell and exchange of other materials with the host environment through exo- and endocytosis (Carlton *et al.*, 2007). This provides support for the notion that while gene duplication events may be random, fixation of duplicate genes is not (Tanaka *et al.*, 2009).

The *Trypanosoma brucei* genome contains a vast pool of extensively expanded variant surface glycoprotein (*VSG*) genes, over 80% of which have been pseudogenized (Berriman *et al.*, 2005). Only a single VSG is expressed at any given time, and multiple rounds of antigen-switching during infection in the same host leads to multiple waves of parasitemia. Antigenic variation allows the parasite to survive long enough to be passed on to a vector or another host. It also allows the parasite to superinfect hosts previously infected with parasites expressing different *VSG* paralogs and have developed paralog-specific immunity (Barry *et al.*, 2005, Stockdale *et al.*, 2008). These features are critical for the success of this species as a parasite and also for its ability to navigate a complex vector-dependent life cycle. The mechanism by which

T. brucei switches between VSG expressions highlights the relevance of gene duplication and gene conversion in the evolution of virulence. A silent *VSG* gene is first duplicated and subsequently recombined into an expression site where it replaces the currently expressed *VSG* gene (Palmer *et al.*, 2007, Stockdale *et al.*, 2008, Gjini *et al.*, 2012). This new VSG then takes over as the dominant surface antigen. The starter population of progeny that switches their VSGs this way then escapes immune surveillance and seed a new cycle of infection.

Antigenic variation as a mechanism for immune evasion is also a prominent feature of *Plasmodium spp* infections and gene duplication plays a crucial role here as well. The *Plasmodium falciparum* genome is rife with expanded gene families including the ~60-member *var* gene family (Templeton, 2009). This gene family encodes surface antigens exported to the red blood cell surface called erythrocyte membrane proteins (PfEMPs) and has been shown to be under strong diversifying selection (Freitas-Junior *et al.*, 2000, Deitsch *et al.*, 2001, Pasternak *et al.*, 2009). PfEMPs have been extensively studied and characterized for their roles in infected red blood cell (iRBC) adhesion to vascular endothelial cells, iRBC rosette formation and antigenic variation (Scherf *et al.*, 2008, Pasternak *et al.*, 2009). These are all immune evasion strategies that are considered to be directly responsible for the pathogenicity of *P. falciparum* infection as well as the difficulty associated with malaria vaccine development (Pasternak *et al.*, 2009). There are a few other multi-gene families in the *Plasmodium* genome such as *stevor* and *Pfmc-2TM* that are also involved in antigenic variation although to a lesser extent than the *var* family (Scherf *et al.*, 2008). Other expanded gene families in *Plasmodium spp.* play various roles in important physiological processes such as metabolism (Bethke *et al.*, 2006), drug resistance (Price *et al.*, 2004, Witkowski *et al.*, 2010) and receptor selection for host cell invasion (Tham *et al.*, 2012).

1.5 GENE EXPANSION IN *TOXOPLASMA GONDII*

Like the protozoan parasites cited above, *T. gondii* also possesses a clutch of expanded gene families which are progressively emerging to be very important in the adaptive evolution of the parasite. By far, the most extensively expanded of these gene families to be identified yet is the surface antigen 1 (SAG1)-related sequence (SRS) superfamily of surface antigens. The *SRS* gene family includes over 160 distinct genes encoding related proteins which are collectively implicated in host cell attachment and invasion (Mineo *et al.*, 1994, Dzierszinski *et al.*, 2000, Carruthers *et al.*, 2007a). In fact the large number of *SRS* genes in *T. gondii* is believed to be associated with the expansive host range of the parasite (Jung *et al.*, 2004). Some SRSs are immunogenic while others are not. It is clear from a number of studies that the canonical member of the SRS family, SAG1 (Kasper *et al.*, 1983), now named SRS29B (Wasmuth *et al.*, 2012), is the immunodominant antigen on the surface of *T. gondii* tachyzoites (Kasper *et al.*, 1983) and is highly immunogenic in both humans (Rodriguez *et al.*, 1985) and animals (Santoro *et al.*, 1986). Recent studies have begun to shed more light on the function of SAG1 in *T. gondii* infection. Wild type parasites induce an acute and lethal form of ileitis in C57BL/6 mice while SAG1-knockout strains fail to induce such a response (Rachinel *et al.*, 2004). Interestingly, while SRS29A and SRS29B drive the balance of infection towards higher virulence in mice, SRS29C, the third member of the three-gene tandemly duplicated locus, appears to negatively regulate virulence (Wasmuth *et al.*, 2012). These findings support the notion that virulence in *Toxoplasma* is a function of various positive and negative interactions which ultimately establish an optimal equilibrium and ensures successful infection as well as transmission.

Another interesting feature of the SRS family is that it is subjected to regulation with some family members exclusively expressed in specific developmental stages of the parasite.

SAG1 and SAG2A are exclusively expressed in tachyzoites while SRS9 and bradyzoite surface antigen 4 (BSR4) are up-regulated during tachyzoite-to-bradyzoite differentiation (Van *et al.*, 2007). SporoSAG is also exclusively expressed in the sporozoite stage (Radke *et al.*, 2004). While it has been suggested that the SRS family has analogous impact on immune evasion as the *var* family in *Plasmodium spp.* and the VSG family in *Trypanosoma spp*., the mechanism involved is clearly different from the classical antigenic variation utilized by these genera (Jung *et al.*, 2004, Boothroyd, 2009). By comparison, it seems that what the SRS family of surface antigens lacks in antigenic variation, it adequately makes up for in abundance, developmental regulation, allelic polymorphism and copy number variation (Jung *et al.*, 2004, Boothroyd, 2009).

There are also examples of tandemly duplicated genes in *T. gondii* which encode proteins that directly challenge the host immune response after infection has been established. The most prominent example is the rhoptry protein 5 (*ROP5*) family of pseudokinases which has been shown by our laboratory and other groups to have a significant impact on virulence in mouse infections (Behnke *et al.*, 2011, Reese *et al.*, 2011b). The *ROP5* locus contains about 4-10 tandemly duplicated genes depending on the strain. It is not clear whether the expansion of this gene family is an adaptive response to more complex immune systems in certain mammalian hosts, but there is significant amount of evidence to suggest that the different members of the *ROP5* family are under strong diversifying selection. For example, the *ROP5* paralogs have evolved subtle variations in their contributions to the virulence phenotype associated with the locus. Intriguingly, the immunity-related GTPases (IRGs) pathway targeted for inactivation by ROP5 is present in mice but not in primates such as humans (Steinfeldt *et al.*, 2010, Howard *et al.*, 2011, Fleckenstein *et al.*, 2012, Niedelman *et al.*, 2012). Perhaps, the fact that mouse-to-cat

(intermediate-to-definitive host) transmission represents an important transition from asexual to sexual phase of the life cycle is an indication of the amount of adaptive pressure on this single locus. Additionally, ROP5 is localized to the host-parasite interface (surface of the parasitophorous vacuole) where they are amenable to the adaptive evolutionary pressure from interactions with different host environments. The specific example of ROP5 being a pseudokinase also suggests that mutations that lead to the loss of primary functions of duplicated genes may be a means of modifying and fine-tuning protein functions to suit a particular context of infection. This is particularly relevant for *T. gondii* which has such an extraordinarily wide host range.

The *ROP5* gene family belongs to a broader family of expanded genes referred to as the *ROP2* superfamily. This includes an extensive array of polymorphic genes with clear homology to protein kinases, although the majority are predicted to be pseudokinases (Peixoto *et al.*, 2010). Members of this superfamily are involved in various host parasite interactions: ROP18 localizes to the surface of the parasitophorous vacuole where it functions cooperatively with ROP5 to phosphorylate and inactivate host IRGs (Fleckenstein *et al.*, 2012, Niedelman *et al.*, 2012); ROP38 intercepts the host MAPK pathway with direct implications for host cell proliferation and apoptosis (Peixoto *et al.*, 2010). Similar to *ROP5*, *ROP38* has undergone further local/tandem duplication and has also been subjected to diversifying selection. Other members of the ROP2 superfamily have maintained their single-copy status since diversifying after the original expansion but in many cases they have developed extensive allelic variations between strains (Boothroyd, 2013). All members of this superfamily are secreted and are considered to have various degrees of direct interaction with the host cell, again underscoring the importance of host-parasite interaction in *T. gondii* genome evolution.

1.6 COMPARATIVE ANALYSIS OF GENE DUPLICATION: DIVERGENT EVOLUTION OF HOST RANGE AND VIRULENCE BETWEEN CLOSELY RELATED SPECIES

While our knowledge of the scope and impact of gene duplication and subsequent expansion in the evolution of *T. gondii* is relatively limited, the above examples clearly point to an important role for duplicated genes in shaping the evolution of the parasite. Moreover, it is becoming increasingly clear that host-parasite interaction is an important driver of adaptive evolution. A comparative analysis of the *ROP5* locus among three closely related protozoan parasites, *Toxoplasma gondii*, *Hammondia hammondi* and *N. caninum* (Figure 1.1), shows that the initial duplication of this locus perhaps occurred prior to the divergence of *N. caninum* from the *T. gondii/ H. hammondi* clade about 28 million years ago; *N. caninum* contains 2 tandem copies of *ROP5* while *T. gondii* and *H. hammondi* contain up to 10 copies (Reid *et al.*, 2012, Walzer *et al.*, 2013). However, although it is unclear whether *ROP5* orthologs in *N. caninum* are functional as virulence genes, available evidence shows that *N. caninum* lacks the ability to phosphorylate and inactivate IRGs, as has been shown to be the function of ROP5. This is consistent with findings from our work showing that *N. caninum* appears to easily succumb to innate immune pressure within the first 24 hours of mouse infection (Figure 1.2). *ROP5* paralogs from *H. hammondi*, on the other hand, have been shown to be just as effective in mediating virulence, and likely through a similar mechanism, as their *T. gondii* orthologs (Walzer *et al.*, 2013). Additionally, it is clear that the *HhROP5* locus has also been under diversifying selection. A similar analysis in *N. caninum* will reveal whether or not immune disruption by way of IRG inactivation is a recently acquired function of the *ROP5* gene family.

One of the most poorly understood questions in pathogen biology (and this is particularly true for parasitic protozoa) is how host range is determined on a molecular level. The above examples provide limited, but encouraging, support for a role for gene duplication, gene family expansion and diversification in the expansion of pathogens into new hosts. For instance, *N. caninum* contains a large repertoire of *SRS* genes, in fact, more than twice as many as can be found in *T. gondii*. However, comparative transcriptome analysis shows only a few *SRS* genes are actually expressed in *N. caninum* (Reid *et al.*, 2012). Relatively, *T. gondii* expresses a lot more surface antigens and it is attractive to think that this has direct implications for host range. Is gene duplication a cause or consequence of host range expansion? What is the impact of gene duplication on adaptive evolution? What is the impact of gene duplication on the evolution of new traits, relative to single nucleotide changes in genes? Comprehensive comparative analyses of gene duplication among protozoan parasites will give a clearer picture of which genes are more susceptible to duplication and how such gene functions impact the parasite's ability to survive different environmental conditions in different hosts. Local or tandem gene duplication is particularly informative since a tandem arrangement of paralogous genes in chromosomal region is indicative of a more recent duplication event (Friedman *et al.*, 2003), and is more likely to provide clues of adaptive evolution of closely-related species.

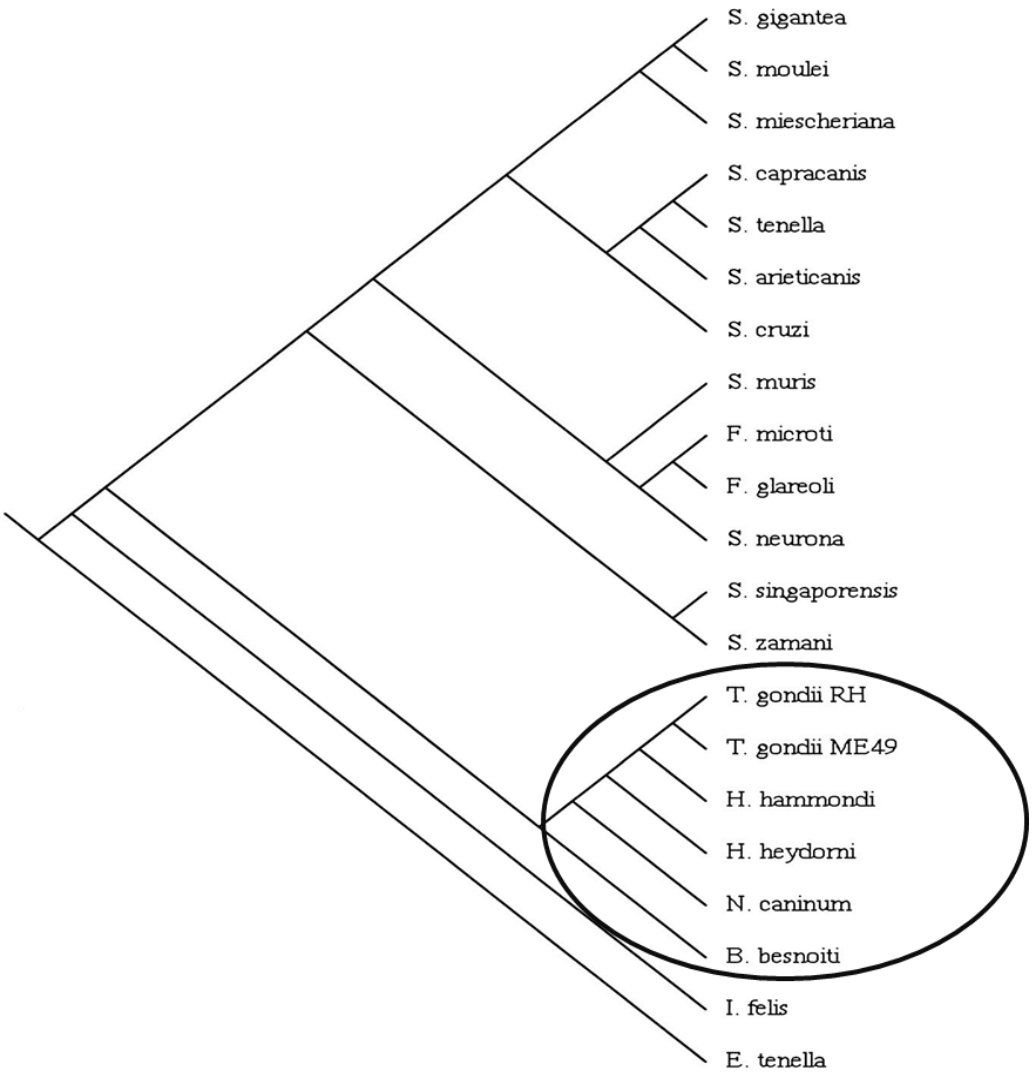
The work described in the following chapters collectively expands existing knowledge on the scope and impact of gene duplication on the evolution of *T. gondii*. The work presented in chapter 2 describes the first comprehensive comparative analyses of tandem gene duplication in the 3 closely related species, *T. gondii*, *N. caninum* and *H. hammondi*. We used read coverage analysis, which was subsequently validated by comparative genomic hybridization (CGH) and Southern blot analysis, to systematically identify and curate all tandemly duplicated gene loci in

these protozoan parasites. Our results show that gene expansion is quite pervasive in these genera, and that there is a significant lack of overlap among these closely related parasites with respect to expanded genes. We also show that the sets of expanded genes in both *N. caninum* and *T. gondii*, mostly uncharacterized in *T. gondii*, are significantly enriched for genes which encode secreted proteins. These findings provide support for the notion that host cell interactions are an important driver of gene expansion and by extension, parasite evolution.

Chapter 3 describes the characterization of the tandemly duplicated gene cluster *EL4*. *EL4* encodes a cluster of duplicated dense granule genes dubbed mitochondrial association factor 1 (*MAF1*). *MAF1* is expanded in *T. gondii* and *H. hammondi* but not in *N. caninum.* Through a collaboration with the laboratory of Dr. John Boothroyd at Stanford University, we identified MAF1 as the parasite factor which mediates the intimate association between host mitochondria and the parasitophorous vacuole which houses the parasite in the host cell. We show that different strains of *T. gondii* have varying copies of *MAF1* and that this locus, much like *ROP5*, has been under strong diversifying selection which has resulted in multiple isoforms of the gene with varying effects on host mitochondrial association. We also propose a possible mechanism by which MAF1 mediates host mitochondrial association.

In chapter 4, I describe preliminary results from the characterization of the expanded locus *EL3*. In the *T. gondii* reference genome, *EL3* occurs on chromosome Ib and is predicted to contain the rhoptry genes *ROP42, 43, 44* and *ROP42L1* which clearly resulted from tandem duplication of an ancestral *ROP42* gene. Our results indicate that this locus is also differentially expanded among *T. gondii* strains. Importantly, this gene cluster seems to have an impact on parasite proliferation and may contribute to regulating the proliferation rate of the parasite during tachyzoite-to-bradyzoite differentiation. To better study the roles of the *ROP42* cluster in

differentiation, we have adapted a tetracycline-inducible expression system to be used for temporal regulation of expression of ROP42 paralogs.



Adapted from Mugridge N B *et al*. Mol Biol Evol 2000

Figure 1.1: Phylogenetic relationships among members of the Apicomplexan family Sarcocystidae.

The subfamily Toxoplasmatinae containing the genera *Toxoplasma*, *Hammondia*, *Neospora* and *Besnoitia* has been circled.

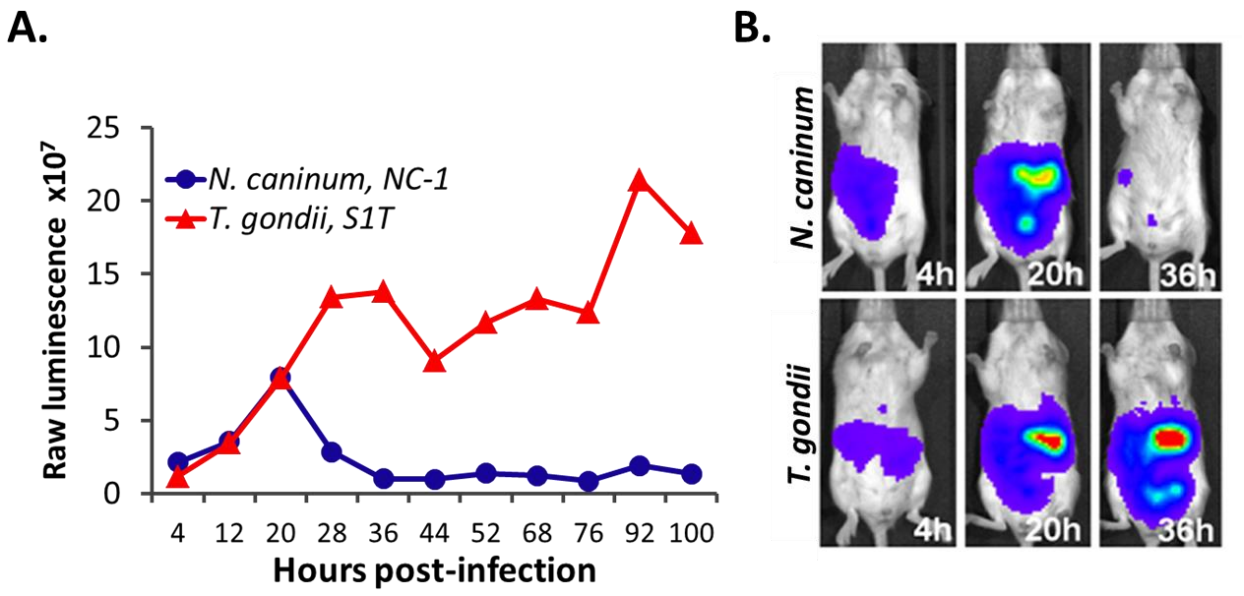


Figure 1.2: Mouse infection profile of *N. caninum* compared with an avirulent strain of *T. gondii*.

Initial increase in parasite burden indicates successful infection. However, *N. caninum* parasite burden drops after 20 hpi while *T. gondii* continues to proliferate. A) Quantitation of bioluminescence imaging (BLI) signal at 8 hour intervals post-infection. B) BLI images showing parasite burden.

2.0 DIFFERENTIAL LOCUS EXPANSION DISTINGUISHES TOXOPLASMATINAE SPECIES AND CLOSELY-RELATED STRAINS OF *TOXOPLASMA GONDII*

Toxoplasma gondii is a zoonotic obligate intracellular parasite that has infected over 30% of the human population, and has a vast intermediate host range compared to its nearest relatives *Neospora caninum* and *Hammondia hammondi*. While these 3 species have highly syntenic genomes (80-99%), in this study we examined and compared species-specific structural variation, specifically at loci that have undergone local (i.e., tandem) duplication and expansion. To do so we used genomic sequence coverage analysis to identify and curate *T. gondii* and *N. caninum* loci which have undergone duplication and expansion ("expanded loci"; EL). The 53 *T. gondii* ELs are significantly enriched for genes with predicted signal sequences and single-exons genes, and genes that are developmentally regulated at the transcriptional level. We validated 24 *T. gondii* ELs using comparative genomic hybridization and these data suggested significant copy-number variation at these loci. High-molecular weight Southern blotting for 3 *T. gondii* ELs revealed that copy-number varies across *T. gondii* lineages and also between members of the same clonal lineage. Using similar methods we identified 64 *N. caninum* ELs which were significantly enriched for genes belonging to the SAG-related surface (*SRS*) antigen family. Moreover, there is significantly less overlap (30%) between the expanded gene sets in *T. gondii* and *N. caninum* than would be predicted by overall overlap of orthologous genes (81%).

Consistent with this finding, only 59% of queried *T. gondii* ELs are similarly duplicated/expanded in *H. hammondi* despite sharing over 99% syntenic loci between them.

2.1 INTRODUCTION

Toxoplasma gondii is a category B biodefense pathogen that can be lethal *in utero* and in immunocompromised humans. This parasite is a candidate bioterrorism agent due to the extreme environmental stability of infective oocysts that could contaminate water or food supplies (de Moura *et al.*, 2006, Lelu *et al.*, 2012). While infections in healthy humans are often benign, the identification of distinct *Toxoplasma* genotypes that are lethal in healthy adults (Darde, 2008) has changed the view of the bioterrorism potential of this pathogen and added to the urgency for the development of new chemotherapeutics and vaccines. *Toxoplasma* is unique among Apicomplexans in its ability to infect, be transmitted by, and cause disease in all warm-blooded animals, a trait which has certainly contributed to its worldwide distribution (Dubey *et al.*, 2003a).

With the advent of whole-genome tiling arrays and most importantly next-generation sequencing technologies it is now possible to more easily examine structural differences in whole genomes both within and between closely-related species. In humans locus expansion and diversification has been linked to psychiatric disorders like autism and schizophrenia (reviewed in (Malhotra *et al.*, 2012)), and susceptibility to a variety of other diseases (reviewed in (Almal *et al.*, 2012)). Locus expansion can also be beneficial. In mammals, expansion and diversification of killer-cell immunoglobulin-like receptor genes is important for recognition of diverse pathogens (Parham *et al.*, 2011). Laboratory studies with bacteria show that adaptation to

selective conditions via gene expansion occurs much more frequently than via point mutation (Kugelberg *et al.*, 2010), and in the field copy-number increases drive drug resistance in *Drosophila* (Schmidt *et al.*, 2010). Phenotypic impact can be driven by gene-dosage, but gene duplication also allows the original copy to maintain its function while duplicated copies are free to change via mutation and selection (Bergthorsson *et al.*, 2007).

Expanded and diversified gene families play important roles in pathogen virulence. Gene family expansions have been linked to virulence in *Candida spp.* (Moran *et al.*, 2011) and *Rickettsia spp.* (Ogata *et al.*, 2005). The *var* family of genes is distributed throughout the *P. falciparum* genome and encodes erythrocyte membrane antigens (PfEMPs) that are under strong diversifying selection (Pasternak *et al.*, 2009). Expanded genes have been linked to virulence, immune evasion (Pasternak *et al.*, 2009), drug resistance (Rottmann *et al.*, 2010) and host range (Hayton *et al.*, 2008) in *Plasmodium spp.*.

Our recent work demonstrates a role for gene duplication and subsequent diversification in *Toxoplasma* host-pathogen interactions. The *T. gondii ROP5* locus contains up to 10 copies depending on the strain, and this locus is essential for parasite virulence (Reese *et al.*, 2011b). Importantly, distinct isoforms from the *ROP5* locus can have synergistic effects on parasite lethality, indicating that individual copies of the *ROP5* gene have evolved subtly distinct functions. The *ROP5* locus also exhibits strain-specific copy number variation (Reese *et al.*, 2011b). The surface antigen-1 related (*SRS*) and rhoptry protein 2 (*ROP2*) superfamilies have duplicated multiple times and tandemly duplicated clusters of genes belonging to this family can be found throughout the genome (Boothroyd *et al.*, 2008). The SRS family has been implicated in immune evasion (Kim *et al.*, 2007), and the single-copy ROP2 superfamily member, ROP18, is a potent virulence factor in mice (Saeij *et al.*, 2006, Taylor *et al.*, 2006).

Less is known about copy-number variation (CNV) between species, although it was recently postulated that differentially duplicated genes and genomic structural variations could contribute to phenotypic differences between chimps and humans, and possibly have played a role in their speciation (Kehrer-Sawatzki *et al.*, 2007). Some support for this hypothesis is found in plants, where species-specific CNV is known to contribute in certain cases to reproductive isolation (Rieseberg *et al.*, 2010).

In this study we used a genome-wide approach to compare the extent of locus expansion across the genomes of *T. gondii*, *H. hammondi* and *N. caninum*. These three species belong to the subfamily *Toxoplasmatinae* (Mugridge *et al.*, 2000) and their genomes have been sequenced revealing a high level of synteny (Reid *et al.*, 2012, Walzer *et al.*, 2013). *T. gondii* and *N. caninum* have distinct intermediate host ranges and different definitive hosts (felines and canines, respectively), while *T. gondii* and *H. hammondi* share the same definitive host (Frenkel *et al.*, 1975). While *T. gondii* has a vast host range that includes birds and is virulent in mice (Saeij *et al.*, 2005a), *H. hammondi* and *N. caninum* cannot infect birds and are avirulent in mice (Dubey *et al.*, 2003b, Collantes-Fernandez *et al.*, 2012). For three *T. gondii* strains (GT1, ME49B7, VEG) and one *N. caninum* strain (NCLIV; (Reid *et al.*, 2012)) we used a manual pipeline to identify and curate all potentially expanded loci, and to compare the degree of overlap between them. This was facilitated by the fact that these genomes have been annotated. The *H. hammondi* reference genome (GenBank Accession AVCM00000000) has not yet been fully assembled into chromosomes nor annotated (Walzer *et al.*, 2013). However, for a subset of expanded loci we were able to determine if they were similarly expanded in *H. hammondi*. Overall, we find that in contrast to the high degree of synteny across these 3 species, there was a

significant lack of overlap in their expanded loci. This suggests an important role for gene expansion in the evolution of these species since their divergence from one another.

2.2 RESULTS:

2.2.1 Fifty-three loci have increased sequence coverage in *T. gondii*

We used a manual identification and curation pipeline to identify putatively expanded loci using sequence read coverage in *T. gondii* (Figure 2.1) and identified 53 loci of high sequence complexity in the nuclear genome of *T. gondii*. Average sequence coverage across the entirety of the 3 currently available Sanger-sequenced genomes differed slightly (median 15X, 19X and 14X for GT1, ME49 and VEG, respectively) due to different numbers of raw sequenced reads (Table 2.3), with 95 to 98% of the raw reads mapping to the ME49B7 genomic assembly using BLAT (Kent, 2002). Normalized sequence coverage across entire chromosomes from all three queried *T. gondii* strains was typically homogeneous, with sporadic patches of increased coverage at certain locations and telomeric sites (Figure 2.2A), indicating that gene duplication/expansion was relatively infrequent. We examined gene expansion at all loci across GT1, ME49 and VEG to estimate copy-number across the three strains (as in Figure 2.2B). Of the 53 expanded loci, only one (Expanded Locus 13, *EL13*; Figure 2.3 and Table 2.1) appeared to be entirely missing in one of the three *T. gondii* strains, in this case the Type I strain (GT1). Otherwise, the remaining 52 loci were conserved in their expanded state in all 3 queried *T. gondii* strains (Table 2.4). However, based on sequence coverage analysis, 22 loci exhibited CNV of ≥ 3 copies across the 3 queried *T. gondii* strains. This list included the *ROP5* locus

(*EL47*, Figure 2.3, Table 2.1) which, based on sequence coverage, we estimate to have ~11 copies in ME49B7 and 6, 4 copies in GT1 and VEG, respectively. This is similar to previously published analyses using high molecular weight Southern blotting of the *ROP5* locus in *T. gondii* strains RH (Type I), ME49 (Type II) and CTG (Type III) (Reese *et al.*, 2011b).

2.2.2 Expanded *T. gondii* loci are enriched for secretory proteins with few exons

Of the 53 expanded loci, 42 were predicted to contain protein-coding genes (www.toxodb.org). In addition, one locus (*EL40*) that did not have a predicted gene model within it does have evidence for being a coding sequence due to the presence of expressed sequence tags that map to this locus (Table 2.4). We anticipate that some of the expanded loci without an associated gene prediction will be transcribed to produce either coding or non-coding RNAs. It should be noted that with few exceptions the number of paralogs predicted in each of these three genome sequences greatly underestimated copy-number (Tables 2.1 and 2.4). This was most certainly due to collapsing of the assembly in regions containing tandemly duplicated clusters of genes that are similar in sequence as has been observed in other genomes (e.g., *Homo sapiens* (Sudmant *et al.*, 2010), and *Trichomonas vaginalis* (Carlton *et al.*, 2007)).

We used existing annotations of the *T. gondii* genome to characterize the nature of the 42 *T. gondii* loci containing predicted genes. We found a significant enrichment for genes predicted to contain N-terminal signal sequences (29 out of 42 compared to the entire predicted proteome; Hypergeometric distribution (HGD) $P=5.1\text{x}10^{-11}$; Table 2.2). In addition, these 42 genes have fewer exons (mean 2.1; median 1) than the rest of the predicted genes in the genome (mean 5.2; median 4). Kolmogrov-Smirnov (KS) analysis revealed a significant ($P=1.4\text{x}10^{-7}$) difference in the exon distribution between these two gene sets reflected by their cumulative distributions

(Figure 2.7). In fact, 26 of the 42 expanded gene-containing loci in *T. gondii* have only 1 exon (a significant enrichment compared to the genome as a whole; P=8.9x10^{-7}; Table 2.2).

2.2.3 Expanded *T. gondii* loci are dominated by genes of unknown function or localization, but also include predicted rhoptry proteins and surface antigens

We examined the degree of annotation of the 42 putatively protein-coding expanded loci using ToxoDB and our own protein family searches. Of these, 29 were annotated only as hypothetical or conserved hypothetical proteins (Table 2.4) and showed little similarity to previously characterized proteins from *T. gondii* or other eukaryotic species based on domain and BLAST analyses available on ToxoDB. We screened all 42 loci for Pfam domains (both A and B), and found 27 with a PFAM-A hit with an expect value ≤ 1 (Table 2.5). There were also multiple loci in this group that encode proteins with unannotated PFAM-B matches. Of those with PFAM-B hits, there were 6 PFAM-B domains that matched to multiple expanded loci, suggesting that genes with similar protein architectures have expanded into multi-locus, multi-gene families (including *EL17* and *EL25* [PFAM-B 2112] and *EL6* and *EL50* [PFAM-B 3349]; Table 2.5). Regardless, the majority of expanded *T. gondii* genes have yet to be characterized in terms of their function or subcellular localization, and contain protein domains that have yet to be annotated and/or characterized.

Of those that were annotated, 5 were predicted rhoptry proteins, some of which have been previously characterized. These are *EL1*, *3*, *16*, *36* and *47* (*ROP4/7*, *ROP42*, *ROP38*, *ROP2/8* and *ROP5*; (Boothroyd *et al.*, 2008); Table 2.1). Based on the HGD (Sokal *et al.*, 2012) this was a significant enrichment in predicted rhoptry proteins over the current annotation of the genome (Table 2; P=7.3x10^{-4}), further implicating the rhoptry proteome as a target for locus expansion

(Boothroyd *et al.*, 2008). Four expanded loci were annotated as members of the SAG1-related family of surface antigens. These are *EL13*, *22*, *37* and *51* (*SRS22*, *SRS26*, *SRS48* and *SRS59*, respectively (Wasmuth *et al.*, 2012)). One was annotated as a dense granule protein (*EL12*; *GRA11*), another as a microneme protein (*EL23*; *MIC17*) and 2 were previously characterized bradyzoite-specific NTPases that had been determined to be found in tandem in the genome (*EL52*, *53*; *NTPases II and I*).

2.2.4 Expanded *T. gondii* loci are enriched for developmentally regulated genes

We used previously published microarray expression data for *T. gondii* strain M4 (Fritz *et al.*, 2012) to quantitatively assess gene expression across multiple *T. gondii* genes during parasite development, focusing on the oocyst to sporocyst and tachyzoite to bradyzoite transition. We found that 13 of the 42 expanded loci contained genes that were significantly up- or down-regulated at the transcriptional level during the tachyzoite to bradyzoite transition *in vitro* and/or *in vivo* (Figure 2.4). Up-regulated genes included the rhoptry proteins *ROP42* and *ROP2/8*, and down-regulated genes included *NTPase I* and a paralog belonging to the *SRS22* family. This enrichment was significant (HGD P=0.0019; Table 2.2) compared to the entire predicted transcriptome assayed by the microarray. Ten of the 42 genes were developmentally regulated during the oocyst-sporozoite transition but this difference was not significant (HGD P=0.07; Table 2.2).

2.2.5 Multiple *T. gondii* expanded loci exhibit within-lineage copy-number variation

Twenty-two of the 53 expanded loci in *T. gondii* showed differences in sampling frequencies between the 3 representative genome strains, suggestive of copy-number variation (as seen previously for *T. gondii ROP5* (Reese *et al.*, 2011b)). Consistent with this, using cnv-seq to statistically assess copy-number variation at all 53 loci (Figure 2.12) we identified 23 that significantly varied in copy number between GT1/VEG and ME49 (6 loci), GT1 and ME49 (12 loci), and VEG and ME49 (5 loci). This list included the *ROP5* locus, for which copy-number has been determined previously in multiple *T. gondii* strains using high-molecular weight Southern blotting (Reese *et al.*, 2011b). We also conducted whole-genome comparative genomic hybridization (CGH) for two distinct members of each of the 3 major lineages of *T. gondii*. Of the 41 gene-coding expanded loci that could be surveyed by microarray, 24 had significantly higher hybridization intensity across the *T. gondii* strains queried (see Materials and Methods for statistical analyses). In addition to this we observed a general correlation between copy-number as estimated by sequence coverage analysis and our CGH data. For example, *EL5* is predicted to have 4-5 copies in the Type I strain GT1, and only 1-2 copies in Type II strain ME49 and Type III strain VEG, and the CGH (Figure 2.8) and cnv-seq (Figure 2.12) analyses reflect this difference. A similar correlation can be found for *EL16* (Figures S2 and S6). These data provide a secondary validation of locus expansion for 24 of the *T. gondii* loci as well as the variation in copy-number observed in the sequence coverage plots. Of the remaining loci, 18 had sufficient CGH data but did not show a significant increase in hybridization intensity (Figure 2.9).

We also identified some loci with CGH intensity profiles suggesting a difference in copy-number between members of the same clonal lineage, most notably *EL30*, for which CGH intensity values were distinct between *T. gondii* strains ME49 and PRU (Figure 2.6B). To

address this further we performed high-molecular weight Southern blotting for *EL30* as well as *EL3* and *EL45* across 6 *T. gondii* strains using restriction enzymes predicted to cut outside of the entire expanded locus. For all 3 loci we observed differences in estimated copy-number both between and within lineages. We observed intra-lineage variation in locus size for *EL3* (*ROP42*) between GT1 and RH as well as between ME49 and PRU, and we estimated that GT1 and RH have 6 and 9 copies, respectively, and that ME49 and PRU have 8 and 6 copies, respectively. We estimated that VEG and CTG have 7 copies (Figure 2.6C). We detected intra-lineage variation for *EL30* and *EL45*, where ME49 and PRU had different-sized loci (Figure 2.6C). Southern blot data for these three loci are generally consistent with the CGH-intensity values, although there are some exceptions. For example for locus *EL3* strain GT1 has a higher CGH intensity than would be predicted based on the Southern blot. This could be due to as-yet unidentified cryptic restriction sites in the locus that are specific to strain GT1. In comparison, for the single-copy locus *AMA1* we did not detect inter- or intra-lineage variation in sequence read coverage, CGH probe intensity, or locus size as estimated by Southern blot (Figure 2.11).

2.2.6 Three *T. gondii* expanded loci are not essential for *in vitro* growth

We successfully knocked out 3 expanded loci (*EL3*, ROP42; *EL6*, and *EL23*) in a virulent Type I background (RHΔ*ku80*:Δ*hxgprt*; (Huynh *et al.*, 2009)). Parasite lines with deletions at each of these loci (Figure 2.13) exhibited no obvious defects in *in vitro* growth, and neither RHΔ*ku80*:Δ*el3* or Δ*el23* parasite clones showed any defects in acute virulence as measured by survival time in mouse infections with 100 tachyzoites. We did not test RHΔ*ku80:*Δ*el6* in mouse virulence assays.

2.2.7 *N. caninum* has a markedly different set of expanded genes that is enriched for members of the SAG1-related surface antigen family

In order to compare gene expansion between *T. gondii* and its close relative *N. caninum*, we first examined gene expansion in *N. caninum* using identical approaches as for *T. gondii*. We identified 65 expanded loci in *N. caninum* (Liverpool Strain; www.toxodb.org), 45 of which contained predicted protein-coding genes. These loci are listed in Table 2.6. The set of *N. caninum* expanded genes was also enriched for genes encoding proteins with signal peptides (34/45) and single-exon genes (31/45) compared to the genome as a whole (HGD P=2.6x10^{-14}, 4.7x10^{-12}, respectively; Table 2.2). Remarkably, nearly half (22/45; 49%) were found to contain a SAG PFAM domain, suggesting that they belong to the SRS family, and this is a significant enrichment over the annotated genome (P=4.1x10^{-12}; Table 2.2). Of the remaining 23 protein-coding expanded *N. caninum* loci, 3 were previously annotated (*ROP4/7*, and *NTPases I* and *II*), 17 had at least one recognized PFAM domain, and 4 were completely unannotated and had no recognizable PFAM domains. While the increased number of *SRS*-family genes in *N. caninum* has been reported previously (Reid *et al.*, 2012), our data indicate that these loci have also been subject to multiple rounds of tandem (i.e., local) duplication.

2.2.8 Distinct sets of genes are expanded in *T. gondii* and *N. caninum*

To determine the degree of overlap between genes that are expanded in *T. gondii* and *N. caninum*, we used BLASTN to identify the syntenic location for each of the 53 *T. gondii* loci described above, and then examined that region of the genome for signatures of gene expansion in *N. caninum*. We found that only 16 of the 53 *T. gondii* loci also had evidence of expansion in

N. caninum (≥ 2-fold higher sequence coverage compared to background; Figure 2.6A). This lack of overlap between *T. gondii* and *N. caninum* at these loci is in contrast to the overall gene-by-gene synteny between the *T. gondii* and *N. caninum* genomes, and this lack of overlap is significant (HGD 16/53 versus 6463/8103 for *T. gondii*: $P=7x10^{-15}$; HGD 16/64 versus 6463/7227 for *N. caninum*; Figure 2.6A). One of the shared expanded loci (Locus *EL15*; annotated in *T. gondii* as *ROP38*) had higher sequence coverage in *N. caninum* while the remaining 15 had higher sequence coverage in *T. gondii*. Of the 37 loci that were uniquely expanded in *T. gondii*, 19 had syntenic orthologs in *N. caninum* but these loci showed no evidence of expansion (sequence coverage was ~1X; e.g., *EL3* and *EL30*, Figure 2.5A). The remaining 18 loci did not have a syntenic ortholog based on the current *T. gondii* annotation based on the clusters of orthologous groups database implemented in ToxoDB (Tatusov *et al.*, 2003).

2.2.9 *T. gondii* and *H. hammondi* share 16 of 27 expanded loci

The *H. hammondi* genome has not been annotated or assembled into chromosomes, preventing a *de novo* analysis of gene duplication and expansion. However we did use the recently published *H. hammondi* genome sequence and raw sequence reads (Walzer *et al.*, 2013) to determine which *T. gondii* loci were similarly expanded in *H. hammondi*. Of the 42 protein-coding loci, 27 had a perfect reciprocal-best-BLAST hit between the *T. gondii* and *H. hammondi*. Of these 27 putative orthologous sequences, we estimated that 16 of these (59%) had more than 1 copy in *H. hammondi*, while data at the remaining 11 loci suggested that they were single-copy genes in *H. hammondi* (Figure 2.6C).

2.3 DISCUSSION

Our previous work has demonstrated a clear and important role for gene duplication in the pathogenesis of *Toxoplasma gondii* (Reese *et al.*, 2011b): we showed that the *ROP5* locus was tandemly expanded in multiple *T. gondii* strains and that this expansion led to diversification of individual copies within the locus. We were therefore interested in identifying other *T. gondii* loci that were tandemly expanded to determine 1) what features were shared among expanded genes and 2) whether these loci were differentially expanded both within the *T. gondii* species and in comparison with its nearest sequenced relatives, *N. caninum* (Reid *et al.*, 2012) and *H. hammondi* (Walzer *et al.*, 2013). Stretches of increased copy number were relatively rare in these genomes, and based on the currently released genome assemblies all of the 42 protein-coding loci harbored multiple tandem duplications of the same gene.

An important finding of this work was that expanded loci exhibit copy-number variation even between members of the same clonal lineage. In Europe and North America, *T. gondii* isolates are dominated by members of 3 main lineages (Types I, II and III) and isolates from within the same lineage appear to be clonal. However, based on Southern blot analysis we show that 3 loci exhibit CNV between members of the same clonal lineage, and other candidate loci with similar within-lineage variation can be identified from our CGH data. We do not assert that members of the same lineage are genetically identical, however based on whole genome comparisons "true" members of the same clonal lineage are more genetically similar to one another than to other strains. For example RH and GT1 strains have only 1,394 SNPs that distinguish them, representing a polymorphism rate of ~0.002% (Yang *et al.*, 2013), compared to a polymorphism rate between lineages that ranges from 1-5% (Boyle *et al.*, 2006). Therefore we find that differences in copy-number at these loci is in contrast to the overall genetic identity of

the clonal strains, suggesting that these loci are changing more rapidly than the rest of the genome. However, we do not discount the impact of single-nucleotide polymorphisms in determining differences between members of the same strain type, which have been identified in RH and GT1 (Yang *et al.*, 2013).

We have shown previously that the raw number of copies does not necessarily track with impact on a particular phenotype (Reese *et al.*, 2011b). ROP5 alleles from Types I and III *T. gondii* strains have a higher contribution to virulence than the Type II alleles, yet the Type II parental strain (ME49) has ~10 copies while Types I (RH) and III (CTG) are estimated to harbor 6 and 4, respectively (Reese *et al.*, 2011b). In this case, the sequence of an individual copy is more important, and therefore the presence or absence of even a single copy of a particular isoform at an expanded locus could have a phenotypic impact. While we cannot yet demonstrate conclusively that these changes are driven by selection, the fact that individual structural changes in the genome are much more rare than individual mutations (Malhotra *et al.*, 2012) certainly points to the possibility that this may be a selection-driven process. Data emerging from the *Toxoplasma gondii* GSCID project (https://sites.google.com/site/*Toxoplasma*gondiigscidproject/) will allow us to rapidly identify other expanded loci that differ between members of the same clonal lineage since a number of these will be sequenced. We also do not know if within-lineage gene expansion occurs during asexual reproduction (which occurs in all intermediate hosts infected by *T. gondii*), sexual reproduction (which occurs only in felines), or both. In the wild *T. gondii* is capable of self-mating (Frenkel *et al.*, 1970) and these expansions could occur during genetic recombination.

We also successfully knocked out three expanded loci (*EL3*, *EL6* and *EL23*) in a highly virulent *T. gondii* strain (RHΔKu80; (Huynh *et al.*, 2009)) and found no defects in their ability to

replicate *in vitro* or for 2 of these loci (*EL3* and *EL23*) no defects in parasite virulence. To date a number of expanded loci encoding secretory proteins such as those encoded in these loci have been deleted, including ROP2/8 (Pernas *et al.*, 2010) and ROP5 (Behnke *et al.*, 2011, Reese *et al.*, 2011b) without any consequences for *in vitro* tachyzoite growth. The fact that these parasite lines show no defects *in vitro* or *in vivo* is not surprising given that all three of these loci are upregulated during the tachyzoite-to-bradyzoite transition (Figure 2.4). Our data, however, show that these loci can indeed be deleted, facilitating future studies on their role during the chronic phase of infection where they are most highly-expressed.

There was a statistically significant lack of overlap between tandemly duplicated loci in *T. gondii* and *N. caninum*. While it is tempting to hypothesize that these differentially expanded gene clusters may be responsible for the phenotypic differences between these species (such as differences in virulence in mouse and the different definitive hosts), the data presented here cannot directly validate this claim. However, there is support for this hypothesis from other pathogenic species. In fungi, comparisons between species with different levels of pathogenesis in humans identified gene duplication and expansion as an important mechanism in the evolution of pathogenicity (reviewed in (Moran *et al.*, 2011)). For example, genomic comparisons between *Candida albicans* and *Candida dubliniensis* found that these species were most highly distinguished by the presence of two highly expanded gene families in pathogenic *C. albicans* compared to non-pathogenic *C. dubliniensis* (Jackson *et al.*, 2009), and these loci are under investigation as virulence factors. Findings from these studies also suggested that gene duplication and subsequent expansion may play a more important role than other mechanisms such as horizontal gene transfer, in the evolution of novel traits in eukaryotic pathogens (Moran *et al.*, 2011).

In both *T. gondii* and *N. caninum*, the expanded gene sets were statistically significantly enriched for genes predicted to encode secreted proteins with fewer exons compared to the genome as a whole. While secreted proteins make up the vast majority of *T. gondii* effectors, the significance of the increased propensity of genes with few exons to duplicate and expand is unknown. One possible explanation could be the fact that introns are free to mutate more freely and that while a single locus could be duplicated, subsequent distinct mutations in the introns of both copies may prevent further expansion of the locus during recombination or genome replication. The other, and not mutually exclusive, possibility is that genes with fewer exons are subjected to stronger selection, since all of the previously characterized *Toxoplasma* secretory proteins known to play roles in pathogenesis, including the single-exon effector genes ROP18 (Saeij *et al.*, 2006, Taylor *et al.*, 2006), ROP5 (Behnke *et al.*, 2011, Reese *et al.*, 2011b), ROP16 (Saeij *et al.*, 2007) and GRA15 (Rosowski *et al.*, 2011) have fewer exons relative to the rest of the genome.

The minimal overlap between the loci that have expanded in *T. gondii* and *N. caninum*, and a similar lack of overlap between genes expanded in *T. gondii* compared to *H. hammondi*, is consistent with what has been reported for other closely-related species with highly syntenic genomes (Carlton *et al.*, 2007). To our knowledge, this is the first comparative analysis of gene expansions across multiple apicomplexan species. In *T. gondii*, the majority of the uniquely expanded loci are of no known function based on primary sequence, while in *N. caninum* the vast majority of the expanded loci are predicted to encode surface antigens belonging to the SRS superfamily. This was reported previously (Reid *et al.*, 2012) although the phenotypic impact of this expansion is unknown.

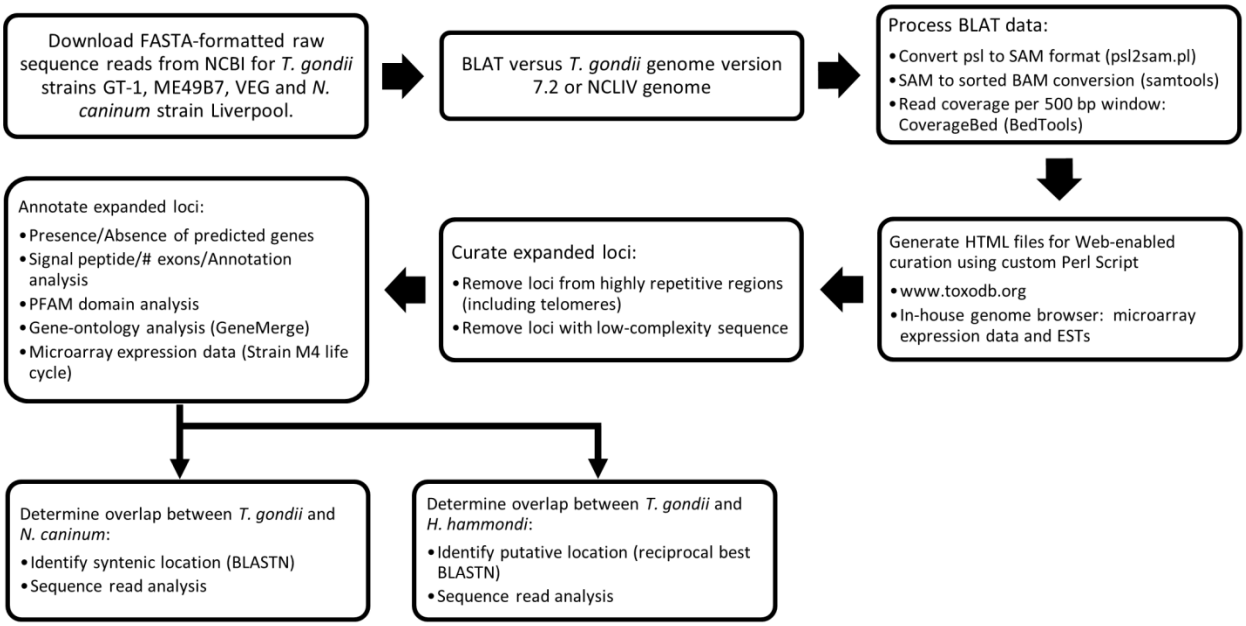


Figure 2.1: Flow chart depicting the pipeline used to identify, curate and annotate expanded loci in *Toxoplasma gondii*, *Neospora caninum* and *Hammondia hammondi*.

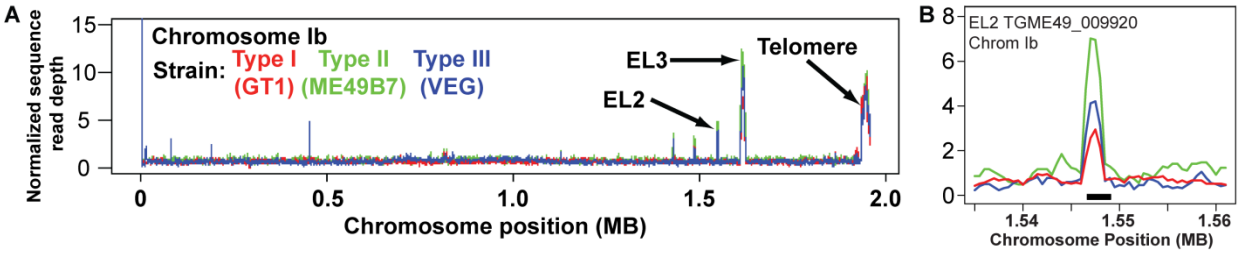


Figure 2.2: Chromosome Ib sequence coverage plot.

A) Normalized sequence coverage plot of chromosome Ib for three T. gondii strains (GT1, ME49B7 and VEG). Coverage data for each 500 bp window was normalized to the average coverage for the entire chromosome for each strain. The locations of EL2 and EL3 (the expanded loci identified on Chr. Ib) and the predicted right arm telomere are indicated. B) Detailed view of normalized sequence read coverage for EL2 for each of the three *T. gondii* strains, showing potential variation in copy number indicated by the read plot. Strains color-coded as in A).

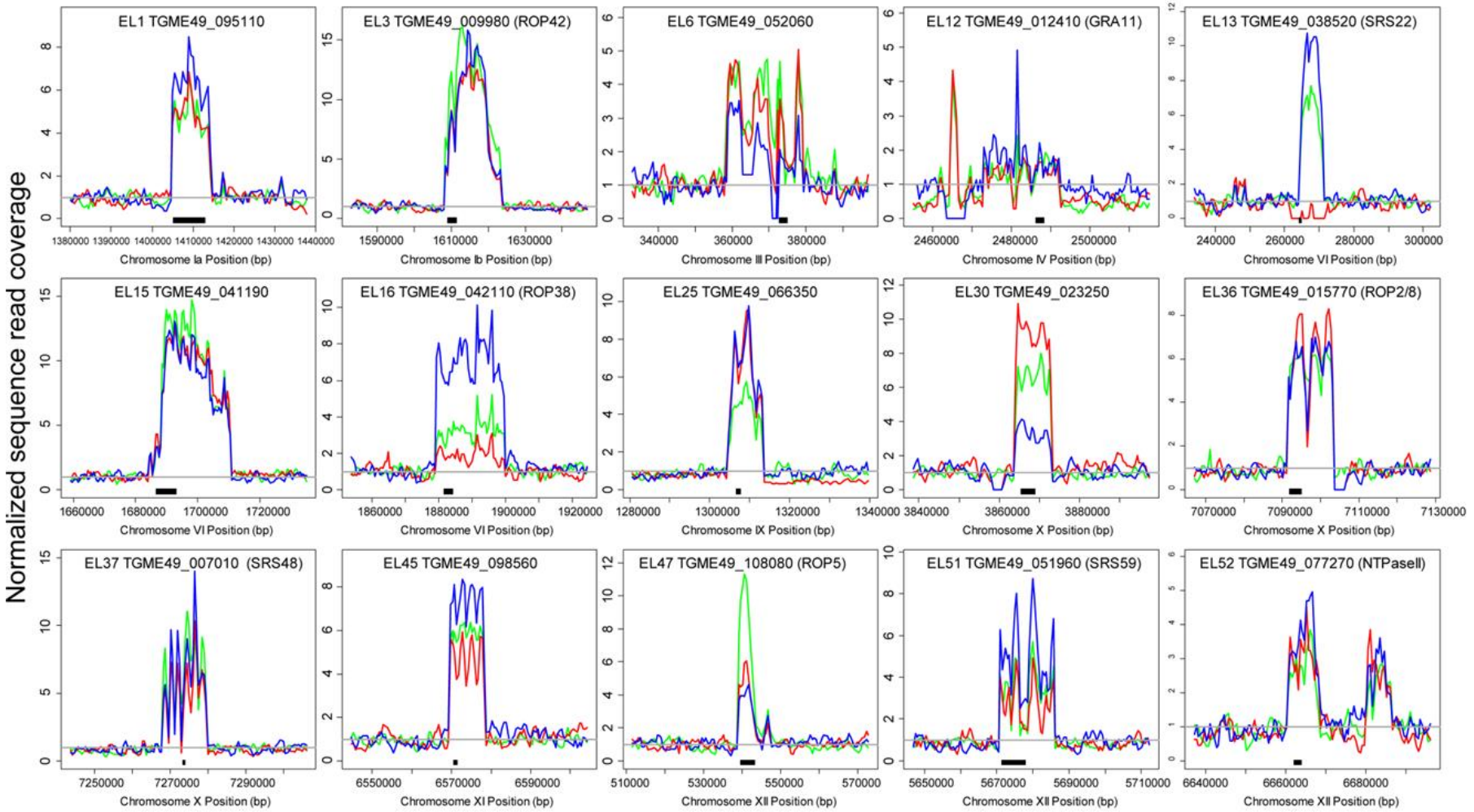
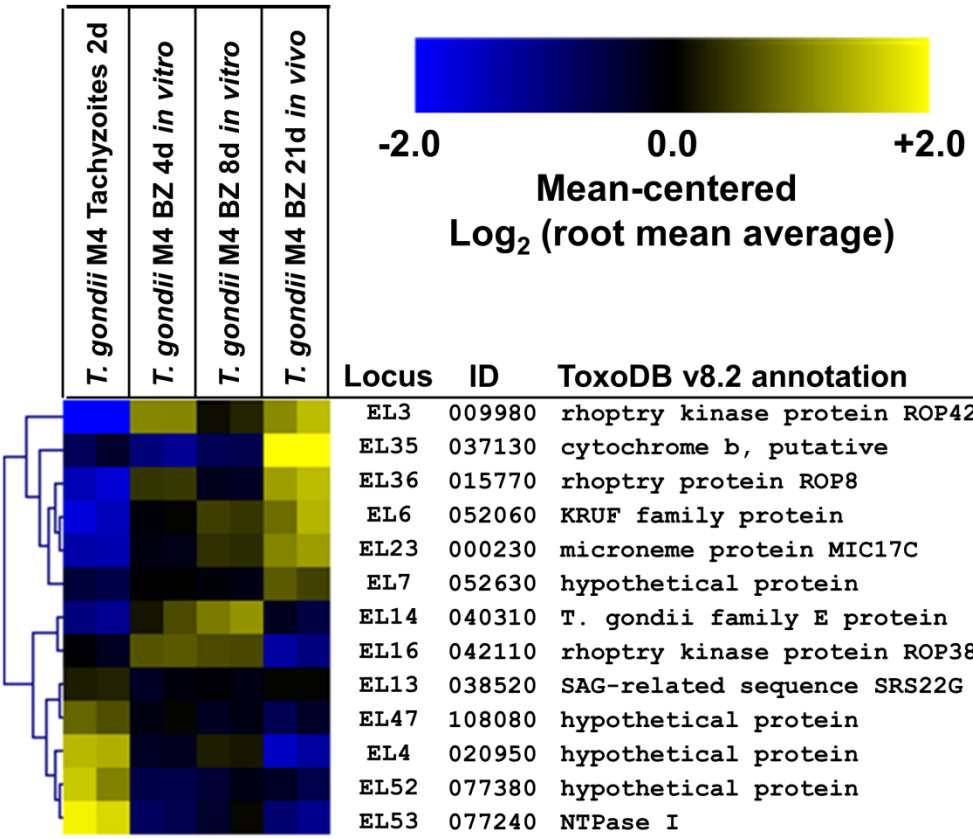


Figure 2.3: Sequence coverage plot for 15 expanded loci in three strain types of *T. gondii*.

Data are normalized to the read coverage in the leftmost 20 Kb flanking the expanded locus, and the grey line indicates normalized sequence coverage of 1. Black bars beneath each plot indicate the location of one of the predicted *T. gondii* isoforms. Information for each locus can be found in Table 2.1.



Locus	ID	ToxoDB v8.2 annotation	Fold-difference BZ-TZ *In vitro*	Fold-difference BZ-TZ *In vivo*
EL3	009980	rhoptry kinase protein ROP42	6.76	14.39
EL35	037130	cytochrome b, putative	-1.08	6.58
EL36	015770	rhoptry protein ROP8	2.45	7.54
EL6	052060	KRUF family protein	4.11	6.64
EL23	000230	microneme protein MIC17C	3.27	5.62
EL7	052630	hypothetical protein	1.33	2.21
EL14	040310	T. gondii family E protein	4.35	1.59
EL16	042110	rhoptry kinase protein ROP38	1.67	-1.95
EL13	038520	SAG-related sequence SRS22G	-1.27	-1.16
EL47	108080	hypothetical protein	-1.90	-2.29
EL4	020950	hypothetical protein	-2.30	-6.98
EL52	077380	hypothetical protein	-3.03	-3.38
EL53	077240	NTPase I	-3.88	-7.03

Figure 2.4: Expression profile for 13 expanded loci in tachyzoites and bradyzoites of *T. gondii* strain M4.

Expression profile for 13 expanded loci in tachyzoites and bradyzoites of *T. gondii* strain M4 (GEO Database Accession: GSE32427) showing an increase or decrease of transcript abundance during the tachyzoite-to-bradyzoite transition *in vitro* and/or *in vivo*.

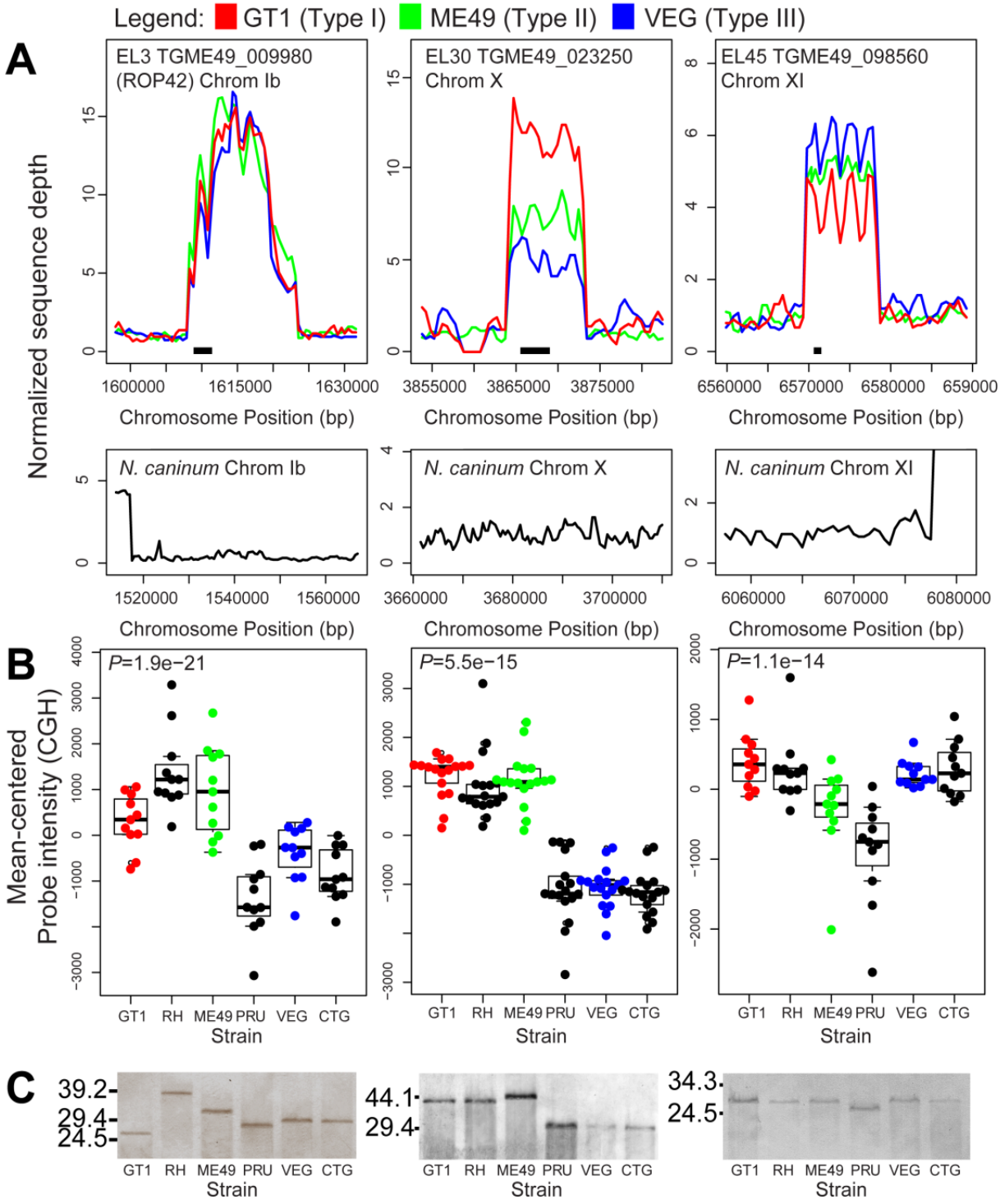

Legend: GT1 (Type I) ME49 (Type II) VEG (Type III)
A
EL3 TGME49_009980 (ROP42) Chrom Ib
EL30 TGME49_023250 Chrom X
EL45 TGME49_098560 Chrom XI
Normalized sequence depth
Chromosome Position (bp)
N. caninum Chrom Ib
N. caninum Chrom X
N. caninum Chrom XI
B
Mean-centered Probe intensity (CGH)
P=1.9e−21
P=5.5e−15
P=1.1e−14
GT1 RH ME49 PRU VEG CTG
Strain
C
39.2
29.4
24.5
44.1
29.4
34.3
24.5

Figure 2.5: Sequence read, comparative genomic hybridization and Southern blot analysis for *EL3, EL30* and *EL45*.

A) Sequence read analysis for three strains of *T. gondii* (top) and *N. caninum* Liverpool (bottom). Data for each strain and locus are normalized to the read coverage in the 20 Kb flanking the expanded locus to the left. B) Comparative genomic hybridization (CGH) across 6 *T. gondii* strains, 2 from each of the 3 canonical lineages. Only microarray probes with perfect matches in GT1, ME49 and VEG were used in the calculations. Boxes span the 1st and 3rd quartiles and contain the median value for all useful probes. *P*-values for significant differences in hybridization intensity compared to the single copy gene AMA1 (grey horizontal line) are at the top of the graph, and individual values are shown in the beeswarm plots. C) Southern blots for 6 *T. gondii* strains using DNA digested with restriction enzymes predicted to cut outside of the repeat locus. Restriction enzymes were BspEI, BglI and NotI, respectively. A similar blot for the single-copy gene *AMA1* can be found in figure 2.11.

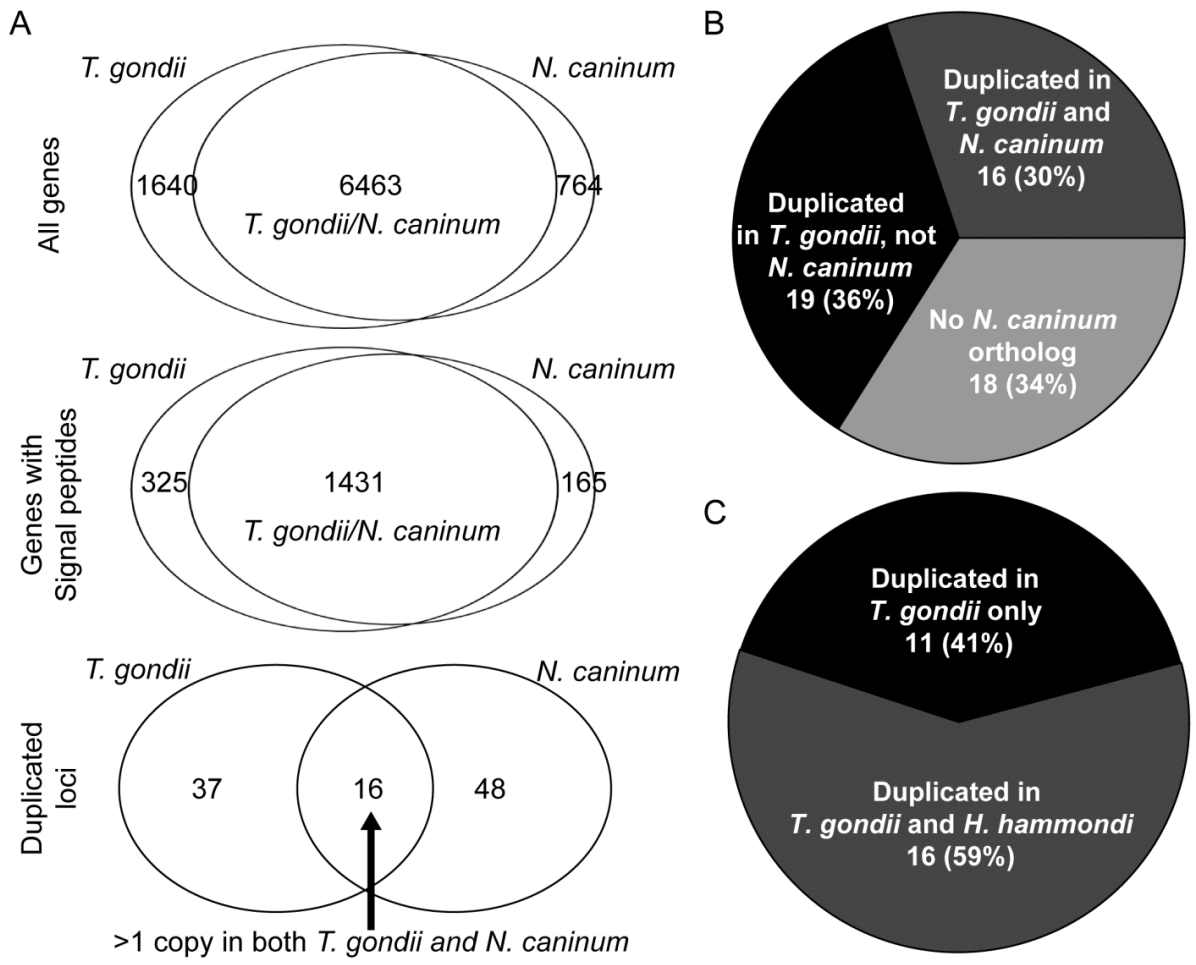


Figure 2.6: Comparison of gene duplication among *T. gondii, N. caninum* and *H. hammondi*.

A) Overall genomic orthologs between *T. gondii* and *N. caninum* predicted genes. Top panel: all predicted genes, middle panel: all predicted genes with predicted signal peptides; bottom panel: expanded loci (53 for *T. gondii* and 64 for *N. caninum*). The top 2 Venn diagrams are based on gene-by-gene synteny, while the bottom panel for expanded loci is based on whether the locus is expanded in both *N. caninum* and *T. gondii* (16 loci), or whether it is absent or present as a single copy in one species (37 and 48 for *T. gondii* and *N. caninum*, respectively). B) Presence and expanded state of all 53 *T. gondii* expanded loci compared to *N. caninum*. Eighteen of the 53 *T. gondii* loci do not have a syntenic ortholog in *N. caninum*, a significant enrichment compared to the entire genome. C) Comparison of the expanded state of *T. gondii* loci for which a clear ortholog was identified in *H. hammondi* HhCatGer041. Of the 27 loci, 11 are uniquely expanded in *T. gondii* compared to *H. hammondi*, while we estimated that *H. hammondi* has at least 2 predicted copies for the remaining 16 loci.

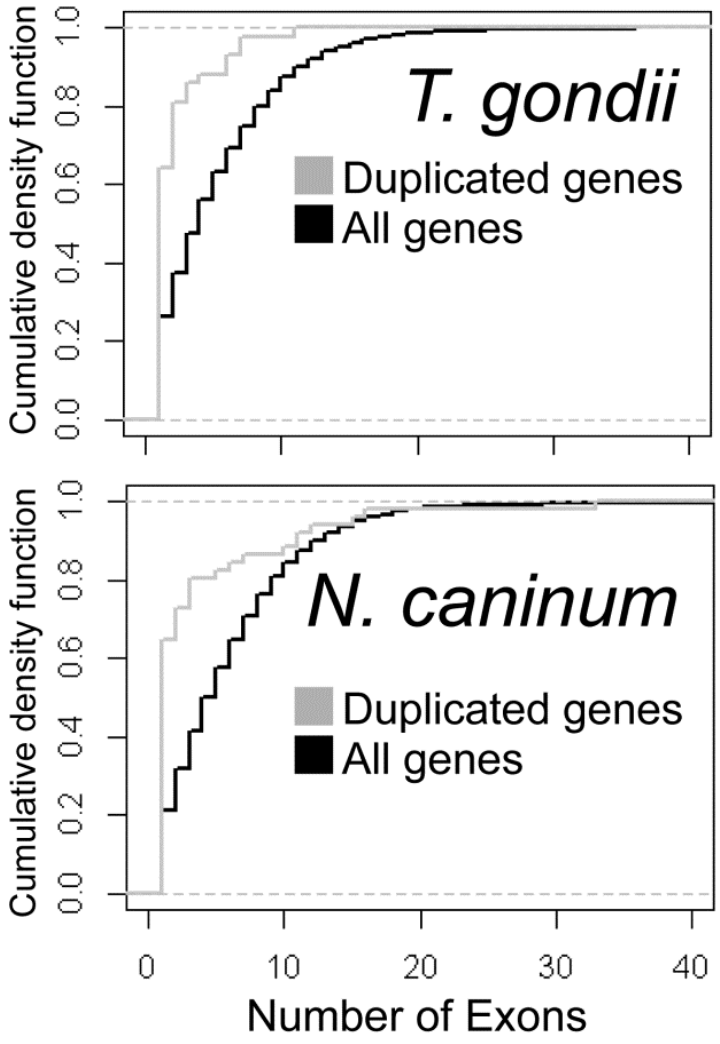


Figure 2.7: Cumulative distribution plots of all genes for *T. gondii* and *N. caninum* and expanded genes.

Cumulative distribution plots of all genes (black line) for *T. gondii* (top) and *N. caninum* (bottom) and expanded genes (grey line) showing the enrichment for genes with fewer exons in the expanded gene sets for both species.

Figure 2.8: Loci with significantly higher CGH probe hybridization intensity compared to *AMAI* (multiple pages).

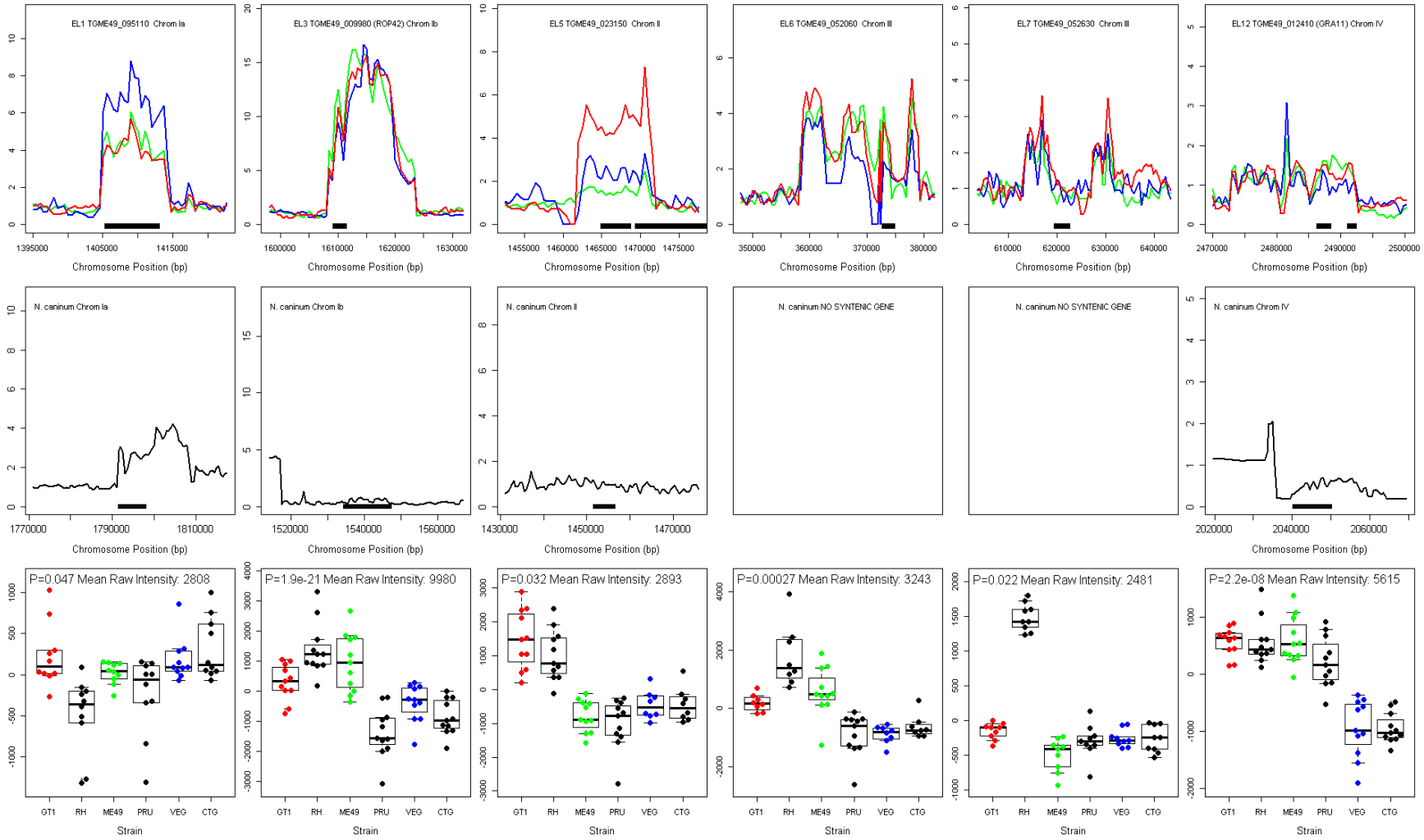


Figure 2.8 continued…

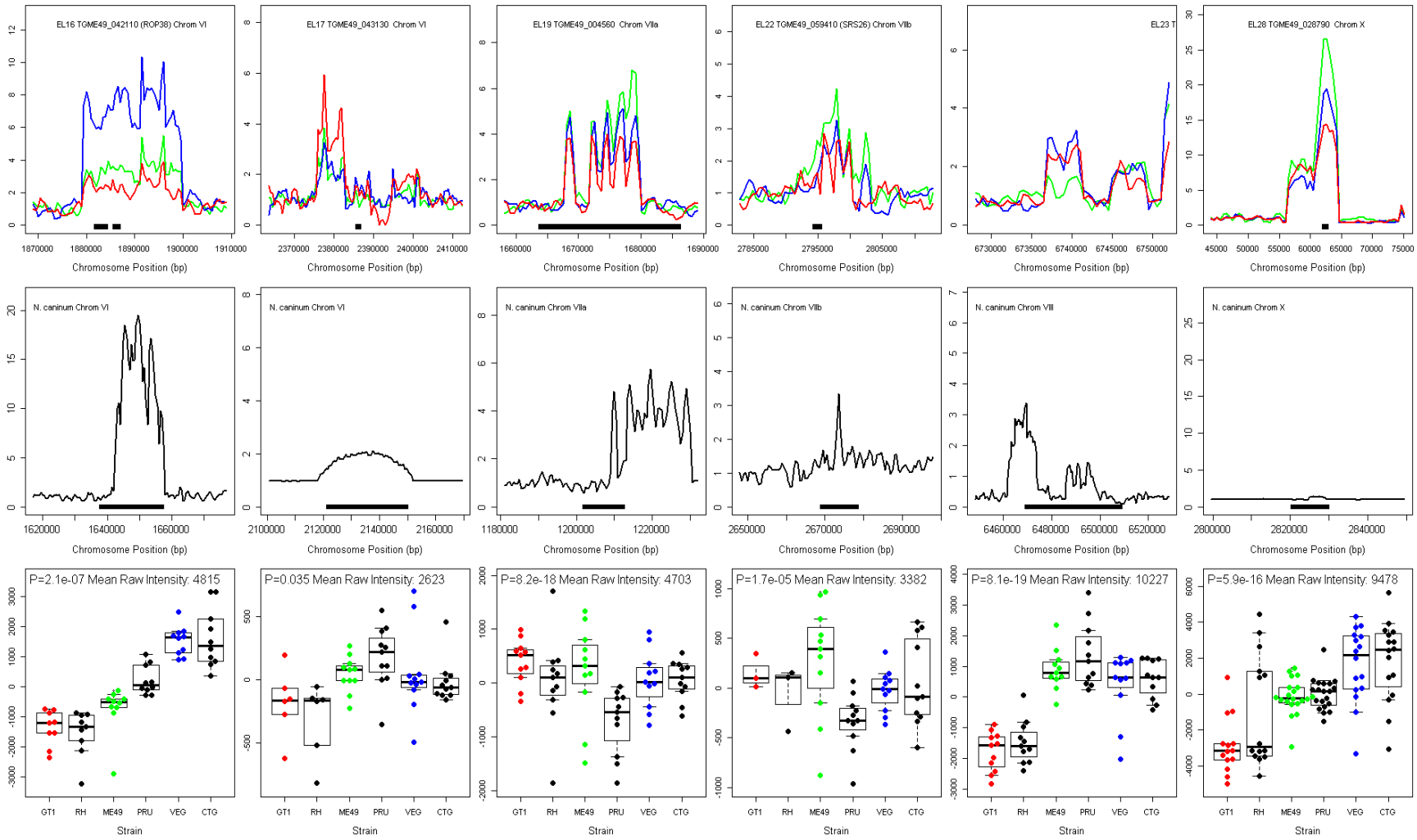


Figure 2.8 continued…

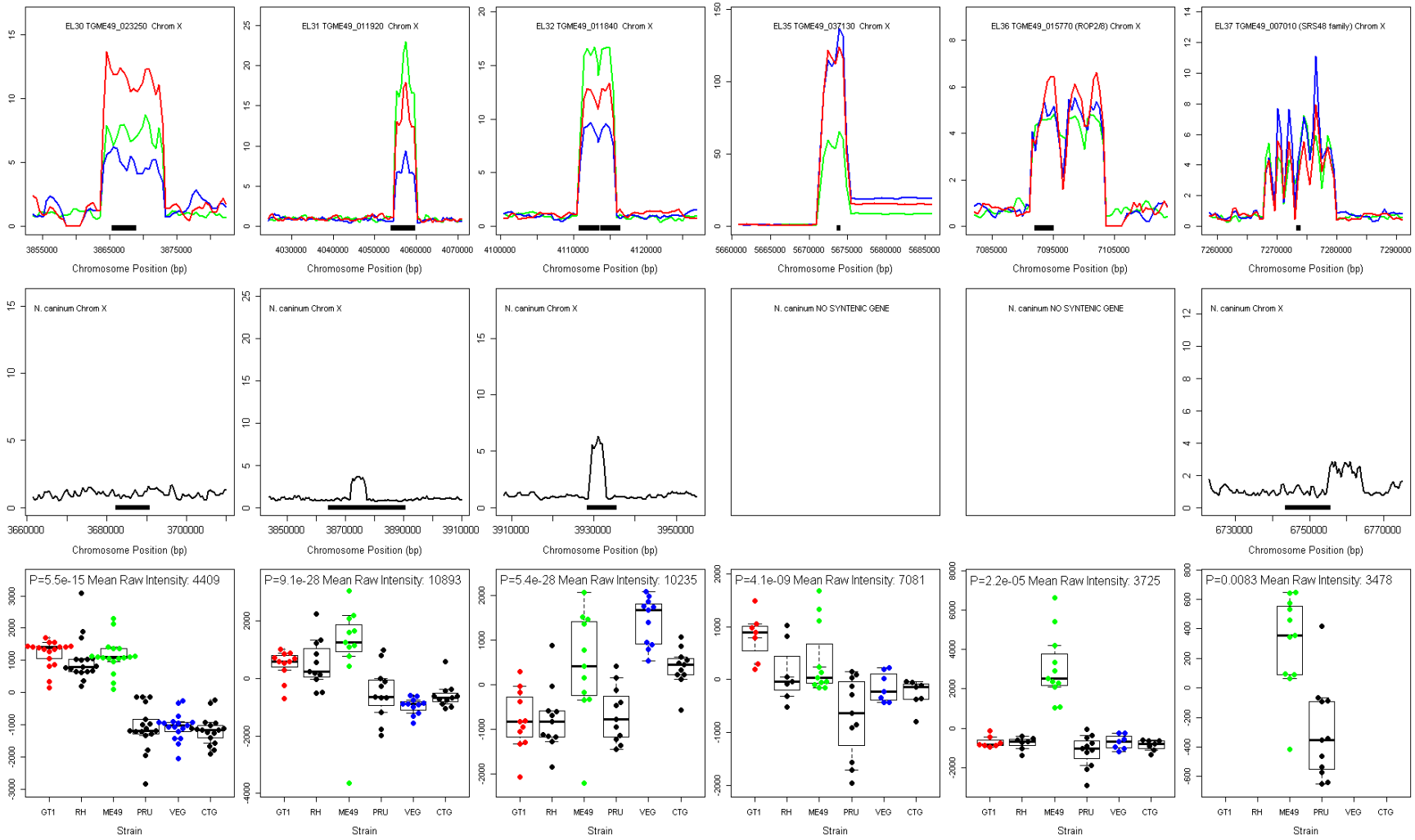


Figure 2.8 continued…

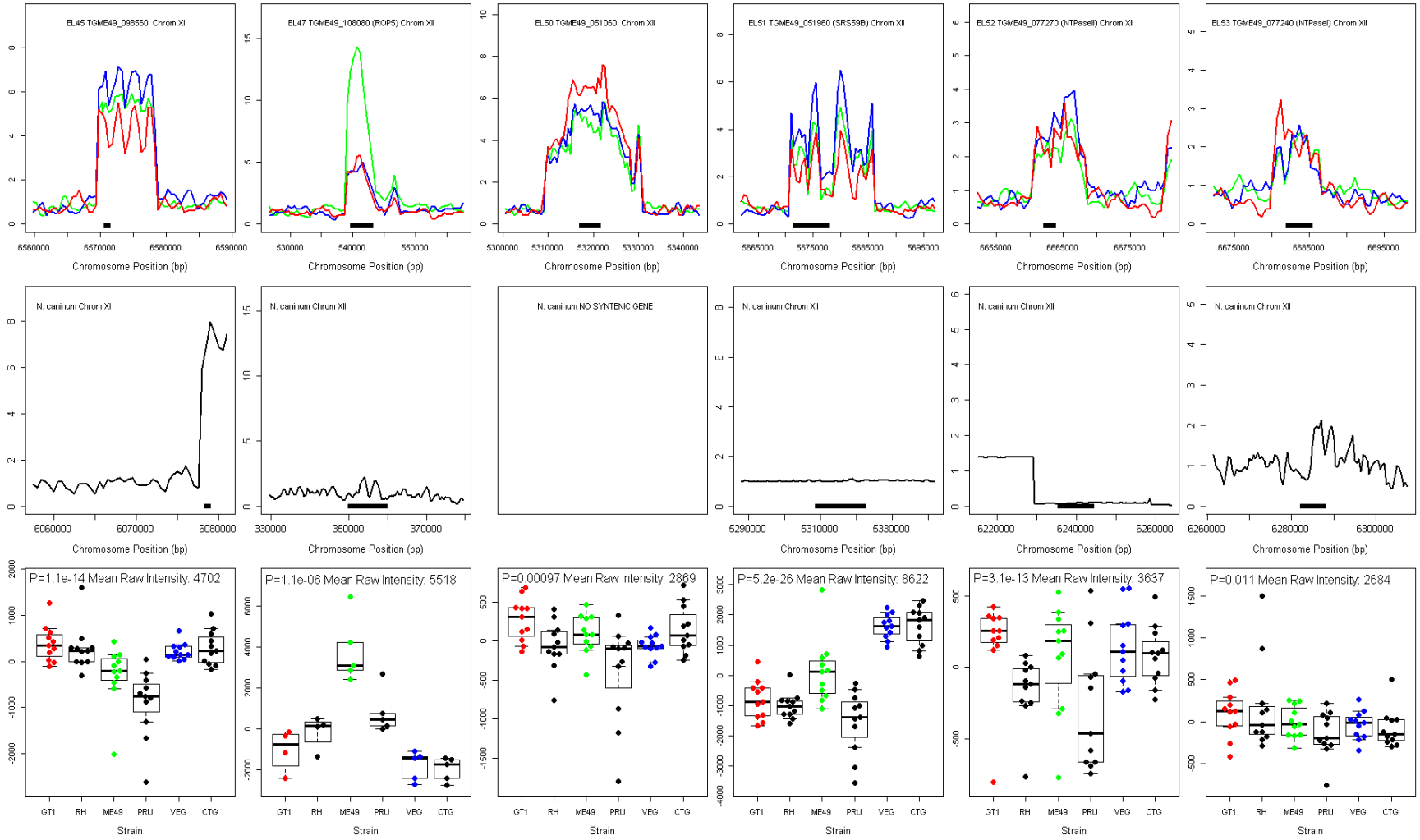


Figure 2.8: Loci with significantly higher CGH probe hybridization intensity compared to *AMAI*.

Sequence read plots for expanded loci containing predicted genes that were queried on the *T. gondii* microarray that had significantly higher ($P \leq 0.05$) CGH intensity compared to the single-copy gene *AMA1*. Plots are shown for three strains of *T. gondii* and one strain (NCLIV) of *N. caninum* (middle panel). Black bars beneath each plot indicate the location of one of the predicted isoforms. All read plot data were normalized for each strain/species to the average read coverage in the 20 Kb flanking the expanded locus to the left. In the CGH plots, boxplots indicate the 1st and 3rd quartiles and contain the median, and data for all useful probes for that gene are shown in the beeswarm overlay. Strains are color-coded as red, green and blue (GT1, ME49 and VEG) or black in the read and CGH plots.

Figure 2.9: Loci containing predicted genes without significantly higher CGH probe hybridization intensities compared to *AMAI*.

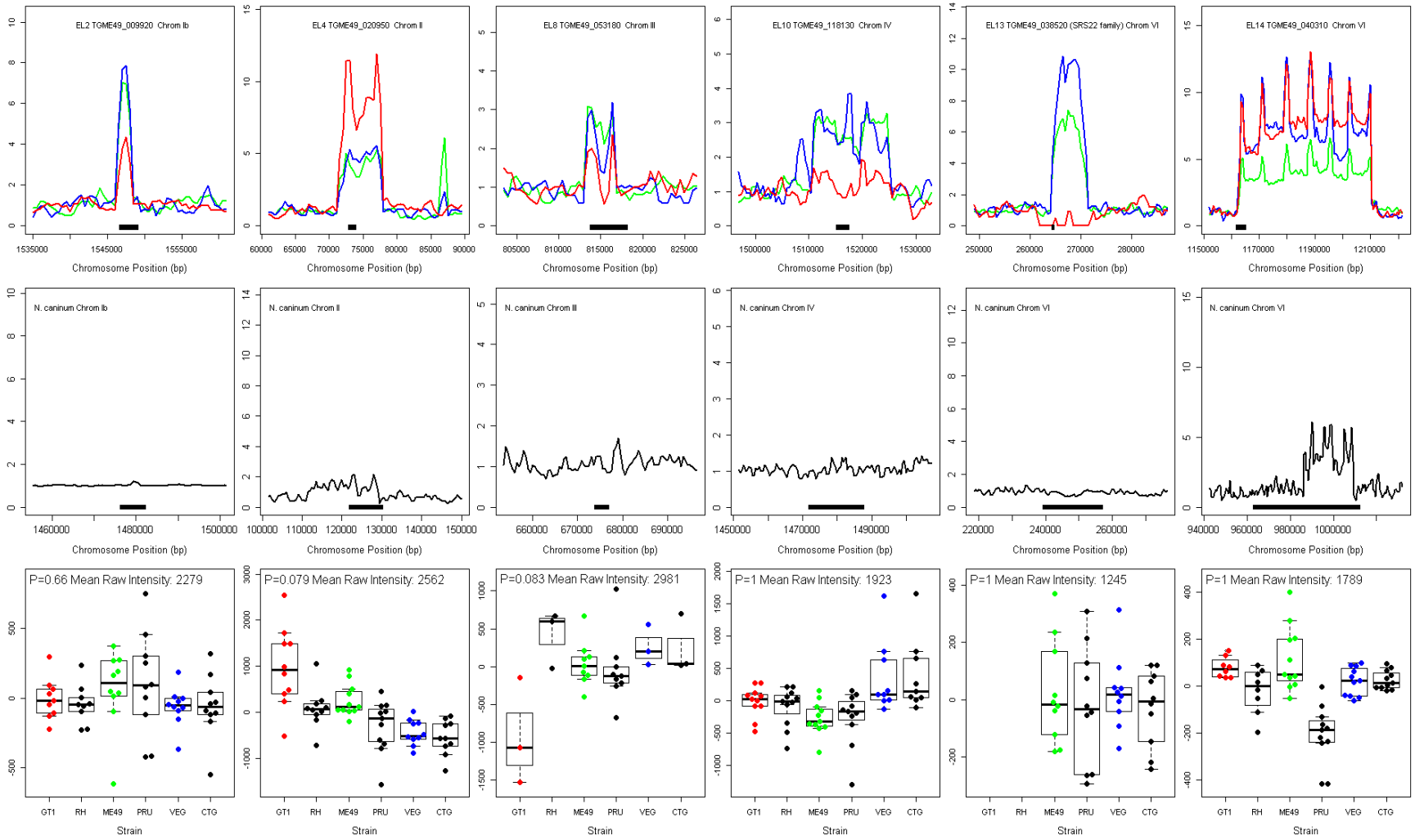


Figure 2.9 continued…

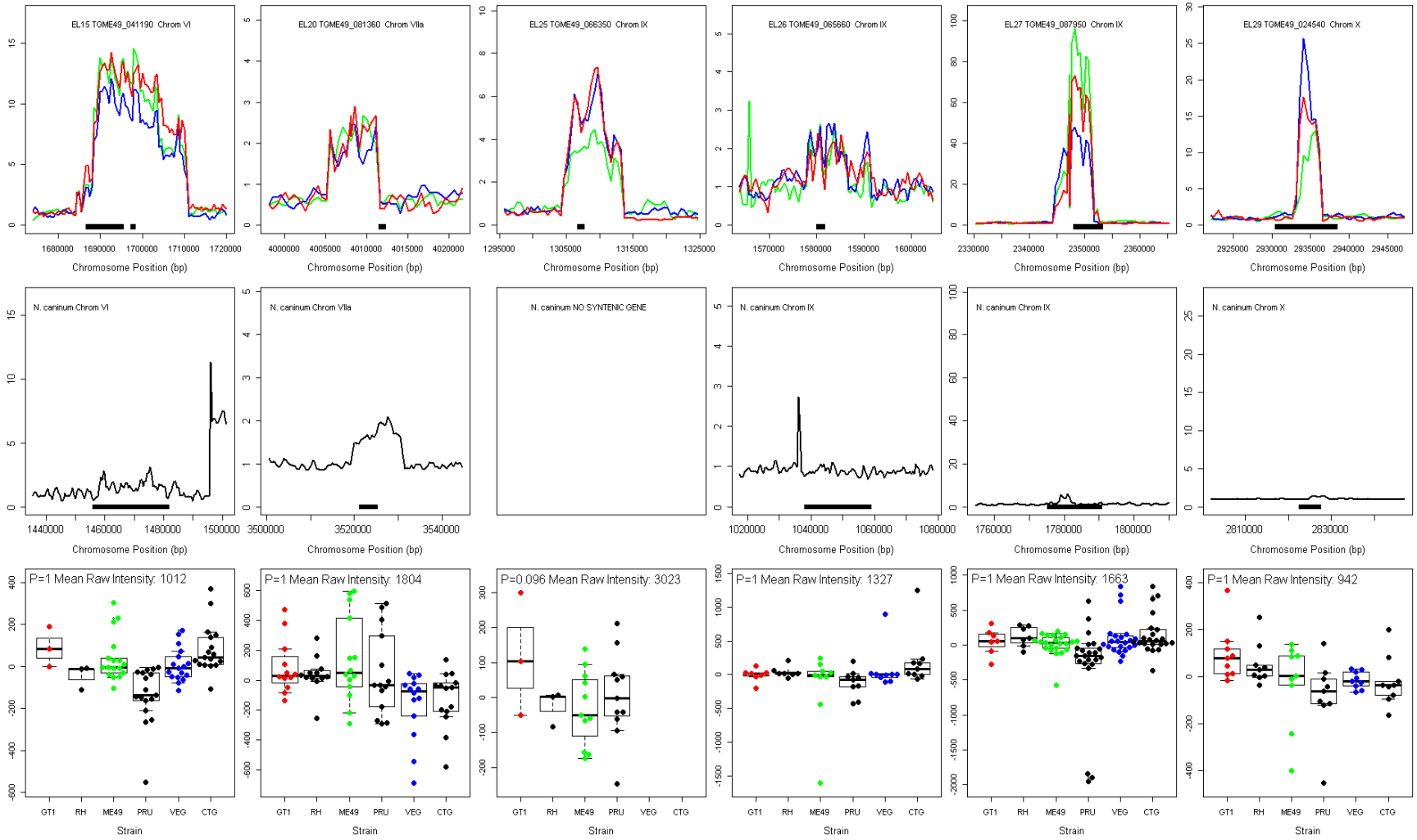


Figure 2.9 continued…

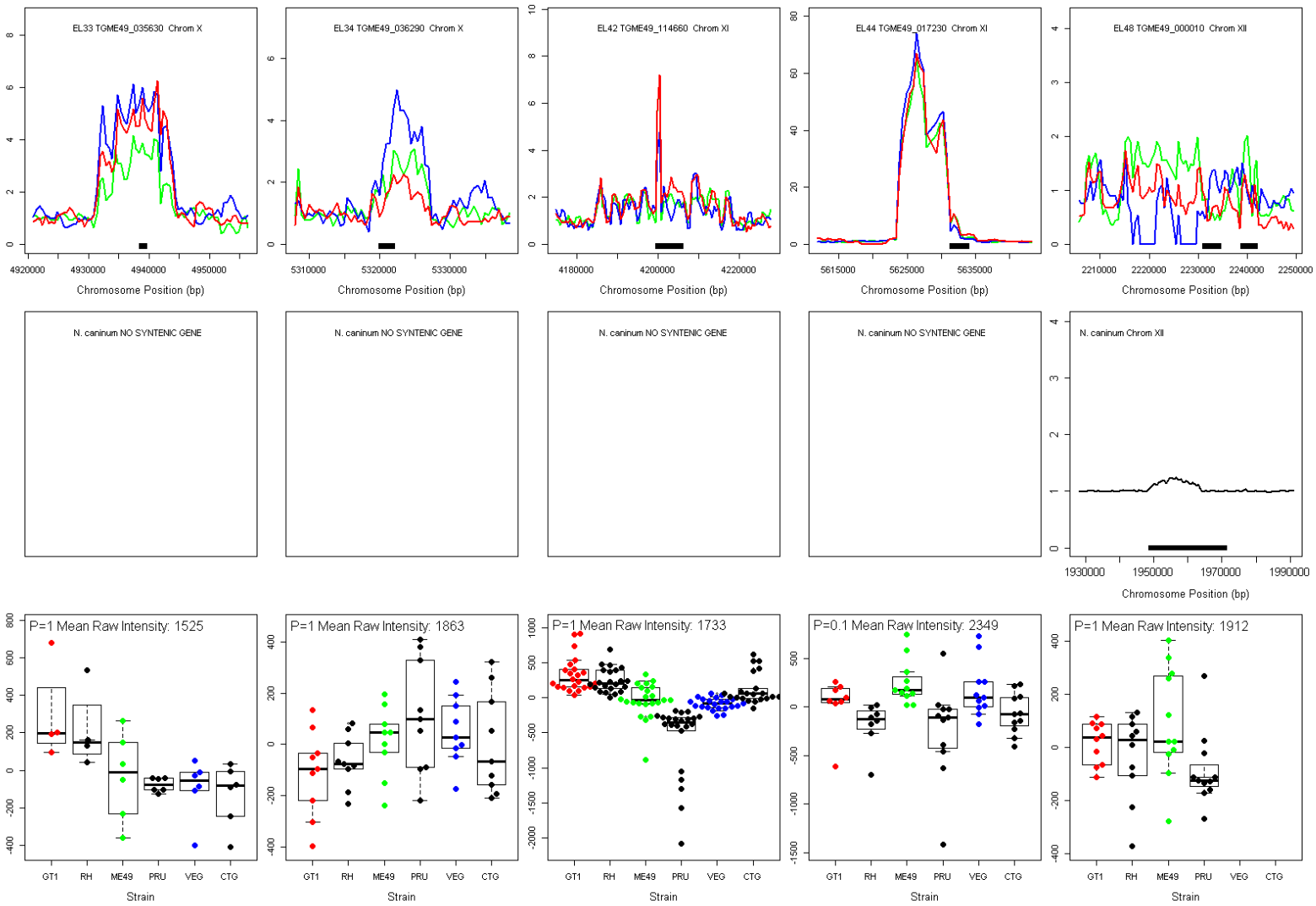
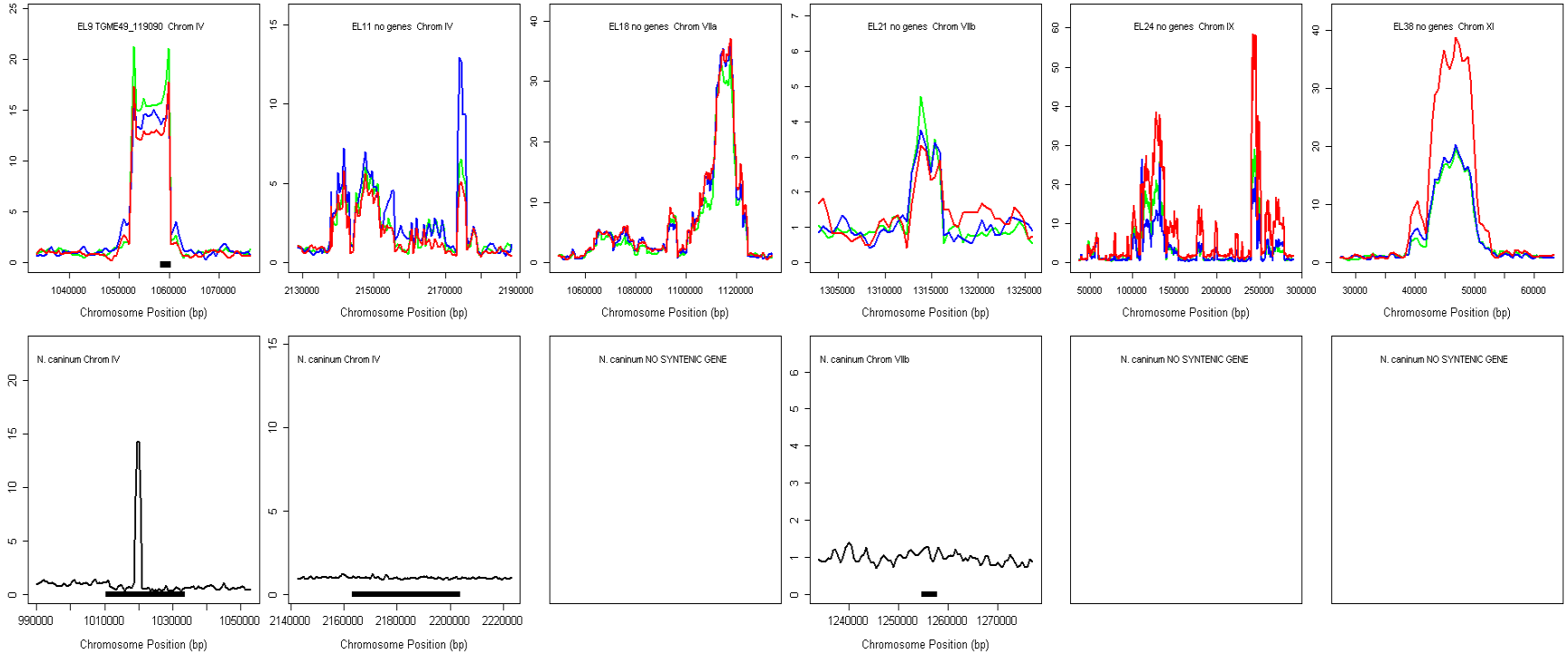


Figure 2.9: Loci containing predictied genes without significantly higher CGH probe hybridization intensities compared to *AMAI*.

Sequence read plots for 18 loci containing predicted genes that were queried on the *T. gondii* microarray that *did not have* significantly higher ($P \leq 0.05$) CGH intensity compared to the single-copy gene *AMA1*. Plots are shown for three strains of *T. gondii* and one strain (NCLIV) of *Neospora caninum* (middle panel). Black bars beneath each plot indicate the location of one of the predicted isoforms. All read plot data were normalized for each strain/species to the average read coverage in the 20 Kb flanking the expanded locus to the left. In the CGH plots, boxplots indicate the 1st and 3rd quartiles and contain the median, and data for all useful probes for that gene are shown in the beeswarm overlay. Strains are color-coded as red, green and blue (GT1, ME49 and VEG) or black in the read and CGH plots.

Figure 2.10: Sequence read plots for 12 loci without CGH data.



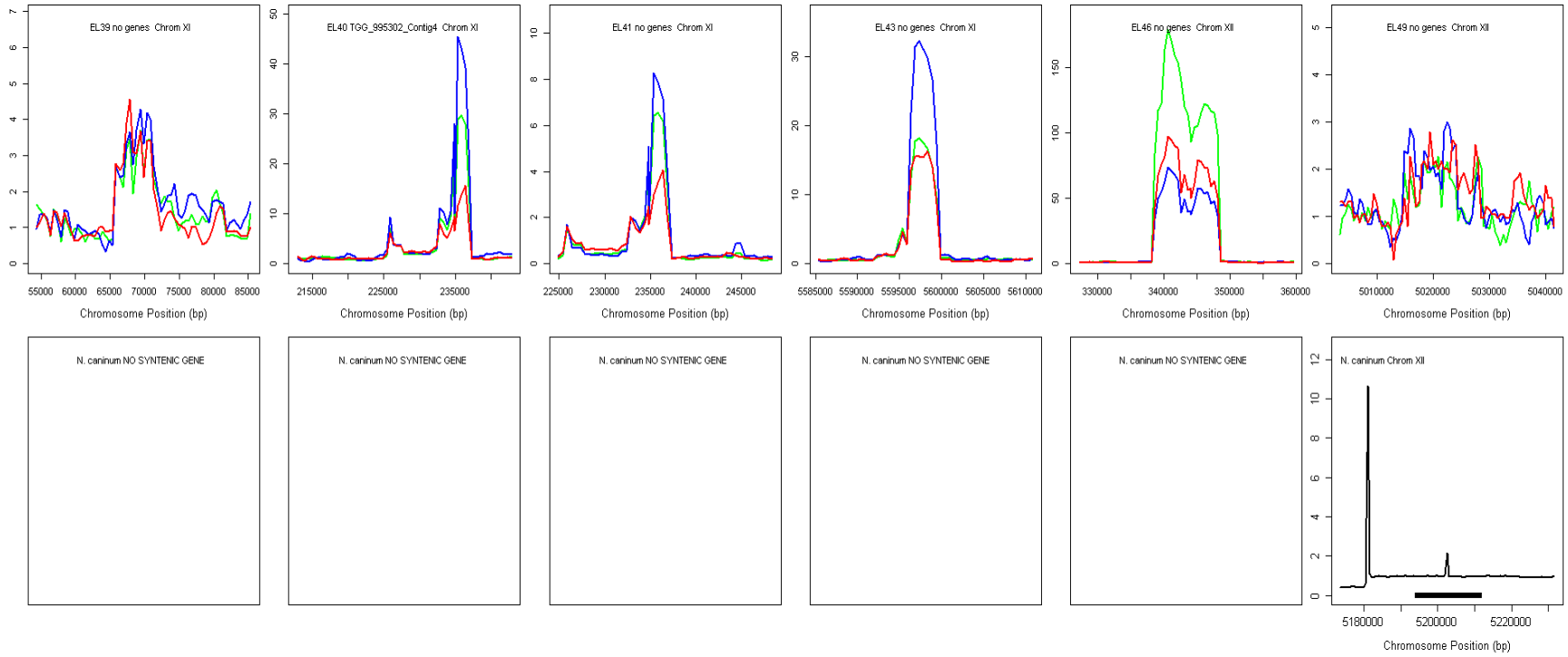


Figure 2.10: Sequence read plots for 12 loci without CGH data.

Sequence read plots for 12 loci without CGH data due to the lack of a predicted gene product in the genome annotation encompassed by the *T. gondii* Affymetrix microarray. Plots are shown for three strains of *T. gondii* and one strain (NCLIV) of *Neospora caninum* (middle panel). All read plot data were normalized for each strain/species to the average read coverage in the 20 Kb flanking the expanded locus to the left. Strains are color-coded as red, green and blue (GT1, ME49 and VEG) or black in the read and CGH plots. Black bars in the *N. caninum* plots indicate the location of the syntenic region as identified by BLASTN.

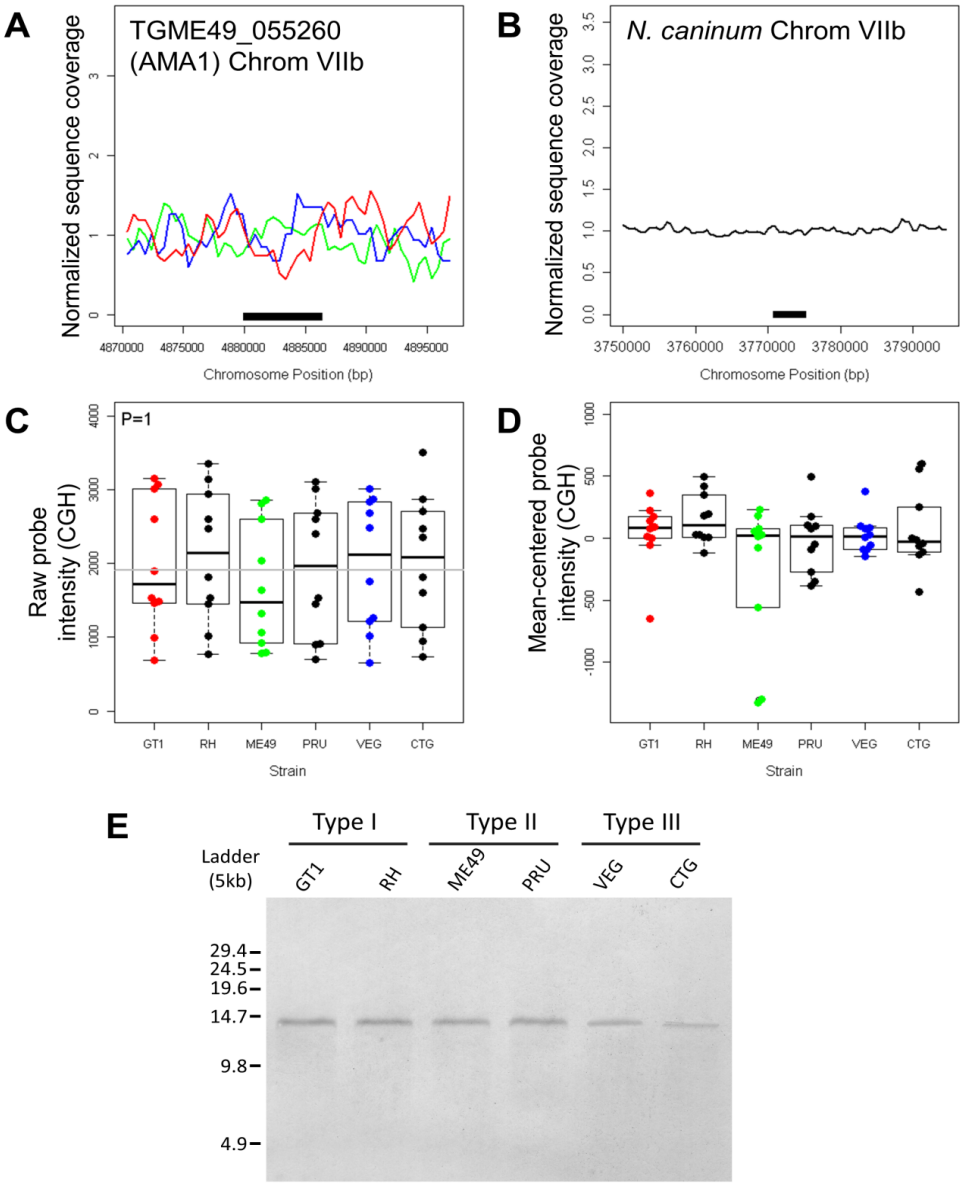


Figure 2.11: Sequence coverage, CGH and Southern blot analysis of the single-copy *AMA1* locus in *T. gondii* and *N. caninum*.

A,B) Sequence coverage analysis in *T. gondii* (A) and *N. caninum* (B) for the *AMA1* locus. C,D) Comparative genomic hybridization data for the AMA1 locus. Both raw (C) and mean-centered (D) hybridization intensity data are shown. E) Southern blot of genomic DNA from 6 *T. gondii* strains digested with ScaI, separated using Pulsed-Field Gel Electrophoresis, and probed with *AMA1*-specific probes.

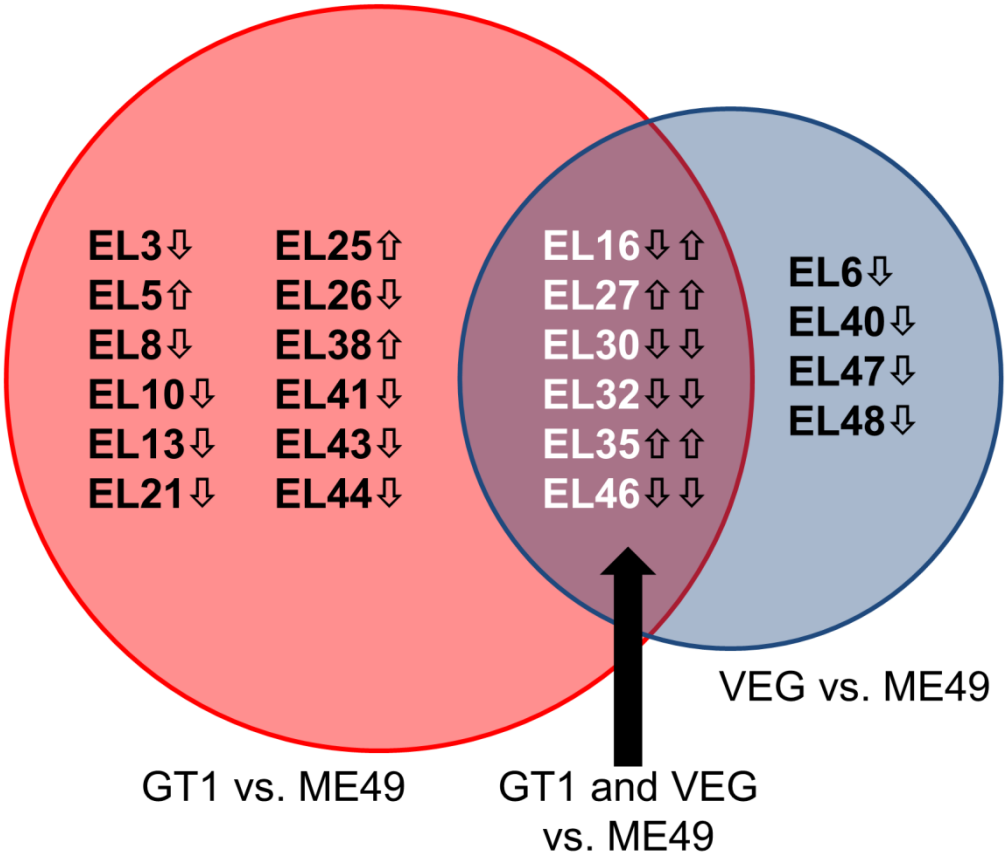


Figure 2.12: Copy-number variation CNV in *T. gondii* GT1 and VEG compared to ME49B7 as determined by cnv-seq.

Loci with statistically-significant evidence for CNV based on sequence coverage analysis are shown. Arrows after each locus name indicate whether the locus had increased or decreased copy number compared to ME49B7. For the overlapping genes, the arrows indicate increased/decreased copy number for GT1 and VEG compared to ME49B7, respectively.

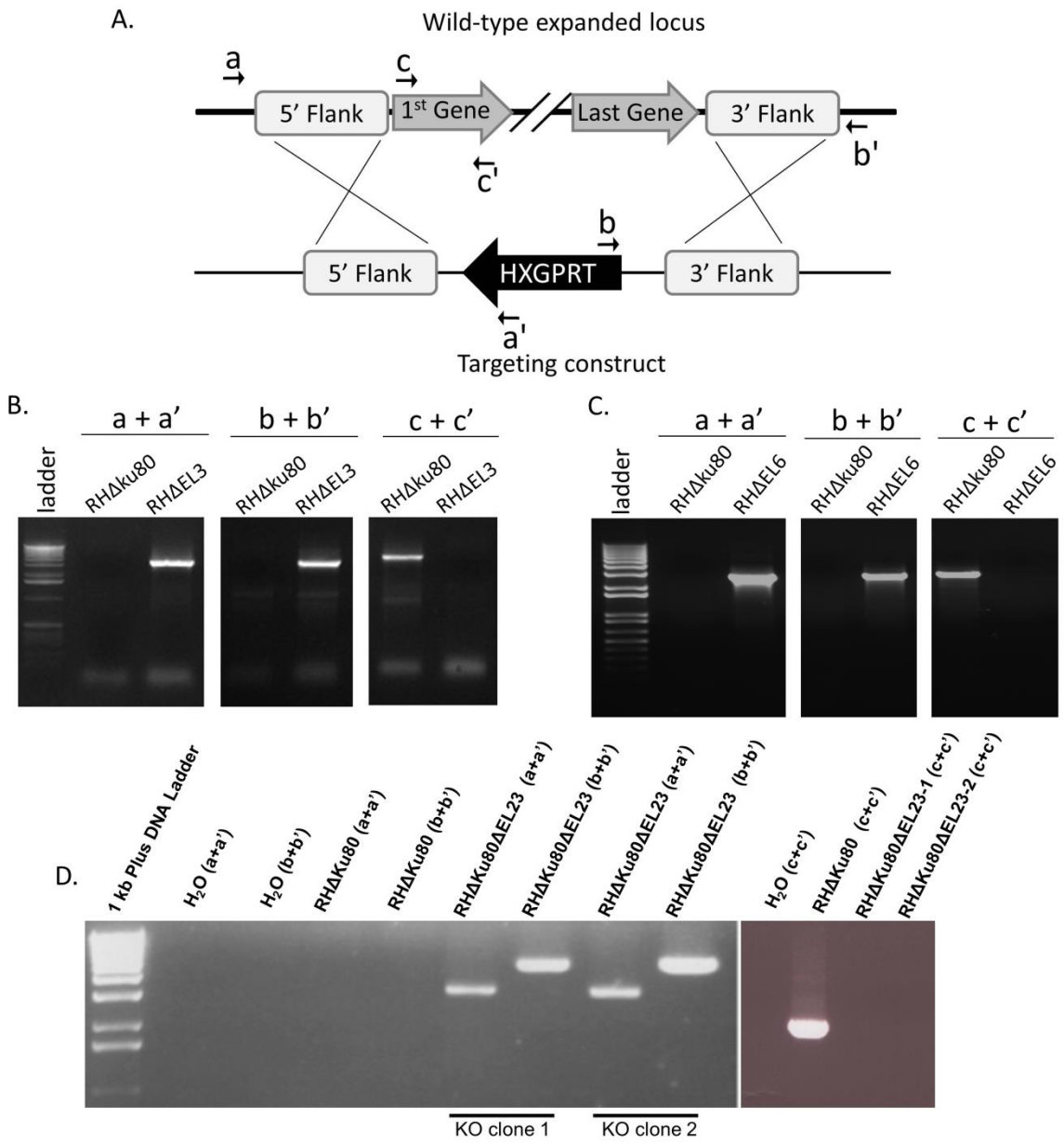


Figure 2.13: Deletion of expanded loci.

A) Schematic depiction of primer locations to validate homologous recombination on both the left (a+a') and right (b+b') flanks of the locus, and to validate the deletion of the gene (c+c'). B-D) Validation of knockouts for *EL3* (B), *EL6* (C) and *EL23* (D) by PCR

Table 2.1: Properties of 18 expanded loci in *T. gondii*, *N. caninum* and *H. hammondi*

					T.g.* copy # by strain			**Number of predicted *T.g.* orthologs by strain†**			***N.c.* ‡**			***H.h.* §**
Locus	**Chr**	**Pos (MB)**	**Gene**	**Annotation**	**I**	**II**	**III**	**I**	**II**	**III**	**Chr**	**Pos (MB)**	**Copy #**	**Copy #**
EL1	Ia	1.41	095110	ROP4/7	5	5	7	1	1	2	Ia	1.79	3	7-8
EL3	Ib	1.61	009980	ROP42	13	13	13	2	1	2	Ib	1.54	1	ND
EL6	III	0.36	052060	KRUF family	4	4	4	2	1	1	NA	NA	0	1
EL12	IV	2.48	012410	GRA11	1	1	2	0	2	1	IV	2.05	1	1
EL13	VI	0.26	038520	SRS22G	0	6	11	4	1	1	VI	0.25	1	ND
EL15	VI	1.68	041190	hypothetical	11	11	11	5	3	3	VI	1.47	2	1
EL16	VI	1.88	042110	ROP38	2	3	8	3	2	1	VI	1.65	18	2
EL22	VIIb	2.79	059410	SRS26A	1	2	1	3	1	3	VIIb	2.67	1	ND
EL23	VIII	6.74	000240	MIC17	3	2	3	2	1	1	VIII	6.49	2	7-8
EL25	IX	1.3	066350	hypothetical	6	4	6	1	1	1	NA	NA	0	ND
EL30	X	3.86	023250	hypothetical	9	6	4	0	1	1	X	3.69	1	ND
EL36	X	7.09	015770	ROP2/8	6	6	6	1	1	2	NA	NA	0	ND
EL37	X	7.27	007010	SRS48	5	5	5	0	1	1	X	6.75	2	ND
EL45	XI	6.57	098560	hypothetical	4	6	7	2	1	3	XI	6.08	8	6-7
EL47	XII	0.54	108080	ROP5	3	14	4	1	1	1	XII	0.35	2	8-9
EL51	XII	5.67	051960	SRS59K	3	3	7	1	1	1	XII	5.32	1	2-3
EL52	XII	6.66	077270	NTPase II	3	3	4	0	1	1	XII	6.24	1	2-3
EL53	XII	6.68	077240	NTPase I	2	2	2	1	1	1	XII	6.29	1	4-5

*Determined using raw sequence read coverage. Type I: GT1; Type II: ME49; Type III: VEG.
†Gene IDs based on *Toxoplasma gondii* strain ME49 sequence, v7.0 (www.toxodb.org).
‡Determined using raw sequence read coverage for *N. caninum* Liverpool strain.
§Determined using raw sequence read coverage for *H. hammondi* CatGER041 strain.

Table 2.2: Bioinformatic properties of duplicated/expanded loci in *Toxoplasma gondii* and *Neospora caninum*

Property	*T. gondii*					*N. caninum*				
	Duplicated genes		All genes		HGD‡ *P*-value	Duplicated genes		All genes		HGD‡ *P*-value
	With property	Total	With property	Total		With property	Total	With property	Total	
Signal peptide*	29	42	1756	8103	5.1E-11	34	45	1596	7227	2.6E-14
Single exon genes*	26	42	2121	8103	8.9E-07	31	45	1514	7227	4.7E-12
Rhoptry proteins*	5	29	47	1756	7.3E-04	2	34	30	1596	0.11
Sag-related surface antigens*	5	29	87	1756	0.010	22	34	218	1596	4.1E-12
Developmentally Regulated: Tachyzoite/Bradyzoite†	13	42	1093	8059	0.0020					
Developmentally Regulated: Oocyst/Sporozoite†	10	42	1337	8059	0.070					

*Based on version 8.2 annotation, toxodb.org; †Based on analysis of complete life cycle transcriptional profile of *T. gondii* strain M4 (Fritz *et al*., 2012).
‡Hypergeometric distribution.

Table 2.3: Mapping statistics of raw reads for *T. gondii* and *N. caninum*

				Sequence coverage	
Species/ Strain	Total Reads	Number of Mapping Reads	Percent mapping	Median	Standard Deviation
T. gondii GT1	707774	696748	98.4	15	171.8
T. gondii ME49	980747	932597	95.1	19	91.0
T. gondii VEG	676028	659342	97.5	14	147.0
N. caninum LIV	551609	529936	96.1	21	665.0

3.0 THE EXPANDED GENE CLUSTER, MAF1, WHICH MEDIATES HOST MITOCHONDRIAL ASSOCIATION IN *TOXOPLASMA GONDII*, IS UNDER STRONG DIVERSIFYING SELECTION

Although gene expansion has been established as a fundamental mechanism for trait evolution, it remains to be thoroughly explored in *T. gondii* and other members of the Apicomplexan phylum. Comparative analysis of gene expansion has revealed that *ROP5*, which is the most potent virulence factor known in *T. gondii*, is highly expanded in *T. gondii* relative to *Neospora caninum*. We therefore hypothesized that such differentially expanded genes will be key to explaining some of the physiological differences between *T. gondii* and its close relatives, and may play important roles in unique interactions between *T. gondii* and its host. We have identified one such differentially expanded locus, dubbed Mitochondrial Association Factor 1, (*MAF1*), as the parasite factor which mediates the close association between host mitochondria and the parasitophorous vacuole during host cell invasion. Host mitochondrial association (HMA) is indeed absent in *N. caninum*. In order to better understand the evolution of the *MAF1* locus, we have performed an extensive characterization of the locus and show that it exhibits significant inter- and intra-lineage copy-number variations among different strains of *T. gondii*. We find that the *MAF1* locus contains 2 main isoforms which have diverged in their ability to mediate HMA. Additionally, we show that the secretion kinetics of MAF1 is consistent with observations that HMA is established at the very early stages of host cell invasion. These

findings, together with results of our analysis of the topology of MAF1 at the parasitophorous vacuole, point to a mechanism of HMA whereby MAF1 is secreted prior to host invasion and gets integrated into the nascent parasitophorous vacuolar membrane with its C-terminus exposed to the host cytosol, where it directly interacts with host mitochondria during invasion.

3.1 INTRODUCTION

Apicomplexan parasites are responsible for many diseases of significant economic importance in humans and animals. *Toxoplasma gondii* belongs to this group of organisms which also includes the malaria parasite, *Plasmodium spp.*. Extraordinary success as a parasite distinguishes *T. gondii* from even its closest relatives, *Hammondia hammondi* and *Neospora caninum*, both of which have comparatively restricted host ranges with little to no pathogenic effect in humans (Goodswen *et al.*, 2013, Walzer *et al.*, 2013). *T. gondii* has the unique ability to infect all nucleated cells in an extremely wide range of intermediate hosts (Sibley, 2003), making this parasite an interesting subject of the evolutionary biology of the apicomplexan phylum. The three predominant clonal lineages of *T. gondii* (Types I, II and III) found in North America and Europe have been shown to differ in many aspects of infection including modulation of host immune-related signaling (Araujo *et al.*, 2003, Saeij *et al.*, 2007, Ong *et al.*, 2011, Rosowski *et al.*, 2011) and virulence (Saeij *et al.*, 2006, Reese *et al.*, 2011b), with a few studies suggesting lineage-related differences in disease outcome (Howe *et al.*, 1995, Grigg *et al.*, 2001b, Khan *et al.*, 2006, Yang *et al.*, 2013). How the parasite's interactions with its numerous host environments may have contributed to the evolution of such an adeptly versatile parasite remains unclear. However, it has been proposed that expansion and subsequent functional fine-tuning of

secreted effectors, such as ROP5 (and indeed the whole ROP2 superfamily) (Reese *et al.*, 2012), allows the parasite to access the right combination of paralogs best suited for a particular "new" host environment it encounters (Boothroyd, 2009). This suggests that the genome structure of *T. gondii* allows it to interact with its hosts in ways that other parasites are not able to.

The basic mode of host cell entry appears to be conserved among members of the Apicomplexan phylum with obligate intracellular life stages (Carruthers *et al.*, 2007a). *T. gondii* enters host cells encased in a host plasma membrane-derived parasitophorous vacuole (PV) which insulates the intracellular parasite from host immune surveillance systems. During or immediately following host cell entry, a very intimate association is established between the PV and host cell organelles such as mitochondria and endoplasmic reticulum (Jones *et al.*, 1972, de Melo *et al.*, 1992, Sinai *et al.*, 1997). While host mitochondrial association (HMA) is not unique to *T. gondii*, and has been extensively studied in other intracellular pathogens including *Legionella pneumophila* and *Chlamydia psittaci* (Horwitz, 1983, Matsumoto *et al.*, 1991, Scanlon *et al.*, 2004), its biological relevance in *T. gondii* and these other pathogens remains unclear partly due to the lack of understanding of the molecular mechanisms involved. In *T. gondii,* HMA has mainly been proposed as a crucial route by which the vacuole-isolated parasite acquires essential metabolites from the host cell (Nakaar *et al.*, 2003, Crawford *et al.*, 2006, Romano *et al.*, 2013). However, recent findings that HMA is a strain-specific trait in *T. gondii* and absent altogether in the closely related *N. caninum* (Pernas *et al.*, 2010) indicates that this trait is dispensable and that its proposed role in nutrient acquisition is not essential. The lack of definition of the nature of the molecular interactions involved in HMA has stalled our understanding of the full impact of this remarkable depiction of host-parasite interaction (Sinai *et al.*, 2001, Pernas *et al.*, 2010).

We recently reported the identification of a tandemly duplicated gene cluster, dubbed Mitochondrial Association Factor 1 (*MAF1*), as the parasite factor which mediates HMA in *T. gondii* (Pernas *et al.*, in revision). The *MAF1* locus was initially identified among the list of genomic loci that are uniquely expanded in *T. gondii* relative to *N. caninum*. *MAF1* encodes a dense granule protein whose expression profile is consistent with the strain-specific nature of the HMA phenotype (no detectable protein in Type II strains). We showed that this gene was necessary and sufficient to induce HMA in a Type II strain, which is naturally devoid of this trait. Additionally, we observed that this interaction had a significant impact on host immune signaling, expanding the possible roles of HMA in *T. gondii* infection and thus opening up a new paradigm for thinking about how *T. gondii* has evolved to interact with its host cell.

In the current study, we have undertaken a detailed characterization of the expanded *MAF1* locus across the 3 *T. gondii* lineages and compared it with *N. caninum*, in order to better understand its evolution. We show that this locus exhibits copy-number variation (CNV) between lineages, as well as between strains of the same lineage, and contains paralogs which are under strong diversifying selection. Additionally, our results suggest a possible mechanism by which MAF1 mediates HMA during host cell invasion.

3.2 RESULTS

3.2.1 *MAF1* shows intra- and inter-lineage copy-number variation

We had previously used sequence read coverage analysis to show that the *MAF1* locus (*EL4*) is uniquely expanded in *T. gondii* but not in *N. caninum* (Chapter 2). Additionally, we observed

differential expansion of the *MAF1* locus between representative strains of the 3 predominant lineages of *T. gondii*. Read coverage analysis predicts gene expansion based on the relative abundance of sequence reads corresponding to a particular locus in the genome (Yoon *et al.*, 2009, Reese *et al.*, 2011b). To experimentally confirm differential expansion of the *MAF1* locus among *T. gondii* strains, we performed a high molecular weight Southern blot analysis comparing the size of the *MAF1* locus in 6 *Toxoplasma* strains. These strains comprised 2 each from the Type I (GT1, RH), Type II (ME49, PRU) and Type III (VEG, CTG) lineages. Genomic DNA isolated from each strain was digested with the ScaI restriction enzyme, which cuts on either side of the entire locus but not within, allowing for a more accurate determination of copy number (Figure 3.1A; Reese *et al.*, 2011b)). Consistent with predictions from the read coverage analysis, we observed varying band sizes ranging from ~45 kb for GT1 to ~27 kb for CTG (Figure 3.1B). Our estimates put *MAF1* copy-number at 7 in GT1 and 5 in all but CTG, which has the smallest locus size with an estimated 3 copies. Copy-numbers were determined based on an estimated repeat-unit size of ~4 kb (toxodb.org) while also accounting for intervening sequences between the ScaI sites and the limits of the repeat. Interestingly, we observed copy-number variation between strains of the same lineage (GT1 vs RH and VEG vs CTG) as previously described for 3 other expanded loci (Chapter 2). No other bands were visible on the blot, indicating that the entire locus was intact for all strains. Moreover, the ScaI sites flanking the expanded locus were conserved in all 6 strains tested (Figure 3.1C), indicating these differences are not due to mutations within the flanking sequences. For comparison, we showed in chapter 2 that the highly conserved, single-copy *AMA1* locus has equal size across all 6 strains tested for *MAF1*.

3.2.2 *MAF1* is present in two major isoforms

To better understand the structure of the *MAF1* locus and to determine the nature of its paralogs, we sequenced 6 PCR-derived *MAF1* clones from each of 3 representative strains from the Type I (RH), II (ME49) and III (CTG) lineages. As expected, we found that this locus encoded multiple diverse isoforms of the *MAF1* gene. In addition to a predicted signal peptide, each isoform was predicted to encode a single transmembrane domain located in the N-terminal region (Figure 3.2A). The major distinguishing feature between the 2 isoforms identified was the presence of a repetitive stretch of 4-7 prolines followed by a serine (P{4:7}S) in some paralogs. This motif was either completely missing or repeated up to 6 times depending on the isoform (Figure 3.2B). For RH, 5 of the 6 clones contained the P{4:7}S motif, while all 6 CTG clones had some form of the repeat motif. Interestingly, of the 6 clones sequenced from ME49, 3 had premature stop codons, and all 3 of these clones were predicted to encode MAF1 isoforms with the P{4:7}S motif. Additionally, 2 other clones which encoded isoforms containing P{4:7}S also contained a 66bp insertion in their 5' UTR. This may be due to a relatively stronger selective pressure to silence the P{4:7}S motif-containing isoforms (referred to hereafter as *MAF1A*) in the Type II strain. Incidentally, *N. caninum* contains a single divergent ortholog of *MAF1* which also lacks any form of the repeat P{4:7}S motif (Figure 3.2B,C).

In order to determine whether copy-number differences and sequence polymorphisms were reflected in protein expression, we performed western blot analysis comparing MAF1 expression in the 6 strains examined in the southern blot analysis. We used a polyclonal anti-MAF1 antibody raised against a canonical *MAF1* paralog from GT1, TGGT1_053770 (Pernas *et al.*, in revision). We observed MAF1 expression in all Type I and III strains tested but not in either of the Type II strains (Figure 3.3A). The Type I strains had similar levels of MAF1

expression between them as did the Type III strains. While this result was consistent with the strain-specificity of the HMA phenotype (Figure 3.4A), it did not reflect the general trend of copy-number variation at the locus. There were no detectable intra-lineage expression differences between the Type I and Type III strains. However, we observed approximately 2-fold higher expression in Type III strains relative to Type I strains (Figure 3.3B). This observation is interesting in light of earlier findings that Type I strains have about twice as much PVM surface area associated with host mitochondria as do Type III strains (Pernas *et al.*, in revision). It is also interesting to note that MAF1 proteins from Type III strains appeared to have a slightly higher molecular weight compared to that of Type I strains. This is consistent with the observation that clones of *MAF1A* sequenced from the CTG strain contained the highest number of the P{4:7}S repeat motif (6 repeats, Figure 3.2B). The highest number of P{4:7}S repeat motifs in a *MAF1A* isoform from the Type I strain was 3. Overall, these results confirm a strong evolutionary selection on this locus which has led to differential expansion, sequence polymorphisms and strain-specific expression difference.

3.2.3 MAF1A and B isoforms differ in their ability to mediate host mitochondrial association

HMA is present as a strain-specific trait in *T. gondii* (lacking in Type II stains) and absent altogether in the closely related *N. caninum* (Figure 3.4, Pernas *et al.*, 2010). Our data indicate that the *MAF1* locus, which is responsible for this phenotype, contains different isoforms which appear to have some significant sequence polymorphisms. In order to test the effect of such polymorphisms on the ability of MAF1 to mediate HMA, we generated HA-tagged clones of the 2 major isoforms, *MAF1A* and *MAF1B* from a Type I strain (GT1), and expressed them in a

Type II strain (TgME49) and also in *N. caninum*. We found that MAF1A was sufficient to mediate HMA in TgME49 and, even more impressively, in *N. caninum* (Figure 3.5). MAF1B on the other hand was unable to mediate HMA in either TgME49 or *N. caninum* although its secretion and localization profiles were similar to that of MAF1A (Figure 3.5). These results indicate that the MAF1 isoforms have significantly diverged in their function upon expansion, as it relates to HMA. To further probe the relevance of the P{4:7}S repeats found exclusively in the MAF1A isoform, we generated *MAF1A* constructs where either all the prolines or the serines in the repeat region were replace with alanines. We found that substituting the repeat residues had no impact on the ability of MAF1A to mediate HMA (Figure 3.6) indicating that other amino acids or structural differences between the two isoforms may be more important for HMA.

3.2.4 HMA in Type II strains has no detectable impact on *in vivo* proliferation and dissemination in mice

We previously showed that MAF1-induced HMA in Type II strains is accompanied by a significant increase in expression of proinflammatory cytokines during *in vitro* MEFs infections. In order to directly examine the impact of MAF1 in an *in vivo* infection system, we infected BALB/C mice with TgME49 wild-type or a MAF1A-complemented line and measured their rates of proliferation and dissemination *in vivo* using bioluminescence imaging (Walzer *et al.*, 2013). We used a sub-lethal dose for a Type II strain (100 parasites) (Saeij *et al.*, 2005b) to allow the mice to survive the full course of the infection and enable us to detect more subtle difference in infection. We observed marginally higher, but statistically insignificant, parasite burdens in infection with TgME49:MAF1A (Figure 3.7A). Results from these experiments showed that

there were no significant observable differences in parasite proliferation and dissemination between the two parasite lines over the course of infection (Figure 3.7B).

We also examined cytokine induction during the course of mouse infection. For this experiment we infected mice with 10^5 parasites in order to fully engage the host immune response and collected tail-blood from infected mice at 48-hour time-points to measure changes in cytokine levels. We sampled 26 cytokines ranging from early pro-inflammatory cytokines such as interleukin 12 (IL12) and interferon gamma (IFN-γ) to late anti-inflammatory cytokines including IL10 and transforming growth factor beta (TGF-β) (Miller *et al.*, 2009, Pifer *et al.*, 2011). Results from these experiments collectively indicate that MAF1A, either through HMA or any other mechanism, does not significantly alter parasite proliferation and dissemination or cytokine induction in mice infections (Figure 3.8). This is in contrast to previous observations where loss of MAF1 in a Type I strain led to decreased levels of several pro-inflammatory cytokines *in vivo* (Pernas et al., in revision).

3.2.5 MAF1 is secreted by extracellular parasites and integrates into the nascent PVM with its C-terminus exposed to the host cytosol

The canonical model for secretion of dense granule proteins suggests that secretion generally occurs post-invasion (Carruthers *et al.*, 1997). Since HMA has been shown to occur during or immediately after invasion (Sinai *et al.*, 2001), it is important to reconcile the kinetics of host mitochondrial association with that of secretion of MAF1, a dense granule protein (Pernas *et al.*, in revision), by determining whether MAF1 secretion begins before, during or after invasion. We therefore tested the hypothesis that MAF1 secretion begins prior to or during invasion. Treatment of extracellular parasites with Cytochalasin-D (CytoD) inactivates the actin/myosin motor

complex, which provides the motile force required to drive host cell penetration, and therefore prevents invasion (Ryning *et al.*, 1978, Dobrowolski *et al.*, 1996, Carruthers *et al.*, 2007a). However, these parasites are still able to engage host cells in productive attachment characteristic of invasion. This attachment is associated with secretion of factors such as MICs (microneme proteins) and RONs and ROPs (rhoptry proteins) which are known to be required for the early stages of invasion (Carruthers *et al.*, 2007a). To determine whether MAF1 is secreted by extracellular parasites prior to invasion, TgME49:HA-MAF1A$_I$ parasites were pretreated with Cyto-D and added to HFFs for 10 min (Dunn *et al.*, 2008). To identify extracellular parasites, fixed cells were stained with anti-SAG1 before permeabilization for MAF1 staining. Anti-HA staining for MAF1 confirmed that indeed, secretion begins prior to invasion. MAF1 was detected in the host cell near the apical end of the extracellular parasite presumably at the point of attachment (Figure 3.9).

MAF1 is predicted to contain a single-pass transmembrane domain covering amino acid 97-116 (Figure 3.2A). In order to begin understanding the mechanism by which MAF1 mediates HMA we determined its topology at the PVM. Based on our observations that the addition of a C-terminal HA-tag inhibits the ability of MAF1 to mediate HMA even though PVM localization is not affected (data not shown), we predicted that the C-terminal domain is important for the interaction with host mitochondria and would be exposed to the host cell cytosol. To determine the topology of MAF1at the PVM, cells were infected with the TgME49:HA-MAF1A$_I$ parasite line expressing an N-terminally HA-tagged MAF1A, and grown for 24hrs followed by permeabilization with digitonin. Low concentrations of digitonin selectively permeabilize the host cell plasma membrane but not the PVM (Beckers *et al.*, 1994, Dunn *et al.*, 2008). Anti-SAG1 staining was used to detect parasite surface and verify selective permeabilization. The

HA-tagged N-terminus of MAF1A was only detected on vacuoles that also stained for SAG1 (Figure 3.10). Together, these results indicate that the N-terminus of MAF1 faces the vacuolar lumen and is only accessible to antibodies when the PVM is permeabilized, and suggest but do not prove, that the C-terminus is exposed to the host cell cytosol.

3.3 DISCUSSION

One of the many ways in which *T. gondii* interacts differently with host cells, compared to its immediate relatives such as *Neospora caninum*, is the close association formed between the PVM of *Toxoplasma* and the host cell mitochondria upon invasion. We recently identified a tandemly duplicated gene cluster, *MAF1*, as the parasite factor that mediates HMA and showed that this interaction represents an important means by which *Toxoplasma gondii* interfaces with the host immune system. We show here that *MAF1* has undergone significant expansion and varies in copy-number as well as expression among the predominant *T. gondii* lineages. The importance of gene expansion in *T. gondii* is typified by the virulence factor ROP5 which is differentially expanded among the different *T. gondii* lineages and has isoforms which vary in sequence and contribute disproportionately to the virulence phenotype association with that locus (Behnke *et al.*, 2011, Reese *et al.*, 2011b). *ROP38* is another example of a tandemly duplicated gene family encoding a secreted protein which also affects host cell function (Peixoto *et al.*, 2010). Much like the *ROP5* and *ROP38* loci, our data indicate that the 2 main isoforms of *MAF1* (A and B) identified here also differ in sequence and in function. While the importance of secreted parasite effectors in establishing a successful infection cannot be overemphasized, the selective duplication of genes encoding secreted proteins in *Toxoplasma,* (and perhaps other

intracellular parasites, (Chapter 2)) supports an attractive model of evolution where gene expansion provides the raw materials needed to develop a finely tuned set of host-parasite effectors with the versatility required to productively survive the wide variety of host environments encountered by the parasite (Boothroyd, 2009).

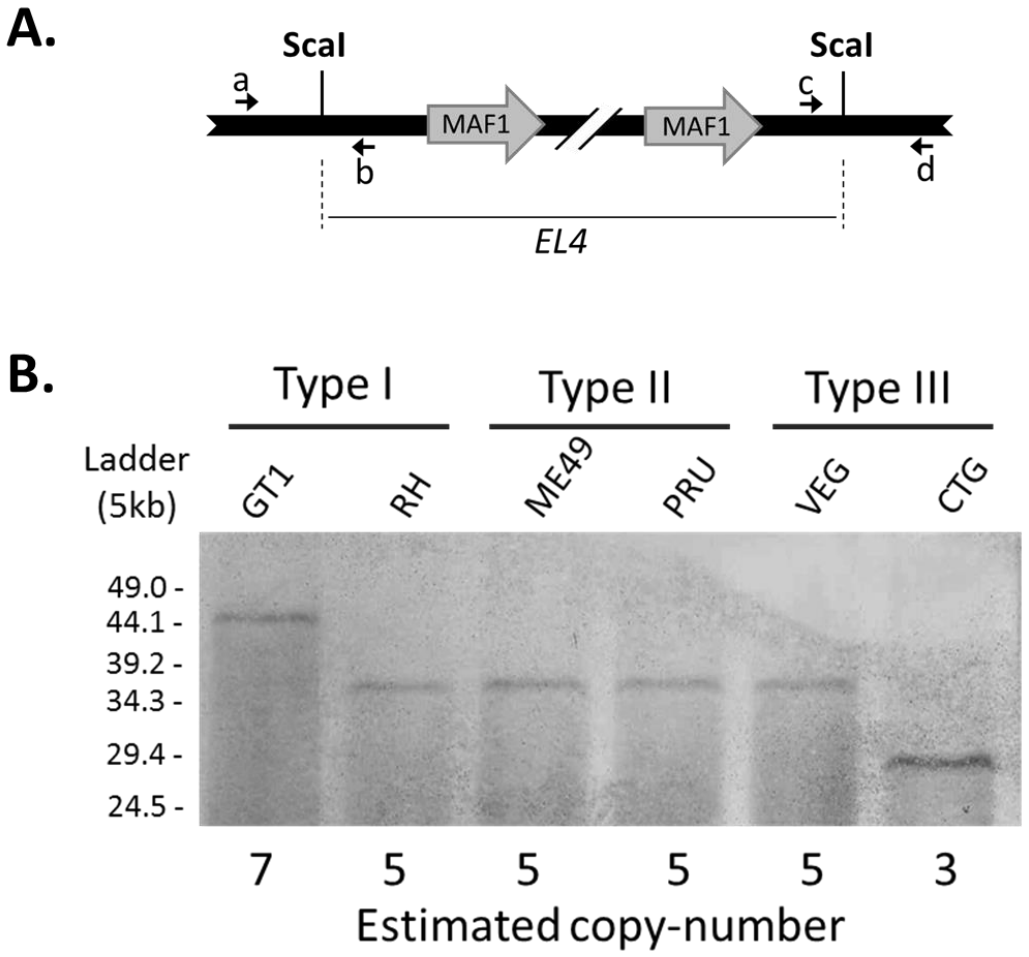
The role of HMA in a vacuole-isolated intracellular organism is very easily predictable based on popular understanding of the functions of mitochondria as the hub of metabolism, and the inherent nutritional limitations of intracellular existence. In *T. gondii* however, such a direct reliance on host mitochondria for metabolic supplies is dispensable, at least in some strains (Pernas *et al.*, in revision). Just like virulence and other immune related phenotypes, the strain-specificity of HMA is yet another example of the flexibility of *Toxoplasma* in interacting with its host. Interestingly, *N. caninum* lacks HMA and carries only a single ortholog of *MAF1* which would be considered a B isoform, suggesting that this trait was acquired later in the evolution of *T. gondii* and might have coincided with the surge in host range expansion. We did not observe any significant MAF1-dependent alteration in virulence either through proliferation and dissemination or through mortality. Moreover, although deletion of the MAF1 locus in a Type I strain abolished HMA, it did not alter virulence either (Pernas *et al.*, in revision). This was not surprising as previous virulence quantitative trait loci (QTL) mapping analyses failed to identify a significant virulence QTL on chromosome II (Saeij *et al.*, 2006, Taylor *et al.*, 2006). What was surprising was the observation in the present study that *in vivo* cytokine induction was not altered in a MAF1-dependent manner. Previous *in vivo* experiments had demonstrated a link between MAF1 expression and production of diverse cytokines. However, those experiments were performed with a *Δmaf1* Type I strain, and cytokine levels were measured in peritoneal exudate cells. With the notion that Type I strains generate a profoundly different immunological response

compared with Type II strains (Robben *et al.*, 2004a, Kim *et al.*, 2006), it is possible that HMA in the Type I background is important for altering host cytokine response while expressing MAF1 and inducing HMA in the Type II background does not have the same impact.

Whether as a route for acquiring metabolites from host cells or as mechanism for modulating host immune signaling, it is conceivable that the vastly adept parasitic system of *Toxoplasma* contains built-in redundancies which allow the parasite to assemble manipulative strategies that are targeted to specific hosts. For example, the IRG pathway targeted for neutralization by ROP5 is missing in humans and other systems where *Toxoplasma* is equally capable of surviving (Niedelman *et al.*, 2012). This suggests that in different host contexts, different sets of effectors will be required for analogous functions, and that ROP5 may be utilized in a different capacity in the different hosts. MAF1-mediated HMA may also follow a similar model. It may carry different degrees of relevance in different hosts and its impact on host cell function, and by extension the parasite's ability to establish a successful infection may be as much a property of the host environment as it is the genotype of the parasite.

Given the well-recognized temporal hierarchy of organelle discharge, and the fact that HMA occurs as early as 1 minute post-invasion, it was initially proposed that the expected kinetics of secretion of the parasite factor involved in this process matched the profile of a rhoptry protein (Carruthers *et al.*, 1997, Sinai *et al.*, 1997, Sinai *et al.*, 2001). However, GRA7 has recently been shown to be secreted into the host cell prior to invasion (Dunn *et al.*, 2008) in a manner nearly identical to MAF1. This contradicts the canonical sequential model for *T. gondii* protein secretion which suggests that dense granule proteins are secreted post invasion. Our findings show that MAF1 is another exception to the rule. The kinetics of MAF1 secretion is consistent with its role as the mediator of HMA. These findings, together with results from

topological analysis of MAF1 relative to the PVM, support a model for MAF1-mediated HMA whereby MAF1 is secreted during invasion and is anchored on the PVM during the formation of the PV. With its C-terminus possibly exposed to the host cytosolic side of the nascent PVM, MAF1 directly interacts with and docks host mitochondria at the surface of the PV. Indeed preliminary results from video microscopy suggest that as the PV is formed at the host plasma membrane and migrates to the perinuclear region, it encounters the abundant host mitochondria which then become physically attached to the PVM, presumably through interaction with MAF1. While it appears that HMA is not an active recruitment process, further modification of MAF1 in the host cell is possible (Treeck *et al.*, 2011) and may be important for association. Residues S394, S397 and S407 in the C-terminus are predicted phosphorylation sites which may be important for MAF1 function. It also remains possible that other PVM- resident parasite proteins play auxiliary roles to fortify this interaction.


A.
ScaI
ScaI
a
b
c
d
MAF1
MAF1
EL4
B.
Type I
Type II
Type III
Ladder (5kb)
GT1
RH
ME49
PRU
VEG
CTG
49.0 -
44.1 -
39.2 -
34.3 -
29.4 -
24.5 -
7
5
5
5
5
3
Estimated copy-number

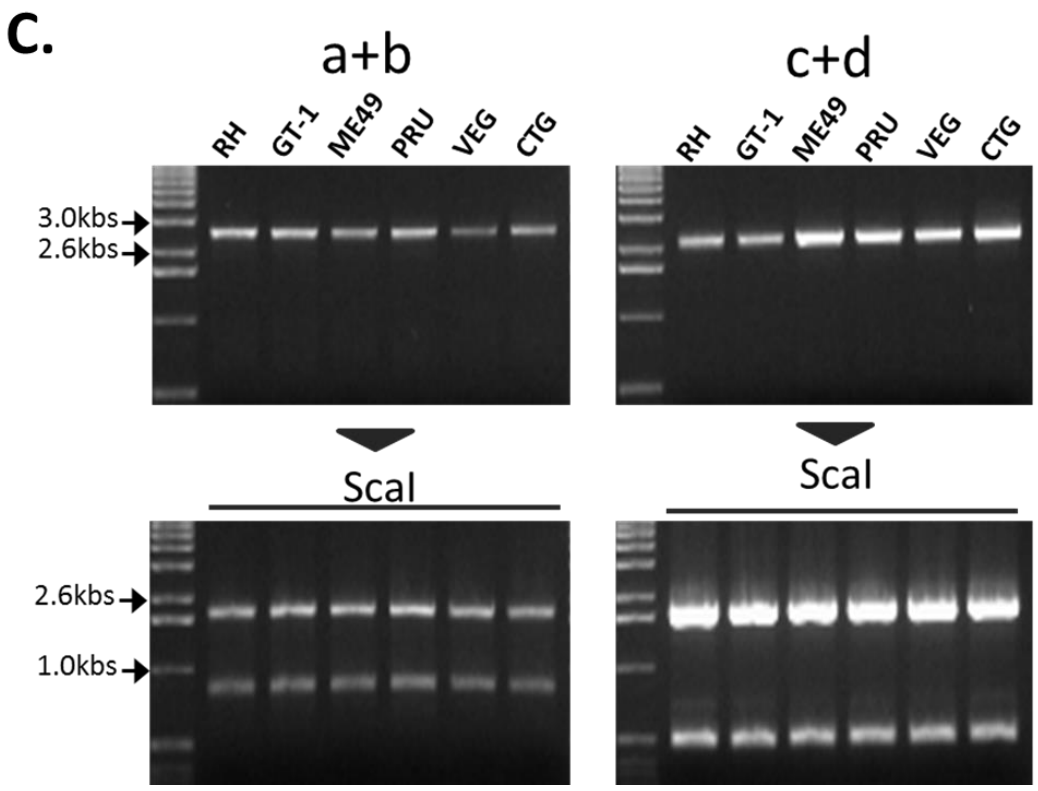


Figure 3.1: The *MAF1* locus is differentially expanded among different strains of *T. gondii.*

(A) Schematic representation of the *MAF1* locus showing ScaI restrictions sites used to determine locus size. a, b, c and d represent primers used to verify location of the ScaI restriction sites. (B) ScaI-digested gDNA from each strain was resolved by PFGE and probed with MAF1-specific probes. The blot shows copy-number variation consistent with predictions from sequence coverage analysis (Pernas *et al.*, in revision). Copy-number for each strain was determined based on the estimated repeat-unit size of ~4 kb. (C) PCR-based diagnostic digest was performed to confirm that the ScaI sites were present at the same predicted locations in all strains.

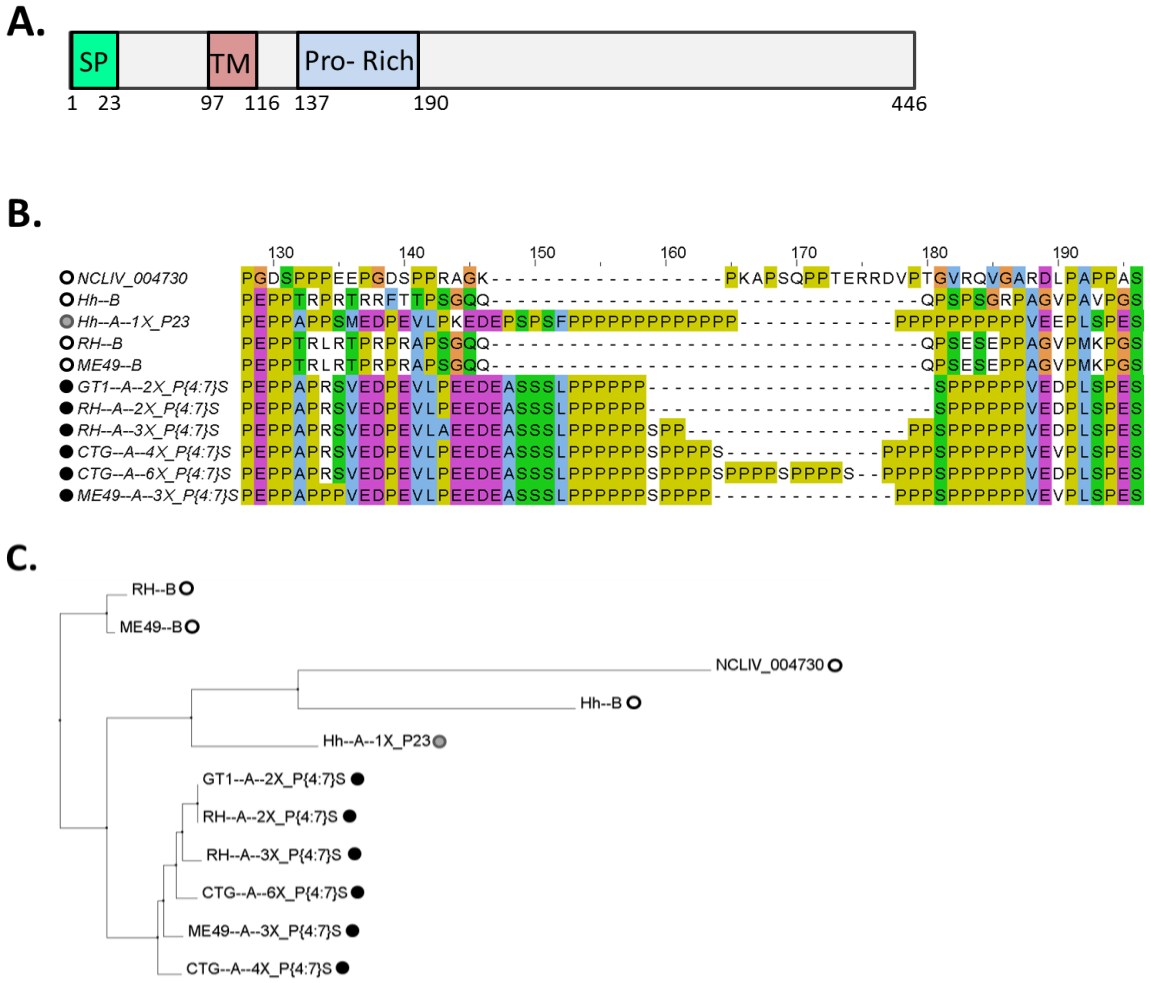


Figure 3.2: The *T. gondii* MAF1 locus harbors two distinct isoforms

(A) Schematic representation of MAF1. Signal peptide (SP) was predicted by SignalP v4.0, transmembrane domain (TM) was predicted by TMHMM v2.0, TM-transmembrane. The proline-rich region (Pro-Rich) stretches from AA137-190. (B) ClustalΩ alignment of the polymorphic proline-rich region for *MAF1* sequences from *T. gondii*, *H. hammondi* and *N. caninum*. GT1--A--2X_P{4:7}S (TGGT1_053770) and NCLIV_004730 were obtained from toxodb.org; all other sequences were obtained from randomly selected clones from PCR amplification of the *MAF1* locus in *T. gondii* and *H. hammondi*. The alignment shows two isoforms (A and B) mainly distinguished by repeats of 4 to 7 prolines followed by a serine, which are present in A isoforms but not in B isoforms. *H. hammondi* encodes an A isoform which contains a stretch of 23 prolines in this region, without serines. (C) Phylogram of the MAF1 amino acid sequences from *T. gondii*, *H. hammondi* and *N. caninum*. ●isoform A, ○isoform B

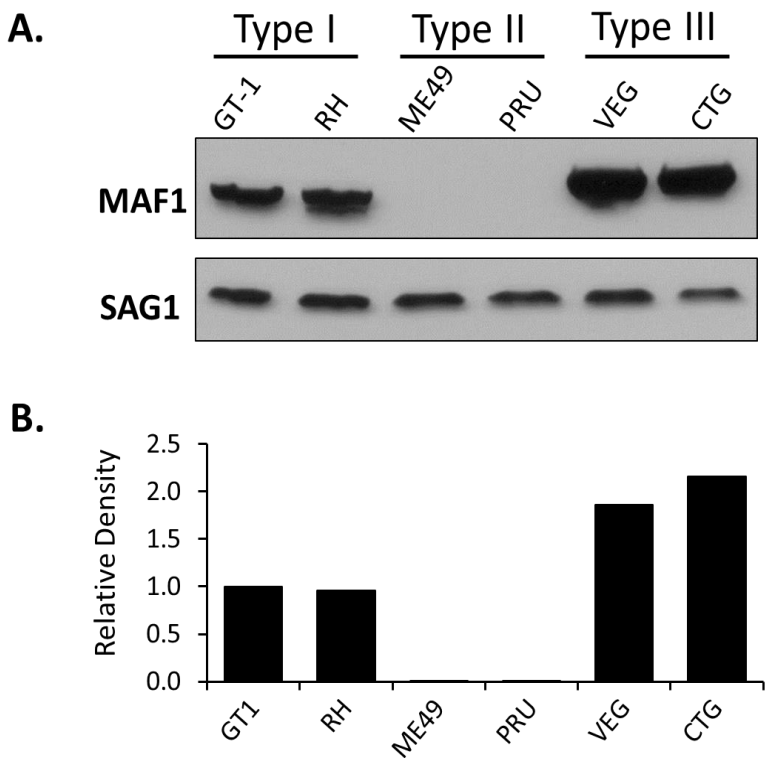


Figure 3.3: *T. gondii* MAF1 expression differs between lineages.

Levels of MAF1 protein were compared among 2 strains each from the 3 predominant lineages of *T. gondii* using polyclonal antibodies against C-terminus of TGGT1_053770. Expression polymorphism of MAF1 correlates with strain-specificity of the host mitochondrial association phenotype. SAG1 is used as a loading control. (B) Densitometric analysis of relative levels of MAF1 in the six strains examined. GT1 was set to 1.

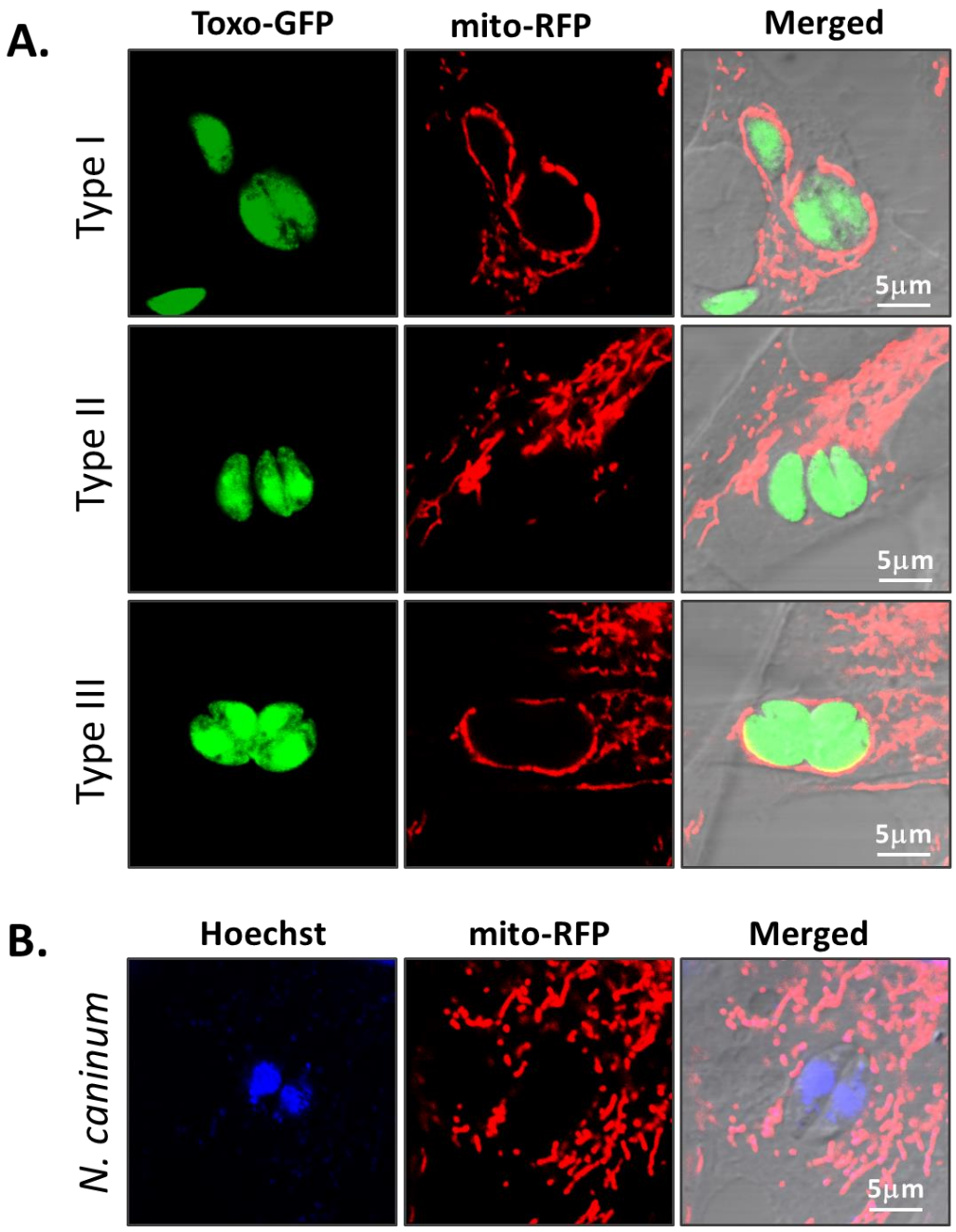


Figure 3.4: Host Mitochondrial Association is a strain-specific phenotype in *T. gondii* and absent in *N. caninum*.

(A) NRK-mitoRFP cells were infected with GFP-expressing *T. gondii* strains. Type II strains are HMA- while Type I and III strains are HMA+. (B) NRK-mitoRFP cells were infected with *N. caninum* strain NC-1. Wild-type *N. caninum* are HMA-. Cells were fixed and counter-stained with Hoechst stain (blue). Scale bar, 5 µm

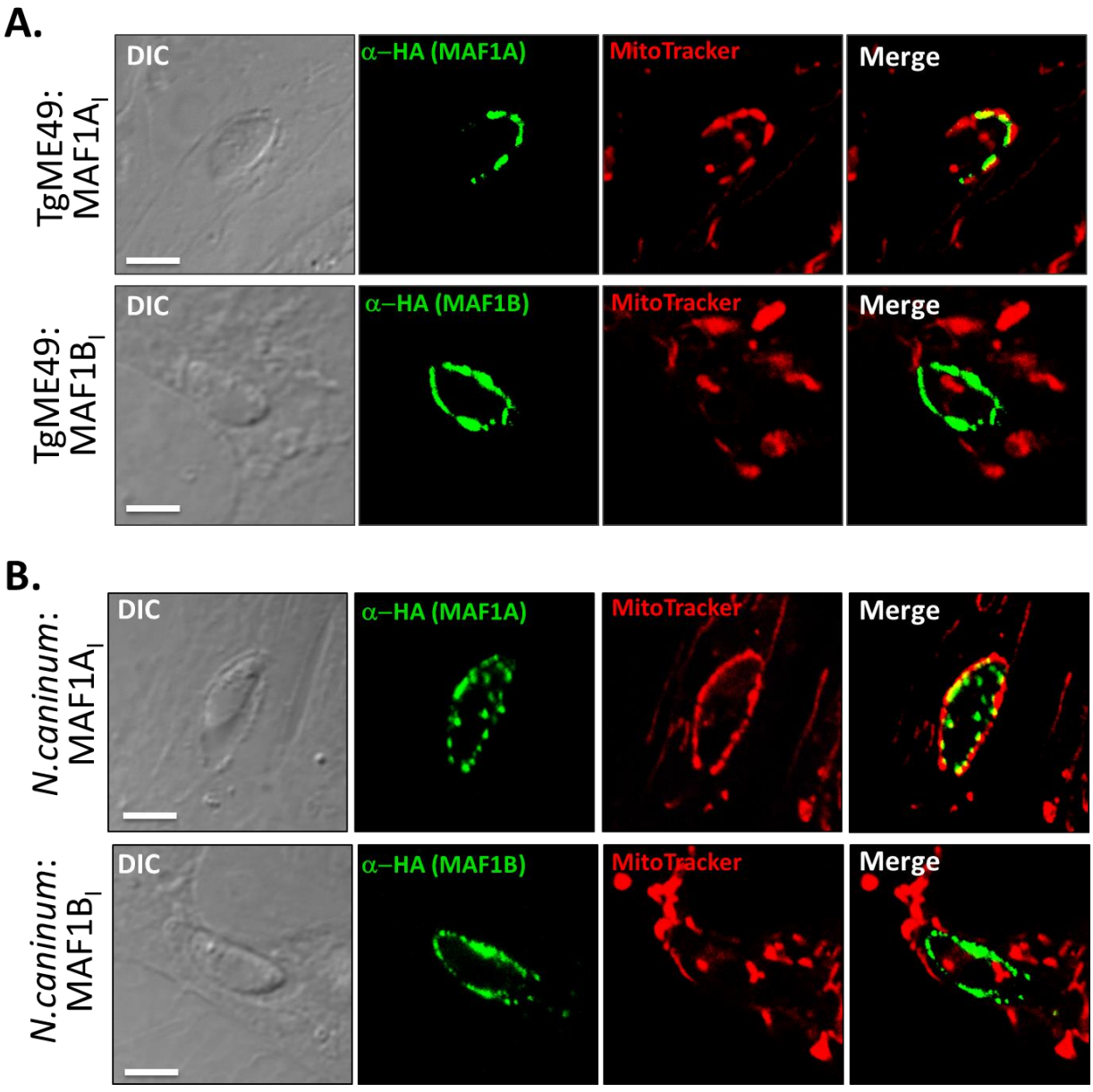


Figure 3.5: MAF1A and B differ in their abilities to complement HMA in *T. gondii* or *N. caninum*.

HFFs were labeled with MitoTracker and infected with parasites ectopically expressing HA-MAF1A or HA-MAF1B cloned from the Type I *T. gondii* strain, GT1. MAF1A but not MAF1B is able to induce HMA in TgME49 (A) and also in *N. caninum* (B). Scale bar, 5.0 μm.

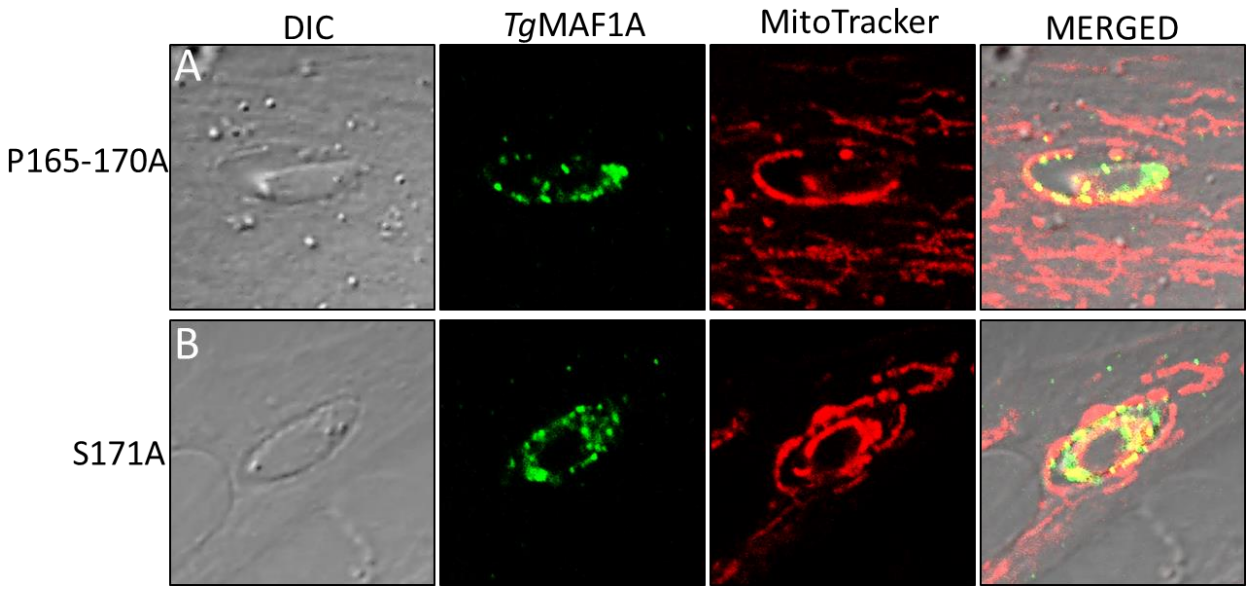


Figure 3.6: Substitution mutations in the poly-proline motif of MAF1A do not affect ability to mediate HMA. Amino acids constituting the P{4:7}S motif in a MAF1A isoform (TGGT1_053770) were replaced with either alanines or glycines and examined for mitochondrial association. A) Substitution of the 12 prolines for alanines. B) Substitution of serine 171 for an alanine.

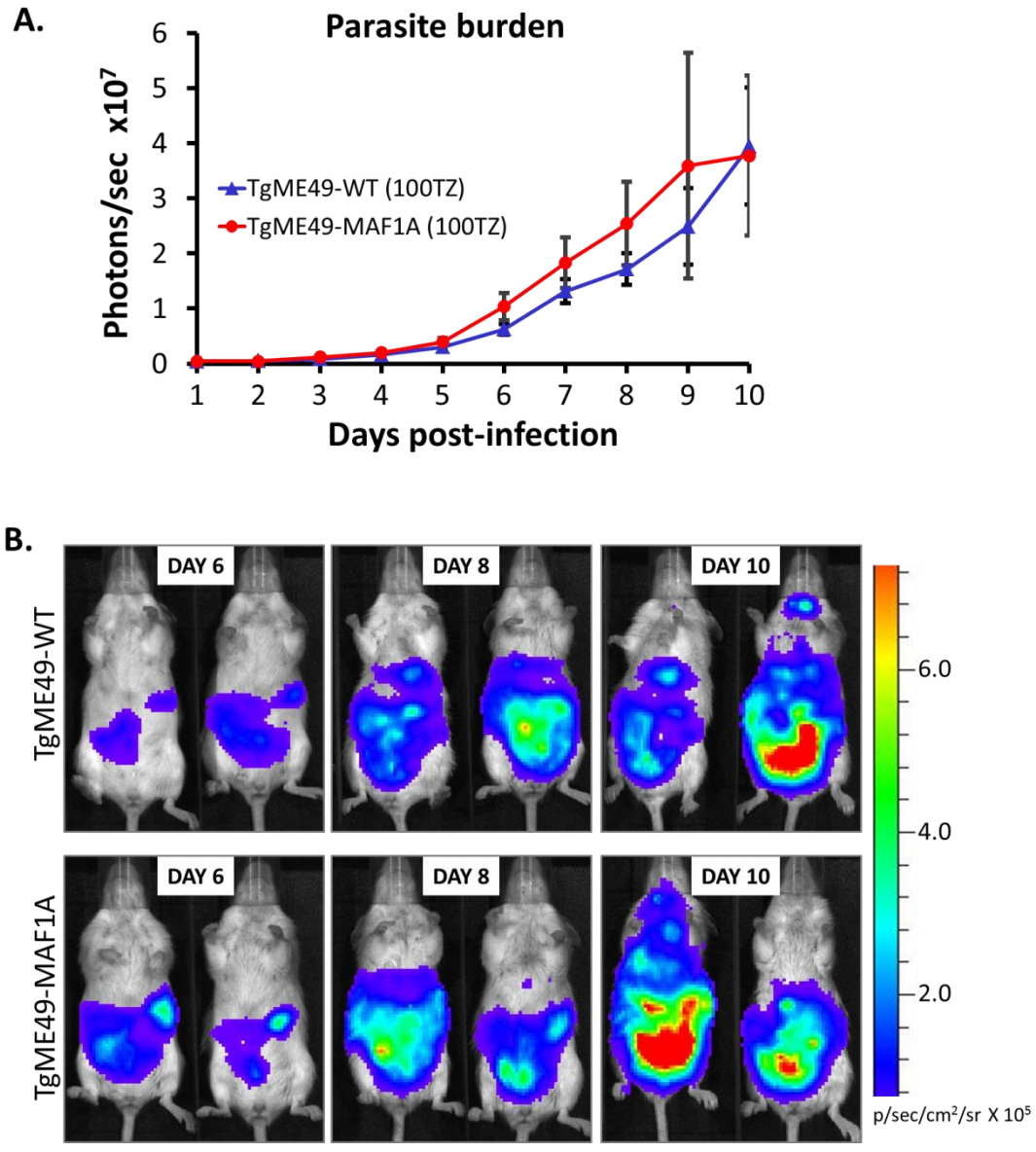


Figure 3.7: HMA has no detectable impact on virulence *in vivo*.

Balb/C mice were infected with 100 tachyzoites of luciferase-expressing TgME49 complemented with MAF1A from Type I GT1 (TgME49-MAF1A, n=5) or empty vector (TgME49-WT, n=4). Infections were performed by intraperitoneal (IP) injections and parasite burden measured by bioluminescence imaging (BLI) at 24-hour time-points. (A) Quantitation of parasite burden by BLI. (B) Representative images of parasite burdens in mice at days 6, 8 and 10 post-infection.

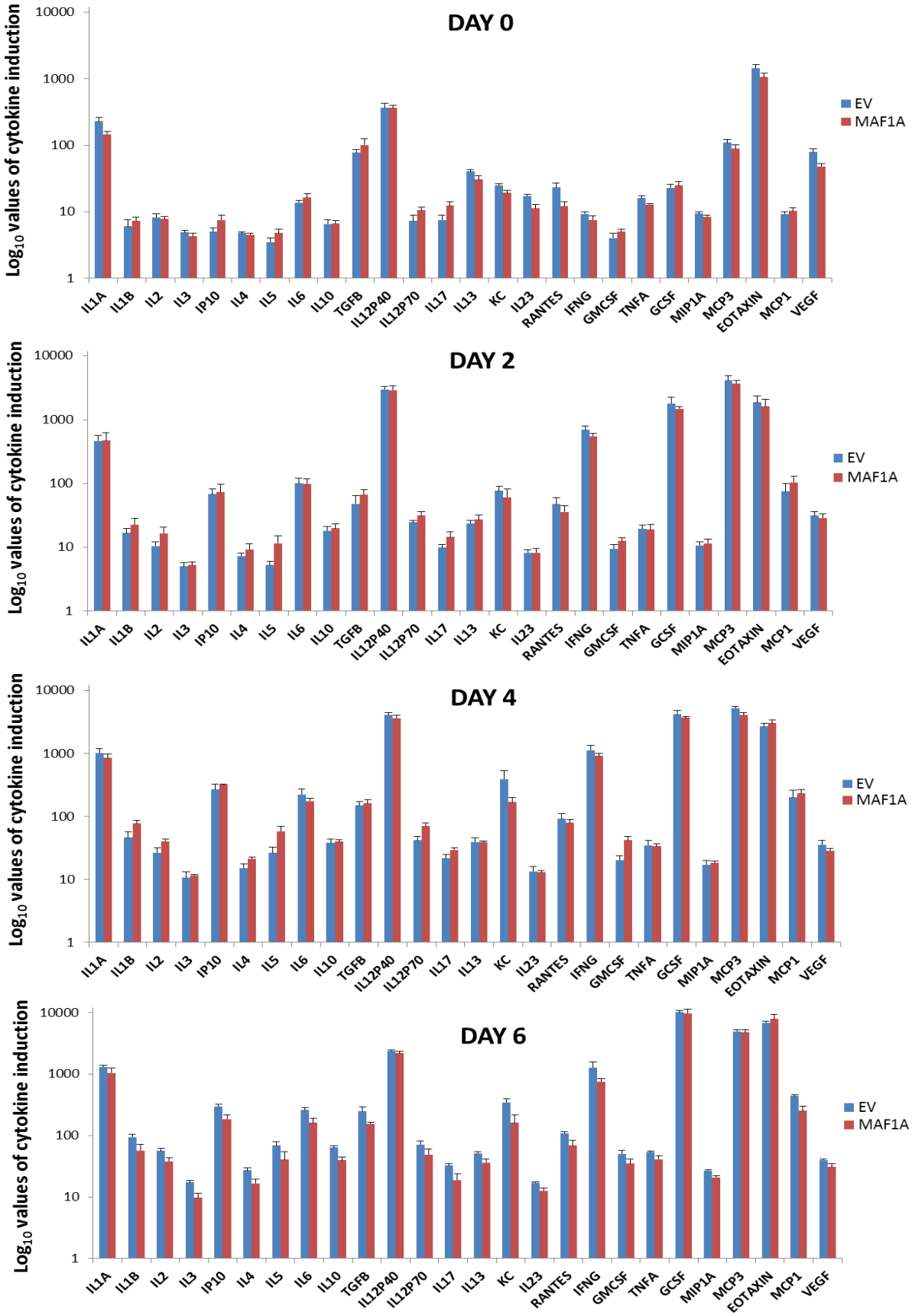

DAY 0
DAY 2
DAY 4
DAY 6
Log10 values of cytokine induction
10000
1000
100
10
1
EV
MAF1A
IL1A
IL1B
IL2
IL3
IP10
IL4
IL5
IL6
IL10
TGFB
IL12P40
IL12P70
IL17
IL13
KC
IL23
RANTES
IFNG
GMCSF
TNFA
GCSF
MIP1A
MCP3
EOTAXIN
MCP1
VEGF

Figure 3.8: HMA in a Type II strain does not impact cytokine response *in vivo*.

Balb/C mice (n=4 per strain) were infected with 10^5 tachyzoites of the Type II strain complemented with MAF1A (MAF1A) or the empty vector control (EV) via intraperitoneal injection. Serum from tail-blood collected at 48-hour intervals were frozen down and thawed immediately prior to measuring cytokine levels using the Luminex mouse 26plex (Millipore, USA) assay. Error bars represent standard error of the mean of serum from 4 mice.

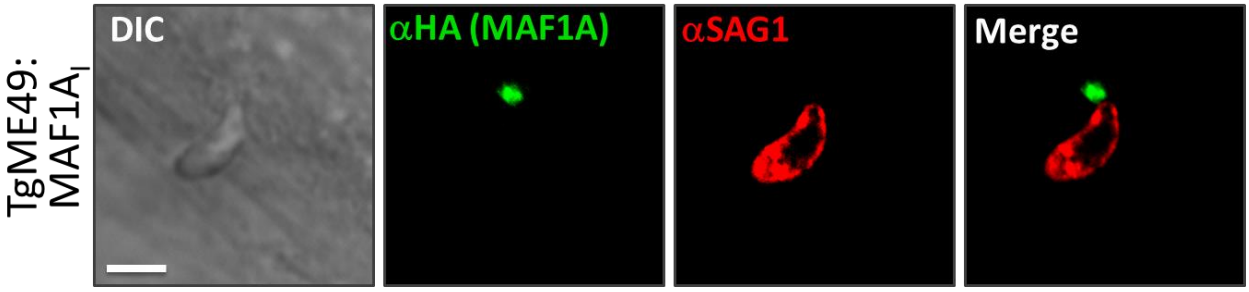


Figure 3.9: MAF1 is secreted by extracellular parasites.

Cytochalasin-D arrested TgME49:HA-MAF1A parasites were added to HFFs for 10 min and stained for MAF1. Extracellular parasites were identified by anti-SAG1 staining prior to permeabilization of the host cells. Scale bar, 5 μm

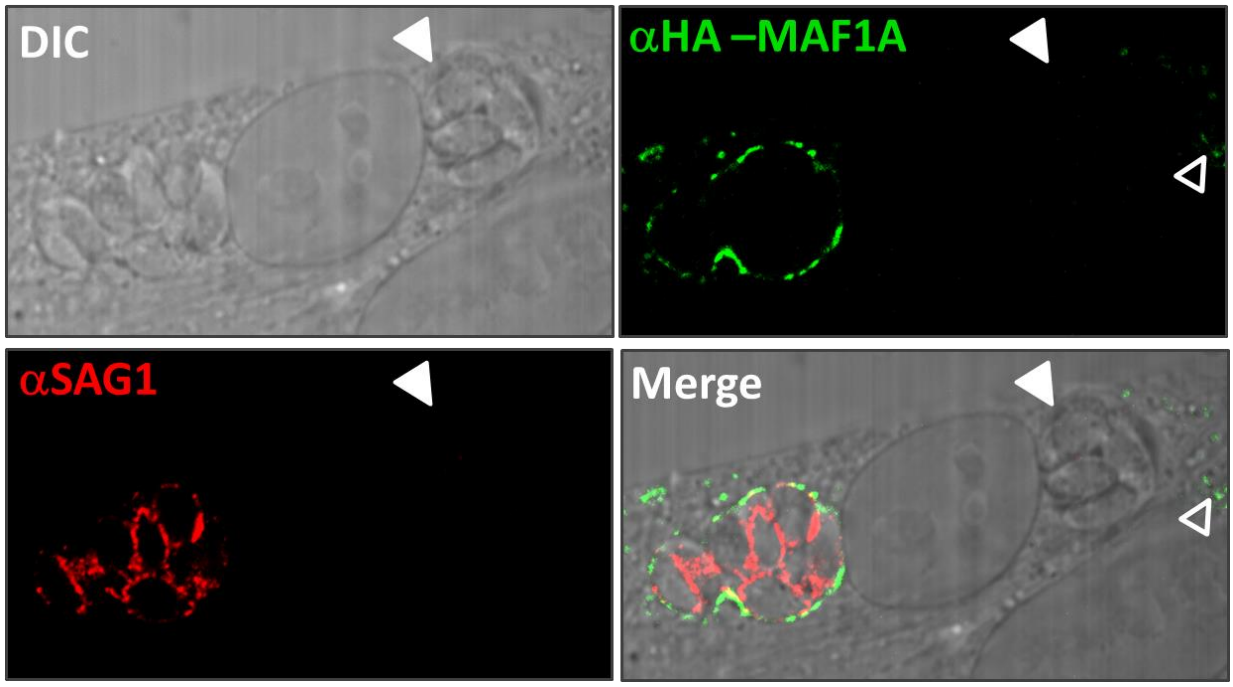


Figure 3.10: The C-terminus of MAF1 is exposed to the host cytosolic side of PVM.

Host cells infected with TgME49:HA-MAF1A parasites were treated with 0.002% Digitonin which selectively permeabiilizes host cell membranes leaving some PVMs intact. Intact vs permeabilized PVM were distinguished by staining the parasite surface protein SAG1. The N-terminal HA-tag of MAF1A was only detectable in vacuoles which also stained for SAG1 indicating those vacuoles were permeabilized. Intact vacuoles (filled arrowheads) did not stain for MAF1 although cytosolic vacuole staining of MAF1 were observed (open arrowheads)

4.0 CHARACTERIZATION OF A DEVELOPMENTALLY REGULATED, TANDEMLY DUPLICATED CLUSTER OF RHOPTRY PROTEINS

Differentiation between the tachyzoite and bradyzoite stages is a major component of *Toxoplasma gondii* pathogenesis and understanding the molecular mechanisms involved is important for the development of effective therapies. While it has been widely demonstrated that bradyzoite differentiation is principally a stress-response mechanism of the parasite, the parasite factors which recognize and transmit these stress signals, as well as effectors which cause morphological and physiological changes associated with differentiation, remain largely unknown. Members of the ROP2 superfamily have been shown to play major roles in acute virulence and pathogenesis but very little is known about their impact on bradyzoite differentiation. In this section, I describe the characterization of a cluster of tandemly duplicated rhoptry proteins (*EL3*), which are induced during bradyzoite differentiation. These proteins belong to the ROP2 superfamily. We show that the *ROP42* locus exhibits differential expansion among strains of different *T. gondii* lineages as well as between members of the same lineage. Overexpression of one member of this family in tachyzoites led to decreased parasite proliferation in both *in vitro* and *in vivo* infections, and we show this phenotype is a specific effect of ROP43 misregulation in tachyzoites. These results suggest a possible role for the ROP42 cluster in the control of proliferation during differentiation.

4.1 INTRODUCTION

During *Toxoplasma gondii* infection, the initial fast growing tachyzoite stage elicits a strong interferon-γ-driven immune response from the host to which the parasite eventually succumbs, differentiating into the slow growing bradyzoite stage. Bradyzoites then become encysted and buried within cells of host tissues such as muscle and brain. Tissue cysts can also be found in visceral organs including kidney and liver (Ferreira da Silva Mda *et al.*, 2008). The ability to undergo tachyzoite-to-bradyzoite differentiation is important for 2 main reasons: 1) Being undetectable by host immune surveillance, tissue cysts can persist for the life of their hosts and allow for lateral transmission between intermediate hosts. In fact, over 60% of all new *T. gondii* infections in European pregnant women can be attributed to ingestion of inadequately prepared meat containing tissue cysts (Cook *et al.*, 2000). 2) While primary *T. gondii* infections are generally asymptomatic, some of the most severe disease outcomes associated with *T. gondii* result from reactivation of latent chronic infections (tissue cysts) in the absence of a competent host immune system. For example, toxoplasmic encephalitis, which emerged as a major opportunistic disease in HIV/AIDS patients in the 1990s, is mainly caused by reactivation of tissue cysts in the central nervous system (Hunter *et al.*, 1994, Carruthers *et al.*, 2007b). Also, toxoplasmic chorioretinitis may result from persistence or reactivation of infection in the eye (Furtado *et al.*, 2013). Recrudescence of acute infections from latent tissue cysts leads to tissue damage which manifests as the lesions associated with the various toxoplasmic diseases. This knowledge notwithstanding, the development of therapies to completely eliminate *T. gondii* in human is constrained, to a large extent, by the dearth of knowledge of the molecular basis of the bradyzoite differentiation process. How the tissue cyst stage is induced, maintained and reactivated under permitting conditions is currently unknown. Furthermore, with some human

populations recording seropositive rates above 90% (Weiss *et al.*, 2000, Montoya *et al.*, 2004), it is becoming increasingly important that the molecular interactions involved in stage differentiation be understood in order to bolster efforts to develop targeted curative therapies.

Through the efforts of many research groups, the morphologically distinct features of bradyzoites have been well documented (Dubey *et al.*, 1998, Berens *et al.*, 2004). It is also well understood, through comparative analyses of gene expression between developmental stages, that tachyzoite-to-bradyzoite differentiation is accompanied by major reprogramming of gene expression involving both transcriptional and post-transcriptional regulatory mechanisms (Singh *et al.*, 2002, Sullivan Jr *et al.*, 2009). Parasite genes whose expression is known to be upregulated during differentiation include those encoding surface antigens such as SRS9 and BAG1 (Buchholz *et al.*, 2011), metabolic enzymes such as ENO1 and LDH2 (Yang *et al.*, 1997, Coppin *et al.*, 2003) and cyst wall proteins such as BPK1 and MCP4 (Buchholz *et al.*, 2011). Indeed, general protein expression is dramatically reduced at the onset of differentiation and this is believed to allow the parasite to shut down expression of tachyzoite-specific proteins in favor of bradyzoite-specific proteins (Sullivan Jr *et al.*, 2009). Such dramatic changes in gene expression are consistent with the substantial morphological and physiological changes that mark differentiation. Importantly, several stress responsive genes are also induced during differentiation suggesting that extracellular or environmental stress is a key trigger for differentiation (Weiss *et al.*, 1998, Echeverria *et al.*, 2005). These exogenous stress factors may be in the form of pro-inflammatory cytokines, nutrient deprivation or high environmental pH and temperature. These findings form the bases for several stress-related bradyzoite induction assays employed in both *in vitro* and *in vivo* studies of differentiation (Ferreira da Silva Mda *et al.*, 2008).

The fact that bradyzoite differentiation is generally considered as a response to environmental stress suggests that sensors and transducers of these stress signals will be key to understanding stage differentiation at the molecular level. It also suggests that bradyzoite-specific proteins that are secreted to the parasite surface or into the host cell may not only play important roles in relaying these signals to induce differentiation but may also be critical for maintaining the latent state of bradyzoites by acting as gate keepers and integrating changes in environmental conditions such as changes in host immune status.

The work presented in this section describes the characterization of a tandemly duplicated cluster of developmentally regulated genes, TgME49_009980 (*ROP42*), _010090 (*ROP43*), _010110 (*ROP44*) and _121700 (*ROP42L1*) (www.toxodb.org). Together with *EL4* described in chapter 3, this expanded locus (*EL3*) was initially identified among the list of genomic loci that were differentially expanded between *T. gondii* and *N. caninum* (Chapter 2). We examined locus structure, expression and subcellular localization, and impact on parasite proliferation *in vitro* as well as *in vivo*. Members of this *ROP42* subfamily collectively belong to the *ROP2* superfamily of rhoptry proteins which also include important virulence genes such as *ROP18, ROP5* and *ROP38* (Saeij *et al.*, 2006, Peixoto *et al.*, 2010, Reese *et al.*, 2011b). We show that all members of this cluster encode developmentally regulated proteins which are secreted and localized to the parasitophorous vacuolar membrane (PVM). While deletion of the entire locus in the Type I background does not show any detectable phenotype, misregulation in tachyzoites appears to impact parasite proliferation *in vitro* as well as *in vivo*. We have subsequently developed a regulatable expression system to enables us to better study the function of this locus by way of inducible stage-specific expression.

4.2 RESULTS

4.2.1 ROP42 and its paralogs encode putative pseudokinase homologs

In chapter 2, we showed that *EL3*, much like *EL4* (MAF1), exhibits differential expansion among lineages as well as between strains of the same lineage. Examination of *EL3* in the *T. gondii* reference genome reveals 3 predicted open reading frames annotated as *ROP42, ROP43, ROP44,* and a fourth unannotated open reading frame which we refer to as *ROP42*-Like *1* (*ROP42L1*). Based on Southern blot analysis we found that for each of 6 strains, *EL3* had a higher copy-number than predicted in the reference genome. We estimated that the RH (Type I) strain contained the most copies (9 copies) and GT1 (Type I) contained the fewest copies (6 copies) (Chapter 2). To examine the nature of the predicted paralogs within the locus, we cloned and sequenced both Type II and III alleles of 3 of the predicted genes (*ROP42*, *43,* and *42L1*). ClustalΩ alignment of these sequences showed that the predicted proteins shared over 98% percent identity in their C-terminal region which is predicted to constitute a kinase domain (Figure 4.1A) as do other members of the *ROP2* superfamily of rhoptry kinases. The *ROP2* superfamily contains both active (ROP18) and inactive (ROP5) kinases. By comparison, the predicted kinase domains of the ROP42 paralogs contained a conserved ATP binding motif (VAIK) and a Mg^{2+} binding motif (DFG) but a mutated HRD catalysis motif (D→K; Figure 4.1A) that would therefore be considered catalytically inactive (Reese *et al.,* 2011a). Phylogenetic analysis of the sequences showed that ROP42L1 was the most divergent between alleles while ROP42 is the least polymorphic (Figure 4.1B).

4.2.2 ROP42 and its paralogs are developmentally regulated and localize to the parasitophorous vacuolar membrane

Expression data from the *T. gondii* genome shows that at least one member of the *ROP42* cluster is upregulated during bradyzoite differentiation. Robust multi-array averages (RMA) values for transcripts from *ROP42* show that its expression is developmentally regulated with peak expression observed in the sporozoite and bradyzoite stages (Figure 4.2A, www.toxodb.org). To characterize thisg this locus for structure and function, we first determined whether all paralogs were expressed in a similar developmentally regulated manner as the "archetypal" ROP42. We focused on tachyzoite-to-bradyzoite differentiation for its significance in pathogenesis as well as the fact that it is a more experimentally tractable process than sporozoite differentiation. A luciferase-based reporter assay was used to measure promoter activities during bradyzoite differentiation for the 3 paralogs sequenced above. Approximately 1 kb of the region immediately upstream of the *ROP42, ROP43* and *ROP42L1* coding sequence was cloned upstream of a firefly luciferase gene to generate constructs in which luciferase expression was driven by the promoter regions of these *ROP42* paralogs (Figure 4.2B). Luciferase expression was then compared between tachyzoites and *in vitro*-induced bradyzoites (Behnke *et al.*, 2008). All 3 promoters showed higher activity in bradyzoites than in tachyzoites, for both Type II (Figure 4.2C) and Type III (Figure 4.2D) alleles. This confirms that all paralogs have similar expression kinetics and that expression is upregulated for all paralogs during tachyzoite-to-bradyzoite differentiation. There were some basal levels of expression in tachyzoites which was consistent with the available RMA expression values for ROP42 and may be an important starting point for the possible roles of these genes in bradyzoite differentiation.

To investigate the possible functions of these genes, we first determined their subcellular localization. The coding sequence of *ROP43* was cloned immediately downstream of the constitutively active GRA1 promoter, and in-frame with a C-terminal HA-tag present on the pGRA1–HA–HPT vector to generate the expression construct pGRA1-ROP43-HA (Figure 4.3A). This vector contains a hypoxanthine phosphoribosyl transferase (*HPT*) gene which allows for selection of stable transfectants in an *HPT*-knockout background by growing transfected parasites in mycophenolic acid (MPA) and xanthine-supplemented media (Chapter 3). Expression of a ~70kDa protein in TgME49 transfected with pGRA-ROP43-HA (ME49:ROP43) was confirmed by anti-HA western blot analysis (Figure 4.3B). This was consistent with the predicted molecular weight for the 646 amino acid (including HA-tag) protein. In an immunofluorescence assay (IFA), ROP43 was detected in the anterior region of the parasite, as expect for a rhoptry protein (colocalizes with ROP7), as well as at the parasitophorous vacuolar membrane (PVM) (Figure 4.3C). This localization pattern indicates that ROP42 paralogs or at least ROP43, like other members of the ROP2 superfamily, is secreted outside of the parasites and gets associated with the parasitophorous vacuolar membrane.

4.2.3 Misregulation of ROP43 in tachyzoites leads to decreased *in vitro* proliferation

Having established the subcellular localization profile of ROP43, we proceeded to examine the effect of misregulation of ROP43 expression on *in vitro* growth properties of tachyzoites. Such temporal misregulation has been used previously to establish the role of the bradyzoite-specific antigen SRS9 in bradyzoite persistence (Kim *et al.*, 2005). Growth rate was first examined using plaque assays. ME49:ROP43 and empty-vector control strains were inoculated onto HFFs at a limiting MOI such that clearly defined plaques representing areas of lysed host cells from an

initial single parasite invasion could be identified after 7 days. We found that 2 independently generated clones of ME49:ROP43 had plaque sizes that were about half that of the control strain at day 7 post-infection (Figure 4.4A). In order to confirm that this slow growth phenotype is a specific effect of ROP43 misregulation, and neither the result of congested secretory channels due to overexpression of ROP43 nor a general artifact of expressing a bradyzoite-specific gene in tachyzoites, plaque sizes of a TgME49 strain expressing GRA1 promoter-driven bradyzoite rhoptry protein 1 (BRP1) were also examined. BRP1 has been reported to be a non-essential bradyzoite-specific rhoptry protein (Schwarz *et al.*, 2005). This strain had a plaque size phenotype similar to that of the control strain (Figure 4.4A) confirming that the slow growth phenotype is a specific effect of ROP43 misregulation in tachyzoites.

To distinguish between increased spontaneous bradyzoite differentiation and decreased replication rate as plausible explanations for the slow growth phenotype, we first compared the rate of differentiation between the ME49:ROP43 strain and the control strains when grown under *in vitro* bradyzoite-inducing conditions. Infected cells were incubated in bradyzoite-induction medium (Buchholz *et al.*, 2011) for 24, 48, or 72 hours. At each time-point, cells were stained for the bradyzoite marker, bradyzoite antigen 1 (BAG1), and the percentage of vacuoles containing BAG1-positive parasites was determined. There was no significant difference in the rate of differentiation between ME49:ROP43 and control strains as measured by BAG1 expression, suggesting that bradyzoite differentiation was not impacted by the misregulation of ROP43 (Figure 4.4B). Next, the rate of replication between the two strains was examined by comparing number of parasites per vacuole after 20 hours of *in vitro* infection. At this time-point the number of parasites per vacuole ranged from 2 to 32. The results of this assay showed that ME49:ROP43 clones replicated at a slower rate compared to wild type. At 20 hpi, only 40% of

ME49:ROP43 vacuoles contained 4 or more parasites versus >70% for the control strain (Figure 4.4C). While there were a number of vacuoles for each strain containing 1 parasite, such vacuoles were eliminated from the analysis since we could not ascertain whether these parasites were dead or alive. Taken together, these results suggested that the slow growth phenotype of the ME49:ROP43 was a result of decreased replication rate with no detectable effect on bradyzoite differentiation.

4.2.4 Misregulation of ROP43 in tachyzoites affects *in vivo* proliferation and dissemination

We next wanted to determine whether the observed *in vitro* phenotype translated into defects *in vivo*. For this, we examined the effects of tachyzoite expression of ROP43 in mouse infections. The TgME49 parental strain used to generate the tachyzoite-expression clones had been engineered to stably express luciferase which allows us to measure parasite burden *in vivo* by non-invasive bioluminescence imaging (Saeij *et al.*, 2005b). This method provides a real-time quantitative readout for the severity of the infection and also monitors the dynamics of parasite dissemination *in vivo*. Parasite-burden and host survival were compared between mouse groups infected with either ME49:ROP43 or empty-vector control strains. Three mice were infected per group. With similar starting parasite loads, we observed a marked decrease in proliferation and dissemination of ME49:ROP43 compared to the control strain (ME49*). Parasite burdens began to diverge from day 5 p.i. and continued until both infections were resolved (Figure 4.5A,B). Additionally, all mice in the ME49:ROP43 group survived the infection while 1 mouse from the control group died at day 12 post-infection (Figure 4.5C). These results together suggest that misregulation of ROP43 has a significant impact on *in vivo* proliferation and dissemination.

4.2.5 Deletion of the *ROP42* locus has no effect on Type I virulence

To further characterize the role of the *ROP42* cluster in parasite physiology, we generated a *ROP42* locus knock-out (KO) strain in which the entire ~38kb locus was deleted. For this purpose we designed a plasmid construct which included 1000-2000 bp of the 5' and 3' flanking regions of the ROP42 locus on either side of an *HPT* selectable maker which was then used to knock-out the ROP42 locus by double homologous recombination. Taking advantage of the Type I RH*Δku80* strain, which has been specifically engineered to favor homologous recombination (Fox *et al.*, 2009, Huynh *et al.*, 2009), *ROP42*-cluster KO was generated in the Type I background and confirmed by diagnostic PCR (Figure 4.6A,B) as well as Southern blot analysis (Figure 4.6C). To determine the effect of loss of the *ROP42* locus on *in vivo* virulence, Balb/C mice were infected with 100 tachyzoites each of either an empty-vector control strain (RH*Δku80*) or *ROP42*-KO strain (RH*Δku80Δrop42*) by IP injection and virulence measured by time of death. There were no apparent differences in virulence that could be attributed to the loss of *ROP42* as both sets of mice survived for approximately the same length of time (Figure 4.6D).

4.2.6 Regulatable ROP43 expression system

The developmental regulation of ROP42 and its paralogs means that gene products are only available at specific developmental stages of the parasite. In order to better understand the functional implications of stage-specific expression of this cluster of genes, we designed a regulatable expression system which could be used to study the effect of selective expression in different developmental stages. The tetracycline-regulatable expression system has been used extensively to control gene expression in many systems and had been previously adapted and

optimized for use in *T. gondii* (Berens *et al.*, 2004, van Poppel *et al.*, 2006). This system consists of two modules. The first is the repressor module (YFP-TetR) for stable expression of a codon-optimized tet repressor under the control of a constitutively active *T. gondii* promoter (Figure 4.7A). In this case the tet repressor is fused to YFP for easy visualization. The second is the tet-regulatable promoter module which was constructed by inserting 4 tet operator (tetO) elements in the highly active *T. gondii* ribosomal protein RPS13 promoter (pRPS13sub(IV)) (Figure 4.7B; van Poppel *et al.*, 2006)). This system has proven to establish a tight control of gene expression using an analogue of tetracycline, anhydrotetracyline (ATc) which is well tolerated by *T. gondii* (Meissner *et al.*, 2001, van Poppel *et al.*, 2006).

The system was adapted for regulatable expression of ROP43 in the following manner. A stable expression strain for YFP-TetR (M:TR) was generated using ME49Δhpt:LUC (MΔ:Luc) as the parental strain with chloramphenicol selection for stable integration. MΔ:Luc is a Type II strain in which *HPT* has been deleted to allow for positive selection of transfectants using MPA/Xanthine, and also expresses a luciferase reporter gene for downstream applications. After confirming expression of YFP-TetR by visual inspection, a chlorophenol red-β-d-galactopyranoside (CPRG) assay (Seeber *et al.*, 1996) was used to confirm its functionality in the Type II background. The M:TR strain as well as the MΔ:Luc parental strain was transfected with the pRPS13(IV)-LacZ plasmid, in which the LacZ reporter gene is expressed in a tet-regulatable manner. Transfected parasites were grown in the presence or absence of ATc after which β-galactosidase (β-gal) expression was determined (Figure 4.7D). MΔ:Luc had ~13% lower CPRG conversion when incubated with ATc than without. This difference, while not significant, indicates a possible minor effect of ATc on the parasites. Incubation of M:TR with ATc showed ~10-fold induction over no ATc incubation. This represents about twice the observed increase in

ROP43 promoter activity during tachyzoite-to-bradyzoite differentiation (Figure 4.2C). Additionally, CPRG conversion in M:TR without ATc is almost as low as the background level in the no-LacZ control indicating that there is only a very minimal amount of leakiness associated with the system. Taken together, these results indicate that TetR is fully functional in the M:TR strain.

Having confirmed the functionality of the M:TR strain, we generated the complementary regulatable ROP43 expression construct by replacing the GRA1 promoter driving ROP43 expression on the pGRA1-ROP43-HA plasmid with the pRPS13(IV) regulatable promoter (Figure 4.7C). This construct was transfected into M:TR and stable transfectants (M:TR-R43i) isolated by MPA/Xanthine selection. ATc-inducible ROP43 expression was confirmed by anti-HA western blot analysis (Figure 4.8A) as well as anti-HA immunofluorescence assay (Figure 4.8B).

In preliminary experiments to examine the impact of ATc-inducible ROP43 expression using the M:TR-R43i strain, we compared parasite proliferation in the presence and absence of ATc in both *in vitro* plaque assays and *in vivo* mouse infections. *In vivo* induction was performed by administering ATc to infected mice via drinking water as previously described (Meissner *et al.*, 2002). Results from these experiments did not reproduce the previously described decreased proliferation phenotype. The proliferation profile of M:TR-R43i in *in vitro* plaque assays (data not shown) and in mouse infections did not change significantly with or with ATc treatment (Figure 4.C).

4.3 DISCUSSION

In Chapter 2, we found a significant enrichment of developmentally regulated genes among the set of genes which were uniquely expanded in *T. gondii* relative to *N. caninum*. The *ROP42* cluster (*EL3*) was for further investigation based on a number of interesting properties of the locus: 1) it shows differential expansion among different *T. gondii* strains as predicted by read coverage analysis (Chapter 2), 2) *ROP42* and its paralogs encode protein kinase homologues that are predicted to be secreted from the parasite into the host cell and 3) one member of the gene cluster is upregulated during tachyzoite-to-bradyzoite (i.e., cyst) differentiation (Buchholz *et al.*, 2011).

While tandem duplication of *EL3* has been previously reported (Peixoto *et al.*, 2010), results from this work provide the first demonstration of differential expansion of this locus among *T. gondii* strains, and also show that this locus has a more complex structure than previously predicted. In the reference genome, *EL3* is predicted to contain 3 paralogs, *ROP42/43/44*, but our results indicate this is an underestimation and that copy-number ranges from 6 to 9 depending on the strain. This is not surprising since such loci containing tandemly arranged repetitive sequences are known to often cause scaffold breaks in genome assembly which lead to loss of sequence information (Carlton *et al.*, 2007). Nonetheless, intra-lineage copy-number variation at these tandemly duplicated loci represents an important distinction between closely related strains and provides another example of the genomic instability associated with tandemly duplicated loci (Bzymek *et al.*, 2001, Dittwald *et al.*, 2013). This locus is also an interesting example of gene conversion through unequal crossing over as *ROP42L1* and *ROP44* appear to be prematurely truncated in Type II and Type III strains, respectively (www.toxodb.org).

From the sequence analysis, it appears that while the ROP42 paralogs, like other members of the ROP2 superfamily (Talevich *et al.*, 2013), may retain a conserved kinase fold in their c-termini, they may be catalytically inactive. While this needs to be confirmed, it is possible that their pseudokinase status may make a significant contribution to their function as has been shown for ROP5 (Reese *et al.*, 2011a) and several other pseudokinases (Reese *et al.*, 2012).

All three paralogs tested, and most likely all remaining paralogs, are upregulated during bradyzoite differentiation. Additionally, our results show that there is a basal background expression of these genes, which is consistent with available RMA values of transcript abundance (www.toxodb.org). It is presently not known whether this basal expression in tachyzoites has any physiological relevance but the possibility exists that this expression profile may support an early role in the differentiation process and perhaps may be involved in committing the parasite to the bradyzoite fate. Further experiments will be required to address this hypothesis, and deletion of the *ROP42* cluster in the more bradyzoite-prone Type II and Type III backgrounds prove important.

Bradyzoite differentiation is marked by a decrease in doubling time from ~6 hours in tachyzoites to about ~16 hours in early bradyzoite and to eventual growth arrest in mature bradyzoites (Jerome *et al.*, 1998, Radke *et al.*, 2001, Radke *et al.*, 2003). With this notion in mind, the slow growth effect of ROP43 expression in tachyzoite provides an intriguing possibility of a role for the *ROP42* cluster in indirectly controlling replication rate during differentiation. In fact, it has been suggested that effective bradyzoite-inducing conditions are those that are able to sufficiently slow down the rate of replication (Fox *et al.*, 2002, Khan *et al.*, 2002). It will be interesting to examine possible functional interactions between the ROP42 cluster and known cell cycle regulators such as TPK2 (Khan *et al.*, 2002).

An interesting observation we made was that the ME49:ROP43 clones appeared to eventually recover from the slow growth phenotype and attain a relatively normal replication rate. This occurred after about 3-4 weeks of passages *in vitro*. We are currently characterizing this phenotype to determine the specific timelines associated with the recovery and also determine whether there are any changes in the dynamics of ROP43 expression at that stage. This will help us properly qualify the slow growth phenotype.

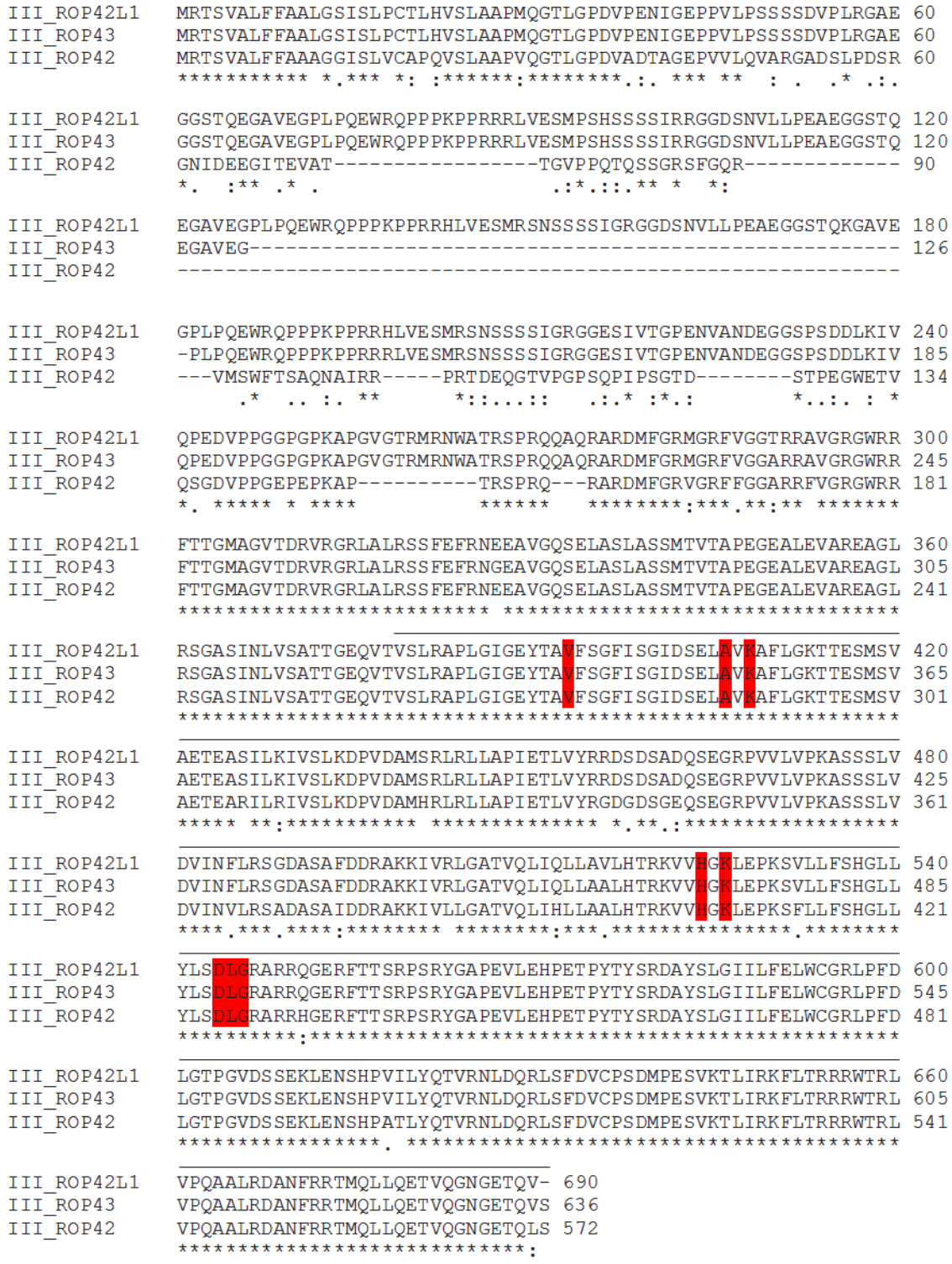
We were able to delete the entire *ROP42* locus in the hypervirulent Type I background. Although we did not observe any detectable effect on parasite proliferation, this was not entirely surprising as this parental RH strain is known to be a notoriously poor tissue cyst former (Jerome *et al.,* 1998, Kirkman *et al.,* 2001). Additionally, the bimodal readout of the mortality assay lacks the sensitivity to detect subtle defects that may be caused by loss of the *ROP42* locus. The hypervirulent nature of the RH strain also could easily obscure such subtle effects. As mentioned above, it may be more informative to examine *ROP42* cluster deletion in Type II and III strains since these are more competent bradyzoite formers.

The tet-inducible expression system provides an excellent tool to investigate developmentally regulated genes by allowing for temporally-controlled expression (van Poppel *et al.,* 2006). We have successfully utilized the tet-inducible system to induce ROP43 expression. Although tet-induction of ROP43 in tachyzoites did not pheno-copy the GRA1-promoter driven expression, we were able to demonstrate a tight regulation of ROP43 expression. While induction may be comparable to endogenous promoter activity, it will need to be compared with GRA1 promoter activity.

In conclusion, we show that the tandemly duplicated locus, *EL3*, is not only differentially expanded between *T. gondii* and *N. caninum*, but also exhibits copy-number variation among

different strains of *T. gondii*. We also provide preliminary evidence for a possible role for ROP43 and its paralogs on bradyzoite differentiation. We have developed an inducible expression system to be used for gene expression in a controlled and regulatable manner. This system will be an excellent tool for further investigating stage-specific expression of ROP43 and its paralogs as it allows one to control the timing and amount of protein expression. This system can also easily be adapted to study other genes.

A.


III_ROP42L1 MRTSVALFFAALGSISLPCTLHVSLAAPMQGTLGPDVPENIGEPPVLPSSSSDVPLRGAE 60
III_ROP43 MRTSVALFFAALGSISLPCTLHVSLAAPMQGTLGPDVPENIGEPPVLPSSSSDVPLRGAE 60
III_ROP42 MRTSVALFFAAAGGISLVCAPQVSLAAPVQGTLGPDVADTAGEPVVLQVARGADSLPDSR 60
III_ROP42L1 GGSTQEGAVEGPLPQEWRQPPPKPPRRRLVESMPSHSSSSIRRGGDSNVLLPEAEGGSTQ 120
III_ROP43 GGSTQEGAVEGPLPQEWRQPPPKPPRRRLVESMPSHSSSSIRRGGDSNVLLPEAEGGSTQ 120
III_ROP42 GNIDEEGITEVAT----------TGVPPQTQSSGRSFGQR---------- 90
III_ROP42L1 EGAVEGPLPQEWRQPPPKPPRRHLVESMRSNSSSSIGRGGDSNVLLPEAEGGSTQKGAVE 180
III_ROP43 EGAVEG---------- 126
III_ROP42L1 GPLPQEWRQPPPKPPRRHLVESMRSNSSSSIGRGGESIVTGPENVANDEGGSPSDDLKIV 240
III_ROP43 -PLPQEWRQPPPKPPRRRLVESMRSNSSSSIGRGGESIVTGPENVANDEGGSPSDDLKIV 185
III_ROP42 ---VMSWFTSAQNAIRR-----PRTDEQGTVPGPSQPIPSGTD--------STPEGWETV 134
III_ROP42L1 QPEDVPPGGPGPKAPGVGTRMRNWATRSPRQQAQRARDMFGRMGRFVGGTRRAVGRGWRR 300
III_ROP43 QPEDVPPGGPGPKAPGVGTRMRNWATRSPRQQAQRARDMFGRMGRFVGGARRAVGRGWRR 245
III_ROP42 QSGDVPPGEPEPKAP----------TRSPRQ---RARDMFGRVGRFFGGARRFVGRGWRR 181
III_ROP42L1 FTTGMAGVTDRVRGRLALRSSFEFRNEEAVGQSELASLASSMTVTAPEGEALEVAREAGL 360
III_ROP43 FTTGMAGVTDRVRGRLALRSSFEFRNGEAVGQSELASLASSMTVTAPEGEALEVAREAGL 305
III_ROP42 FTTGMAGVTDRVRGRLALRSSFEFRNEEAVGQSELASLASSMTVTAPEGEALEVAREAGL 241
III_ROP42L1 RSGASINLVSATTGEQVTVSLRAPLGIGEYTAWFSGFISGIDSELAVKAFLGKTTESMSV 420
III_ROP43 RSGASINLVSATTGEQVTVSLRAPLGIGEYTAWFSGFISGIDSELAVKAFLGKTTESMSV 365
III_ROP42 RSGASINLVSATTGEQVTVSLRAPLGIGEYTAWFSGFISGIDSELAVKAFLGKTTESMSV 301
III_ROP42L1 AETEASILKIVSLKDPVDAMSRLRLLAPIETLVYRRDSDSADQSEGRPVVLVPKASSSLV 480
III_ROP43 AETEASILKIVSLKDPVDAMSRLRLLAPIETLVYRRDSDSADQSEGRPVVLVPKASSSLV 425
III_ROP42 AETEARILRIVSLKDPVDAMHRLRLLAPIETLVYRGDGDSGEQSEGRPVVLVPKASSSLV 361
III_ROP42L1 DVINFLRSGDASAFDDRAKKIVRLGATVQLIQLLAVLHTRKVVHGKLEPKSVLLFSHGLL 540
III_ROP43 DVINFLRSGDASAFDDRAKKIVRLGATVQLIQLLAALHTRKVVHGKLEPKSVLLFSHGLL 485
III_ROP42 DVINVLRSADASAIDDRAKKIVLLGATVQLIHLLAALHTRKVVHGKLEPKSFLLFSHGLL 421
III_ROP42L1 YLSDLGRARRQGERFTTSRPSRYGAPEVLEHPETPYTYSRDAYSLGIILFELWCGRLPFD 600
III_ROP43 YLSDLGRARRQGERFTTSRPSRYGAPEVLEHPETPYTYSRDAYSLGIILFELWCGRLPFD 545
III_ROP42 YLSDLGRARRHGERFTTSRPSRYGAPEVLEHPETPYTYSRDAYSLGIILFELWCGRLPFD 481
III_ROP42L1 LGTPGVDSSEKLENSHPVILYQTVRNLDQRLSFDVCPSDMPESVKTLIRKFLTRRRWTRL 660
III_ROP43 LGTPGVDSSEKLENSHPVILYQTVRNLDQRLSFDVCPSDMPESVKTLIRKFLTRRRWTRL 605
III_ROP42 LGTPGVDSSEKLENSHPATLYQTVRNLDQRLSFDVCPSDMPESVKTLIRKFLTRRRWTRL 541
III_ROP42L1 VPQAALRDANFRRTMQLLQETVQGNGETQV- 690
III_ROP43 VPQAALRDANFRRTMQLLQETVQGNGETQVS 636
III_ROP42 VPQAALRDANFRRTMQLLQETVQGNGETQLS 572

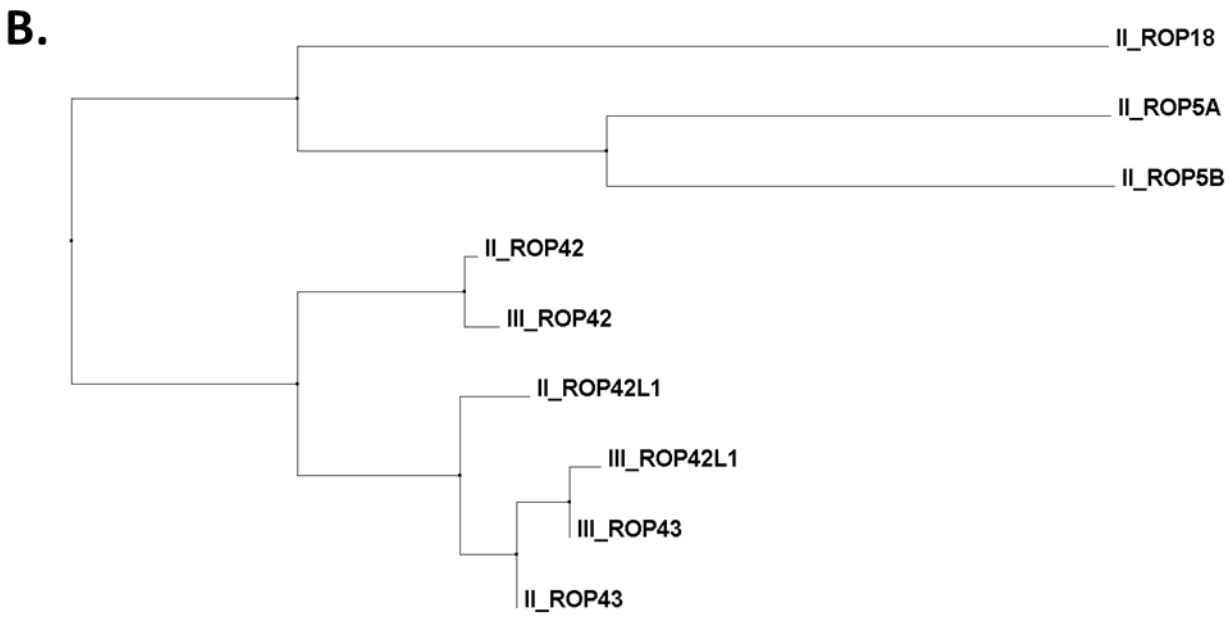


Figure 4.1: ROP42 and its paralogs encode putative pseudokinase homologs.

A) ClustalΩ alignment of primary sequences of the Type III alleles of ROP42, ROP43 and ROP42L1. The predicted *pseudo*kinase domain is indicated by the line. Residues predicted to comprise the 3 functional motifs are highlight in red. B) Phylogenetic analysis of ROP42, ROP43 and ROP42L1 from both Type II and III strains. The tree was generated from ClustalΩ alignment of the sequences using neighbor joining with percent identity. Type II ROP18 and 2 Type II ROP5 paralogs were used as outgroups.

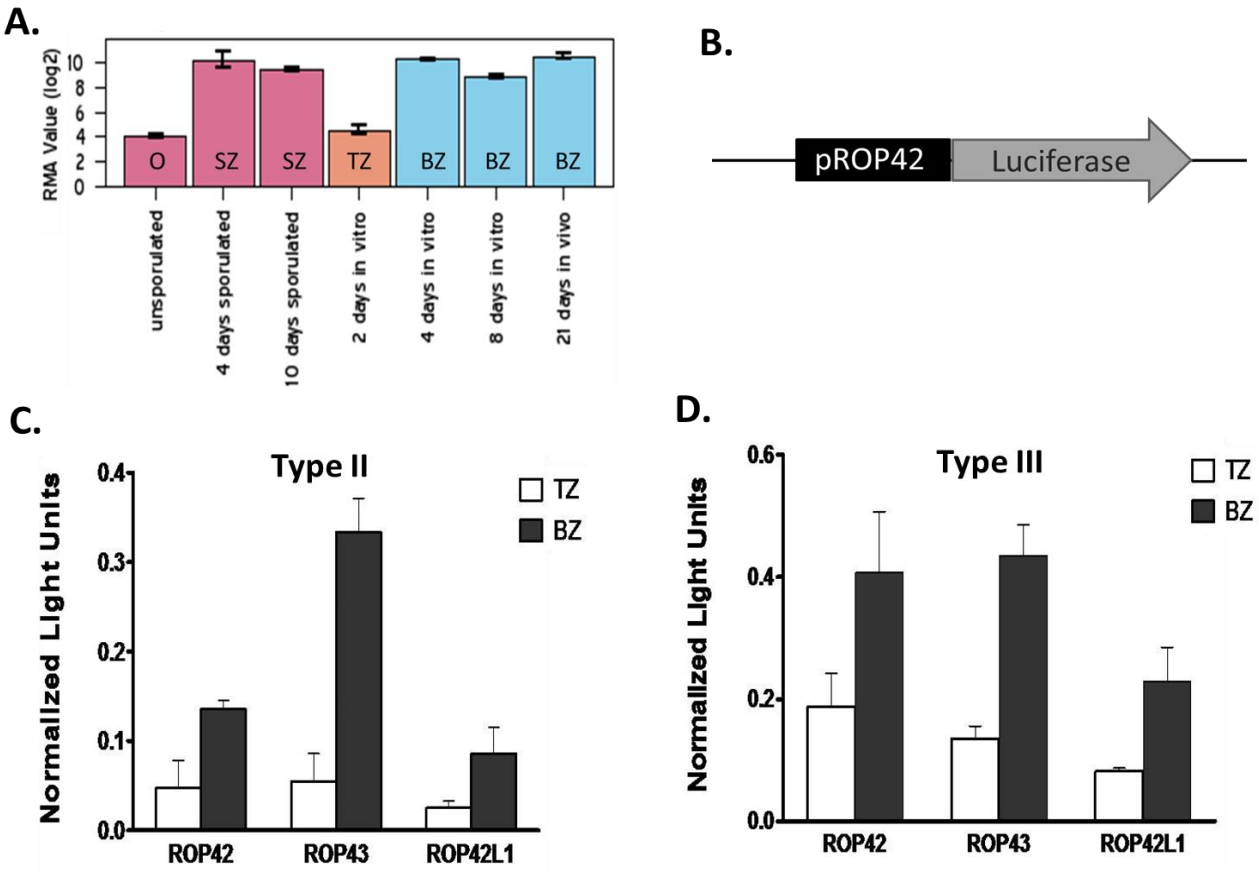


Figure 4.2: All ROP42 paralogs are developmentally regulated.

A) Robust multi-array averages of ROP42 transcript abundance in oocysts (O), sporozoites (SZ), tachyzoites (TZ) and bradyzoites (BZ) based on data collected by Fritz et al. (Gajria *et al.*, 2008, Buchholz *et al.*, 2011, Fritz *et al.*, 2012). B) Schematic of luciferase reporter construct to test promoter activity of ROP42 paralogs. C, D) Promoter activity of ROP42 paralogs as determined in dual luciferase assay. C, Type II alleles tested in Type II background; D, Type III alleles tested in Type III background.

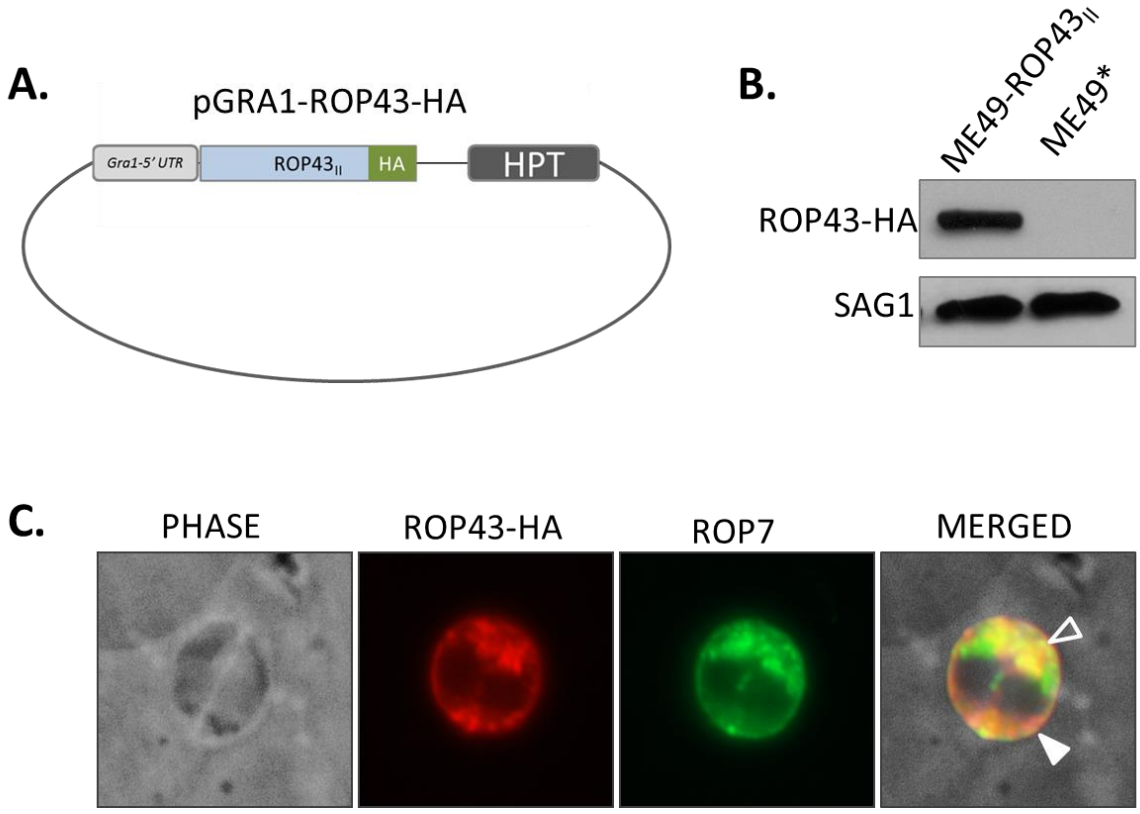


Figure 4.3: ROP43 is secreted and localizes to the parasitophorous vacuolar membrane.

A) The coding sequence of the Type II allele of ROP43 was cloned downstream of a GRA1 promoter and in-frame with a C-terminal HA-tag using the pGRA1-HA-HPT vector. B) Anti-HA western blot analysis showing ROP43-HA expression in the transfected Type II strain, ME49. The molecular weight is ~70kDa, which is consistent with predictions based on the length of the amino acid sequence. C) Anti-HA immunofluorescence assay showing subcellular localization of ROP43. The protein can be detected in the rhoptries (open arrowhead; colocalizes with ROP7) as well as on the parasitophorous vacuolar membrane (filled arrowhead).

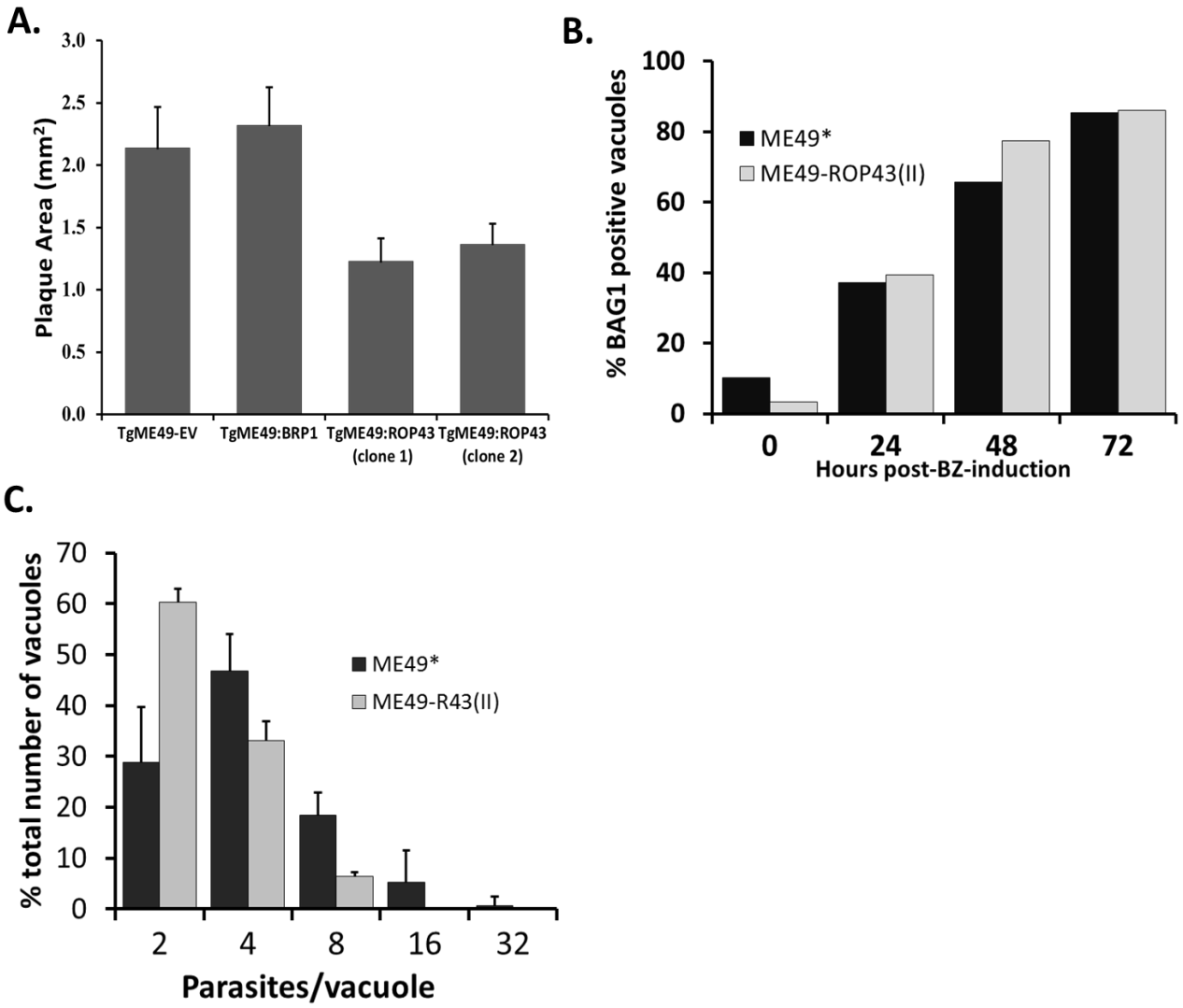
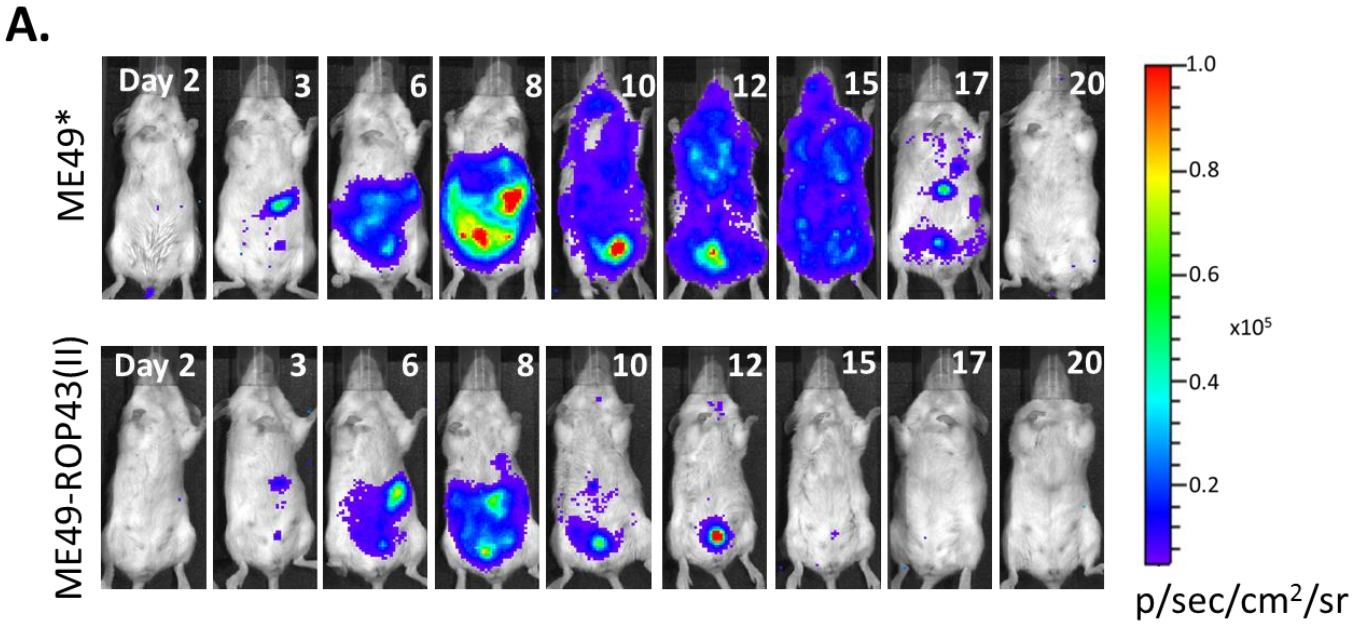
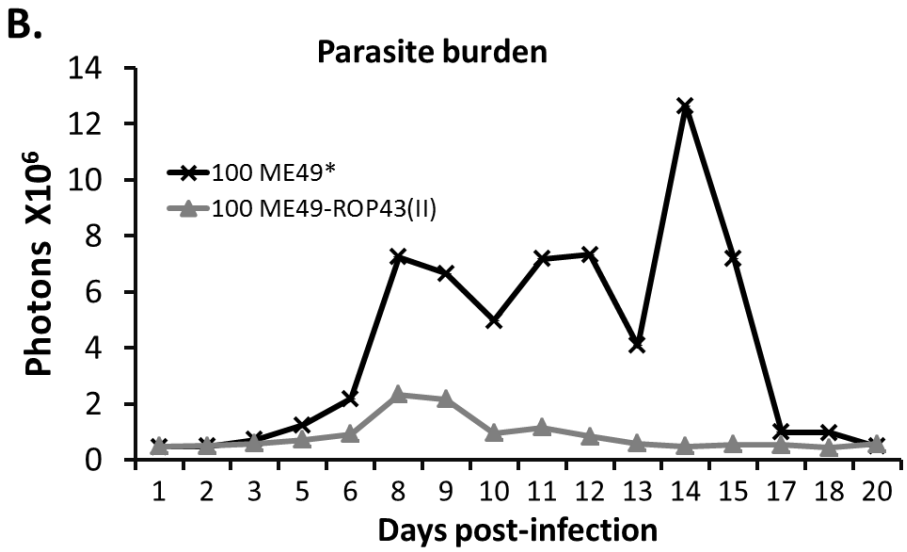


Figure 4.4: Misregulation of ROP43 in tachyzoites lead to a decreased proliferation rate *in vitro*.

A) Quantitation of day 7 plaque sizes in ME49* (empty vector control), ME49-BRP1 and 2 independent clones of ME49-ROP43. B) Rate of tachyzoite-to-bradyzoite differentiation compared between ME49* and ME49-ROP43. C) Number of parasites per vacuole compared between ME49* and ME49-ROP43 at 20 hpi.


A.
ME49*
Day 2
3
6
8
10
12
15
17
20
ME49-ROP43(II)
Day 2
3
6
8
10
12
15
17
20
1.0
0.8
0.6
x10^5
0.4
0.2
p/sec/cm²/sr


B.
Parasite burden
Photons X10^6
14
12
10
8
6
4
2
0
100 ME49*
100 ME49-ROP43(II)
1 2 3 5 6 8 9 10 11 12 13 14 15 17 18 20
Days post-infection

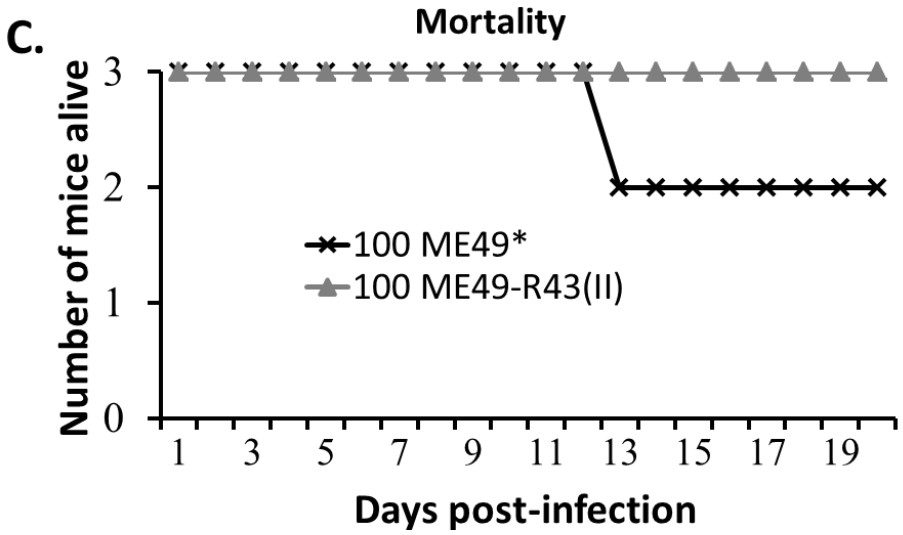


Figure 4.5: Overexpression of ROP43 in tachyzoites leads to decreased *in vivo* proliferation and dissemination.

BALB/C mice were infected with either ME49:ROP43 or ME49* (empty vector control) by IP injections at a dose of 100TZ. A) BLI images showing the proliferation and dissemination of the parasites over the course of the infection. B) Quantitation of photon emission representing parasite burden. Values represent an average of 3 mice per group. C) Survival curve showing mortality over the course of the infection.

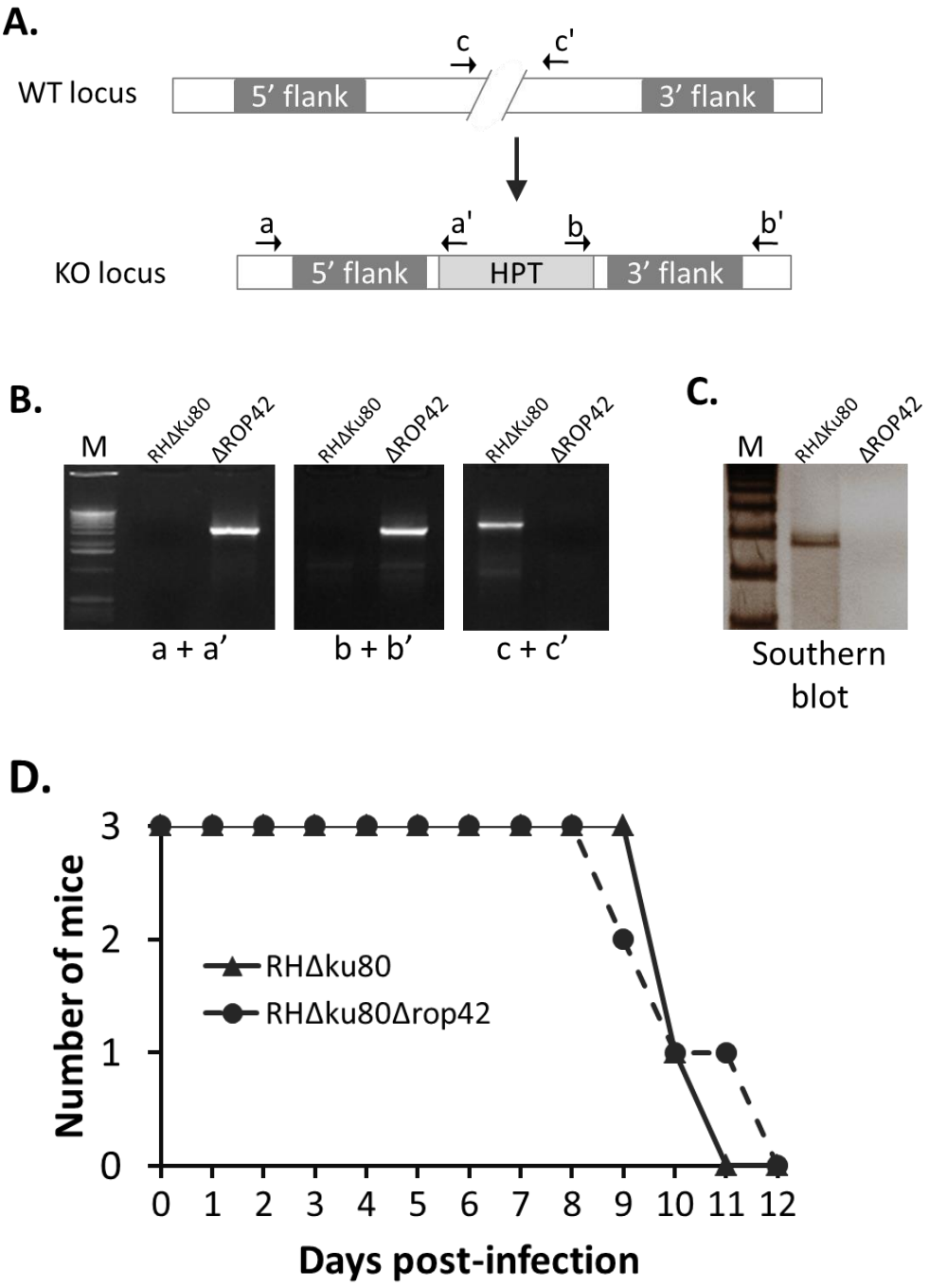
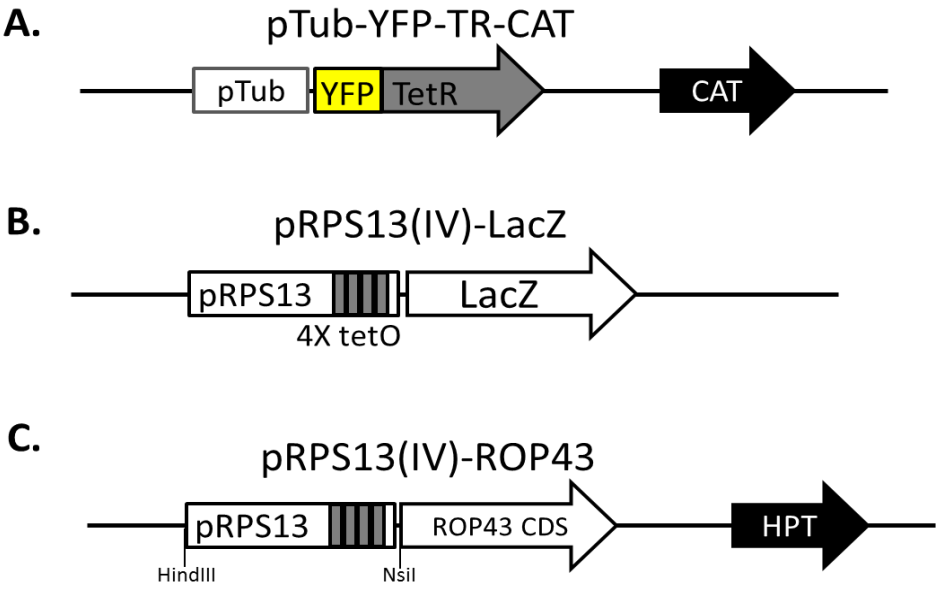


Figure 4.6: Deletion of *EL3* does not alter virulence in Type I RH strain.

A) Schematic depiction of *EL3* deletion by double homologous recombination. Deletion of *EL3* was confirmed by B) diagnostic PCR and C) Southern blot analysis. D) Mice (n=3) were infected with 100 parasites each and survival time determined.


A.
pTub-YFP-TR-CAT
pTub
YFP
TetR
CAT
B.
pRPS13(IV)-LacZ
pRPS13
LacZ
4X tetO
C.
pRPS13(IV)-ROP43
pRPS13
ROP43 CDS
HPT
HindIII
NsiI

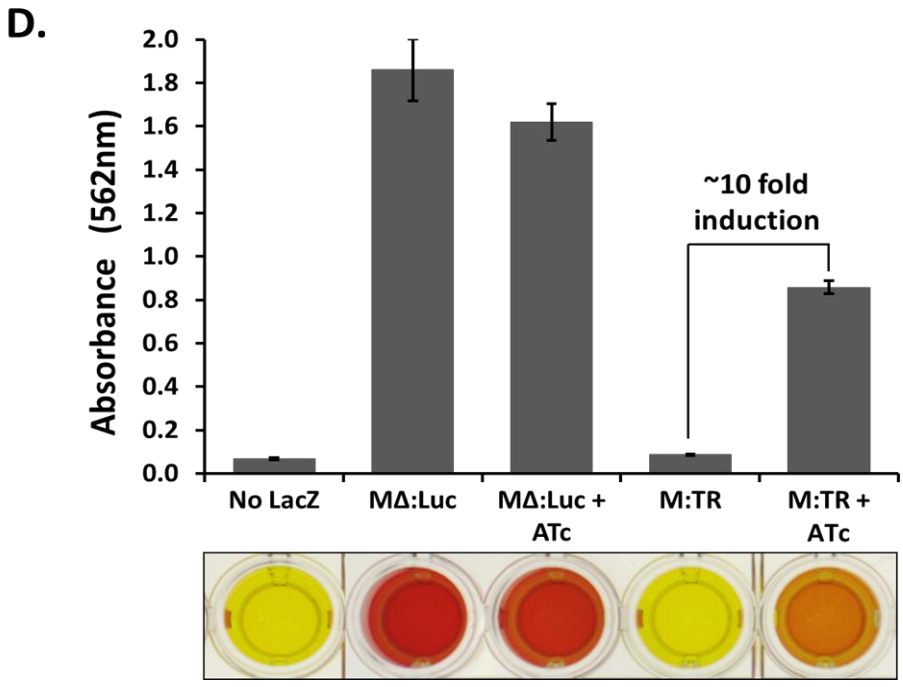

D.
Absorbance (562nm)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
~10 fold induction
No LacZ
MΔ:Luc
MΔ:Luc + ATc
M:TR
M:TR + ATc

Figure 4.7: Tet-inducible expression system.

The tet-inducible expression system is made up of two modules: A) a plasmid encoding a tet-repressor under the control of a *T. gondii* promoter and B) a tet-operator element inserted into a different *T. gondii* promoter to generate a tet-regulatable *T. gondii* promoter. C) Tet-regulatable ROP43 construct was generated by cloning the regulatable promoter from B. into the pGRA1-ROP43-HA plasmid (Figure 4.3A) to replace the GRA1 promoter. D) The M:TR strain was generated by stably expressing the pTub-YFP-TR-CAT plasmid in the Type II background. The functionality of YFP-TR was tested in a CPRG assay.

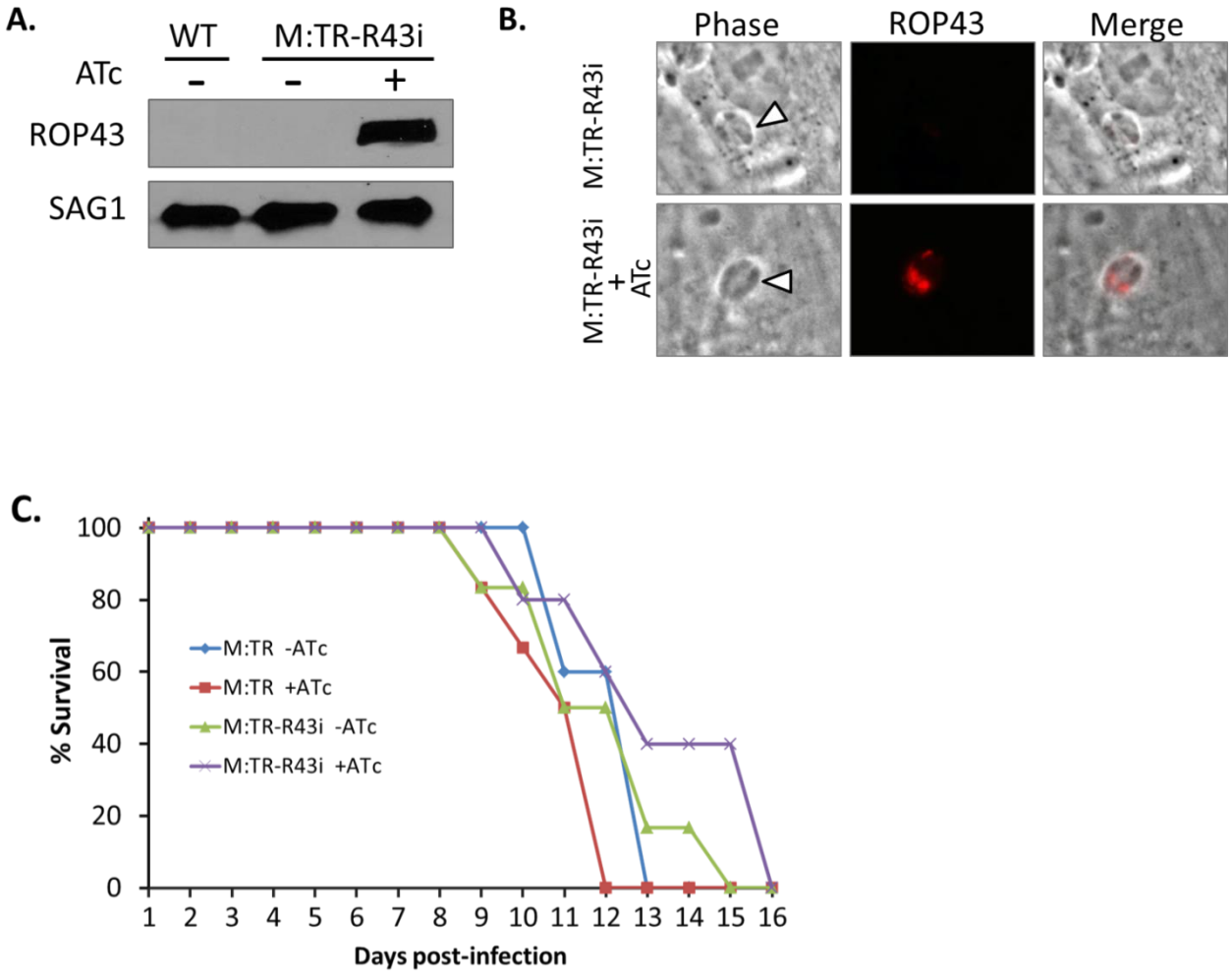


Figure 4.8: Tet-induced ROP43 expression does not significantly affect *in vivo* proliferation.

A) Western blot analysis showing ROP43 induction with anhydro-tetracycline (ATc). B) IFA showing proper expression and localization of ATc-induced ROP43. C) Survival curve showing mortality upon infection with M:TR or MT:R-R43i, and with or without ATc treatment.

5.0 CONCLUSIONS AND FUTURE DIRECTIONS

Gene duplication, ranging from small scale single gene duplications to large scale whole genome duplication events, has long been posited to be a major tool for differentiation and evolution (Ohno, 1970). However, its contribution to shaping the close evolutionary relationships between parasites and their hosts is yet to be fully assessed. Important lessons could be learnt about what factors, intrinsic and extrinsic, drive gene expansion and the overall impact of such genomic structural changes on host-parasite interactions. The focus of the work described here is to provide better understanding of the scope and impact of gene duplication on the evolution of the apicomplexan parasite *T. gondii*. Findings from this study expand existing knowledge on the nature and depth of tandem gene duplication as well as some specific roles performed by duplicated genes in *T. gondii*, as it relates to the parasite's ability to successfully adapt to its host environments.

In chapter 2, we used comparative genomics approaches to compare tandem gene duplication between the closely related coccidian parasite species, *T. gondii, N. caninum* and *H. hammondi*. Our focus on tandem gene duplication, as opposed to transpositive duplication, is informed by reports that tandemly arrayed genes tend to be products of more recent duplication events and may be more informative with respect to divergence of closely related species.

Increasingly high depth and coverage achievable by next generation sequencing techniques make read coverage analysis one of the most sensitive methods for identifying

tandemly duplicated loci. We find that gene duplication in these species is more prevalent than previously thought and that there is very little overlap between their duplicated gene profiles. This is an intriguing finding considering the fact that these species share a high degree of synteny among them (Reid *et al.*, 2012). Our results also indicate that the sets of tandemly duplicated genes in these species include a significantly large number of genes encoding secretory proteins which are presumably involved in direct or indirect interactions with the host. This observation evokes the underlying principles of adaptive evolution and points to an important role for the host environment in driving evolution by gene duplication. The kinds of genes that are uniquely expanded between *T. gondii* and *N. caninum* provide interesting hypotheses about the impact of duplicated genes on the evolution of virulence and host range. For example, uniquely expanded genes in *T. gondii* are enriched for developmentally regulated genes. While both species are able to undergo stage differentiation (Vonlaufen *et al.*, 2004, Goodswen *et al.*, 2013), it is not clear whether they are equally capable. It is conceivable that *T. gondii* would be more reliant on its differentiation abilities in establishing such a wide host range, and therefore would be more likely to build up or, when they occur randomly, retain genomic modifications which would impact its capacity to differentiate from the immunologically vulnerable tachyzoites stage into what would be considered to be its principal transmissible stage, the bradyzoite stage. Further analysis of the list of uniquely expanded genes between *T. gondii* and *N. caninum* will potentially provide insights into more of the morphological and physiological differences between the two species.

A large number of the tandemly duplicated genes that were identified turned out to be uncharacterized. Characterizing candidate genes with interesting expression and localization features using similar approaches as those applied to the *MAF1* and *ROP42* loci will potentially

uncover more interesting functions performed by duplicated genes, and help paint a more complete picture of the relevance of gene duplication in the parasite's evolution. It will also be interesting to incorporate host range analysis in order to directly test the hypothesis regarding gene duplication and host range evolution.

In chapter 3, we characterized the expanded locus, *EL4*, for both structure and function. In a parallel study, we had identified locus *EL4* to be associated with the host mitochondrial association (HMA) trait (Pernas *et al.*, in revision). We confirmed the gene encoded by this locus as the parasite factor responsible for HMA and named it mitochondrial association factor 1 (*MAF1*). We show that this locus contains two main isoforms of *MAF1*, and that these isoforms differ in their ability to mediate HMA. This locus has evolved differently not only between *T. gondii* and *N. caninum*, but also among different strains of *T. gondii*. Copy-number ranges from 3 in a Type III stain to 7 in a Type I strain. In *N. caninum* however, only a single divergent copy can be identified. Interestingly, this difference extends to the HMA trait mediated by this locus: present in *T. gondii*, absent in *N. caninum*, and it appears that the presence or absence of HMA is not directly related to copy-number but rather to the nature of the isoforms present. It is not clear whether HMA was lost in *N. caninum* or gained in *T. gondii*, but clearly, the framework for this trait is conserved in both species as a single *T. gondii MAF1A* gene is sufficient to induce HMA in *N. caninum*.

This work represents the first conclusive identification of the *T. gondii* gene involved in the phenomenon of host mitochondria association. To put in context, ROP2 was initially proposed as the parasite factor responsible for HMA but was later found to be dispensable for this trait (Pernas *et al.*, 2010). The biological relevance of HMA is still an ongoing debate. A few lines of evidence have suggested that this may be a mechanism by which the parasite sequesters

essential metabolites from the host cell. Interesting results from our experiments also show that there may be an immune response component to the function of this trait. While the evidence implicating HMA in immune modulation is relatively new compared to nutrient acquisition, there is an increasing amount of evidence highlighting the important role of host mitochondria in innate response to infections (reviewed in (West *et al.*, 2011)). Moreover, there is no evidence to suggest that these two proposed functions are mutually exclusive, or that either one is more predominant.

Prospectively, a more detailed characterization of HMA in different *T. gondii* strains in different host systems will be required to better understand the full biological relevance of this trait. It will also be important to establish a detailed cellular mechanism of HMA in order to better understand the full biological impact of the trait. This will require the identification of host factors involved in this process. Moreover, it appears that HMA alters the morphology of the associating mitochondria and there are currently ongoing experiments in the lab to examine the effect of this interaction and the resultant morphological changes on membrane integrity and physiology of host mitochondria.

Lastly, chapter 4 describes preliminary characterization of the expanded locus *EL3*. This locus contains a cluster of developmentally regulated rhoptry genes, three of which had been previously annotated in the *T. gondii* reference genome as *ROP42, 43, 44*. We identified an open reading frame encoding a fourth paralog which we referred to as *ROP42L1*. Our data indicate that there are more copies than the four predicted, and that copy-number varies among strains. Tandemly duplicated loci are notoriously difficult to assemble correctly by genome assembly programs and usually lead to scaffold breaks with loss of sequence data (Carlton *et al.*, 2007,

Peixoto *et al.*, 2010). This has been a recurring theme for most of the expanded loci we have identified and will need to be addressed in order to get a better picture of the genome structure.

Results from our initial functional analysis of the *ROP42* family suggest a role in regulating proliferation, perhaps during stage differentiation. First, it will be important to determine whether this is true for all paralogs since only *ROP43* has been tested. Second, deletion of the *ROP42* locus in a bradyzoite-prone Type II or III strain will be more useful for evaluating loss-of-function effects of the locus. Using an inducible expression system as has been described for ROP43 will allow for controlled expression in different developmental stages in order to better investigate the relevance of stage-specific expression.

Taken together, this research shows that gene duplication is a major distinguishing feature between the highly successful parasite *T. gondii* and its close relatives such as *N. caninum* and *H. hammondi*. We present structural and functional characterization of two examples of duplicated gene loci that have evolved differentially between these species. With these examples, in addition to some previously reported tandemly duplicated loci with significant impact on parasite biology (Peixoto *et al.*, 2010, Reese *et al.*, 2011b), it is becoming apparent that duplicated genes may provide explanations for some of the important physiological differences between these species. This work raises the profile of gene duplication as a tool for differentiation and evolution and provides the basis for further analysis of the roles of duplicated genes in the evolution of *T. gondii*.

6.0 MATERIALS AND METHODS

6.1 PARASITE STRAINS AND HOST CELLS MAINTENANCE

All parasite strains used in this study were maintained by regular passage of tachyzoites from freshly lysed human foreskin fibroblast (HFF) onto new HFF monolayers and grown at 37^{o}C in 5% CO_2. HFF and NRK-mitoRFP cells (a kind gift from Dr. Jennifer Lippincott-Schwartz, NIH, Bethesda, MD) (Mitra *et al.*, 2010) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FCS, 2 mM glutamine, and 50 µg/mL each of penicillin and streptomycin

6.2 SEQUENCE READ ALIGNMENTS

Raw sequence reads for *T. gondii* strains GT-1, ME49 and VEG were downloaded from the NCBI trace archive in fasta format. *T. gondii* and *N. caninum* reads were aligned to reference genomes using BLAT with the following settings: -fastMap -minIdentity=95 -minScore=90 (Kent, 2002). Following conversion of the BLAT output file (psl format) to SAM format using the psl2sam.pl script within the BlAT distribution, the SAM file was converted to a sorted BAM file using Samtools (Li *et al.*, 2009). Sequence coverage was determined in each 500 bp window using coverageBed, distributed with BEDtools (Quinlan *et al.*, 2010). Raw *H. hammondi*

sequence reads were aligned to the assembled contigs for *H. hammondi* strain HhCatGer041 using bowtie2 (Langmead *et al.,* 2012) and sequence coverage was determined in 500 bp windows across each contig using Samtools mpileup (Li *et al.,* 2009) and a custom Perl script.

Output was uploaded in R statistical software as well as a locally-run genome browser to generate whole chromosome and locus-specific plots and to facilitate manual curation of the expanded loci. For locus-specific plots data for each strain were normalized to the local sequence coverage for that strain of the 20 Kb upstream of the putatively expanded locus. Reads for all three strains were aligned to the ME49 genomic assembly (version 7.0; ToxoDB).

6.3 COVERAGE ANALYSIS, CURATION AND EXPRESSION ANALYSIS

Genome coverage plots were generated using the data above to construct chromosome-specific files that linked directly to ToxoDB or our own in-house genome browser. For visual inspection, peaks of coverage that were at least 3-fold above background were curated as follows. We removed loci containing highly repetitive sequence (such as di- and tri-nucleotide repeats). To begin determining whether the locus was tandemly expanded or if the increased sequence coverage was due to the presence of an identical (or nearly identical) gene somewhere else in the genome, we parsed the BLAT output to identify sequence reads that mapped to different chromosomes or to a location on the same chromosome that was at least 25 kb away from the putatively expanded locus. For the remaining loci the chromosomal region was examined for the presence or absence of predicted genes using ToxoDB. For those regions containing predicted genes in *T. gondii* or *N. caninum*, we examined the current annotation of the locus in ToxoDB and our own genome browser and collected evidence for gene duplication based on the

occurrence of multiple predicted genes in the same locus with high identity. For the curated expanded gene sets we performed enrichment studies for a variety of features, including the number of predicted exons and the presence/absence of a predicted signal peptide. PFAM domain analyses on predicted proteins from each locus were run locally.

6.4 COMPARATIVE GENOMIC HYBRIDIZATION

For each of the 6 strains tested, parasites were grown in HFFs, released from host cells by needle passage, washed 1X by centrifugation at 800 x g, and filtered through 5.0 µm PVDF syringe filters (Millipore). Genomic DNA was harvested from purified parasites using DNAzol (Invitrogen) according to the manufacturer's protocol and treated with RNase A to remove RNA contamination. After confirmation of purity by gel electrophoresis, genomic DNA was sheared as follows: 1 µg of gDNA was added to 750 µl of shearing buffer (TE pH 8.0 and 10% glycerol) and 1 µl of 20 µg/µl glycogen in a pre-chilled nebulizer (Invitrogen) on ice. Nebulization was performed at 40 psi for 3 min using pressurized nitrogen. The size range of the resultant gDNA fragments was confirmed to be 200-600bp by gel electrophoresis. DNA fragments were precipitated using 100% isopropanol. Biotin-labeling of DNA fragments was performed using the BioPrime Array CGH Genomic Labeling System (Invitrogen) according to the manufacturer's protocol. The 10X dCTP nucleotide mix was used with Biotin-dCTP. Purification of labeled fragments was performed with the BioPrime purification module. For each strain, 2µg of labeled DNA was hybridized to the Affymetrix ToxoGeneChip.

6.5 COMPARATIVE GENOMIC HYBRIDIZATION DATA ANALYSIS

Raw Affymetrix data files for each strain were analyzed in R statistical software using the "affy" module (cran.us.r-project.org). Individual probe-level data were normalized using the "constant" method and raw intensity values were exported using the expression chip definition file for the ToxoGeneChip (Bahl *et al.*, 2010). For each probe sequence from the *T. gondii* Affymetrix GeneChip, the number of occurrences of that 24 bp sequence in the raw genomic sequence reads for *T. gondii* GT1, ME49B7 and VEG was calculated, and probes not present in a particular strain were not used in subsequent calculations. This resulted in some strains having no useful probes for a given expanded locus (e.g., *EL48*, Figure 2.9). To determine whether a locus showed significantly higher hybridization intensity, which is indicative of locus duplication/expansion, data from all 6 strains were pooled and compared to data from the single-copy gene AMA1 using a Bonferroni-corrected Students' T-test (Dunn, 1961). For display purposes and visual inspection of between- and within-lineage variation in CGH hybridization intensity, values for each probe were mean-centered.

6.6 IDENTIFICATION OF LOCI WITH STRAIN-SPECIFIC COPY-NUMBER VARIATION USING CNV-SEQ

We used CNV-seq (Xie *et al.*, 2009) to compare sequence coverage at the 53 curated loci for all 3 *T. gondii* strains. In addition to default settings we used the following parameters: --genome-size=65000000 --minimum-windows-required=1. CNV-seq empirically calculated the minimum window size based on sequence coverage for each strain, which was 6158 bp for GT1 versus

ME49B7 and 6607 bp for VEG versus ME49B7. Copy-number variation was determined at significance threshold of $P<0.0001$.

6.7 HIGH MOLECULAR WEIGHT SOUTHERN BLOTTING

Genomic DNA was isolated from 6 *T. gondii* strains (2 representative strains from each of the 3 clonal lineages: Types I, II and III). For each strain 20 µg of genomic DNA was digested overnight (*EL3*, BspEI; *EL30*, BglI; *EL45*, NotI; *EL4*, ScaI and *AMA1*, SacI) and resolved by Pulsed Field Gel Electrophoresis using the CHEF-DR III system (Bio-Rad; Run parameters: 6.0 kV, 120°, 1.3 sec for 8 hrs, 2.7 sec for 7 hrs). Resolved fragments were transferred onto nylon membrane (BIO-RAD) and probed with DIG-labeled (Roche), locus-specific probes made from PCR-generated DNA fragments as per the manufacturer's protocol. Primers used for generating locus-specific probes are listed in Table 6.1

6.8 GENERATION OF EXPRESSION CONSTRUCTS AND TRANSGENIC PARASITES

Chapter 3: For generation of an N-terminally hemagglutinin (HA) tagged MAF1 expression construct (pMAF1), the promoter (~1.5kb sequence upstream of start codon) was cloned into the pGra-HA_HPT vector (Coppens *et al.*, 2006) at the HindIII and NsiI sites. For the reverse primer, HA-tag was fused immediately downstream of the region encoding the predicted signal peptide. The coding region downstream of the signal peptide of MAF1 up to the stop codon was

cloned into the NsiI and PacI sites. All primers used are listed in Table 6.1. . For pMAF1B, the coding sequence for *MAF1B* was amplified from GT1 cDNA and cloned downstream of the TGGT1_053770 promoter on the pMAF1A plasmid, replacing the *MAF1A* coding sequence. HA-tag was introduced immediately after the predicted signal peptide using the SOE-PCR technique (Heckman *et al.*, 2007). Transgenic parasite lines were generated by transfecting TGME49Δ*hpt* (MΔLuc) and NC-1Δ*hpt* parental strains with 50 µg each of HindIII-linearized pMAF1A or pMAF1B. Stable-expression lines were isolated by selection in MPA/Xanthine followed by limiting dilution in 96-well plates.

Chapter 4: For promoter luciferase constructs used in promoter-activity assays, ~1000bp of upstream sequences (i.e ~-1000 to +3) of individual paralogs were PCR-amplified from ME49 and VEG genomic DNA and cloned into the pENTR/D-Topo vector (Invitrogen) and subsequently transferred into the pDest-firefly vector by LR cloning. ROP43 misregulation construct was generated by PCR-amplification of ROP43 coding region from genomic DNA followed by cut-and-paste cloning into the pGRA1-HA-HPT vector using the restriction enzymes Nsi I and Nco I. Parental strains were transfected with 50ug of the plasmid and selected in MPA/Xanthine medium followed by limiting dilution. Primers used for cloning are listed in Table 6.1.

6.9 IMMUNOFLUORESCENCE ASSAYS AND CONFOCAL MICROSCOPY

Host cells were seeded on 12mm coverslips in 24-well plates containing coverslips and grown to about 80% confluency. NRK-mitoRFP cells were infected with *N. caninum* or *T. gondii* strains expressing GFP and incubated for 8hrs. MitoTracker staining was performed as follows: growth

medium on HFF monolayer was replaced with DMEM containing MitoTracker (Red CMXRos - Invitrogen) at a 30nM concentration and incubated for 30 min at 37^{o}C. Cells were then washed with PBS, infected with parasites in prewarmed DMEM and incubated for 4hrs at 37^{o}C. After incubation the infected cells were washed with PBS, fixed with 3% PFA in PBS for 15 min, and blocked in PBS containing 5% BSA and 0.2% Triton-X100 for permeabilization. For NRK-mitoRFP infected cells, coverslips were fixed with 3% PFA and either mounted directly or Hoechst-stained prior to mounting followed by visualization. Selective permeabilization of host cell plasma membrane with digitonin was performed as described elsewhere (Beckers *et al.*, 1994). Infected cells were grown for 24hrs and the most consistent results were obtained with 0.002% of digitonin. Immunostaining was performed using rat monoclonal anti-HA (3A10 clone, Roche) at 1:1000 for HA-tagged MAF1 and mouse monoclonal anti-SAG1 antibody at 1:2000 for SAG1. Images were captured on the FV1000 Olympus confocal and processed with ImageJ.

6.10 WESTERN BLOT ANALYSIS

Parasites were filtered away from host cell debris and lysed in 1X SDS lysis buffer. Proteins were resolved by SDS-PAGE, transferred onto nitrocellulose membrane and blocked for 1 hour in 5% (w/v) milk in TBS-T. Primary antibody incubation was performed in blocking buffer for 45min followed by 3 washes in TBS-T. Anti-HA and anti-MAF1 antibodies were used at 1:1000 while anti-SAG1 was used at 1:2000. Secondary antibody incubation was performed with Horseradish peroxidase-conjugated secondary antibodies to the respective primary antibodies, in blocking buffer for 45min. Bands were visualized with SuperSignal West Pico chemiluminescent substrate (Thermo Scientific). Densitometric analysis was performed using ImageJ.

6.11 *IN VIVO* CYTOKINE INDUCTION ASSAY

Tail-blood for Day 0 was collected immediately prior to infection. Four mice per group were infected with either empty-vector control (EV) or MAF1A-complemented ME49 (MAF1A) at a dose of 10^5 parasites/mouse via intraperitoneal injection. At each subsequent sampling time-point, parasite burden was measured in each mouse using bioluminescence imaging to confirm successful infection, after which mice were tail-bled to collect about 100 µl of whole blood. The blood was allowed to clot for 1 hour at 4^{o}C followed by microcentrifugation at 10,000 rpm for 10 min at 4^{o}C. The supernatant (serum) was collected and stored at -80^{o}C. Samples were thawed on ice immediately prior to measurement of cytokine levels on the Luminex mouse 26plex platform (Millipore, USA; service provided by the Luminex Core facility at the Hilman Cancer Center, University of Pittsburgh).

6.12 DUAL LUCIFERASE ASSAY

Promoter activities were tested by dual luciferase assay (Promega) as described elsewhere (Behnke *et al.*, 2008). Briefly, intracellular parasites were syringe lysed and $2x10^7$ each transfected with a combination of 40µg of promoter-firefly-luciferase plasmid and 20µg of pTub-renilla-luciferase plasmid. After electroporation, each sample was split equally into 2 T25

flasks seeded with HFFs and incubated for about 8 hours to allow parasites to recover and invade. For each transfection, the medium in one flask was replaced with bradyzoite induction medium (Buchholz *et al.*, 2011) while the medium on the other was replaced with regular growth medium. Flasks were then incubated at 37^{o}C for 48 hours. After incubation, cells were scraped, pelleted and resuspended in 100µl of 1X lysis buffer. To measure luciferase activity, 100µl of dual luciferase reagent was added to 20µl of each sample and light readings taken on the Centro LB 960 Microplate Luminometer. Firefly luciferase activity was first measured followed by Renilla luciferase activity. Three reactions were setup for each extract. Firefly luciferase readings were normalized to Renilla luciferase readings to account for variability in transfection efficiencies and then compared between bradyzoite-induced samples and tachyzoite samples.

6.13 *IN VITRO* GROWTH ASSAYS

Plaque assays were performed in 12-well plates seeded with HFFs. Intracellular parasites were syringe-lysed and inoculated onto 1 week old HFF monolayers at 500 parasites per well. Infected plates were incubated at 37^{o}C for 7 days undisturbed. Each infected well was washed with PBS, fixed with 100% ethanol for 5min and stained with crystal violet for another 5min. Wells were subsequently washed 3 times with PBS and air-dried. Plaques were visualized with an inverted microscope and images taken at 20X magnification. Plaques sizes (area of lysed host cells) were measured with ImageJ. To compare rates of bradyzoite differentiation, 12-mm coverslips were seeded with HFFs in 24-well plates and infected with freshly syringe-lysed parasites at a multiplicity of infection (MOI) of 0.5 and incubated at 37^{o}C for 4 hours to allow parasites to invade. The media was then replaced with bradyzoite induction media and incubated at 37^{o}C for

24, 48 or 72 hours, after which the cells were fixed with 3% paraformaldehyde, permeabilized with 0.2% Triton-100X and stained for BAG1 expression. Mouse anti-BAG1 antibodies was donated by Dr. Peter Bradley, UCLA. The percentage of BAG1-positive vacuoles were determine for a total of 5 randomly selected fields per coverslip and 2 coverslips were analyzed per strain for each time point. To compare intracellular multiplication rates, 24-well plates containing 12-mm coverslips were seeded with HFFs and infected with freshly syringe-lysed parasites at an MOI of 2. After 20 hours of incubation, cells were fixed, permeabilized and stained for SAG1, which outlines individual parasites. For 10 randomly selected 40X fields, the number of vacuoles and the number of parasites per vacuole were counted. Two coverslips were analyzed per strain.

6.14 TET-REGULATABLE EXPRESSION

pTub-YFP-TR-CAT and pRPS13(IV)-LacZ plasmids were obtained from Dick Schaap (van Poppel *et al.*, 2006). In order to generate the M:TR strain, 50ug of pTub-YFP-TR-CAT was linearized by NotI digest and transfected into ME49Δhpt:LUC (MΔ:Luc). Transfectants were selected by incubation in 20uM chloramphenicol in cDMEM for 4 passages followed by limiting dilution in a 96-well plate. Clones which were YFP-positive were selected and tested for TetR activity. To generate the pRPS13(IV)-ROP43 construct, the RPS13(IV) promoter was PCR-amplified from the pRPS13(IV)-LacZ plasmid and cloned into the pGRA1-ROP43-HA plasmid in place of the GRA1 promoter. This places the RPS13(IV) promoter upstream of the ROP43-HA coding sequence. The primers used for cloning are listed in Table 6.1.

6.15 CPRG ASSAY

Ten million tachyzoites of either MΔ:Luc or M:TR were each transfected with 30ug of pRPS13(IV)-LacZ. Five million transfected parasites were inoculated onto HFF seeded in a 24-well plate and incubated with or without 1ug/ml ATc in cDMEM for 24 hours. Wells were washed once with PBS and 300ul of CPRG assay buffer (100mM Hepes, 1mM MgSO4, 1% Triton 100-X, 5mM dithiothreitol) added. The plate was incubated at 50^{o}C for 1 hr. Lysates were clarified by microcentrifugation at 10,000RPM for 5min. 50ul of clarified lysate from each sample was diluted to 200ul with CPRG substrate buffer (1mM CPRG in assay buffer) and incubated for 16hrs at 37^{o}C for substrate conversion. Absorption at 562nm was measured using a microplate reader.

6.16 ANIMAL EXPERIMENTS

All mice experiments were performed with 4-8wk old BALB/C mice. All animal procedures in this study meet the standards of the American Veterinary Association and were approved locally under IACUC protocol #12010130.

Table 6.1: Primers used in this study

Forward (5'-3') Primer	Reverse (5'-3') Primer	Purpose
CCGgctagcCTACCAGTGAGAGGGCGAAG	CCCaagcttCACCTCTGGCTTGACTCTCC	Amplify left flank for EL3 knockout
TTActgcagCGAAGTGATTCGAGTGACGA	ATAgcggccgcCGTTGGGAAGAAGAGACGAG	Amplify right flank for EL3 knockout
TGTCTCTGTGGATCATCTATTCGT	AAACGCACTACTAAAGCGAAACTT	Query left flank for EL3 knockout
GTAGTCCGGGACGTCGTACGGGTAGA	GCAACCGTTTTGATAGATATAGGC	Query right flank for EL3 knockout
CCGatgcatATGA GGACGTCTGT TGCCCTGTT	GTCCAATTTccatggAACTTGGGTTTCA CCATTCCCTTGAACG	Cloning/Tagging and Southern blot probe for EL3
TCCAAGCTGCAAGCCATGACGTC	AGGCTTCGCTCGGTGAGGTCGGCTT	Verify BspEI site at left flank of EL3
CTACACGCAGTTCTGTGTTAGACAC	TTGCGTGGCTGTTGCATGCGTG	Verify BspEI site at right flank of EL3
TATTGAACTCCTACGTCGTTGTGT	TATCCGCTCACAATTCCACACCGTGC AATCTTATGTGAAAATAG	Amplify left flank for EL6 knockout
CTATTTTCACATAAGATTGCACGGT GTGGAATTGTGAGCGGATA	GGGTAAATCGAGTGATCTAGGAGAC CACCATCACCACGACTACA	Amplify HPT cassette for EL6 knockout
TGTAGTCGTGGTGATGGTGGTCTC CTAGATCACTCGATTTACCC	ATAAGAGGGGAGTTGCACAGTTAC	Amplify Right flank for EL6 knockout
AGTTGGCACATATTCTAGTGCTGA	AAGGTTTCGTGCTGATCATAACTT	Query left flank for EL6 knockout
GAACTACGGTGTTTGTTCCTTTCT	ATGCCTTTGGTCTTGATCTAATGT	Query right flank for EL6 knockout
AAACGCTCTAGGACAGATGCTC	TATCCGCTCACAATTCCACACGCTAT CAAGAGACACGAACAG	Amplify left flank for EL23
CTGTTCGTGTCTCTTGATAGCGTGT GGAATTGTGAGCGGATA	GAAAATATGTGTGGCAGCTCTGCCA CCATCACCACGACTACA	Amplify HPT cassette for EL23 knockout
TGTAGTCGTGGTGATGGTGGCAGA GCTGCCACACATATTTTC	CTGGGGACTTCACTTACAGGAC	Amplify right flank for EL23 knockout
CAGTTTTACGGGTTCTTCGATTAT	ACAAGAAAGTCATCAGCGTTGTAG	Query left flank for EL23 knockout
GCGTATGACATCCACAGAACTTAC	GGTCGAGAGGACAGAGTTAGTAGC	Query right flank for EL23 knockout
GATTCCCTACCGCAACTACATGG	TGGAGACAGCTGACAACCAG	Southern blot probe for EL30
GTTCatgcatATGGCCTTCTTCAGCAC CAGACGT	TGGCccatggCGCCGAAGAGGCATCTT GCAGCTGGCGAAT	Southern blot probe for EL45
CCGAAGTTGATGGCACATTATACC GG	CCGGCGAAGTAGCAGCCTCCTCCTA G	Southern blot probe for AMA1
CACCaagcttAGGGAT ACGAACAACT CGCTTATTAA G	GTCttaattaaTCAGTCCAGCATGCTAGC CAGATACG	Southern blot probe for EL4 (MAF1)

CAGCTTCCTGCACTCTCTCTGAGG	CGAAACAGTACAGACATATGACC	Verify ScaI site at left flank of EL4
CGAACAGGTCGAACGCGTTGTGCGAG	TGGCAGCAAGATGGACATGCAC	Verify ScaI site at right flank of EL4
CCaagcttCTGCGACGTGATCGTGGCAA	CCatgcat**CGCGTAGTCCGGGACGTCGTACGGGTA**ACCGGCGGTCAG	Primers for cloning N-term HA-tagged pMAF1A; promoter with signal sequence
CCatgcat**GGTGGT**CTAGGCAGTCAGATGTCGG	CCttaattaaTC AGTCCAGCATGCTAGCCAG	Primers for cloning N-term HA-tagged pMAF1A; region downstream of signal sequence
atgcatATGTGG CGCATCTGGAGATGCC	GTCttaattaaTCA GTCGCCTTGCGGAAACTTGTACTTCC	Primers for cloning MAF1B from cDNA
CACCaagcttAGGGAT ACGAACAACTCGCTTATTAA G	TCGCGTAGTCCGGGACGTCGTACGG GTAACCGGCGGTCAGAGCGCCCAGC	Amplify MAF1A promoter from pMAF1A
CGACGTCCCGGACTACGCGATGCA TGGTGGTCTAGGCAGTCAGATGTCGG	GTCttaattaaTCAGTCGCCTTGCGGAAA CTTGTACTTCC	Amplify MAF1B ORF for pMAF1B (N-term tagged MAF1B under MAF1 promoter

Restriction sites are in lowercase, and nucleotides 5' to restriction sites are to promote restriction enzyme recognition. Regions used for splicing by overlap extension PCR are underlined.

APPENDIX A

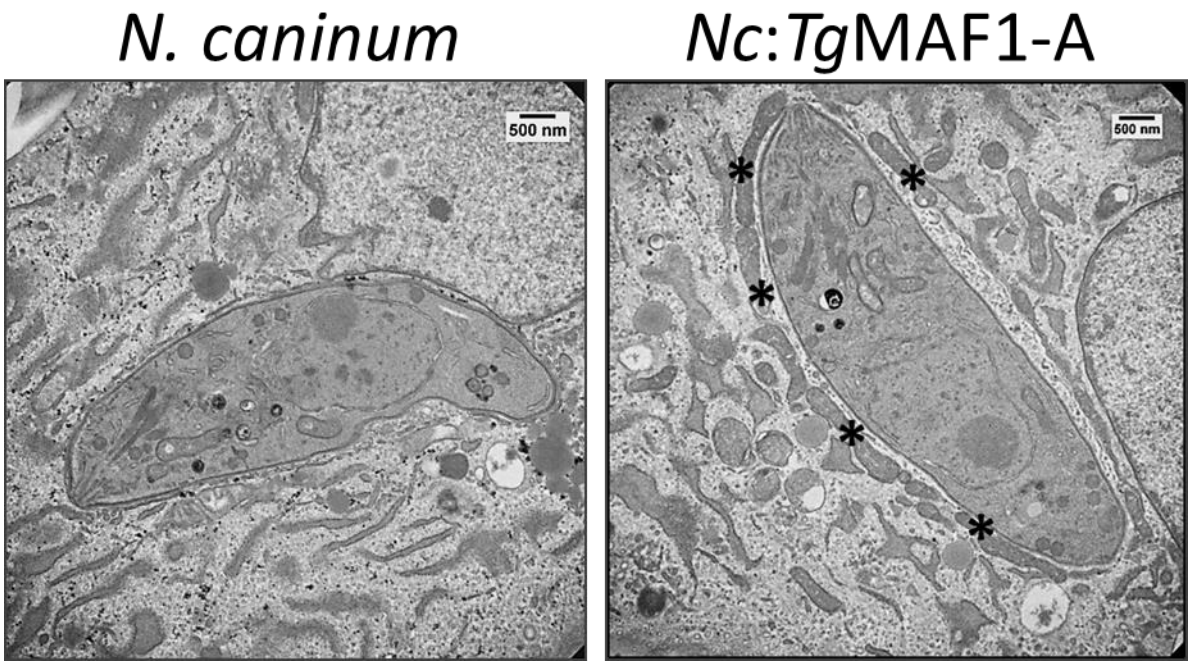


Figure A1: Transmission electron micrographs showing HMA in *N. caninum* expressing *T. gondii* MAF1A. Asterisks show host mitochondria associated with the parasitophorous vacuolar membrane.

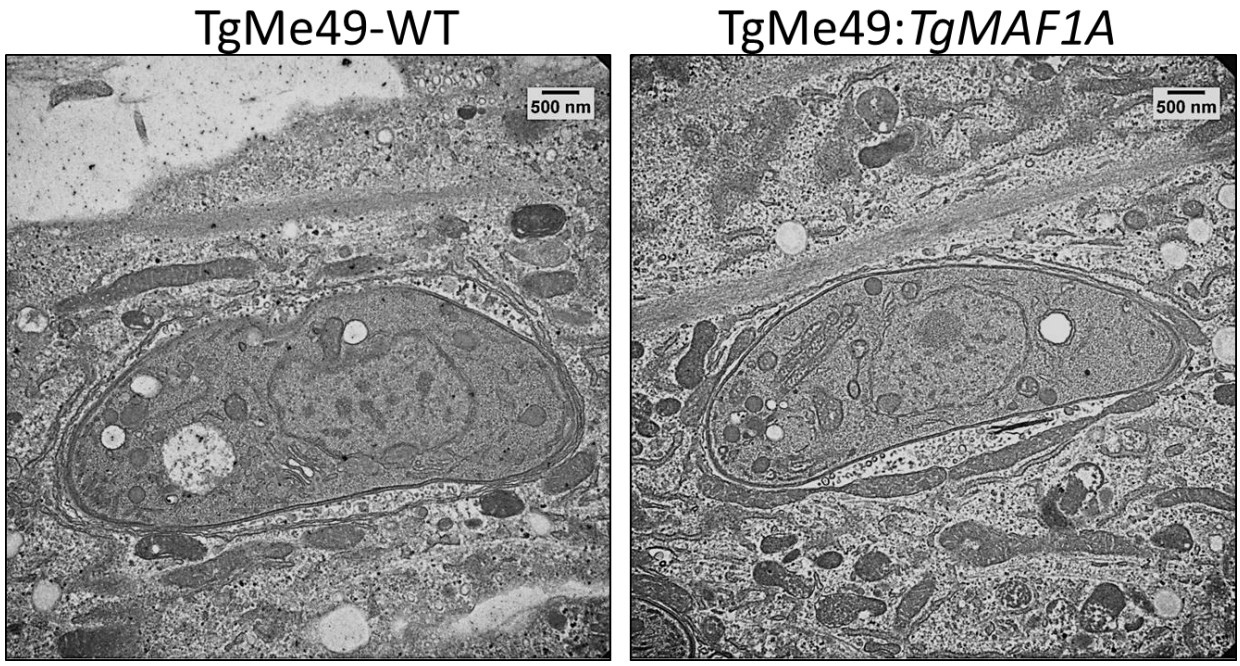


Figure A2: Transmission electron micrographs showing HMA in *T. gondii* Type II ME49 strain expressing *T. gondii* MAF1A from a Type I strain. Asterisks show host mitochondria associated with the parasitophorous vacuolar membrane.

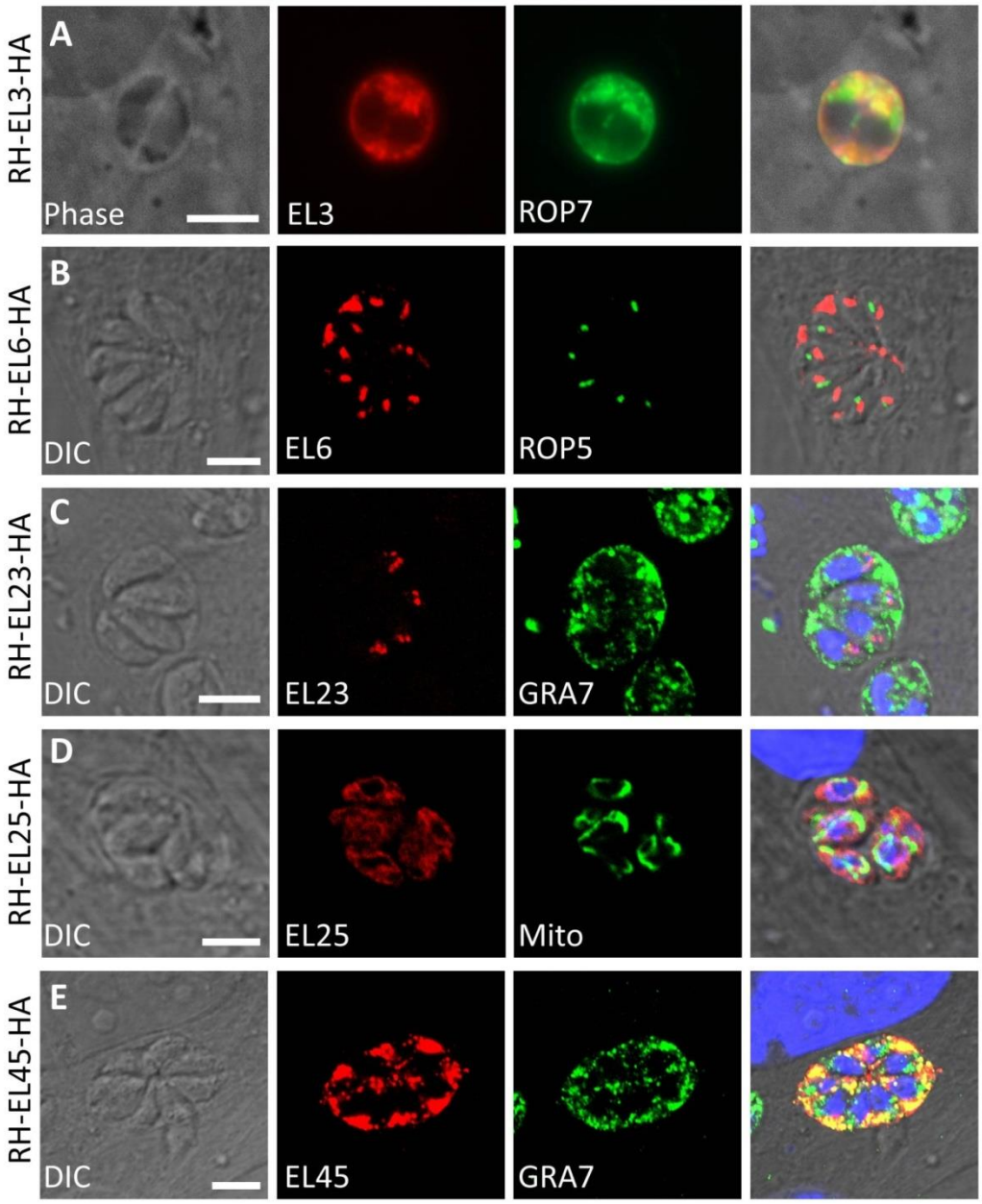


Figure A3: Immunofluorescence assays of *T. gondii* RH parasites transfected with GRA1 promoter-driven HA-tagged constructs for each of the indicated loci. Parasites were stained with anti-HA as well as the indicated secondary antibody to illustrate co-localization (or lack thereof) with the indicated marker. Images for locus *EL3* were taken using epifluorescence microscopy and the rest of the images were taken with a confocal microscope. Scale bar: 5 µm

BIBLIOGRAPHY

Almal, S.H. and Padh, H. (2012). Implications of gene copy-number variation in health and diseases. *Journal of human genetics* **57,** 6-13.

Araujo, F.G. and Slifer, T. (2003). Different strains of Toxoplasma gondii induce different cytokine responses in CBA/Ca mice. *Infection and immunity* **71,** 4171-4174.

Bahl, A., Davis, P.H., Behnke, M., Dzierszinski, F., Jagalur, M., Chen, F., *et al.* (2010). A novel multifunctional oligonucleotide microarray for Toxoplasma gondii. *BMC genomics* **11,** 603.

Barry, J.D., Marcello, L., Morrison, L.J., Read, A.F., Lythgoe, K., Jones, N., *et al.* (2005). What the genome sequence is revealing about trypanosome antigenic variation. *Biochemical Society transactions* **33,** 986-989.

Beckers, C.J., Dubremetz, J.F., Mercereau-Puijalon, O. and 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. *The Journal of cell biology* **127,** 947-961.

Behnke, M.S., Khan, A., Wootton, J.C., Dubey, J.P., Tang, K. and Sibley, L.D. (2011). Virulence differences in Toxoplasma mediated by amplification of a family of polymorphic pseudokinases. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 9631-9636.

Behnke, M.S., Radke, J.B., Smith, A.T., Sullivan, W.J., Jr. and White, M.W. (2008). The transcription of bradyzoite genes in Toxoplasma gondii is controlled by autonomous promoter elements. *Molecular microbiology* **68,** 1502-1518.

Berens, C. and Hillen, W. (2004). Gene regulation by tetracyclines. *Genetic engineering* **26,** 255-277.

Bergthorsson, U., Andersson, D.I. and Roth, J.R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proc Natl Acad Sci U S A* **104,** 17004-17009.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., *et al.* (2005). The genome of the African trypanosome Trypanosoma brucei. *Science* **309,** 416-422.

Bethke, L.L., Zilversmit, M., Nielsen, K., Daily, J., Volkman, S.K., Ndiaye, D., *et al.* (2006). Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of Plasmodium falciparum. *Molecular and biochemical parasitology* **150,** 10-24.

Boothroyd, J.C. (2009). Expansion of host range as a driving force in the evolution of Toxoplasma. *Memorias do Instituto Oswaldo Cruz* **104,** 179-184.

Boothroyd, J.C. (2013). Have it your way: how polymorphic, injected kinases and pseudokinases enable toxoplasma to subvert host defenses. *PLoS pathogens* **9,** e1003296.

Boothroyd, J.C. and Dubremetz, J.F. (2008). Kiss and spit: the dual roles of Toxoplasma rhoptries. *Nature reviews. Microbiology* **6,** 79-88.

Boyle, J.P., Rajasekar, B., Saeij, J.P., Ajioka, J.W., Berriman, M., Paulsen, I., *et al.* (2006). Just one cross appears capable of dramatically altering the population biology of a eukaryotic pathogen like Toxoplasma gondii. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 10514-10519.

Buchholz, K.R., Fritz, H.M., Chen, X., Durbin-Johnson, B., Rocke, D.M., Ferguson, D.J., *et al.* (2011). Identification of tissue cyst wall components by transcriptome analysis of in vivo and in vitro Toxoplasma gondii bradyzoites. *Eukaryotic cell* **10,** 1637-1647.

Bzymek, M. and Lovett, S.T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 8319-8325.

Carlton, J.M., Hirt, R.P., Silva, J.C., Delcher, A.L., Schatz, M., Zhao, Q., *et al.* (2007). Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. *Science (New York, N.Y.)* **315,** 207-212.

Carruthers, V. and Boothroyd, J.C. (2007a). Pulling together: an integrated model of Toxoplasma cell invasion. *Current opinion in microbiology* **10,** 83-89.

Carruthers, V.B. and Sibley, L.D. (1997). Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts. *European journal of cell biology* **73,** 114-123.

Carruthers, V.B. and Suzuki, Y. (2007b). Effects of Toxoplasma gondii infection on the brain. *Schizophrenia bulletin* **33,** 745-751.

Collantes-Fernandez, E., Arrighi, R.B., Alvarez-Garcia, G., Weidner, J.M., Regidor-Cerrillo, J., Boothroyd, J.C., *et al.* (2012). Infected dendritic cells facilitate systemic dissemination and transplacental passage of the obligate intracellular parasite Neospora caninum in mice. *PloS one* **7,** e32123.

Cook, A.J., Gilbert, R.E., Buffolano, W., Zufferey, J., Petersen, E., Jenum, P.A., *et al.* (2000). Sources of toxoplasma infection in pregnant women: European multicentre case-control

study. European Research Network on Congenital Toxoplasmosis. *BMJ (Clinical research ed.)* **321,** 142-147.

Coppens, I., Dunn, J.D., Romano, J.D., Pypaert, M., Zhang, H., Boothroyd, J.C. and Joiner, K.A. (2006). Toxoplasma gondii sequesters lysosomes from mammalian hosts in the vacuolar space. *Cell* **125,** 261-274.

Coppin, A., Dzierszinski, F., Legrand, S., Mortuaire, M., Ferguson, D. and Tomavo, S. (2003). Developmentally regulated biosynthesis of carbohydrate and storage polysaccharide during differentiation and tissue cyst formation in Toxoplasma gondii. *Biochimie* **85,** 353-361.

Crawford, M.J., Thomsen-Zieger, N., Ray, M., Schachtner, J., Roos, D.S. and Seeber, F. (2006). Toxoplasma gondii scavenges host-derived lipoic acid despite its de novo synthesis in the apicoplast. *The EMBO journal* **25,** 3214-3222.

Darde, M.L. (2008). Toxoplasma gondii, "new" genotypes and virulence. *Parasite* **15,** 366-371.

de Melo, E.J., de Carvalho, T.U. and de Souza, W. (1992). Penetration of Toxoplasma gondii into host cells induces changes in the distribution of the mitochondria and the endoplasmic reticulum. *Cell structure and function* **17,** 311-317.

de Moura, L., Bahia-Oliveira, L.M., Wada, M.Y., Jones, J.L., Tuboi, S.H., Carmo, E.H., *et al.* (2006). Waterborne toxoplasmosis, Brazil, from field to gene. *Emerging infectious diseases* **12,** 326-329.

Deitsch, K.W., Calderwood, M.S. and Wellems, T.E. (2001). Malaria. Cooperative silencing elements in var genes. *Nature* **412,** 875-876.

Dittwald, P., Gambin, T., Gonzaga-Jauregui, C., Carvalho, C.M., Lupski, J.R., Stankiewicz, P. and Gambin, A. (2013). Inverted low-copy repeats and genome instability--a genome-wide analysis. *Human mutation* **34,** 210-220.

Dobrowolski, J.M. and Sibley, L.D. (1996). Toxoplasma invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. *Cell* **84,** 933-939.

Dubey, J.P. (2009). History of the discovery of the life cycle of Toxoplasma gondii. *International journal for parasitology* **39,** 877-882.

Dubey, J.P., Graham, D.H., Dahl, E., Hilali, M., El-Ghaysh, A., Sreekumar, C., *et al.* (2003a). Isolation and molecular characterization of Toxoplasma gondii from chickens and ducks from Egypt. *Vet Parasitol* **114,** 89-95.

Dubey, J.P., Lindsay, D.S. and Speer, C.A. (1998). Structures of Toxoplasma gondii tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clinical microbiology reviews* **11,** 267-299.

Dubey, J.P. and Sreekumar, C. (2003b). Redescription of Hammondia hammondi and its differentiation from Toxoplasma gondii. *International journal for parasitology* **33,** 1437-1453.

Dunn, J.D., Ravindran, S., Kim, S.K. and 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. *Infection and immunity* **76,** 5853-5861.

Dunn, O.J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association* **56,** 52-64.

Dzierszinski, F., Mortuaire, M., Cesbron-Delauw, M.F. and Tomavo, S. (2000). Targeted disruption of the glycosylphosphatidylinositol-anchored surface antigen SAG3 gene in Toxoplasma gondii decreases host cell adhesion and drastically reduces virulence in mice. *Molecular microbiology* **37,** 574-582.

Echeverria, P.C., Matrajt, M., Harb, O.S., Zappia, M.P., Costas, M.A., Roos, D.S., *et al.* (2005). Toxoplasma gondii Hsp90 is a potential drug target whose expression and subcellular localization are developmentally regulated. *Journal of molecular biology* **350,** 723-734.

Fares, M.A., Keane, O.M., Toft, C., Carretero-Paulet, L. and Jones, G.W. (2013). The roles of whole-genome and small-scale duplications in the functional specialization of Saccharomyces cerevisiae genes. *PLoS genetics* **9,** e1003176.

Ferreira da Silva Mda, F., Barbosa, H.S., Gross, U. and Luder, C.G. (2008). Stress-related and spontaneous stage differentiation of Toxoplasma gondii. *Molecular bioSystems* **4,** 824-834.

Fleckenstein, M.C., Reese, M.L., Konen-Waisman, S., Boothroyd, J.C., Howard, J.C. and Steinfeldt, T. (2012). A Toxoplasma gondii pseudokinase inhibits host IRG resistance proteins. *PLoS biology* **10,** e1001358.

Fox, B.A. and Bzik, D.J. (2002). De novo pyrimidine biosynthesis is required for virulence of Toxoplasma gondii. *Nature* **415,** 926-929.

Fox, B.A., Ristuccia, J.G., Gigley, J.P. and Bzik, D.J. (2009). Efficient gene replacements in Toxoplasma gondii strains deficient for nonhomologous end joining. *Eukaryotic cell* **8,** 520-529.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., *et al.* (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. *Nature* **407,** 1018-1022.

Frenkel, J.K. and Dubey, J.P. (1975). Hammondia hammondi: A new coccidium of cats producing cysts in muscle of other mammals. *Science* **189,** 222-224.

Frenkel, J.K., Dubey, J.P. and Miller, N.L. (1970). Toxoplasma gondii in cats: fecal stages identified as coccidian oocysts. *Science* **167,** 893-896.

Friedman, R. and Hughes, A.L. (2003). The temporal distribution of gene duplication events in a set of highly conserved human gene families. *Molecular biology and evolution* **20,** 154-161.

Fritz, H.M., Buchholz, K.R., Chen, X., Durbin-Johnson, B., Rocke, D.M., Conrad, P.A. and Boothroyd, J.C. (2012). Transcriptomic analysis of toxoplasma development reveals many novel functions and structures specific to sporozoites and oocysts. *PLoS One* **7,** e29998.

Furtado, J.M., Winthrop, K.L., Butler, N.J. and Smith, J.R. (2013). Ocular toxoplasmosis I: parasitology, epidemiology and public health. *Clinical & experimental ophthalmology* **41,** 82-94.

Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., *et al.* (2008). ToxoDB: an integrated Toxoplasma gondii database resource. *Nucleic acids research* **36,** D553-556.

Gazzinelli, R.T., Wysocka, M., Hayashi, S., Denkers, E.Y., Hieny, S., Caspar, P., *et al.* (1994). Parasite-induced IL-12 stimulates early IFN-gamma synthesis and resistance during acute infection with Toxoplasma gondii. *Journal of immunology (Baltimore, Md. : 1950)* **153,** 2533-2543.

Gjini, E., Haydon, D.T., Barry, J.D. and Cobbold, C.A. (2012). The impact of mutation and gene conversion on the local diversification of antigen genes in African trypanosomes. *Mol Biol Evol* **29,** 3321-3331.

Goodswen, S.J., Kennedy, P.J. and Ellis, J.T. (2013). A review of the infection, genetics, and evolution of Neospora caninum: from the past to the present. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **13,** 133-150.

Grigg, M.E., Bonnefoy, S., Hehl, A.B., Suzuki, Y. and Boothroyd, J.C. (2001a). Success and virulence in Toxoplasma as the result of sexual recombination between two distinct ancestries. *Science (New York, N.Y.)* **294,** 161-165.

Grigg, M.E., Ganatra, J., Boothroyd, J.C. and Margolis, T.P. (2001b). Unusual abundance of atypical strains associated with human ocular toxoplasmosis. *The Journal of infectious diseases* **184,** 633-639.

Hayton, K., Gaur, D., Liu, A., Takahashi, J., Henschen, B., Singh, S., *et al.* (2008). Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of Plasmodium falciparum invasion. *Cell host & microbe* **4,** 40-51.

Heckman, K.L. and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature protocols* **2,** 924-932.

Horwitz, M.A. (1983). Formation of a novel phagosome by the Legionnaires' disease bacterium (Legionella pneumophila) in human monocytes. *The Journal of experimental medicine* **158,** 1319-1331.

Howard, J.C., Hunn, J.P. and Steinfeldt, T. (2011). The IRG protein-based resistance mechanism in mice and its relation to virulence in Toxoplasma gondii. *Current opinion in microbiology* **14,** 414-421.

Howe, D.K. and Sibley, L.D. (1995). Toxoplasma gondii comprises three clonal lineages: correlation of parasite genotype with human disease. *The Journal of infectious diseases* **172,** 1561-1566.

Hunter, C.A. and Remington, J.S. (1994). Immunopathogenesis of toxoplasmic encephalitis. *The Journal of infectious diseases* **170,** 1057-1067.

Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. *PLoS biology* **2,** E206.

Huynh, M.H. and Carruthers, V.B. (2009). Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. *Eukaryotic cell* **8,** 530-539.

Jackson, A.P., Gamble, J.A., Yeomans, T., Moran, G.P., Saunders, D., Harris, D., *et al.* (2009). Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome research* **19,** 2231-2244.

Jerome, M.E., Radke, J.R., Bohne, W., Roos, D.S. and White, M.W. (1998). Toxoplasma gondii bradyzoites form spontaneously during sporozoite-initiated development. *Infection and immunity* **66,** 4838-4844.

Jones, J.L., Dargelas, V., Roberts, J., Press, C., Remington, J.S. and Montoya, J.G. (2009). Risk factors for Toxoplasma gondii infection in the United States. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **49,** 878-884.

Jones, T.C. and Hirsch, J.G. (1972). The interaction between Toxoplasma gondii and mammalian cells. II. The absence of lysosomal fusion with phagocytic vacuoles containing living parasites. *The Journal of experimental medicine* **136,** 1173-1194.

Jung, C., Lee, C.Y. and Grigg, M.E. (2004). The SRS superfamily of Toxoplasma surface proteins. *International journal for parasitology* **34,** 285-296.

Kasper, L.H., Crabb, J.H. and Pfefferkorn, E.R. (1983). Purification of a major membrane protein of Toxoplasma gondii by immunoabsorption with a monoclonal antibody. *J Immunol* **130,** 2407-2412.

Kehrer-Sawatzki, H. and Cooper, D.N. (2007). Structural divergence between the human and chimpanzee genomes. *Human genetics* **120,** 759-778.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. *Genome research* **12,** 656-664.

Khan, A., Jordan, C., Muccioli, C., Vallochi, A.L., Rizzo, L.V., Belfort, R., Jr., *et al.* (2006). Genetic divergence of Toxoplasma gondii strains associated with ocular toxoplasmosis, Brazil. *Emerging infectious diseases* **12,** 942-949.

Khan, F., Tang, J., Qin, C.L. and Kim, K. (2002). Cyclin-dependent kinase TPK2 is a critical cell cycle regulator in Toxoplasma gondii. *Molecular microbiology* **45,** 321-332.

Kim, L., Butcher, B.A., Lee, C.W., Uematsu, S., Akira, S. and Denkers, E.Y. (2006). Toxoplasma gondii genotype determines MyD88-dependent signaling in infected macrophages. *Journal of immunology (Baltimore, Md. : 1950)* **177,** 2584-2591.

Kim, S.K. and Boothroyd, J.C. (2005). Stage-specific expression of surface antigens by Toxoplasma gondii as a mechanism to facilitate parasite persistence. *Journal of immunology (Baltimore, Md. : 1950)* **174,** 8038-8048.

Kim, S.K., Karasov, A. and Boothroyd, J.C. (2007). Bradyzoite-specific surface antigen SRS9 plays a role in maintaining Toxoplasma persistence in the brain and in host control of parasite replication in the intestine. *Infection and immunity*.

Kirkman, L.A., Weiss, L.M. and Kim, K. (2001). Cyclic nucleotide signaling in Toxoplasma gondii bradyzoite differentiation. *Infection and immunity* **69,** 148-153.

Kugelberg, E., Kofoid, E., Andersson, D.I., Lu, Y., Mellor, J., Roth, F.P. and Roth, J.R. (2010). The tandem inversion duplication in Salmonella enterica: selection drives unstable precursors to final mutation types. *Genetics* **185,** 65-80.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods* **9,** 357-359.

Lelu, M., Villena, I., Darde, M.L., Aubert, D., Geers, R., Dupuis, E., *et al.* (2012). Quantitative estimation of the viability of Toxoplasma gondii oocysts in soil. *Applied and environmental microbiology* **78,** 5127-5132.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078-2079.

Lu, J., Peatman, E., Tang, H., Lewis, J. and Liu, Z. (2012). Profiling of gene duplication patterns of sequenced teleost genomes: evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. *BMC genomics* **13,** 246.

Lynch, M. and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science (New York, N.Y.)* **290,** 1151-1155.

Malhotra, D. and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148,** 1223-1241.

Matsumoto, A., Bessho, H., Uehira, K. and Suda, T. (1991). Morphological studies of the association of mitochondria with chlamydial inclusions and the fusion of chlamydial inclusions. *Journal of electron microscopy* **40,** 356-363.

Meissner, M., Brecht, S., Bujard, H. and Soldati, D. (2001). Modulation of myosin A expression by a newly established tetracycline repressor-based inducible system in Toxoplasma gondii. *Nucleic acids research* **29,** E115.

Meissner, M., Schluter, D. and Soldati, D. (2002). Role of Toxoplasma gondii myosin A in powering parasite gliding and host cell invasion. *Science (New York, N.Y.)* **298,** 837-840.

Miller, C.M., Boulter, N.R., Ikin, R.J. and Smith, N.C. (2009). The immunobiology of the innate response to Toxoplasma gondii. *International journal for parasitology* **39,** 23-39.

Mineo, J.R. and Kasper, L.H. (1994). Attachment of Toxoplasma gondii to host cells involves major surface protein, SAG-1 (P30). *Experimental parasitology* **79,** 11-20.

Mitra, K. and Lippincott-Schwartz, J. (2010). Analysis of mitochondrial dynamics and functions using imaging approaches. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* **Chapter 4,** Unit 4 25 21-21.

Montoya, J.G. and Liesenfeld, O. (2004). Toxoplasmosis. *Lancet* **363,** 1965-1976.

Moran, G.P., Coleman, D.C. and Sullivan, D.J. (2011). Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. *Eukaryotic cell* **10,** 34-42.

Mugridge, N.B., Morrison, D.A., Jakel, T., Heckeroth, A.R., Tenter, A.M. and Johnson, A.M. (2000). Effects of sequence alignment and structural domains of ribosomal DNA on phylogeny reconstruction for the protozoan family sarcocystidae. *Molecular biology and evolution* **17,** 1842-1853.

Nakaar, V., Ngo, H.M., Aaronson, E.P., Coppens, I., Stedman, T.T. and Joiner, K.A. (2003). Pleiotropic effect due to targeted depletion of secretory rhoptry protein ROP2 in Toxoplasma gondii. *Journal of cell science* **116,** 2311-2320.

Niedelman, W., Gold, D.A., Rosowski, E.E., Sprokholt, J.K., Lim, D., Farid Arenas, A., *et al.* (2012). The rhoptry proteins ROP18 and ROP5 mediate Toxoplasma gondii evasion of the murine, but not the human, interferon-gamma response. *PLoS pathogens* **8,** e1002784.

Ogata, H., Renesto, P., Audic, S., Robert, C., Blanc, G., Fournier, P.E., *et al.* (2005). The genome sequence of Rickettsia felis identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol* **3,** e248.

Ohno, S. (1970) Evolution by Gene Duplication, Springer.

Ong, Y.C., Boyle, J.P. and Boothroyd, J.C. (2011). Strain-dependent host transcriptional responses to Toxoplasma infection are largely conserved in mammalian and avian hosts. *PLoS One* **6,** e26369.

Palmer, G.H. and Brayton, K.A. (2007). Gene conversion is a convergent strategy for pathogen antigenic variation. *Trends in parasitology* **23,** 408-413.

Parham, P., Norman, P.J., Abi-Rached, L. and Guethlein, L.A. (2011). Variable NK cell receptors exemplified by human KIR3DL1/S1. *J Immunol* **187,** 11-19.

Pasternak, N.D. and Dzikowski, R. (2009). PfEMP1: an antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite Plasmodium falciparum. *Int J Biochem Cell Biol* **41,** 1463-1466.

Pays, E., Van Meirvenne, N., Le Ray, D. and Steinert, M. (1981). Gene duplication and transposition linked to antigenic variation in Trypanosoma brucei. *Proceedings of the National Academy of Sciences of the United States of America* **78,** 2673-2677.

Peixoto, L., Chen, F., Harb, O.S., Davis, P.H., Beiting, D.P., Brownback, C.S., *et al.* (2010). Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses. *Cell host & microbe* **8,** 208-218.

Pernas, L. and Boothroyd, J.C. (2010). Association of host mitochondria with the parasitophorous vacuole during Toxoplasma infection is not dependent on rhoptry proteins ROP2/8. *International journal for parasitology* **40,** 1367-1371.

Pernas, L.F., Adomako-Ankomah, Y., Shastri, A.J., Treck, M., Boyle, J.P. and Boothroyd, J.C. (in revision). Toxoplasma alters host immune response via recruitment of host mitochondria. *PLoS biology*.

Pifer, R. and Yarovinsky, F. (2011). Innate responses to Toxoplasma gondii in mice and humans. *Trends in parasitology* **27,** 388-393.

Price, R.N., Uhlemann, A.C., Brockman, A., McGready, R., Ashley, E., Phaipun, L., *et al.* (2004). Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* **364,** 438-447.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841-842.

Rachinel, N., Buzoni-Gatel, D., Dutta, C., Mennechet, F.J., Luangsay, S., Minns, L.A., *et al.* (2004). The induction of acute ileitis by a single microbial antigen of Toxoplasma gondii. *J Immunol* **173,** 2725-2735.

Radke, J.R., Gubbels, M.J., Jerome, M.E., Radke, J.B., Striepen, B. and White, M.W. (2004). Identification of a sporozoite-specific member of the Toxoplasma SAG superfamily via genetic complementation. *Molecular microbiology* **52,** 93-105.

Radke, J.R., Guerini, M.N., Jerome, M. and White, M.W. (2003). A change in the premitotic period of the cell cycle is associated with bradyzoite differentiation in Toxoplasma gondii. *Molecular and biochemical parasitology* **131,** 119-127.

Radke, J.R., Striepen, B., Guerini, M.N., Jerome, M.E., Roos, D.S. and White, M.W. (2001). Defining the cell cycle for the tachyzoite stage of Toxoplasma gondii. *Molecular and biochemical parasitology* **115,** 165-175.

Reese, M.L. and Boothroyd, J.C. (2011a). A conserved non-canonical motif in the pseudoactive site of the ROP5 pseudokinase domain mediates its effect on Toxoplasma virulence. *The Journal of biological chemistry* **286,** 29366-29375.

Reese, M.L. and Boyle, J.P. (2012). Virulence without catalysis: how can a pseudokinase affect host cell signaling? *Trends in parasitology* **28,** 53-57.

Reese, M.L., Zeiner, G.M., Saeij, J.P., Boothroyd, J.C. and Boyle, J.P. (2011b). Polymorphic family of injected pseudokinases is paramount in Toxoplasma virulence. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 9625-9630.

Reid, A.J., Vermont, S.J., Cotton, J.A., Harris, D., Hill-Cawthorne, G.A., Konen-Waisman, S., *et al.* (2012). Comparative genomics of the apicomplexan parasites Toxoplasma gondii and Neospora caninum: Coccidia differing in host range and transmission strategy. *PLoS pathogens* **8,** e1002567.

Richards, F.O., Jr., Kovacs, J.A. and Luft, B.J. (1995). Preventing toxoplasmic encephalitis in persons infected with human immunodeficiency virus. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **21 Suppl 1,** S49-56.

Rieseberg, L.H. and Blackman, B.K. (2010). Speciation genes in plants. *Annals of botany* **106,** 439-455.

Robben, P.M., Mordue, D.G., Truscott, S.M., Takeda, K., Akira, S. and Sibley, L.D. (2004a). Production of IL-12 by macrophages infected with Toxoplasma gondii depends on the parasite genotype. *Journal of immunology (Baltimore, Md. : 1950)* **172,** 3686-3694.

Robben, P.M. and Sibley, L.D. (2004b). Food- and waterborne pathogens: you are (infected by) what you eat! *Microbes and infection / Institut Pasteur* **6,** 406-413.

Rodriguez, C., Afchain, D., Capron, A., Dissous, C. and Santoro, F. (1985). Major surface protein of Toxoplasma gondii (p30) contains an immunodominant region with repetitive epitopes. *European journal of immunology* **15,** 747-749.

Romano, J.D., de Beaumont, C., Carrasco, J.A., Ehrenman, K., Bavoil, P.M. and Coppens, I. (2013). Fierce competition between Toxoplasma and Chlamydia for host cell structures in dually infected cells. *Eukaryotic cell* **12,** 265-277.

Rosowski, E.E., Lu, D., Julien, L., Rodda, L., Gaiser, R.A., Jensen, K.D. and Saeij, J.P. (2011). Strain-specific activation of the NF-kappaB pathway by GRA15, a novel Toxoplasma gondii dense granule protein. *The Journal of experimental medicine* **208,** 195-212.

Rottmann, M., McNamara, C., Yeung, B.K., Lee, M.C., Zou, B., Russell, B., *et al.* (2010). Spiroindolones, a potent compound class for the treatment of malaria. *Science* **329,** 1175-1180.

Ryning, F.W. and Remington, J.S. (1978). Effect of cytochalasin D on Toxoplasma gondii cell entry. *Infection and immunity* **20,** 739-743.

Saeij, J.P., Boyle, J.P. and Boothroyd, J.C. (2005a). Differences among the three major strains of Toxoplasma gondii and their specific interactions with the infected host. *Trends in parasitology* **21,** 476-481.

Saeij, J.P., Boyle, J.P., Coller, S., Taylor, S., Sibley, L.D., Brooke-Powell, E.T., *et al.* (2006). Polymorphic secreted kinases are key virulence factors in toxoplasmosis. *Science* **314,** 1780-1783.

Saeij, J.P., Boyle, J.P., Grigg, M.E., Arrizabalaga, G. and Boothroyd, J.C. (2005b). Bioluminescence imaging of Toxoplasma gondii infection in living mice reveals dramatic differences between strains. *Infection and immunity* **73,** 695-702.

Saeij, J.P., Coller, S., Boyle, J.P., Jerome, M.E., White, M.W. and Boothroyd, J.C. (2007). Toxoplasma co-opts host gene expression by injection of a polymorphic kinase homologue. *Nature* **445,** 324-327.

Santoro, F., Charif, H. and Capron, A. (1986). The immunodominant epitope of the major membrane tachyzoite protein (P30) of Toxoplasma gondii. *Parasite immunology* **8,** 631-639.

Scanlon, M., Leitch, G.J., Visvesvara, G.S. and Shaw, A.P. (2004). Relationship between the host cell mitochondria and the parasitophorous vacuole in cells infected with Encephalitozoon microsporidia. *The Journal of eukaryotic microbiology* **51,** 81-87.

Scherf, A., Lopez-Rubio, J.J. and Riviere, L. (2008). Antigenic variation in Plasmodium falciparum. *Annual review of microbiology* **62,** 445-470.

Schmidt, J.M., Good, R.T., Appleton, B., Sherrard, J., Raymant, G.C., Bogwitz, M.R., *et al.* (2010). Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genet* **6,** e1000998.

Schwarz, J.A., Fouts, A.E., Cummings, C.A., Ferguson, D.J. and Boothroyd, J.C. (2005). A novel rhoptry protein in Toxoplasma gondii bradyzoites and merozoites. *Molecular and biochemical parasitology* **144,** 159-166.

Seeber, F. and Boothroyd, J.C. (1996). Escherichia coli beta-galactosidase as an in vitro and in vivo reporter enzyme and stable transfection marker in the intracellular protozoan parasite Toxoplasma gondii. *Gene* **169,** 39-45.

Sibley, L.D. (2003). Toxoplasma gondii: perfecting an intracellular life style. *Traffic* **4,** 581-586.

Sibley, L.D. and Ajioka, J.W. (2008). Population structure of Toxoplasma gondii: clonal expansion driven by infrequent recombination and selective sweeps. *Annual review of microbiology* **62,** 329-351.

Sinai, A.P. and Joiner, K.A. (2001). The Toxoplasma gondii protein ROP2 mediates host organelle association with the parasitophorous vacuole membrane. *The Journal of cell biology* **154,** 95-108.

Sinai, A.P., Webster, P. and Joiner, K.A. (1997). Association of host cell endoplasmic reticulum and mitochondria with the Toxoplasma gondii parasitophorous vacuole membrane: a high affinity interaction. *Journal of cell science* **110 (Pt 17),** 2117-2128.

Singh, U., Brewer, J.L. and Boothroyd, J.C. (2002). Genetic analysis of tachyzoite to bradyzoite differentiation mutants in Toxoplasma gondii reveals a hierarchy of gene induction. *Molecular microbiology* **44,** 721-733.

Sokal, R.R. and Rohlf, F.J. (2012) Biometry: the principles and practice of statistics in biological research New York, W.H. Freeman.

Steinfeldt, T., Konen-Waisman, S., Tong, L., Pawlowski, N., Lamkemeyer, T., Sibley, L.D., *et al.* (2010). Phosphorylation of mouse immunity-related GTPase (IRG) resistance proteins is an evasion strategy for virulent Toxoplasma gondii. *PLoS biology* **8,** e1000576.

Stockdale, C., Swiderski, M.R., Barry, J.D. and McCulloch, R. (2008). Antigenic variation in Trypanosoma brucei: joining the DOTs. *PLoS biology* **6,** e185.

Su, C., Evans, D., Cole, R.H., Kissinger, J.C., Ajioka, J.W. and Sibley, L.D. (2003). Recent expansion of Toxoplasma through enhanced oral transmission. *Science (New York, N.Y.)* **299,** 414-416.

Sudmant, P.H., Kitzman, J.O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., *et al.* (2010). Diversity of human copy number variation and multicopy genes. *Science (New York, N.Y.)* **330,** 641-646.

Sullivan Jr, W.J., Smith, A.T. and Joyce, B.R. (2009). Understanding mechanisms and the role of differentiation in pathogenesis of Toxoplasma gondii: a review. *Memorias do Instituto Oswaldo Cruz* **104,** 155-161.

Talevich, E. and Kannan, N. (2013). Structural and evolutionary adaptation of rhoptry kinases and pseudokinases, a family of coccidian virulence factors. *BMC evolutionary biology* **13,** 117.

Tanaka, K.M., Takahasi, K.R. and Takano-Shimizu, T. (2009). Enhanced fixation and preservation of a newly arisen duplicate gene by masking deleterious loss-of-function mutations. *Genetics research* **91,** 267-280.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al.* (2003). The COG database: an updated version includes eukaryotes. *BMC bioinformatics* **4,** 41.

Taylor, S., Barragan, A., Su, C., Fux, B., Fentress, S.J., Tang, K., *et al.* (2006). A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen Toxoplasma gondii. *Science (New York, N.Y.)* **314,** 1776-1780.

Templeton, T.J. (2009). The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, Plasmodium falciparum. *Molecular and biochemical parasitology* **166,** 109-116.

Tham, W.H., Healer, J. and Cowman, A.F. (2012). Erythrocyte and reticulocyte binding-like proteins of Plasmodium falciparum. *Trends in parasitology* **28,** 23-30.

Treeck, M., Sanders, J.L., Elias, J.E. and Boothroyd, J.C. (2011). The Phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii Reveal Unusual Adaptations Within and Beyond the Parasites' Boundaries. *Cell Host Microbe* **10,** 410-419.

van Poppel, N.F., Welagen, J., Duisters, R.F., Vermeulen, A.N. and Schaap, D. (2006). Tight control of transcription in Toxoplasma gondii using an alternative tet repressor. *International journal for parasitology* **36,** 443-452.

Van, T.T., Kim, S.K., Camps, M., Boothroyd, J.C. and Knoll, L.J. (2007). The BSR4 protein is up-regulated in Toxoplasma gondii bradyzoites, however the dominant surface antigen recognised by the P36 monoclonal antibody is SRS9. *International journal for parasitology* **37,** 877-885.

Vonlaufen, N., Guetg, N., Naguleswaran, A., Muller, N., Bjorkman, C., Schares, G., *et al.* (2004). In vitro induction of Neospora caninum bradyzoites in vero cells reveals differential antigen expression, localization, and host-cell recognition of tachyzoites and bradyzoites. *Infection and immunity* **72,** 576-583.

Walzer, K.A., Adomako-Ankomah, Y., Dam, R.A., Herrmann, D.C., Schares, G., Dubey, J.P. and Boyle, J.P. (2013). Hammondia hammondi, an avirulent relative of Toxoplasma gondii, has functional orthologs of known T. gondii virulence genes. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 7446-7451.

Wasmuth, J.D., Pszenny, V., Haile, S., Jansen, E.M., Gast, A.T., Sher, A., *et al.* (2012). Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of Toxoplasma virulence. *mBio* **3**.

Weiss, L.M. and Kim, K. (2000). The development and biology of bradyzoites of Toxoplasma gondii. *Frontiers in bioscience : a journal and virtual library* **5,** D391-405.

Weiss, L.M., Ma, Y.F., Takvorian, P.M., Tanowitz, H.B. and Wittner, M. (1998). Bradyzoite development in Toxoplasma gondii and the hsp70 stress response. *Infection and immunity* **66,** 3295-3302.

Wendte, J.M., Miller, M.A., Lambourn, D.M., Magargal, S.L., Jessup, D.A. and Grigg, M.E. (2010). Self-mating in the definitive host potentiates clonal outbreaks of the apicomplexan parasites Sarcocystis neurona and Toxoplasma gondii. *PLoS genetics* **6,** e1001261.

West, A.P., Shadel, G.S. and Ghosh, S. (2011). Mitochondria in innate immune responses. *Nature reviews. Immunology* **11,** 389-402.

Witkowski, B., Nicolau, M.L., Soh, P.N., Iriart, X., Menard, S., Alvarez, M., *et al.* (2010). Plasmodium falciparum isolates with increased pfmdr1 copy number circulate in West Africa. *Antimicrobial agents and chemotherapy* **54,** 3049-3051.

Xie, C. and Tammi, M.T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC bioinformatics* **10,** 80.

Yang, N., Farrell, A., Niedelman, W., Melo, M., Lu, D., Julien, L., *et al.* (2013). Genetic basis for phenotypic differences between different Toxoplasma gondii type I strains. *BMC genomics* **14,** 467.

Yang, S. and Parmley, S.F. (1997). Toxoplasma gondii expresses two distinct lactate dehydrogenase homologous genes during its life cycle in intermediate hosts. *Gene* **184,** 1-12.

Yoon, S., Xuan, Z., Makarov, V., Ye, K. and Sebat, J. (2009). Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome research* **19,** 1586-1592.