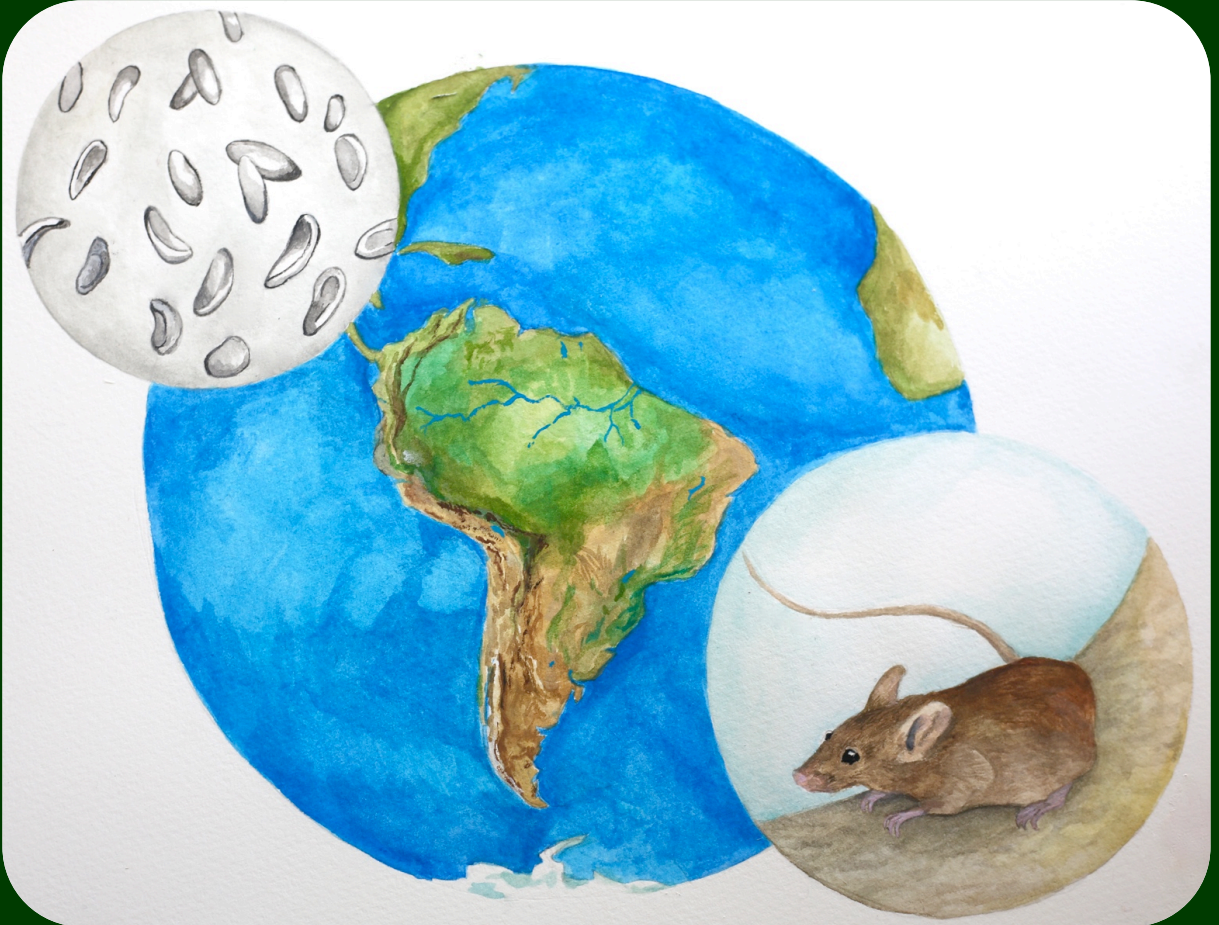


# Immunity-related GTPases (IRGs) in the house mouse and the parasite *Toxoplasma gondii* in South America

Martha Catalina Alvarez Meneses



Dissertation presented to obtain the Ph.D degree in Integrative Biology and Biomedicine  
Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,  
January, 2021



UNIVERSIDADE  
**NOVA**  
DE LISBOA

Oeiras, January, 2021

## Immunity-related GTPases (IRGs) in the house mouse and the parasite *Toxoplasma gondii* in South America

Martha Catalina  
Alvarez Meneses




ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal  
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

**[www.itqb.unl.pt](http://www.itqb.unl.pt)**

# Immunity-related GTPases (IRGs) in the house mouse and the parasite *Toxoplasma gondii* in South America

Martha Catalina Alvarez Meneses

Dissertation presented to obtain the Ph.D degree in Integrative Biology and Biomedicine  
Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by:  INSTITUTO  
GULBENKIAN  
DE CIÊNCIA

Oeiras, January, 2021



**Cover:** Artistic representation of *Toxoplasma gondii* tachyzoites (top left), a Brazilian *Mus musculus domesticus* (bottom right) and the South American continent (center) by Dr. Katya Mack (2020).

## **Declaration**

This thesis is the result of the work that I developed as a PhD student at the Instituto Gulbenkian de Ciência, supervised by Professor Jonathan Howard and co-supervised by Dr. Luis Teixeira under the scope of the Integrative Biology and Biomedicine Programme from January 2015 to December 2020. Most of the work was carried out at IGC except for field work (samples collection) in different regions in Portugal, the isolation of cells from wild-derived mice at the University of California, USA and the expression of *Irgb2-b1* alleles in *CIM<sub>Irgb2-b1KO</sub>* at Institute of Virology, Medical Center University of Freiburg, Germany. Author contributions are briefly described in a paragraph after the first page of each chapter. Our laboratory was supported by central funds of the Instituto Gulbenkian de Ciência, the Sonderforschungsbereiche 670 and 680 and Schwerpunkt 1399 of the Deutsche Forschungsgemeinschaft.

## **Summary**

Adaptation to local pathogens is critical for co-evolution within the context of host-pathogen interactions. Parasites constitute a constant threat to the immune system of a host and when a virulent parasite overcomes the immunological response and kills the host, it might look like the parasite has won the battle. However, this is far from being true. Although, it is in the interest of both host and parasite to prolong their encounter, when a parasite kills its current host the chances to pass to a new host are remarkably reduced. Virulence in this case can be considered as a failure of co-adaptation. The ubiquitous intracellular Apicomplexan parasite *Toxoplasma gondii* has an unusual life cycle, where it reproduces sexually inside of its definite host (true cats). Cats release in their feces oocysts that sporulate in soil and become highly infectious for foraging animals, like the house mouse (*Mus musculus*). Avirulent *T. gondii* strains are controlled by the immunity-related GTPases (IRG proteins) mouse resistance system. However, virulent strains secrete polymorphic kinases and pseudokinases (ROP18 and ROP5 proteins) that inactivate IRG proteins and lead to mouse death. Previous works have reported the existence of wild mice that can counteract virulent *T. gondii* strains by the production of the highly polymorphic IRG proteins (Irgb2-b1) that inhibit the ROP kinases. Here we focused our work on the ecological and evolutionary relationship between *M. musculus* and its recent confrontation against highly virulent South American *T. gondii* strains.

We began our study by analyzing the transcriptome of IFN $\gamma$  induced cell lines from wild-caught and wild-derived European (Portuguese) and South American (Brazilian) mice. We found a high prevalence of the Irgb2-b1<sub>PWK</sub> allele in Brazilian mice, whereas European mice display a more diverse set of Irgb2-b1 alleles. By looking for signs of bottleneck events or selection pressure in the Brazilian mouse population that could explain the high

prevalence of the *Irgb2-b1*<sub>PWK</sub> allele, we found high genetic diversity outside the *Irgb2-b1* locus as well in other genetic markers along the genome.

We then hypothesized that the *Irgb2-b1*<sub>PWK</sub> allele might have the ability to confer resistance against infection with highly virulent South American *T. gondii* strains. We confirmed the high expression of the *Irgb2-b1*<sub>PWK</sub> allele, as well as its capacity to efficiently load onto the parasitophorous vacuole membrane (PVM) of Brazilian cells infected with local *T. gondii* strains. Our data showed that expression of the *Irgb2-b1*<sub>PWK</sub> protein in CIM<sub>Irgb2-b1KO</sub> cells confers full resistance to type I strains and virulent Brazilian strains. We also demonstrated that European mice (carriers of the *Irgb2-b1*<sub>PWK</sub> allele) resist the RH strain but are susceptible to South American *T. gondii* strains. On the other hand, Brazilian mice are significantly more resistant to South American strains than their European counterparts.

Finally, by using a whole-genome analysis, we found that the three Brazilian *T. gondii* strains used for this study display high genetic similarity. Although differences along the genome can be found, the ROP18 and ROP5 genes were identical among these strains. We identified the ROP5 isoform types present in these strains as well as their phylogenetic relationships with other strains. By the development of a PCR-RFLP based protocol, we managed to accurately genotype those Brazilian strains, which we were not able to do by the standard protocols given the high similarity among them.

In conclusion, our results suggested that Brazilian mouse populations have been exposed to selection pressure on standing variation, possibly associated with fitness advantage (resistance) to highly virulent local *T. gondii* strains.

## **Resumo**

A adaptação a patógenos locais é crítica para a coevolução das interações patógeno-hospedeiro. Os parasitas constituem uma ameaça contínua ao sistema imunológico do hospedeiro: quando um parasita virulento ultrapassa a resposta imunológica e mata o hospedeiro, o parasita ganha a batalha, aparentemente. No entanto, este desfecho não é vantajoso para o parasita. O hospedeiro e o parasita têm vantagens ao prolongar seu encontro, e quando um parasita mata o hospedeiro, as hipóteses de passar para um novo hospedeiro reduzem-se consideravelmente. Assim, a virulência pode ser considerada uma falha de co-adaptação. O parasita apicomplexo *Toxoplasma gondii* (intracelular e ubíquo) tem um ciclo de vida incomum, reproduzindo-se sexualmente dentro de seu hospedeiro definitivo (qualquer felídeo verdadeiro). Os felídeos defecam, libertando nas fezes os oocistos de *T. gondii* que esporulam no solo, e que são altamente infecciosos para animais forrageiros, como o rato doméstico, também conhecido como murganho (*Mus musculus*). As estirpes avirulentas de *T. gondii* são controladas por um sistema de resistência do rato doméstico, as GTPases relacionadas com imunidade (proteínas IRG). No entanto, estirpes virulentas do parasita secretam cinases polimórficas e pseudocinases (proteínas ROP18 e ROP5) que inativam as proteínas IRG e levam à morte do animal. Trabalhos anteriores relataram a existência de ratos selvagens capazes de neutralizar estirpes virulentas de *T. gondii* através da produção de proteínas IRG altamente polimórficas (Irgb2-b1) que inibem as cinases ROP. No presente estudo, focamos o nosso interesse na relação ecológica e evolutiva entre *M. musculus* e seu recente confronto com estirpes Sul-americanas de *T. gondii* altamente virulentas.

Iniciamos o nosso estudo com a análise do transcriptoma de linhas celulares induzidas por IFN $\gamma$ , que foram previamente estabelecidas a partir de ratos selvagens capturados na Europa (portugueses) e América do Sul

(brasileiros), e de ratos derivados de selvagens. Encontrámos uma prevalência predominante do alelo *Irgb2-b1<sub>PWK</sub>* em ratos domésticos brasileiros, enquanto ratos domésticos europeus apresentam um conjunto mais diverso de alelos *Irgb2-b1*. No despiste de eventos de gargalo ou pressão selectiva na população brasileira de ratos domésticos, que poderiam explicar a notável prevalência do alelo *Irgb2-b1<sub>PWK</sub>*, encontrámos uma elevada diversidade genética fora do locus *Irgb2-b1*, bem como noutros marcadores genéticos ao longo do genoma.

De seguida, formulámos a hipótese de que o alelo *Irgb2-b1<sub>PWK</sub>* poderia conferir resistência contra a infecção com estirpes sul-americanas de *T. gondii* altamente virulentas. Confirmámos os elevados níveis de expressão do alelo *Irgb2-b1<sub>PWK</sub>*, bem como sua capacidade eficiente de recrutamento para a membrana do vacúolo do parasita (PVM) em células de ratos brasileiros derivados de selvagens, infectadas com *T. gondii* isolados localmente. Os nossos dados mostraram que a expressão da proteína *Irgb2-b1<sub>PWK</sub>* em células CIM<sub>*Irgb2-b1*KO</sub> confere resistência total a estirpes do tipo I e estirpes brasileiras virulentas de *T. gondii*. Também demonstrámos que os ratos domésticos europeus (portadores do alelo *Irgb2-b1<sub>PWK</sub>*) resistem à estirpe RH, mas são susceptíveis às estirpes sul-americanas de *T. gondii*. Por outro lado, os ratos domésticos brasileiros são significativamente mais resistentes às estirpes sul-americanas do que os ratos domésticos europeus.

Finalmente, recorrendo a uma análise do genoma total, descobrimos que as três estirpes brasileiras de *T. gondii* usadas para este estudo apresentam alta similaridade genética. Embora tenhamos identificado diferenças ao longo do genoma, os genes *ROP18* e *ROP5* são idênticos entre essas estirpes. Identificamos as diferentes isoformas de *ROP5* presentes nessas estirpes, bem como as suas relações filogenéticas com outras estirpes de *T. gondii*. Na ausência de protocolos padrão que permitissem distinguir as

variantes brasileiras, devido à sua similaridade genética, desenvolvemos um protocolo baseado em PCR-RFLP. Este novo protocolo permitiu genotipar com precisão e distinguir as variantes brasileiras de *T. gondii*.

Em conclusão, os nossos resultados sugerem que as populações brasileiras de ratos domésticos foram expostas à pressão de seleção numa variação contínua, possivelmente associada à vantagem da aptidão (resistência) contra estirpes locais de *T. gondii* altamente virulentas.

## **Acknowledgments**

When I decided to move from Colombia to Portugal more than 5 years ago, I was not fully aware of all the challenging but also wonderful moments that pursuing a Ph.D. would bring, but I am very happy that I made the decision. I would like to use this section to thank everyone who supported and trusted me during this long journey.

First and foremost, I would like to thank my supervisor **Jonathan Howard**, who showed interest in my work the first time I contacted him. **Jonathan**, thank you for choosing me to join your lab and for believing I had the necessary skills to work on a project together. Your unconditional support, not only as a scientist but also as a wonderful mentor and human being allowed me to successfully complete the challenging process of pursuing a Ph.D. I will always appreciate that you pushed me to do better and think harder. You also encouraged me to discuss my work with the most important scientists in our field and you stimulated me to do many exciting and diverse things: from catching wild mice in the field to perusing highly complex bioinformatics approaches for complex data analyses. Thanks for showing me the importance to be passionate about science and for teaching me how to stay focused and never losing my scientific objectives. I really appreciate that you listened to my “crazy” ideas about new projects, collaborations, and possible experiments. You always patiently guided me towards what was the right thing to do. I am now looking forward to applying everything I learned from you on the long and exciting journey ahead.

None of the experiments for my Ph.D. would have been possible without the outstanding support from the current and past members of the **Howard Lab in Lisbon**. **Claudia**, you are the best lab manager, wild-mice handler supporter, colleague, and friend that a Ph.D. student could ever ask for. Thank you for the uncountable hours we spent together planning



experiments and later trying to make sense out of them. I'm sure you are one of the few people in the world who understands how hard, terrifying but also fun is to work with wild-derived mice. Thanks for all the "special times" we had together working with these animals. I will always appreciate your unconditional support and advice during these last 5 years. **Ana Lina**, thank you for sharing this PhD journey with me and for being such a good friend. You always inspired me to be more committed and to give my best to everything I had to do both outside and inside the lab. You showed me that nothing can stop you when you want to pursue your dream. Also, thank you for helping me to improve my Portuguese; I know it is not always easy listening to someone struggling with a new language (Muito obrigada!). **Joana**, you always had a word of advice for me and you cheered me to continue, even though everything looked terrible with my experiments and I was about to give up. I would like to thank you not only for your scientific feedback, but to always have a friendly smile and to open some free time on your schedule to talk about life. **Carolina**, you were the first person to show me some of the protocols in the lab, and even these days I try to follow your advice when I do them. Thank you for always making me laugh at your jokes about my "Mexican" origin.

I want to also thank the members of the **Howard lab in Cologne** who welcomed me to their lab when I was just a newbie on this long Ph.D. journey. **Ben**, you were my "mouse sensei", I never imagined myself catching wild mice in farms around the great Lisbon area, but thanks to your enthusiasm and love for science, it is one of the things I enjoyed the most about my Ph.D. Thank you for always getting some free time to talk to me and give me your honest feedback, many times I felt you were one of the few people in the lab that could truly understand my struggles. You will always be a role model for me. **Tao**, thanks to you I can call myself now a "bioinformatics enthusiast". I know it is not easy to explain many of the very complex things you do, but you did it for me and I could not be more grateful

to you because of that. I will always appreciate your patience as a mentor, especially when it looked like I would never be able to understand how to do some of the analysis. Thank you also for your unconditional friendship and for all the great times we had together. **Tobi**, it has been a pleasure to collaborate with you and the members of your wonderful lab. Thank you for your helpful comments and the very engaging discussions we had about my data. **Steffy**, **Jelena**, and **Helen** thank you for your scientific feedback during our joint lab meetings, those were crucial at the beginning of my project.

I would also like to thank my co-supervisor, **Luis Teixeira**. Thank you so much for accepting the co-supervision of my project, even though it was a bit outside of the scope of your research. You were always willing to provide useful comments, help, and suggestions for experiments and analysis. I will always appreciate the opportunity you gave me to actively participate in lab meetings, retreats, and in engaging discussions with the members of your lab. It was always a bit challenging but very exciting whenever I had to share my latest results with your lab. They were always willing to help with bioinformatics and statistic analysis that would improve my project and make my data significantly more complete. I would also like to say thanks to the present and past members of the HMI lab, especially to **Rita**, **Ines**, **Elves**, **Gonçalo**, **Sergio**, **Rupinder**, **Ana**, **Marta**, **Miguel**, **Filipe**, and **Catarina** for being such great colleagues in the lab but also good friends.

At IGC, I would also like to thank my thesis committee members **Ivo Chelo** and **Carlos Penha Gonçalves** for the very fruitful discussions we had together. I appreciate your willingness to meet to talk not only about my project but also about life. Special thanks to the IGC Ph.D. program for accepting me onto the IBB2015 program and to its former director **Élio Sucena** and current director **Jorge Carneiro**. I will always be in debt to **Manuela Cordeiro** and **Ana Aranda** for everything you did for me these

years; you both were like second moms for us the Ph.D. students at IGC. My dear **IBBs 2015**, you were wonderful colleagues during our initial very challenging six months of classes and you are now my friends for life. Thank you.

Outside of IGC, I would like to thank our collaborators from the University of California Berkeley, **Michael Nachman** and **Ellen Robey**. I will always appreciate that you have opened your labs for me and allowed me to successfully obtain some of the cell lines that were crucial for the development of this project. Your lab members were also very kind to me and some of them are now good friends of mine.

También quisiera agradecerles que a todas esas personas que hicieron mi vida más fácil y amena dentro y fuera del instituto. A mis compatriotas y amigas muy especiales, **Lina y Vero**, ustedes fueron las primeras en darme todo su apoyo cuando llegué aquí. Me hicieron sentir como en casa y eso es algo que nunca voy a olvidar. Gracias por su amistad incondicional, me alegro haber hecho parte también de muchas etapas importantes en sus vidas. A mi querida “Morenita”, **Luna**, gracias por tu amistad y tus consejos, no solo científicos si no también personales. Le doy gracias a la vida por haberme permitido conocer personas tan especiales como tu que me ayudaron a no desistir en estos largos años del PhD.

Muchas gracias también a mis amigos de la Universidad de los Andes en Colombia que hoy en día continúan siendo mis amigos y que me ayudaron mucho en estos años del PhD a no desconectarme de mi país. **Margarita**, tantos años de amistad especial. Fue realmente único poder hablar contigo en algunos momentos casi a diario sobre nuestros logros, pero también las dificultades que tuvimos haciendo nuestros doctorados. Gracias por tu apoyo y amistad incondicionales. **Andrés**, a pesar de no hablar tan

seguido, creo que siempre fuiste una gran inspiración para continuar con mi PhD. Gracias por siempre preguntar como iban las cosas

El más grande agradecimiento va para mi familia, a quienes dedico esta tesis. Papito y mamita, **Martha** y **Mauricio**, gracias por su apoyo incondicional durante todos estos años que he estado lejos de casa, no fue fácil tomar la decisión de vivir tan lejos de ustedes, pero el poder escuchar lo orgullosos que ustedes están de mi cada día me permitió seguir adelante. Gracias por todos los sacrificios que hicieron por mi que me permitieron llegar al lugar a donde estoy hoy, no sería nadie sin ustedes. Gracias a mi hermano, **Diego**, por todo tu apoyo y confianza en mi, siempre recuerda que estoy muy orgullosa de ti. A mis abuelitas, especialmente a mi abuela **Lilia**, que soñaba con el día en que llegara a su casa con la buena noticia que ya había terminada mi doctorado, no alcanzaste a estar aquí para verlo, pero sé que desde donde estés vas a estar muy orgullosa de mí.

Também gostaria de agradecer às minhas famílias adotivas aqui em Portugal, as famílias **Amaral** e **Trindade**, nomeadamente a **Luisa**, a **Sandra** e o **João**. Vocês acolheram-me de maneira incondicional. Obrigada por fazerem-me sentir como outro membro da família, pela alegria que sentem com cada uma das minhas conquistas, mas também pelo apoio que deram-me para que eu pudesse superar as dificuldades desta fase da minha vida.

Finalmente, me gustaría agradecerle con todo mi corazón a mi novio **Aristides**, solo tu sabes lo importante que fue para mi tenerte a mi lado durante estos últimos años. Tu amor y tu apoyo me ayudó a no desistir en este duro proceso que fue terminar el PhD y escribir la tesis. Solo espero superar esta etapa para ver los nuevos retos que la vida nos va a traer juntos. Soy la persona más afortunada por tenerte en mi vida, gracias.

## **List of abbreviations**

APS	Ammonium peroxydisulfate
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
Canx	Calnexin gene
CLR	C-type lectin receptors
CO <sub>2</sub>	Carbon dioxide
DC	Dendritic cells
ddH <sub>2</sub> O	Double-distilled water
DMEM	Dulbecco's modified Eagle's medium
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FCS	Fetal calf serum
FLUAV	Influenza A viruses
FSC	Forward scatter
GBP	Guanylate-binding protein
GDP	Guanosine diphosphate
GED	GTPase effector domain
GFG	Gene-for-gene model
GTP	Guanosine triphosphate
GVIN	Large inducible GTPases
HFFs	Human foreskin fibroblasts
HI-FBS	Heat-inactivated fetal bovine serum
HMGB1	High-mobility group box 1
IFN $\gamma$	Interferon (IFN)-gamma
iNOS	Nitric oxide synthase
IRG	Immunity-related GTPases
ISG	Interferon-stimulated gene
MA	Matching-allele model
MCMV	Murine cytomegalovirus
MHC	Major Histocompatibility Complex
MIC	Micronemal proteins
MJ	Moving junction
MLST	Multilocus genes sequencing
MOI	Multiplicity of infection
MS	Microsatellite markers
Mx1	Myxovirus resistance protein 1
NEAA	Nonessential amino acids
NK	Natural killer cells
NLPR	Pyrin domain PYD-containing proteins
NLR	NOD-like receptors
ORF	Open reading frame

P/W buffer	Permeabilization/wash buffer
PBD	Peptide-binding domain
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PCR-RFLP	PCR-Restriction Fragment Length Polymorphism
PFA	Paraformaldehyde
PRF	Profilin-like protein
PRR	Pattern recognition receptors
PV	Parasitophorous vacuole
PVDF	Polyvinylidene difluoride
PVM	Parasitophorous vacuole membrane
QTL	Quantitative Trait Locus
RCF	Relative centrifugal force
RLR	RIG-I-like receptors
RON	Rhoptry neck protein
ROP	Rhoptry bulb protein
RT	Room temperature
SAG1	Toxoplasma major surface antigen 1
SDS	Sodium dodecylsulfate
SSC	Side scatter
<i>T. gondii</i> , Tg	<i>Toxoplasma gondii</i>
TEMED	Tetramethylethylenediamine
TLR	Toll-like receptors
TRAP	Thrombospondin-related anonymous protein
UV	Ultraviolet light
WD	Working dilution
WT	Wild type
w/o	Without

# **Table of contents**

<b>1. Introduction</b>	<b>1</b>
1.1. Co-evolution and local adaptation in host-pathogen interactions	2
1.2. Mechanisms of host resistance to pathogens	4
1.3. <i>Toxoplasma gondii</i>	8
1.3.1. Discovery	8
1.3.2. Classification	8
1.3.3. Life cycle	10
1.3.4. Ecological impact	12
1.3.5. Population genetics	15
1.3.5.1. Genotyping	15
1.3.5.2. Geographic distribution	16
1.3.6. Host invasion and virulence mechanisms	18
1.3.6.1. Host cell invasion	18
1.3.6.2. ROP proteins: role in pathogenicity	21
1.4. Host immune response	23
1.4.1. Innate immunity to <i>T. gondii</i> in mice	23
1.4.2. Interferon-inducible GTPases	26
1.4.2.1. Immunity-related GTPases	30
1.4.2.2. Immunity-related GTPases in mice and cell-autonomous resistance against <i>T. gondii</i> .	34
1.4.2.3. IRG inactivation by virulent <i>T. gondii</i> strains	35
1.4.2.4. IRG diversity in wild mice and resistance to virulent <i>T. gondii</i> strains	36
1.5. The house mouse world radiation	40
1.5.1. Expansion of the <i>M. m. domesticus</i> subspecies	42
1.5.2. Appearance of the laboratory and wild-derived mouse strains	43
1.6. Thesis aims.	44
<b>2. Diversity of the IRG system in South American and European <i>Mus musculus</i></b>	<b>65</b>
2.1. Abstract	67
2.2. Introduction	68
2.3. Materials and methods	70
2.3.1. Ethical and collection permits	70
2.3.2. <i>Mus musculus</i> sample collection	70

2.3.3.	Preparation and immortalization of tissue culture lines from wild-caught and wild-derived mice .....	73
2.3.4.	Isolation of RNA from immortalized DDCs for Transcriptome analysis using Illumina sequencing.....	76
2.3.5.	Isolation of genomic DNA from mouse tissues.....	77
2.3.6.	Amplification of IRG genes .....	77
2.3.7.	Nanopore sequencing Irgb2-b1 gene .....	78
2.3.8.	PCR and sequencing of mitochondrial DNA (D-loop).....	78
2.3.9.	Phylogenetic analysis of mitochondrial DNA (D-loop) .....	80
2.3.10.	Transcriptome analysis and IRG genes genotyping .....	80
2.3.11.	Individual analysis of SNPs in IRG genes alleles .....	81
2.3.12.	Expression levels of IRG genes (Transcriptome-based analysis) .....	81
2.4.	Results.....	82
2.4.1.	Polymorphisms in the IRG genes in wild <i>Mus musculus</i> .....	82
2.4.1.1.	IRG alleles on chromosome 11 in Portuguese mice .....	82
2.4.1.2.	IRG alleles in Brazilian and other North and South American mice .....	85
2.4.2.	Genetic diversity within the carriers of the Irgb2-b1 <sub>PWK</sub> allele.....	87
2.4.2.1.	SNPs in the Irgb2-b1 <sub>PWK</sub> allele.....	88
2.4.2.2.	Variations in the Irgm1 and Irgd genes among the Irgb2-b1 <sub>PWK</sub> allele carriers.....	89
2.4.2.3.	Haplotypes in the carriers of the Irgb2-b1 <sub>PWK</sub> allele based on SNPs in the Irgm1, Irgb2-b1 and Irgd genes. ....	91
2.4.3.	Maternal phylogeny of wild and wild-derived <i>Mus musculus</i> .....	95
2.4.3.1.	Brazilian <i>Mus musculus domesticus</i> ancestry .....	96
2.4.4.	Irgb2-b1 expression in wild mice is allele dependent .....	99
2.4.5.	Irgb2-b1 expression levels are independent of other IRGs .....	101
2.4.6.	Polymorphisms in the IRGB2 <sub>PWK</sub> protein subunit.....	103
2.4.7.	Polymorphisms in the IRGB6 protein of mice carriers of the Irgb2 <sub>PWK</sub> allele .....	106
2.5.	Discussion .....	108
2.6.	Acknowledgments .....	111
2.7.	Supplementary material.....	112
<b>3.</b>	<b>IRG-mediated resistance in wild <i>Mus musculus</i> against South American <i>Toxoplasma gondii</i> .....</b>	<b>120</b>
3.1.	Abstract .....	122
3.2.	Introduction .....	123



3.3. Materials and methods .....	124
3.3.1. Ethical permits .....	124
3.3.2. Mice .....	125
3.3.3. Cell lines .....	125
3.3.4. SDS-polyacrylamide gel electrophoresis (PAGE) and Western blot .....	126
3.3.4.1. Generation of IFN $\gamma$ -induced cell lysates for SDS-PAGE .....	126
3.3.4.2. Estimation of protein concentration by BCA assay .....	126
3.3.4.3. SDS-PAGE .....	127
3.3.4.4. Western blot (BioRad Semi-dry transfer system) .....	127
3.3.5. Propagation of <i>Toxoplasma gondii</i> strains .....	128
3.3.6. Immunofluorescence microscopy .....	128
3.3.7. Quantification of IRG proteins loading .....	129
3.3.8. Quantification of host cell necrosis and rate of infection in response to <i>T. gondii</i> strains .....	131
3.3.9. Quantification of inhibition of parasite replication .....	134
3.3.10. Mouse survival assay .....	135
3.4. Results .....	136
3.4.1. IRG protein expression in Brazilian mouse cells .....	136
3.4.2. Vacuolar intensities of IRGB2-B1 in Brazilian cells infected with different virulent <i>T. gondii</i> strains .....	137
3.4.3. In vitro IRG-mediated control of Brazilian <i>T. gondii</i> strains by Irgb2-b1 <sub>PWK</sub> carriers .....	140
3.4.4. The Irgb2-b1 <sub>PWK</sub> allelic form confers resistance against the Eurasian highly virulent RH-YFP strain <i>in vitro</i> .....	143
3.4.5. In vitro resistance to virulent Brazilian <i>T. gondii</i> strains is mediated by the Irgb2-b1 <sub>PWK</sub> allelic form .....	145
3.4.6. Mice carrying the Irgb2-b1 <sub>PWK</sub> allelic form can control the growth of Eurasian <i>T. gondii</i> strains in vivo .....	146
3.4.7. European carriers of the Irgb2-b1 <sub>PWK</sub> allelic form display an intermediate resistance against infection with some Brazilian <i>T. gondii</i> strains .....	148
3.4.8. Resistance to South American <i>T. gondii</i> strains relies on host genetic background and local adaptation .....	150
3.4.8.1. Peruvian mice are susceptible to the infections with Eurasian Type I and Brazilian <i>T. gondii</i> strains .....	150
3.4.8.2. Brazilian mice display different levels of resistance against virulent Type I and sympatric <i>T. gondii</i> strains .....	151
3.4.9. European and Brazilian Irgb2-b1 <sub>PWK</sub> allelic form carriers resist differently to the infection against highly virulent Brazilian <i>T. gondii</i> strains ..	153

3.5. Discussion .....	154
3.6. Acknowledgments .....	158
<b>4. Genetic characterization of South American <i>Toxoplasma gondii</i> strains .</b>	<b>162</b>
4.1. Abstract .....	164
4.2. Introduction .....	165
4.3. Materials and methods .....	167
4.3.1. Cell culture maintenance and propagation of <i>Toxoplasma gondii</i> strains .....	167
4.3.2. Separation and purification of <i>Toxoplasma gondii</i> tachyzoites .....	168
4.3.3. Genomic DNA isolation .....	169
4.3.4. Validation of <i>Toxoplasma gondii</i> tachyzoites purification .....	170
4.3.5. PCR-RFLP <i>Toxoplasma gondii</i> genotyping .....	171
4.3.6. <i>Toxoplasma gondii</i> whole genome sequencing .....	172
4.3.7. Genome and SNP calling analyses .....	172
4.3.8. PCR-RFLP design for Brazilian <i>Toxoplasma gondii</i> strains .....	173
4.3.9. ROP5 and ROP18 gene sequences analysis .....	173
4.4. Results .....	174
4.4.1. Commonly used PCR-RFLP markers for <i>T. gondii</i> genotyping are not suitable for the distinction of some Brazilian strains. ....	174
4.4.2. Brazilian <i>T. gondii</i> strains AS28, BV and N can be differentiated by their distinct genetic backgrounds .....	175
4.4.3. A new PCR-RFLP assay allows the genotyping of the Brazilian AS28, BV, and N strains. ....	178
4.4.4. New PCR-RFLP assay allows the distinction of Brazilian <i>T. gondii</i> strains using DNA extracted from tissue samples. ....	181
4.4.5. Genetic diversity in virulence factors of the Brazilian strains AS28, BV, and N. ....	182
4.4.5.1. ROP5 isoforms in Brazilian strains display characteristics of highly virulent genetic backgrounds. ....	182
4.4.5.2. The AS28, BV, and N strains are carriers of identical ROP18 sequences .....	183
4.5. Discussion .....	185
4.6. Acknowledgments .....	188
4.7. Supplementary material .....	190
<b>5. General discussion .....</b>	<b>200</b>
5.1. Summary of the findings .....	202
5.2. High prevalence of the Irgb2-b1 <sub>PWK</sub> in South American house mice and the rapid adaptation to local parasites. ....	205

5.3. Genotype–genotype interactions: <i>lrgb2-b1</i> alleles and ROP5/ROP18 isoforms. ....	211
5.4. Concluding remarks and future perspectives .....	214
<b>6. Key resources table.....</b>	<b>221</b>

## List of figures

<b>Figure 1.1.</b> Phylogeny of <i>Toxoplasma gondii</i> .....	10
<b>Figure 1.2.</b> Life cycle of <i>Toxoplasma gondii</i> .....	12
<b>Figure 1.3.</b> Integrated model for effector protein secretion by <i>Toxoplasma gondii</i> tachyzoites into the host cell .....	20
<b>Figure 1.4.</b> Interleukin 12 and Interferon- $\gamma$ (IFN $\gamma$ ) production in the infection with <i>T. gondii</i> .....	24
<b>Figure 1.5.</b> The IFN-inducible GTPase superfamily in mice .....	27
<b>Figure 1.6.</b> IRG genes genomic location in inbred mouse strains .....	31
<b>Figure 1.7.</b> Polymorphism at the protein level in IRG genes (Chromosome 11 and Chromosome 18) of inbred mouse strains.....	37
<b>Figure 1.8.</b> Susceptibility and resistance of mouse genotypes against type I virulent strains of <i>T. gondii</i> .....	40
<b>Figure 1.9.</b> Worldwide distribution of <i>Mus musculus</i> subspecies.....	41
<b>Figure 2.1.</b> Mouse sampling sites in Portugal.....	72
<b>Figure 2.2.</b> Mouse sampling sites in Brazil. ....	72
<b>Figure 2.3.</b> Mouse diaphragm dissection .....	76
<b>Figure 2.4.</b> Map of the mouse mitochondrial genome.....	79
<b>Figure 2.5.</b> Polymorphism of IRG proteins on chromosome 11 in Portuguese mice .....	84
<b>Figure 2.6.</b> Polymorphism of IRG proteins on Chromosome 11 in Brazilian mice ..	86
<b>Figure 2.7.</b> Schematic representation of the <i>Irgb2-b1</i> gene.....	89
<b>Figure 2.8.</b> Examples of the alleles found in the <i>Irgm1</i> gene in mice that are carries of the <i>Irgb2-b1</i> PWK allele .....	90
<b>Figure 2.9.</b> Examples of the alleles found in the <i>Irgd</i> gene in mice that are carries of the <i>Irgb2-b1</i> <sub>PWK</sub> allele.....	91
<b>Figure 2.10.</b> Maternal phylogeny of wild and wild-derived <i>Mus musculus domesticus</i> .....	97
<b>Figure 2.11.</b> Expression levels of the <i>Irgb2-b1</i> gene are allele dependent.....	101
<b>Figure 2.12.</b> <i>Irgb2-b1</i> expression levels are not dependent on other IRG genes. ....	102
<b>Figure 2.13.</b> Polymorphism for the IRGB2 protein subunit in three mouse lines. ....	105
<b>Figure 2.14.</b> Polymorphisms in the IRGB6 protein.....	107
<b>Figure 3.1.</b> Quantification of IRG proteins in IFN $\gamma$ -stimulated mouse cells infected with <i>T. gondii</i> .....	131
<b>Figure 3.2.</b> Gating strategy for detection by flow cytometry of host cell necrosis and rate of infection in <i>T. gondii</i> infected cells .....	133
<b>Figure 3.3.</b> Gating strategy for the inhibition of parasite replication assay by flow cytometry. ....	135
<b>Figure 3.4.</b> IRG proteins expression in Brazilian cell lines .....	137
<b>Figure 3.5.</b> Representative fluorescent images of <i>T. gondii</i> strains RH $\Delta$ ta $\Delta$ hxgprt, AS28, BV and N derived vacuoles.....	139

<b>Figure 3.6.</b> Higher IRGB2-B1 loading intensity in IFN $\gamma$ induced Brazilian cells infected with local <i>T. gondii</i> strains .....	140
<b>Figure 3.7.</b> Brazilian <i>T. gondii</i> strains are controlled by Irgb2-b1 <sub>PWK</sub> allelic form carriers .....	143
<b>Figure 3.8.</b> Irgb2-b1 <sub>PWK</sub> is involved in the resistance against highly virulent <i>T. gondii</i> strains <i>in vitro</i> .....	144
<b>Figure 3.9.</b> Irgb2-b1 <sub>PWK</sub> controls replication of Brazilian <i>T. gondii</i> strains .....	146
<b>Figure 3.10.</b> PWD mice are highly resistance to type I virulent <i>T. gondii</i> strains ..	148
<b>Figure 3.11.</b> PWD mice are susceptible to the Brazilian virulent <i>T. gondii</i> strains AS28 and BV, but display an intermediate resistance against the N strain .....	149
<b>Figure 3.12.</b> Peruvian mice are susceptible to virulent <i>T. gondii</i> strains from Brazil .....	151
<b>Figure 3.13.</b> Brazilian mice display different levels of resistance against virulent Type I and local <i>T. gondii</i> strains .....	152
<b>Figure 3.14.</b> Brazilian mice have a higher resistance to Brazilian <i>T. gondii</i> strains than European PWD mice. ....	154
<b>Figure 4.1.</b> Validation of <i>T. gondii</i> tachyzoites purification by CF-11 cellulose column. ....	171
<b>Figure 4.2.</b> Multiplex multilocus nested PCR-RFLP (Mn-PCR-RFLP) analysis of <i>Toxoplasma gondii</i> samples using two gene markers .....	175
<b>Figure 4.3.</b> PCR amplification of <i>Toxoplasma gondii</i> samples using four gene markers .....	178
<b>Figure 4.4.</b> PCR-RFLP of <i>Toxoplasma gondii</i> samples using four gene markers. ....	180
<b>Figure 4.5.</b> PCR-RFLP of <i>Toxoplasma gondii</i> samples using two gene markers to distinguish N strain in infected mouse brains .....	181
<b>Figure 4.6.</b> Isoforms of the ROP5 gene in Brazilian <i>Toxoplasma gondii</i> strains ...	183
<b>Figure 4.7.</b> ROP18 gene in the AS28, BV and N <i>Toxoplasma gondii</i> strains. ....	184
<b>Figure 4.8.</b> Phylogenetic analysis of ROP18 sequences from multiple <i>T. gondii</i> strains. ....	185
<b>Supplemental Figure 2.1.</b> Maternal phylogeny of wild and wild-derived <i>Mus musculus</i> species in the world. ....	114
<b>Supplemental Figure 4.1.</b> Phylogenetic analysis of ROP5 sequences from AS28 <i>T. gondii</i> strain .....	211

## **List of tables**

<b>Table 2.1.</b> Mice sampled for this study. ....	73
<b>Table 2. 2.</b> Haplotypes in the <i>Irgb2-b1<sub>PWK</sub></i> allele carriers based on SNPs in the <i>Irgm1</i> , <i>Irgb2-b1</i> and <i>Irgd</i> genes. ....	93
<b>Table 4.1.</b> SNPs between AS28, BV, N and ME49 <i>Toxoplasma gondii</i> strains. ...	176

# **Chapter 1**

---

## **1. Introduction**

## 1.1. Co-evolution and local adaptation in host-pathogen interactions

Co-evolution is the process of reciprocal, adaptive genetic change in two or more species. This process can occur between any interacting populations (e.g. plant and herbivore, prey and predator, etc.). However, in host-pathogen systems, co-evolution is expected to be remarkably important due to the close nature of the association between host and microbes, as well as the strong selective pressures imposed on each other<sup>1</sup>. Host-pathogen interactions play an important role in many ecological and evolutionary processes<sup>2</sup>. They can lead to long-term coevolutionary associations, affecting the pace of molecular evolution as well as the rates of diversification and speciation, for both the host and the parasite<sup>3</sup>.

For more than 70 years theoretical studies have tried to explain how co-evolutionary processes between hosts and parasites occur in natural populations<sup>4,5</sup>. Many of those studies helped to build the principles of the Red Queen hypothesis. This hypothesis states that under rapid antagonistic coevolution between hosts and their parasites, the generation of new and rare host genotypes through sex can be advantageous for the host because most of the current parasite population will not be adapted to these new genotypes<sup>6</sup>. Plant-pathogen systems were the first ones to be studied in this context, leading to models such as “Gene-for-gene” (GFG) and “Matching-allele” (MA)<sup>1</sup>.

Some of the first experiments to understand host-pathogen interactions were done by Flor (1951) involving the rust pathogen *Melampsora lini* and its host plant *Linum usitatissimum*<sup>1,7</sup>. These experiments showed that resistance to *Melampsora lini* is due to the simultaneous presence of an R gene (Resistance) in *Linum usitatissimum* and a matching avirulence (Avr) gene in *Melampsora lini*. The absence of either the R gene or the Avr gene



results in disease<sup>8</sup>. The GFG model was largely built on the foundations of Flor's work on the rust-plant interaction. GFG is based on the concept of genetic recognition leading to incompatible disease outcomes, which is the typical recognition of the host immune system<sup>3</sup>.

Experiments in plant-pathogen systems have shown that the GFG model governs many of the mechanisms of the plant immune response against biotrophic and hemibiotrophic pathogens<sup>3</sup>. However, since plants and other hosts are in a continuous arms race with their pathogens, GFG systems such as the R genes in plants have evolved to acquire new recognition specificities to pathogens and at the same time, those pathogens find new ways to counteract this new recognition machinery. Therefore, any particular single source of resistance (e.g. allele, protein conformation) is only effective for a certain period of time<sup>8</sup>. Additionally, selection in GFG systems is frequency-dependent. This means that a parasite allele that matches the most common host resistance allele will show a decrease in frequency. Interestingly, rare host alleles have a selective advantage in this context, because they will avoid resistance and so, their frequency will increase and initiate selection for a currently rare virulence allele<sup>9</sup>. In general, for each gene involved in host resistance, there is another gene for avirulence (non-infectivity) in the pathogen. One parasite genotype might have "universal virulence" and infect all existing host genotypes<sup>10</sup>. However, the outcome of the interaction uniquely relies on the combination of alleles at certain loci<sup>1</sup> and it assumes that non-recognition (by the host) of the pathogen's gene products results in infection<sup>3</sup>.

In the MA model, in contrast, a compatible interaction given by the genetic recognition between the host and the pathogen is required for resistance. This model is based on self/non-self recognition systems, where a match between host and pathogen genotypes is necessary for infection to be effective<sup>3</sup>. The concept of "Universal virulence" is not possible under the

assumptions of this model and polymorphisms in genes involved in virulence or resistance are easily maintained by negative frequency-dependent selection<sup>10</sup>. The MA model has been widely applied to understand the effect of host-parasite co-evolution on sex and recombination<sup>11,12</sup>, speciation<sup>13</sup> and local adaptation<sup>14,15</sup>.

Co-evolution occurs during local adaptation. Local adaptations are sympatric interactions that are more compatible for survival between both a host and a parasite, than allopatric combinations<sup>1</sup>. These interactions rely on the genetic composition of local populations, which presumably differ through time. For example, a certain allele can be lost due to oscillatory genetic dynamics, unless populations are constantly connected by a high gene flow<sup>9</sup>. Systems such as trematodes and snails<sup>16</sup>, trematodes and fish<sup>17,18</sup>, bacterium or microsporidians and their host the crustacean *Daphnia magna*<sup>19,20</sup> display variation in the genetic composition according to their geographic location. In these interactions, populations of the parasite are best adapted to their local host populations, suggesting a very fast process of co-adaptation<sup>9</sup>. In a certain geographic location, virulent and avirulent parasites can infect multiple hosts. However, when a susceptible host encounters a virulent parasite, its immune system gets overcome and dies, but this may be a Pyrrhic victory because under certain conditions the parasite may reduce its probability of transmission to a new host. Therefore, a highly virulent pathogen is a clear example of co-adaptation failure<sup>1,21,22</sup>. It is in the interest of both, hosts and parasites, to prolong their encounter and avoid death. One of the first authors to conjecture about this phenomenon was Haldane in 1949, who stated that an intense and fluctuating selection pressure imposed by parasites leads to polymorphism in host protein associated with resistance<sup>5,23</sup>.

## **1.2. Mechanisms of host resistance to pathogens**

Host resistance against infectious diseases is due to a multiplicity of mechanisms. Innate and adaptive immunity both aim to achieve reduction in pathogen loads, while mechanisms of disease tolerance aim for a reduction in pathogen-induced pathology. Both resistance mechanisms contribute to the preservation of host homeostasis<sup>24</sup> and both are subject to genetic variation, ultimately affecting the trade-offs imposed on host and pathogen's fitness<sup>25</sup>.

Tolerance and resistance are difficult to uncouple due to the several genes and pathways that intertwined between them. However, both of them aim for a balance between collateral damage of host tissues and control of the pathogen. In the context of this work, disease tolerance is a mechanism of protection that does not impact negatively on pathogen loads<sup>25</sup>. This concept was first introduced to explain how plants survive fungal pests, making a clear distinction between disease tolerance-associated mechanisms and resistance mechanisms<sup>26</sup>. It is still not clearly understood how the mechanisms underlying disease tolerance work, however, many of the events involve evolutionarily conserved stress and damage responses, such as tissue damage control to limit disease severity<sup>27</sup>. On the other hand, resistance mechanisms are those that aim to protect the host by reducing pathogen load<sup>24</sup>. The ability to tolerate or control the pathogen must be conserved in the host; however, the pathogen should also retain its infectivity and ability to limit the host immune response, in order to find a balance between them.

Defense genes products constitute some of the first barriers for pathogen control in the host and display signatures of host and parasite interactions, leading to rapid evolution at the protein level because of the action of positive selection<sup>28</sup>, usually accompanied by polymorphism. These events might be caused by the succession of selective sweeps associated with the presence of a new resistance allele. Persistent dynamic polymorphisms with

fluctuations in allele frequencies are followed by positive selection in changing directions, due, for example, to negative frequency-dependent selection<sup>1,28</sup>.

The Major Histocompatibility Complex (MHC) in mammals is an example of positive selection and balancing selection in the context of host-pathogen interactions. The MHC system was first discovered in inbred mouse strains (referred to as the histocompatibility-2 (H-2) complex) when transplanted tumors displayed rejection according to tissue compatibility associated with the mouse strain<sup>29</sup>. However, the MHC system is found in all jawed vertebrates, including humans.<sup>29–31</sup> At the genome level, the MHC system is characterized by a number of closely linked genes encoding glycoproteins with a distinctive structure that bind self and non-self peptides in intracellular membrane-bounded compartments and deliver them to the surface, where T cells and natural killer (NK) cells screen them<sup>31,32</sup>. In natural populations, MHC molecules are extremely polymorphic, with hundreds of allelic variants segregating. This high level of polymorphism is widely agreed to be the result of a molecular arms race, where pathogens evade immune recognition and hosts survive based on their specific MHC alleles. Polymorphic sites for MHC molecules are located for the most of them in the peptide-binding domain (PBD), specifically in the peptide-binding sites (PBSs), where multiple amino acids interact directly with bound peptides<sup>31</sup>. PBSs display a striking excess of nonsynonymous amino acid substitutions consistent with episodes of strong positive selection through pathogen resistance<sup>33,34</sup>. Differences in peptide-binding profiles in MHC alleles affect susceptibility to infectious diseases. Classical HLA alleles have been reported to display positive genetic associations for infectious diseases such as HIV/AIDS, Hepatitis, Leprosy, Tuberculosis, Malaria, Leishmaniasis, and Schistosomiasis<sup>35</sup>.

Another example of proteins involved in resistance to pathogens is the Mx

dynamin-like GTPases. These high molecular weight intracellular GTPases are important antiviral effector proteins induced by type I and type III interferons (IFN)<sup>36</sup>. They control a wide range of viruses by the inhibition of the early stages of viral replication in vertebrates, from fish to mammals<sup>37</sup>. The first Mx protein family member identified was the mouse Mx1 (myxovirus resistance protein 1) in the A2G inbred mouse strain. Mice carrying the Mx<sup>+</sup> allele synthesize a 72-kDa nuclear protein that confers resistance to lethal doses of influenza A viruses (FLUAV) and other Orthomyxoviruses, whereas mice homozygous for the Mx<sup>-</sup> allele cannot synthesize the protein rendering them susceptible to the infection. This resistance phenotype is fully dependent on the single Mx1 gene encoded on chromosome 16. Mx<sup>-</sup> and Mx<sup>+</sup> alleles are equally frequent. In wild mice, however, there is a selective advantage for heterozygous individuals in the wild<sup>36,38</sup>. Several laboratory mouse strains have crippled Mx alleles due to deletions and/or mutations that destroyed the corresponding protein's function<sup>36,39</sup>. It is thought that in the wild, the Mx1 gene protects feral mice from influenza-like infections<sup>37</sup>, however, these mice also have a second Mx GTPase (Mx2), which in contrast to Mx1 (a nuclear protein) is a cytoplasmic protein that inhibits viruses with a cytoplasmic replication such as rhabdoviruses and bunyaviruses<sup>40</sup>. Inbred mouse strains including A2G are carriers of a defective Mx2 gene. This pattern of expression of the active Mx-1 allele in the wild is consistent with there being a cost associated with Mx-1 expression, a cost compensated in the heterozygote by the advantage conferred by the protein in viral resistance. The situation thus resembles that of the sickle cell anemia allele, HbS, which confers resistance in humans to malaria but carries an intrinsic selective disadvantage through sickle cell anemia in the absence of the malaria parasite.

Humans also express two Mx GTPases, both cytoplasmic and referred to as MxA and MxB (or Mx1 and Mx2, respectively). Human MxA confers wide antiviral protection against viruses irrespective of their intracellular

replication. It was for a long time thought that the MxB protein even though it was expressed, did not have demonstrable antiviral activity<sup>40</sup>, however, recent studies have shown its broad activity against retrovirus (HIV-1), as well as herpes virus infections (HSV-1, HSV-2, and KSHV)<sup>41,42</sup>.

### **1.3. *Toxoplasma gondii***

#### **1.3.1. Discovery**

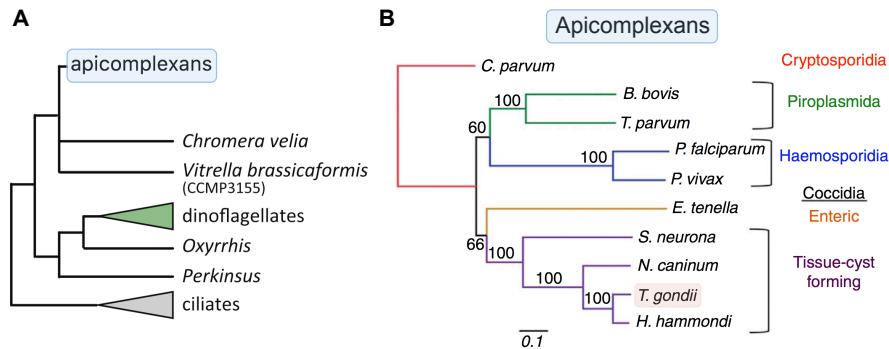
*Toxoplasma gondii* is an obligate intracellular protozoan parasite. Two scientific groups, Nicolle and Manceaux in Tunis and Splendore in Brazil, simultaneously discovered it in 1908. Nicolle and Manceaux identified this parasite in tissues from a hamster-like mammal, the *Ctenodactylus gundi*<sup>43</sup>, whereas Splendore found it in a rabbit<sup>44</sup>. Initially believed to be a *Leishmania*, they soon realized it was a new parasite species and named it based on its characteristic curved morphology (mod. L. *toxos*= arc or bow, *plasma*= life) and host (*Ctenodactylus gundi*, initially misspelled *Ctenodactylus gondi*)<sup>45</sup>.

#### **1.3.2. Classification**

*T. gondii* belongs to the protozoan phylum Apicomplexa. This ancient phylum (approximately 700-900 Myr)<sup>46</sup> is very close to dinoflagellates and ciliates within the Alveolata superphylum<sup>47</sup> (**Figure 1.1, panel A**). Apicomplexans are characterized by the presence of an apical complex, which is an assembly of organelles located at the anterior end of the cell and responsible for the secretion of proteins involved in host cell invasion<sup>47</sup>. Most Apicomplexans also possess an essential non-photosynthetic plastid-like organelle, the apicoplast<sup>48</sup>. This organelle likely evolved through secondary endo-symbiosis of eukaryotic red algae, which was later enslaved as a plastid and surrounded by two more host cell membranes<sup>47</sup>. The apicoplast has a small circular genome (35 kb) that encodes for

important biochemical pathways that control the parasite metabolism and makes it a good drug target<sup>48,49</sup>. However, not all Apicomplexans possess an apicoplast, for example, the *Cryptosporidium* genus does not have it. The apicoplast in these organisms is likely to have been secondarily lost<sup>47,50</sup>.

There are approximately 5000 Apicomplexan species, including parasites of insects and mollusks, as well as many important human and domestic animal pathogens. This adaptation to various vertebrate hosts started more than 400 Myr ago as a result of changes in host range and molecular virulence mechanisms<sup>51</sup>. Apicomplexans can be split into different groups based on their host specificity and phylogeny (**Figure 1.1, panel B**). The Cryptosporidia are the most distantly related subgroup, containing only the genus *Cryptosporidium* spp., which is an important intestinal parasite for humans and livestock. The Haematozoans (Haemosporidia and Piroplasmida) are parasites of blood cells that rely on blood-sucking invertebrates (mosquitos, ticks, etc) for transmission and complete their sexual stages in a vertebrate host. Important human and domestic animal parasites such as *Plasmodium*, *Babesia*, and *Theileria* species belong to this group<sup>52</sup>. The Coccidians (Enteric and Tissue-cyst forming) are parasites that firstly colonize the intestinal tract and are usually transmitted through a fecal-oral route. *T. gondii* belongs to the tissue-cyst forming coccidians, which can be distinguished from the Enteric coccidians due to the alternating two-host life cycle<sup>51</sup>.



**Figure 1.1.** Phylogeny of *Toxoplasma gondii*. A. Consensus tree of Alveolata. Modified from<sup>47</sup>. B. Phylogenetic tree of apicomplexans. The conserved DEAD box helicase protein (TGME49\_249810, OrthoMCL OG5126701) was used to build the tree. The Neighbour-joining tree algorithm with bootstrap values was applied. Distance equals 0.1 amino acid substitutions/site. Taken from reference<sup>51</sup>.

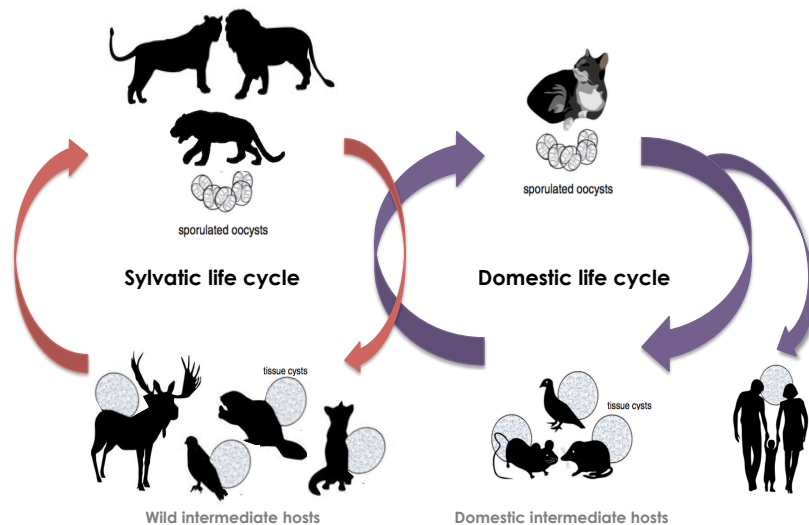
### 1.3.3. Life cycle

*T. gondii* is a parasite with a complex life cycle that includes multiple sexual and asexual stages (**Figure 1.2.**). The sexual cycle is restricted to the members of the *Felidae* family, including all Felinae (e.g. cougar, lynx, leopard, etc.), Pantherinae (e.g. jaguar, lion, tiger, etc.) and the domestic cat (*Felis silvestris catus*)<sup>53–55</sup>. Recent studies have shown that feline-specificity for *T. gondii* sexual cycle development is associated with the presence of a host-derived lipid in the feline intestine, the linoleic acid. Felines are the only mammals that cannot process linoleic acid due to the lack of the delta-6-desaturase enzyme, resulting in an excess that promote the sexual development of the parasite<sup>56</sup>.

Cats become infected with the parasite by the ingestion as prey of an infected intermediate host (warm-blooded animals) carrying encysted parasites (bradyzoite stage). The bradyzoite (Gr. brady = slow) is a slowly replicating stage that forms intracellular tissue cysts surrounded by a dense wall. Once bradyzoites are ingested by a cat, pepsin and the acid digestion in the feline stomach promote their release from the cyst to the intestinal



epithelium. Parasites invade the epithelial cells and differentiate into five morphologically distinct types of schizonts (Type A – E). Two stages divide by endodyogeny (binary fission) (Type A and B), which then are followed by three rounds of schizogony that divide by endopolygeny (Type C–E)<sup>45</sup>. The whole process of transition between stages last approximately 2 days and leads to the development of merozoites. Merozoites proliferate and differentiate within epithelial enterocytes into macrogametes (oocytes) and microgametes (spermatozoa). These two stages fuse to produce a diploid zygotic nucleus (unsporulated oocysts), which are surrounded by thick, impermeable, and highly resistant walls. Two sporocysts are formed as the results of a meiotic division, which after a second mitotic division, each of them gives rise to four haploid sporozoites. Unsporulated oocysts are shed in the feces to the environment, where ambient air and proper temperature promotes their sporulation and maturation. Matured oocysts carry eight haploid sporozoites as a result of an extra mitotic division that presumably happens during the maturation process<sup>57</sup>. They are highly infectious and can be stable up to 18 months in unfavorable conditions until they get to a new host<sup>56</sup>. The asexual cycle happens in warm-blooded animals upon consumption of food or water contaminated with oocysts. The parasite disseminates throughout the host by cycles of intracellular invasion, replication, escape and reinfection (tachyzoite phase) and, following the induction of host immunity, differentiates into tissue cysts (bradyzoite phase).



**Figure 1.2.** Life cycle of *Toxoplasma gondii*. *T. gondii* can be transmitted through both sylvatic and domestic cycles. In the sylvatic cycle, parasites (sporulated oocysts) are excreted by definitive hosts (wild felines) into the environment with the feces, where they are ingested by foraging wild animals (intermediate hosts). In the domestic cycle, definitive hosts are the domestic cats, and these become infected by eating intermediate hosts associated with human habitation such as small mammals (mostly rodents) and birds that become infected by inadvertently foraging food contaminated with oocysts excreted by domestic cats. Once in the intermediate host, the parasite transitions to an intracellular proliferative stage (tachyzoites), that given the proper conditions (immune response and host genetics) might be controlled by the host and transition to a slowly replicating stage (bradyzoites). Bradyzoites form tissue cysts in non-replicative cells such as neurons in the brain and muscle cells that persist for the life of the intermediate host. Intermediate hosts in the cat's food chain are considered evolutionarily relevant for the parasite (e.g. house mouse), whereas infected animals outside that food chain, like humans, are irrelevant for the parasite transmission (dead-end hosts). Modified from reference <sup>58</sup> and reference <sup>59</sup>.

#### 1.3.4. Ecological impact

The subfamily *Felinae* has evolved into one of the world's most successful carnivore families since the late Miocene (10-11 Myr), inhabiting all the continents except Antarctica<sup>60</sup>. *T. gondii* has adapted to be efficiently transmitted in felids by carnivorism and in other hosts by the fecal-oral (oocysts) route<sup>45</sup>. Therefore, the adaptation of *T. gondii* to felid-like

carnivores was crucial to the spread of the parasite all over the world since only wild and domestic felids can shed *T. gondii* oocysts to the environment. Overall, young cats shed a higher number of oocysts, however, domestic cats, in general, can shed approximately one billion oocysts typically during 1 to 2 weeks after infection<sup>61,62</sup>. The high resistance of the oocyst wall allows them to disseminate through watersheds and ecosystems, as well as persist in soil, vegetables, and seafood<sup>62</sup>, which has a tremendous impact on the ecology, epidemiology, and public health measurements associated with *T. gondii* infections<sup>63</sup>.

Toxoplasmosis is usually mild or asymptomatic, however, in some cases, it can cause serious disease and death in humans and animals. Several human toxoplasmosis outbreaks have been associated with the presence of oocysts in drinking water. For example, oocysts shed by cougars were the source of a large waterborne outbreak in Victoria, Canada in 1995<sup>45,64</sup>, as well as the contamination of several municipal water sources. These events of water contamination can result in outbreaks, specially in developing countries such as Brazil (25 outbreaks recorded in Brazil over the past 50 years)<sup>62,65-67</sup>. Cases of severe and acute Toxoplasmosis have been reported in wild animals too. Recently, environmental studies have shown an increase in the presence of *T. gondii* oocysts in the marine environment, which makes mollusk-eating mammals especially susceptible to infection with *T. gondii*<sup>68</sup>. Cetaceans, phocids, otariids, walruses, sirenians, and sea otters can show symptoms of toxoplasmosis<sup>62</sup>. The high rate of infection in the southern sea otters in California is now the classic example of the ecological impact of *T. gondii* in natural marine mammals. Sea otters get infected with *T. gondii* via the ingestion of oocysts that accumulate in coastal filter-feeding mollusks such as clams from contaminated freshwater run-off<sup>63</sup>. Up to 70% of these otters are infected with *T. gondii*, which contributes to their mortality in more than 17%<sup>63</sup>. Moreover, all sea otters that died due to *T. gondii* as a primary cause of death were infected with the atypical Type

X genotype, which is reported to be highly virulent in these sea mammals. Those *T. gondii* strains were also found in terrestrial felids from watersheds bordering the sea otter range, demonstrating a direct link between water contamination with oocysts and the *T. gondii* prevalence in marine animals<sup>63</sup>.

Cat domestication is dated back to the advent of agriculture in the Middle East around 11000 years ago<sup>69</sup>. These cats began their commensal association with humans by feeding on the rodent pests that infested grain stores of the first farmers<sup>59,70</sup>. Along with cats, another species became also very abundant, the house mouse (*Mus musculus*), due to easier access to food sources such as stored grains in human settlements. Nowadays, domestic cats are the most abundant felids in the world. Only in the USA, for example, the number of pet cat ownerships has increased from 50 million to 90 million animals in the past 20 years. Also, the number of colonies of free-roaming cats has increased due to the creation of feeding stations by animal welfare activists<sup>71</sup>. The number of wild felids was always low because of their position at the top of the food chain and is now being drastically reduced worldwide by environmental degradation.

In urban areas, the house mouse is one of the main prey species for domestic cats<sup>58,72,73</sup>. Mice are foraging animals, which makes them susceptible to the ingestion of oocysts spread in cat feces<sup>58</sup>. Urban *M. musculus* in the UK, for example, were shown in one study to have a prevalence of *T. gondii* infections (PCR based test) of 59%<sup>73</sup>. This study has not been confirmed in other urban settings and other studies in the rest of the world have shown lower levels of *T. gondii* prevalence in natural *M. musculus* populations. In the USA, for example, seroprevalences in the range from 0-3% have been reported in wild *M. musculus*<sup>74</sup>. Recent studies in Africa (Senegal) have shown values closer to 5%<sup>75</sup>. Other animals with significant infection rates for *T. gondii* might be also part of the domestic cat

diet, animals like the European field mouse *Apodemus* spp., voles, shrews, pigeons, etc.<sup>58,76–79</sup>. However, the house mouse is certainly an important reservoir and an excellent candidate for an evolutionary significant host for *T. gondii*<sup>58</sup>.

### 1.3.5. Population genetics

#### 1.3.5.1. Genotyping

*T. gondii* is the only species of the *Toxoplasma* genus<sup>80,81</sup>. It has a 65 Mb genome that is predicted to encode approximately 8000 genes located on 13 chromosomes (previously annotated chromosomes VIIb and VIII are in fact a single chromosome)<sup>51,82</sup>. First studies in the population structure of this parasite revealed three clonal lineages (type I, II, and III) which share a common ancestor (10,000 years ago<sup>83</sup>) in isolates from human and domestic animals<sup>84–86</sup>. Type I strains are lethal to most laboratory mouse strains, causing acute infections and death within 2 weeks ( $LD_{50}$ = 1 parasite). Type II strains are intermediately virulent ( $LD_{50}$ =  $10^2$  -  $10^4$  parasites) and type III strains are mostly non-virulent to laboratory mice ( $LD_{50}$ =  $>10^5$  parasites)<sup>59</sup>. While the lethality of type II and type III strains is reported over a wide range, it is clear that the genetics of different mouse strains, and probably also different husbandry, contribute to the outcome of experimental infection.

Studies using multilocus genotyping (Polymerase chain reaction (PCR) - Restriction Fragment Length Polymorphism (PCR-RFLP) markers and microsatellite (MS) markers) and whole-genome analysis in diversified samples showed a much more complex genetic structure in *T. gondii*<sup>87–91</sup>. These analyses allowed a more accurate strain allelic composition analysis and the detection of genomic recombinations<sup>92</sup>. PCR-RFLP genotyping is based on 10 genomic markers and one apicoplast marker<sup>93</sup> and the most recent MS analysis involves 15 markers along 11 different chromosomes<sup>94</sup>.

The combination of PCR-RFLP patterns and microsatellites is translated into a specific number for each genotype according to the codification adopted by the ToxoDB database (<http://www.toxodb.org/toxo/>). Some strains do not belong to any of the clonal lineages defined by the MS genotyping; therefore, they are named “atypical” or “unique genotype strains”<sup>92</sup>. Another strategy used for *T. gondii* genotyping is the combination of multilocus sequence typing (MLST) and Whole Genome Sequencing, which clusters strains distributed around the world into 16 haplogroups that belong to six ancestral groups (clades)<sup>51,91</sup>. A table with the correspondence between the different strategies for *T. gondii* genotyping can be found in reference<sup>92</sup>.

#### 1.3.5.2. Geographic distribution

*T. gondii* has subpopulation structures in different geographic locations. The Northern Hemisphere, for example, is characterized by a strong clonal structure, where type II strains are the most prevalent and circulate in domestic and wild animals in Europe and North America. Nevertheless, strains from any of the three clonal types can be found as well circulating in different environments<sup>95</sup>. Besides the three clonal types, a recent study showed that North American wild animals are carriers of a fourth clonal lineage (type X or 12), which is highly virulent in laboratory mice<sup>90</sup>. A possible explanation for the strong clonality in the Northern Hemisphere points to the very limited number of felids in those territories, which limits the chances for sexual recombination of the parasite<sup>89,92</sup>. For example, only 3 species of felids are present in Europe, whereas at least 8 species of wild felids coexist in South America<sup>92</sup>.

In East Asia and North Africa, there is a prevalence of types II and III isolates. However, the type I strains are the second most common genotype in Asian countries<sup>96</sup>. Regional clonal lineages are also present in these two continents, such as the Chinese 1 in China or Africa 1 and Africa 3 in Africa

(they are not related to the classical clonal lineages and are considered as new major haplogroups)<sup>92,96–98</sup>. Current genetic population structures in these regions indicate a constant circulation from Asia to Europe, and East or North Africa promoted by migration pathways (birds, rodents, felids, and human activities)<sup>96</sup>.

South America is considered an exceptional place in terms of *T. gondii* genetic diversity. The most recent common ancestor of the modern *T. gondii* populations appeared in South America approximately 1.5 Myr ago, possibly in the area that corresponds nowadays to the Colombian territory<sup>99</sup>. This emergence, however, is much more recent than the appearance of the original ancestral *T. gondii*, which is calculated to have happened approximately 11 Myr ago<sup>99</sup>. The antique presence of *T. gondii* in a sylvatic environment with several definitive hosts possibly favored the diversification of alleles by sexual replication, as well as the appearance of a higher number of mutations caused by genetic drift<sup>92,100</sup>. All South American *T. gondii* strains (atypical or exotic) can be classified within the 6 major clades that contain a total of 16 distinct haplogroups<sup>91</sup>. They are virulent in naïve laboratory mice<sup>101–103</sup> and have been associated with a high prevalence in humans<sup>104</sup>, as well as to severe forms of primary (i.e. not due to vertical transmission to a fetus) ocular and cerebral toxoplasmosis<sup>105–107</sup>.

*T. gondii* domestic strains from different geographic locations share a highly conserved monomorphic version of the Chromosome Ia (ChrIa) and show signs of a recent genetic bottleneck. This version of the ChrIa is associated with the successful expansion of strains from the three major clonal lineages in North America and Europe<sup>89,108</sup>. Moreover, it is suggested that the predominance of ChrIa in an anthropized environment is the result of an advantage in transmission in domestic cats and therefore it potentially contributes to the spread of pathogenicity determinants<sup>109</sup>. Interestingly, some South American strains also maintain regions of the monomorphic

Chr1a, although they are otherwise genetically different populations (a possible recent introgression of the Chr1a from northern strains into South American ones)<sup>110</sup>. Recently, new studies have shown that the fixation of the monomorphic Chr1a took place very recently, within only 10,000 yrs<sup>111</sup>.

### **1.3.6. Host invasion and virulence mechanisms**

*T. gondii* intracellular behavior entails the successful co-option of a vast range of host cells, where each of them defines its own species-specific and cell type-specific biology<sup>112</sup>. For this, *T. gondii* has developed a battery of secretory proteins capable of modulating the host immune response to establish a successful infection. However, not all of these are considered virulence factors. Parasite virulence factors are crucial to determine disease severity, but this does not mean that all of them are essential for parasite transmission and survival<sup>113</sup>.

#### **1.3.6.1. Host cell invasion**

The relationship between *T. gondii* tachyzoites and non-phagocytic, fibroblastic mammalian cell lines is the best-studied *T. gondii*-host cell interaction. This interaction starts when a tachyzoite invades the host cell, a process largely mediated by the parasite itself<sup>112,114</sup> (**Figure 1.3.**). Once a tachyzoite binds its apical end to a host cell the invasion process starts. This process is mediated by two sets of secretory organelles located at the apical end of the parasite, the micronemes, and the rhoptries. Micronemes are numerous small rice-shaped organelles that secrete micronemal proteins (MICs)<sup>112</sup>. MIC proteins include both adhesins involved in the recognition of glycoconjugates on the host cell surface (MIC1 – MIC8 and AMA1)<sup>108</sup> and transmembrane proteins connected to an acto-myosin based gliding motility motor located underneath the parasite's plasma membrane by thrombospondin-related anonymous proteins (TRAPs)<sup>113</sup>. Rhoptries are



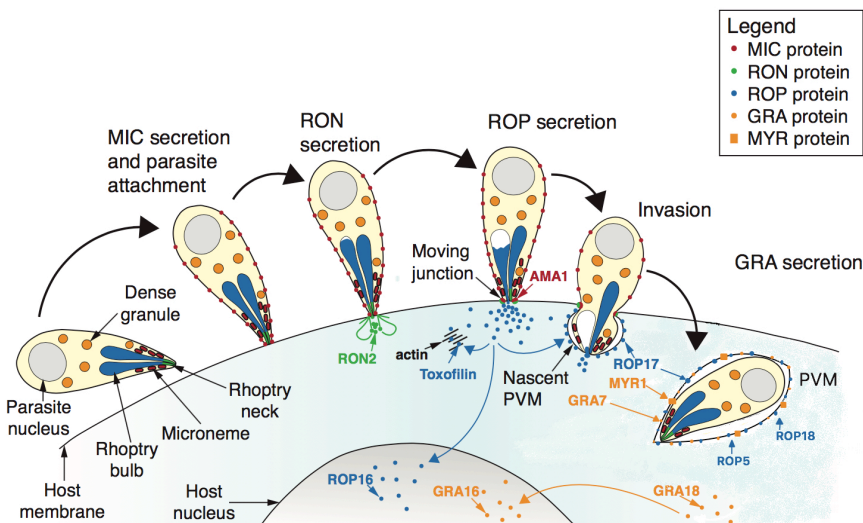
club-shaped organelles that secrete proteins into the host cells not only during the invasion process; they play a very important for the establishment of the parasite's replicative niche (inside of the parasitophorous vacuole (PV)) and the manipulation of the host cell functions<sup>108,115</sup>. Each rhoptry can be partitioned into two regions, an apical region, which is very narrow and contains the rhoptry neck proteins (RONs) and a second wide basal bulb region that contains the rhoptry bulb proteins (ROPs)<sup>112</sup>.

Secreted proteins are crucial for the initial attachment to the host membrane as well as for the establishment of the moving junction (MJ). The MJ is a tight connection between the parasite and the host cell membrane that is formed during the invasion process<sup>116</sup>. The MJ starts at the apical end of the parasite and moves progressively to the posterior part while it enters the host cell<sup>116</sup>, which helps to propel the parasite into a vacuole that will surround the parasite, the parasitophorous vacuole (PV)<sup>114,117</sup>. The MJ is also involved in determining the biochemical composition of the PV membrane (PVM) (e.g. access to GPI-anchored or raft-associated multipass transmembrane proteins and intercellular adhesion molecules<sup>118,119</sup>). The presence of these specific proteins will later determine parasite development and the non-fusogenic nature of the PV<sup>114,116</sup>. Thus, the PVM, even a mere 20 seconds after its formation, is already distinctive in its properties and can no longer be considered representative of the plasma membrane<sup>115,120</sup>.

RON proteins are often conserved among Apicomplexans and play a crucial role in the early stages of the invasion process as anchoring points for the moving junction located into the host membrane (RON2 protein along with other RON partners and AMA1)<sup>108,116</sup>. After the secretion of the RON proteins end, ROP proteins start to be released into the host cell. These proteins may be either translocated into the host nucleus, like ROP16 protein<sup>121</sup>, associated with early stages and the mature PVM, like ROP18<sup>122</sup>

and ROP5<sup>123,124</sup> proteins or to establish an association with cytosolic actin, like Toxofilin<sup>125</sup>.

Another group of proteins that also participate during and soon after the invasion process are the GRA proteins. These proteins are secreted by a set of spherical secretory organelles called dense granules, which are distributed along the parasite's body. A subset of the GRA proteins translocates across the PVM and form associations with ROP proteins, like the ROP18/ROP5/GRA7 complex<sup>126</sup> or for the PV maturation and nutrient acquisition (e.g. GRA17, GRA23<sup>127</sup>, etc.). Other GRA proteins can also localize at the PVM (GRA15<sup>128,129</sup>) and co-opt host cell signaling pathways or only partially interact with the host cell cytosol while they remain in the PV (GRA6<sup>130</sup>). GRA proteins can also be translocated directly to the host cell nucleus (e.g. GRA16, GRA18, TgIST, GRA24, HCE1/TEEGR, and GRA28<sup>131–133</sup>), where they can affect host-cell signaling pathways.



**Figure 1.3.** Integrated model for effector protein secretion by *Toxoplasma gondii* tachyzoites into the host cell. Invasion and establishment of the Parasitophorous Vacuole Membrane (PVM). The first proteins to be secreted during the *T. gondii* invasion are from the micronemes, the micronemal proteins (MICs), which become anchored to the parasite plasma membrane and mediate parasite attachment to the host cell. Once the parasite gets

strongly attached to the host cell, rhoptries discharge their contents. Rhoptry neck proteins (RONs) are secreted first into the host cell; this process goes along with the establishment of the moving junction. Subsequently, rhoptry bulb proteins (ROPs) are secreted and either translocate to the host nucleus (e.g. ROP16), associate with the nascent or fully formed PVM (e.g. ROP17), or associate with cytosolic actin (i.e. toxofilin). Secretion of dense granule proteins (GRAs) into the host also begins at an early stage during parasite invasion. Once the parasites have fully penetrated the host cell, the PVM gets loaded by GRA proteins (e.g. GRA7) and ROPs (e.g. ROP5, ROP18) that help to neutralize host defense proteins. Image modified from reference <sup>112</sup>.

#### *1.3.6.2. ROP proteins: role in pathogenicity*

ROP proteins are determinants of pathogenicity. There are at least 30 ROP proteins encoded in the *T. gondii* genome and all of them belong to the rhoptry 2-kinase family protein superfamily<sup>134</sup>. Some of them have undergone local tandem duplication, locus expansion events and are under strong selection pressure. Members of the ROP2 superfamily are characterized by the presence of an N-terminal domain that encodes membrane-interacting amphipathic helices. Their C-terminal domain encodes either a functional kinase domain or pseudokinase domain<sup>135</sup>.

Most ROP genes that determine virulence in mice were identified using classical genetics. For example, two strains that differ in virulence were fed to cats, upon crossing, the haploid progeny were used to challenge naïve mice and the parasite loci associated with virulence defects identified by QTL (Quantitative Trait Locus) mapping. However, nowadays, new tools for genetic engineering such as the CRISPR/Cas9 technology are slowly replacing the classical genetic strategies<sup>136,137</sup>. Some of the first classical analyses between the highly virulent GT1 strain (Type I) and the avirulent CTG (Type III) strain revealed that a single QTL on Chromosome VIIa could be responsible for differences in virulence between these two strains<sup>108,138</sup>. Finer genome mapping and transcriptional analysis identified the ROP18 gene as a major determinant of pathogenicity. The highly polymorphic

ROP18 gene encodes a serine/threonine kinase; the kinase activity, as well as the capacity to anchor to membranes (amphipathic helix domain) are essential for the virulence-enhancing properties of the protein<sup>139</sup>. ROP18 polymorphism is also responsible for differences in virulence between the Type II ME49 strain and the Type III CTG strain<sup>108,122</sup>. Type III strains under-express the ROP18 gene due to a 2.1kb insertion in the ROP18 promoter, whereas Type I and Type II strains display normal levels of expression<sup>122,138</sup>. South American *T. gondii* strains are carriers of ROP18 alleles that resemble those found in Type I strains, which is correlated with the high virulence seen in these strains in laboratory mice<sup>103,139</sup>.

Most ROP proteins contain kinase-like domains, although many others lack an obvious catalytic triad<sup>134,140</sup>. One of those proteins is the ROP5 protein, which the substitution of a key catalytic aspartate and the non-canonical structure of the bound ATP predicted it to be a pseudokinase<sup>123</sup>. Along with the ROP18 kinase, these two proteins are responsible for differences in acute virulence in laboratory mice. The ROP5 locus consists of a family of 4 to 10 tandem duplicates (isoforms) of highly polymorphic genes<sup>141</sup>. Type I strains have approximately 6 copies, type II strains 10 copies, and type III 4 copies<sup>123</sup>. These isoforms can be classified into three major groups: ROP5A, ROP5B, and ROP5C. Polymorphisms in the ROP5 gene in type I and III strains encode virulence alleles, while the ones found in type II strains encode combinations that are less virulent<sup>102,123,142</sup>. Moreover, almost all South American *T. gondii* strains encode combinations of the ROP18/ROP5 that are considered extremely virulent in naïve laboratory mice<sup>103,141</sup>.

The ROP16 protein is another member of the ROP family that plays an important role in virulence. ROP16 phosphorylates STAT3/STAT6, which suppresses NFkB and therefore lowers inflammation (attenuation of IL-12 signaling and Th2 response)<sup>143,144</sup>. ROP16 alleles in the type I and type III strains are highly similar and have a strong phosphorylation activity,

whereas the type II strains have an inactive form associated with a single polymorphic residue<sup>108,145</sup>.

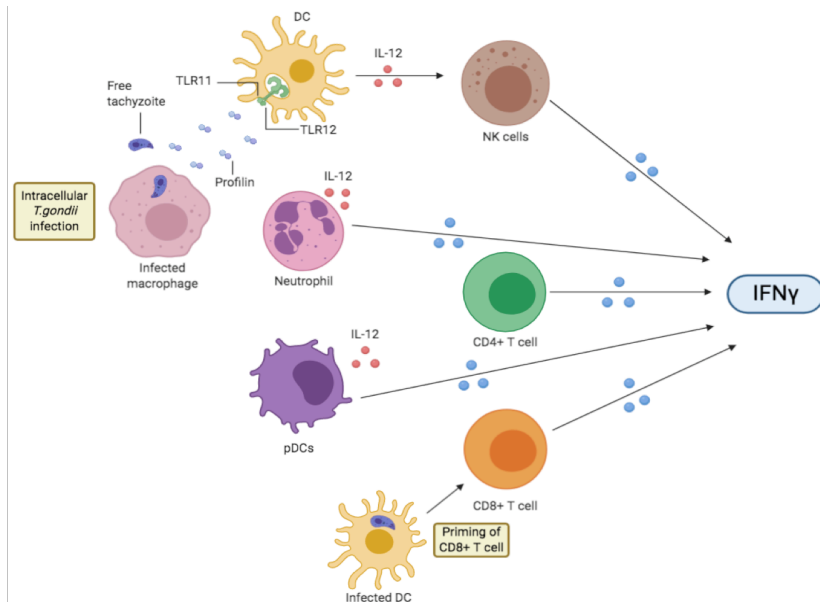
## **1.4. Host immune response**

### **1.4.1. Innate immunity to *T. gondii* in mice**

During the initial infection with *T. gondii* in mice, the parasite rapidly disseminates from the point of entry (e.g. the peritoneal cavity in artificial intraperitoneal injections (i.p) or from the intestinal mucosa in natural oral infections) and disseminates to distant organs such as the brain<sup>108,137,146</sup>. One of the first barriers to control the proliferation and dissemination of *T. gondii* is the innate immune system. This system is armed with a variety of innate pattern recognition receptors (PRRs) (e.g. Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), NOD-like receptors (NLRs), and C-type lectin receptors (CLRs)) that recognize multiple structures conserved among microorganisms or material released by damaged and infected cells<sup>147</sup>. The sensed molecules are named pathogen-associated or damage-associated molecular patterns (PAMPs or DAMPs)<sup>147,148</sup>.

In mice, there are at least two independent pathways able to sense *T. gondii* material to activate immunological signaling cascades. One of them involves the chemokine (cysteine-cysteine motif) receptor 5 (CCR5) by the recognition of *T. gondii*-derived ligands<sup>149</sup> and a second one where TLRs are involved. TLRs are activated upon recognition of PAMPs<sup>148</sup> and encompass an ancient multigene family, where each TLR subfamily has evolved to detect essential components from invading pathogens, maintaining them under selective pressure to preserve their specificity<sup>150</sup>. The TLR-signaling pathway is essential for the control of *T. gondii*, mice deficient in components of this pathway such as IRAK4 and MyD88 show an impaired resistance to the parasite<sup>150–152</sup>. The interaction of *T. gondii* released factors with innate immune cells stimulates the production of the

Th1-polarizing cytokine interleukin-12 (IL-12) by macrophages<sup>153</sup>, neutrophils<sup>154,155</sup>, plasmacytoid DCs<sup>156,157</sup>, and CD8 dendritic cells (DCs)<sup>158,159</sup> (Figure 1.4). TLR engagement also leads to the activation of NF- $\kappa$ b, which induces the transcription of other pro-inflammatory genes such as IL-1 $\beta$ , IL-18, nitric oxide synthase (iNOS) and some NLRs.



**Figure 1.4.** Interleukin 12 and Interferon- $\gamma$  (IFN $\gamma$ ) production in the infection with *T. gondii*. The production of Interferon- $\gamma$  (IFN $\gamma$ ) is critical for the activation of several immune pathways in the control of *Toxoplasma gondii* infections. Multiple cells are able to produce IFN $\gamma$ . Natural Killer (NK) cells produce this cytokine in a Toll-like receptor 11 (TLR11)-mediated manner when dendritic cells (DCs) recognize *T. gondii*'s profilin. Plasmacytoid dendritic cells (pDCs) also produce large quantities of IFN $\gamma$  dependent on TLR11 and present parasite antigens to other cells. CD4+ T cell-derived and CD8+ T cells can also produce IFN $\gamma$ . CD4+ T cell-intrinsic myeloid differentiation primary-response protein 88 (MYD88) pathway regulates T helper 1 (TH-1) cell response to produce IFN $\gamma$ . *T. gondii* antigens sensing by infected DCs mediate CD8+ T cells activation. Neutrophils can also be important innate sources for IFN $\gamma$ . Image inspired in Figure 2 from reference<sup>160</sup> and made in Biorender.

One of the major pathways for the induction of IL-12 is initiated by the detection of the actin-binding profilin-like protein (PRF) by TLR11 (mouse

Chromosome 4) and TLR12 (mouse Chromosome 14)<sup>161–163</sup>. Upon activation with PRF, TLR11 and TLR12 form heterodimers involved in signaling in myeloid DCs<sup>161</sup>. TLR11 and TLR12 are necessary for the activation of macrophages CD11c+ and DCs, whereas TLR12 alone is sufficient for the activation and the production of Type I IFN in plasmacytoid DCs<sup>150</sup>. Among all TLR genes, TLR11 and TLR12 are the most divergent ones and there is some evidence that both genes are relatively new and have rapidly evolved due to positive selection imposed by pathogens<sup>164</sup>. TLR11/12 genes encode functional proteins in several groups of mammals such as rodents, lagomorphs, horses, rhinos, elephants, manatees, etc. However, in humans, TLR11 is a pseudogene (contains three stop codons) located in Chromosome 1 and TLR12 does not exist. These genes are also absent or pseudogenised in orcas, dogs, and cats<sup>58</sup>. It is far from clear whether the distribution of functional TLR11/12 between groups corresponds to the risk of such mammals from *T. gondii*. Polymorphisms in TLRs in humans and domestic animals have been found to affect resistance to pathogens; however, studies in wild mice have reported that polymorphisms in TLR11 and TLR12 only play a minor role in the resistance to *T. gondii*<sup>165</sup>.

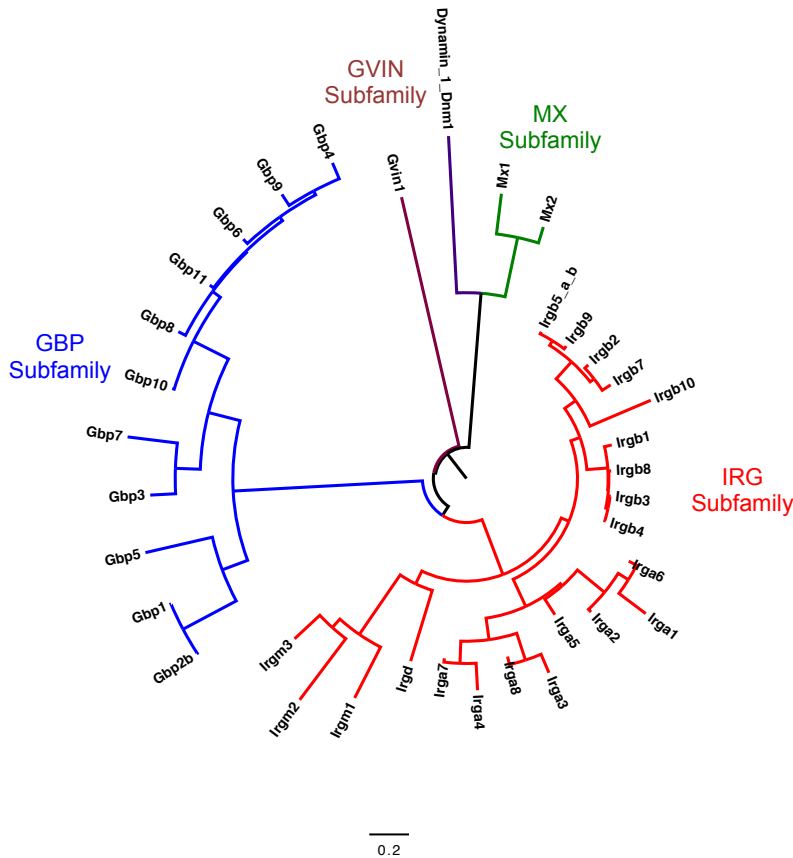
Once IL-12 is in the system, it induces the production of IFN $\gamma$ , initially from NK cells<sup>166,167</sup>, and subsequently from activated CD8<sup>168</sup> and CD4<sup>169</sup> T cells (**Figure 1.4.**). Activation of the IFN $\gamma$  signaling pathway requires phosphorylation of the STAT1 transcription factor and the posterior translocation to the nucleus for the induction of interferon-stimulated genes (ISGs)<sup>170,171</sup>. Even though the activation of STAT1 is quite efficient, if infection occurs earlier than the activation signal, *T. gondii* can block STAT1, inhibiting the cell-autonomous IFN $\gamma$  response and consequently induction of key interferon-inducible resistance proteins such as the MHC proteins, iNOS and the IFN $\gamma$ -inducible GTPases<sup>173,172</sup>.

*T. gondii* sensing has been also attributed to a group of cytosolic NLRs, the NOD, leucine-rich repeat (LRR), and pyrin domain (PYD)-containing proteins (NLPRs), specifically the NLRP1 and NLRP3. These proteins are involved in the formation of multimolecular protein complexes called inflammasomes, which regulate the activity of caspase-1 and the maturation and release of IL-1 $\beta$ <sup>173,174</sup>. Interestingly, sterile immunity in rats is been linked to the *Nlrp1* gene, where macrophages from sterile-immune rats rapidly undergo NLRP1-mediated pyroptosis, which comes along with the maturation and release of IL-1 $\beta$  and IL-18<sup>55,175</sup>.

#### **1.4.2. Interferon-inducible GTPases**

Proteins encoded by ISGs are known for their capacities to enhance pathogen detection and innate immune signaling, as well as for the restriction of intracellular microorganisms such as viruses, bacteria and parasites<sup>176</sup>. According to the Interferome database (v.2.01), which catalogs gene expression profiling studies in the context of IFN stimulation, there are currently 9768 (Genes with p-value <0.05 and fold change  $\geq 2$ ) identified ISGs induced by Type II IFN in the human genome and more than 4336 in the house mouse genome<sup>177</sup>. There are different types of ISGs, among them; the IFN-inducible GTPase superfamily is the most prominent. This superfamily encompasses four subfamilies: the 72–82 kDa Myxoma proteins (Mx), the 65 kDa Guanylate-binding proteins (GBPs) – formerly called p65 GTPases, the 200–285 kDa very large inducible GTPases (VLIGs or GVINs) and the 47 kDa immunity-related GTPases (IRGs) – formerly called p47 GTPases<sup>178</sup> (**Figure 1.5.**).





**Figure 1.5.** The IFN-inducible GTPase superfamily in mice. Unrooted phylogenetic tree based on a nucleotide multiple sequence alignment of several members of the four IFN-inducible GTPase subfamilies: IRGs, GBPs, GVINS and MX proteins. A Jukes-Cantor genetic distance model along with a neighbor-joining building method with no outgroup were computed in the software Mega 6. A bootstrap resampling method with 500 replicates was run as well. Scale=0.2 % differences. Made for this study.

The first IFN-inducible GTPases to be discovered were the mouse Mx1 and human GBP1 more than 30 years ago<sup>179,180</sup>. IRGs were described later in the 1990s<sup>181–187</sup>. Their origins can be traced back to before the radiation of the cephalochordates. Proteins with unambiguous similarities to IRG proteins can be found in the primitive lancelets (Cephalochordate)<sup>188,189</sup>. GVINS, on the other hand, emerged solely on vertebrates<sup>188</sup>. GBPs and IRGs have undergone a dynamic evolution in the vertebrate lineage. For

example, functional GBP gene copies are present in most vertebrate genomes, including mice and humans. However, events of gene loss can be also found, such as the presence of GBPs in tetrapods and zebrafish but they are lost in pufferfishes, sticklebacks, and Japanese rice fishes<sup>188</sup>. There are seven GBP genes and one pseudogene so far identified in the human genome, all encoded within a single gene cluster in Chromosome 1 with close orthologues present in most anthropomorphic primates<sup>188,190,191</sup>. In contrast, mice possess 11 GBP genes and two pseudogenes encoded in two clusters, one in Chromosome 3 and another in Chromosome 5<sup>192</sup>.

IRG proteins have also undergone dramatic episodes of gene duplication, gene conversion, expansion, and loss across vertebrate species. For example, IRG genes are present in all fish genomes reported to today, but variations in copy numbers are frequently found (e.g. Zebrafish harbors 11 IRG genes, whereas pufferfishes only 2)<sup>188,189,193</sup>. Recent birds appear to have lost the IRG gene family entirely, but preliminary results from our lab have shown the presence of one IRG-like gene in the ancient group of flightless birds (Palaeo) that includes ostriches, tinamus and kiwis (Howard et al., unpublished data). This new data might represent a reevaluation of the hypotheses around the loss of these genes in birds, such as the lack of a selective pressure to maintain them or the high fitness cost that IRGs may have on these species. In mammals, IRG genes have an erratic presence, where monotremes (platypus), marsupials (opossum), some ungulates (elephant), insectivores (hedgehog), primates (lemur), dogs and small rodents (mouse, rat, guinea pig, hamster) have them, but pigs, bats, horses, domestic cats, Old and New World monkeys, and hominines do not<sup>55,189,194</sup>. In fact, humans and other hominines express only a gene fragment related in sequence and chromosomal location to other IRGs known as IRGM. This locus has been genetically associated with Crohn's disease, an inflammatory bowel disease of uncertain etiology<sup>194,195</sup>.

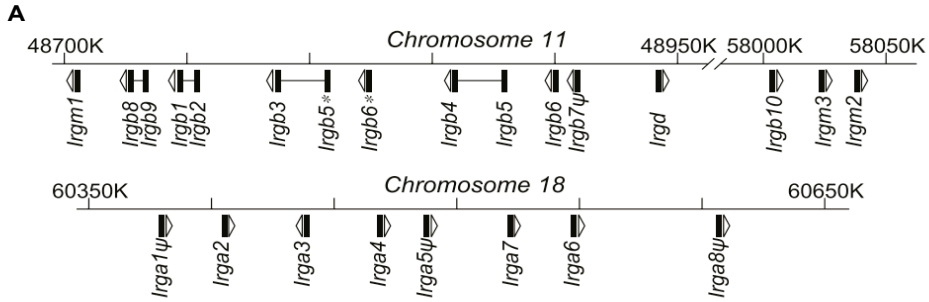
GBP, Mx, and IRG proteins are grouped into the large dynamin-like GTPases superfamily. These proteins can be distinguished from the small Ras-like GTPases and other regulatory GTPases (e.g.  $\alpha$ -subunits of heterotrimeric G-proteins) due to their structural and biochemical similarities, as well as their role in host resistance against intracellular pathogens<sup>187,196,197</sup>. Despite differences in their primary sequences, all dynamins have a distinct large globular N-terminal GTPase domain ( $\approx 200$  amino acids) and share the presence of a middle domain and a C-terminal helical GTPase effector domain (GED). These latter domains, most clearly defined for Mx and GBP proteins, are involved in the oligomerization and regulation of GTPase activity. Catalytic activity stimulated by self-oligomerization of the protein needs a highly coordinated interaction between the GTPase domain, the middle domain, and the GED. The GTPase domain contains the four classical GTP-binding motifs (G1-G4), which are involved in the binding and hydrolysis of the guanine-nucleotide. These motifs are highly conserved, however, some proteins like the GBPs have a diverse G4 motif. Additionally, dynamins and IRG proteins have a low GTP-binding affinity and some of them can interact with lipid membranes<sup>197</sup>. In the particular case of the IRGs, the only IRG protein so far from which the crystal structure has been resolved is Irga6. *In vitro* experiments have shown that this myristoylated protein forms GTP-dependent oligomeric complexes that cooperatively hydrolyze guanosine triphosphate (GTP) to guanosine diphosphate (GDP)<sup>198</sup>. This process works in a cyclical way, where the binding of GTP by the GTPase initiates a conformational change that drives a forward step in the cycle, whereas the hydrolysis of GTP to GDP terminates the forward activity and reverts the GTPase to its inactive state<sup>199</sup>.

The oligomerization is specifically dependent on the formation of symmetrical homodimers via an interface involving the nucleotide-binding

site in the G-domain. GTP is then hydrolyzed in trans, where two Irga6 proteins act as mutual GTPase activating proteins<sup>200</sup>. Consistently, the specific rate of GTP hydrolysis by Irga6 is cooperative. Generally, the IFN $\gamma$ -inducible GTPases have low affinities for guanine nucleotides, in the micromolar range. Unusually, Irga6 has a higher affinity for GDP (~1 $\mu$ M) than for GTP (~15 $\mu$ M). Unpublished work reported by our lab has shown that Irgd and Irgb6 do not share this attribute of Irga6, do not oligomerize in the presence of GTP and do not show cooperative GTP hydrolysis activity<sup>201</sup>.

#### *1.4.2.1. Immunity-related GTPases*

IRG genes are strongly induced via Type II IFN (IFN $\gamma$ ) and to a lesser extent by Type I IFN (IFN- $\alpha/\beta$ ) upon infection in mice<sup>181–187</sup>. The approximately 420 amino acids and 47 kDa units of these proteins are typically encoded by a single large exon, however, in some cases, the initial methionines are encoded with a few N-terminal residues on short 5' exons<sup>202</sup>. IRG proteins contain a Ras-like G domain and a helical domain combining N- and C-terminal elements<sup>203</sup>. In mice, where these proteins have experienced the most significant gene expansion and increase in allelic variation, there are about 21 intact IRG coding units in the reference genome of the laboratory strain C57BL/6 (BL6). They are encoded in two adjacent clusters on Chromosome 11 and in one cluster on Chromosome 18 (**Figure 1.6**). By the end of 1998, at least six members of the IRG family were identified and individually named, however, when the whole family was identified by 2005, a unified nomenclature was established based on sequence homology using the generic name IRG<sup>189</sup>. The *M. musculus* current genome assembly GRCm38 (mm10) from the Genome Reference Consortium has not properly attributed this unified nomenclature to some of the genes that encode the IRG proteins.



**Figure 1.6.** IRG genes genomic location in inbred mouse strains. Linear order of IRG gene clusters on Chromosome 11 and Chromosome 18 in the laboratory mouse strain C57BL/6.

IRGs can be classified within two major functional and sequence subfamilies, the “GKS” proteins (IRGA, IRGB, and IRGD) and the “GMS” proteins (IRGM). The “GKS” subfamily is characterized by the presence of a canonical and highly conserved GxxxxGKS/T in the P-loop of the nucleotide-binding site of the G-domain (G1), whereas the “GMS” subfamily contains an altered GxxxxGMS motif<sup>181,187,189</sup>. Even though it has been reported in other GTPases that mutations in the G1 motif lysine lead to impaired nucleotide-binding properties, there is no evidence for this in the GMS proteins<sup>197,204,205</sup>. Indeed there is no direct evidence that IRGM proteins retain the ability to hydrolyse GTP; it is probable that by this criterion they are pseudogenes since the positive charge on lysine in the GKS motif contributes to stabilization of the gamma-phosphate of GTP before hydrolysis. So far, at least 12 IRG proteins have an amino-terminal myristoylation signal identified, which potentially allows them to target proteins of endomembranes and plasma membrane systems<sup>199,206,207</sup>. However, Irga6 is the only protein where lipidation of the N-terminal has been demonstrated *in vivo*. The usual, probably dynamic, subcellular localization of Irga6 is in the endoplasmic reticulum (ER) (approximately 60 %) and the cytosol (approximately 40 %) in a GDP bound form<sup>189,199,206,208</sup>. GMS proteins are localized on endomembranes from the endolysosomal system, ER, mitochondria, and lipid droplets<sup>209–212</sup>. This specific subcellular localization of the IRGs relies on negative regulation by GMS proteins

functioning as guanine nucleotide dissociation inhibitors (GDIs) to control the nucleotide-bound state of the GKS proteins and retain them in the inactive state. This type of regulation is essential to avoid GKS proteins binding to endomembranes or/and getting prematurely activated<sup>193</sup>.

There is a distinct group of four proteins within the IRGBs of the GKS subfamily (Irgb2-b1, Irgb5-b3, Irgb5-b4, and Irgb9-b8), the “tandem” proteins. They consist of two apparently canonical 47 kDa GKS units that are encoded on adjacent coding exons and expressed as “tandem” proteins with a molecular weight of approximately 94 kDa<sup>202</sup>. Not all tandem proteins are expressed in all mice, some mouse strains display a strong expression, whereas others barely express them<sup>21,55,202</sup>. Finally, there is an additional member of the IRG superfamily encoded in Chromosome 7 called Irgc or Cinema (IRGC). This IRG homolog is constitutively expressed in testis, is not inducible by IFN, and is highly conserved in all mammals. In some other mammalian species (e.g. pig, dog, shrew, opossum, and bat) IRGC is closely linked to a homologous sequence, “IRGC-like” or IRGC2, which is also expressed in the testis and not IFN $\gamma$ -inducible<sup>189,194</sup>. Irgc has ancient roots with no obvious phylogenetic relationship with other IRGs and so far it has not shown pathogen control properties as other IRGs do<sup>189</sup>.

IRG proteins play an important role in the resistance against a wide variety of intracellular pathogens. This conferred resistance is cell-autonomous based and has been characterized in different types of cells, such as fibroblasts, astrocytes, and macrophages. Multiple members of the IRG protein family have shown a proven microbe control activity. One of them is the Irgm1 protein; Irgm1 KO mice are susceptible to a wide range of intracellular protozoa (*Leishmania major*, *Trypanosoma cruzi*, and *T. gondii*) and intracellular bacteria (*Listeria monocytogenes*, *Salmonella typhimurium*, *Mycobacterium avium*, *M. tuberculosis*, and *Chlamydia trachomatis*)<sup>213–219</sup>.

Some authors have proposed some mechanisms of action for *Irgm1* that include the enhancement of phagosome maturation (lysosome fusion) and induction of autophagy in mycobacterial immunity to destruct pathogen-containing vacuoles<sup>217,220</sup>. However, these explanations obscured important literature on *Irgm1* deficiency and activity that point in an entirely different direction. Infections in *Irgm1*-deficient mice have demonstrated that these animals enter into immune failures. Rather than the absence of *Irgm1*, it is the rest of the IFN response that is causing the problem. The GTPase cycle of the effector IRG proteins cannot be properly controlled in the absence of *Irgm1*. This results in the formation of large IRG protein aggregates in certain cell types that can cause in the mouse a generalized lymphopenia. Losing *Irgm3* and *Irgm1* together causes rapid clearance of the aggregates and relieves the cytopathic phenotype<sup>221</sup>.

*Irgd* KO mice are susceptible to *T. gondii* but resistant to *L. monocytogenes*, *S. typhimurium*, *M. tuberculosis*, and Murine cytomegalovirus (MCMV)<sup>213</sup>. Additionally, *Irgm3* KO mice are also susceptible to the infection with *T. gondii* and *C. trachomatis* but showed resistance to *T. cruzi*, *L. monocytogenes*, *S. typhimurium*, *M. avium*, *M. tuberculosis* and MCMV<sup>222</sup>. *Irgb10* has also been involved in the growth restriction of *Francisella novicida*<sup>223</sup>, as well as for *C. trachomatis* and *C. psittaci*, but it does not control the mouse-adapted strain *C. muridarum*<sup>224,225</sup>. *Irgb10* deficiency is also reported to have no impact on the control of avirulent *T. gondii* strains<sup>226</sup>. Additional members of the IRGA and IRGB families are also required for control of pathogen growth. *Irga6* KO mice are susceptible to some avirulent *T. gondii* strains and *C. trachomatis*<sup>206,227</sup> in contrast to *Irgb10*-deficient mice<sup>226</sup>. Absence of the two GMS regulator IRG proteins, *Irgm1* and *Irgm3* results in what is effectively an IRG-effector null phenotype, highly susceptible to avirulent *T. gondii*. IFN $\gamma$ -induced cells from such mice are unable to control growth of *T. gondii*, or the microsporidian fungal pathogen *Encephalitozoon cuniculi*<sup>228</sup> or the parasite *Neospora*

*caninum*<sup>229</sup>. Unexpectedly, another intracellular Apicomplexan parasite distantly related to *T. gondii*, namely *Plasmodium berghei* is not controlled by the IRG system<sup>230</sup>. Further studies in intracellular pathogens such as *Hammondia hammondi* and *Cryptosporidium spp* are still needed to continue unraveling the mechanisms underlying the ability of IRG proteins to attack membranes in vacuole-forming Apicomplexan parasites.

#### 1.4.2.2. Immunity-related GTPases in mice and cell-autonomous resistance against *T. gondii*.

When an interferon-stimulated cell is infected with an avirulent *T. gondii* strain, the recently formed PVM, due to remodeling by the parasite, is now permissive for the binding of proteins like IRGs, unlike the plasma membrane from which it came from. It is unclear what the relevant changes are, but they happen fast. IRG proteins initiate a rapid accumulation onto the PVM few minutes post parasite invasion<sup>231,232</sup>. The timing until IRG proteins fully load is highly variable for individual vacuoles, but it increases on average up to about 90 minutes. This massive accumulation of multiple IRG proteins occurs hierarchically and cooperatively, being Irgb6 and Irgb10 the pioneers, loading first, followed by Irga6 and Irgd<sup>232</sup>. It is thought that this cooperative behavior contributes to the genetic non-redundancy of the individual proteins<sup>194</sup>. The proposed mechanism of action for these proteins states that monomeric GKS proteins in a cytosolic GDP-bound state reach the PV by diffusion. Once on the PVM, GKS become activated by binding GTP cooperatively and towards the end of loading, PVM disruption starts<sup>210,231,233–235</sup>. Disruption occurs by an abrupt local break in the PVM integrity, which peels back the PVM and leaves the parasite exposed to the cytosol<sup>194,236</sup>. Right before the total disruption occurs, the typical banana-shaped vacuole starts to round up as if the membrane surface area surrounding the vacuole is being actively reduced<sup>236</sup>. The PVM in this case appears to be extensively folded into ruffles, probably associated with a vesiculation process that leads to the membrane bursting by build-up



tension<sup>231,233,237</sup>. The parasite exposed to the cytosol dies within approximately 20 minutes and the host cell experiences necrotic death about one hour after<sup>236</sup>. The cause of death of *T. gondii* is so far unknown, as well as the specific trigger for the host cell death. Both processes are characterized by the loss of plasma membrane integrity and, for the cellular necrosis, the release of the chromatin remodeling protein High-mobility group box 1 (HMGB1), which is distinct from processes such as apoptosis or classical pyroptosis<sup>236</sup>.

#### 1.4.2.3. IRG inactivation by virulent *T. gondii* strains

While the IRG system efficiently controls the infection with avirulent *T. gondii* strains, infections with virulent strains are associated with a dramatic failure of this system. For example, in IFN $\gamma$  stimulated cells infected with type I virulent *T. gondii* strains, the initial loading of the PVM with the IRG proteins is remarkably reduced<sup>232,235</sup> and the failure of the IRG system results in the death of the susceptible host in few days due to an acute infection. *T. gondii* strains have evolved several polymorphic effector proteins to block the IRG system in order to preserve the integrity of the PVM and to continue with their replication process. These polymorphic effector proteins are the pseudokinase ROP5 and the active kinase ROP18, which both localize to the external surface of the PVM<sup>138,238</sup> after initial secretion into the cytosol at the point of entry. These proteins form a complex necessary for the efficient phosphorylation of conserved threonine residues of the switch I loop of the nucleotide-binding domain of effector IRGs such as Irga6 and Irgb6<sup>239,240</sup>. Avirulent type II and type III strains can also express the ROP5 and ROP18 proteins<sup>122,141</sup>, however, they have allelic forms with a lower expression and various insertions and deletions in the promoter<sup>139,241</sup> that fail to restrict IRG protein accumulation. Interestingly, for an efficient phosphorylation of the IRG proteins, they must be kept in the inactive GDP-bound conformation, which is facilitated by the direct binding of the ROP5 pseudokinase<sup>141,242</sup>.

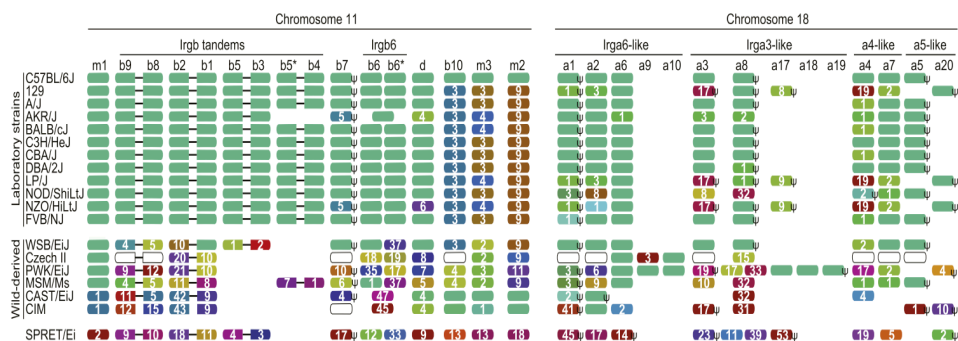
ROP5 is also directly associated with the GRA7 protein, which is not only essential for efficient phosphorylation of Irga6, but it also regulates along with ROP5 the kinase activity of ROP18<sup>126,243,244</sup>. Another active ROP protein with demonstrable phosphorylation activity for Irga6 and Irgb6 is ROP17. Even though it is able to phosphorylate both proteins, ROP17 shows a stronger affinity for Irgb6, and unlike ROP18, kinase activity is independent of ROP5<sup>245</sup> (ROP5/ROP18/GRA7-ROP17 complex). All together, this shows that a successful load of IRG proteins on the PVM is vital for host survival<sup>246</sup>. However, it also begs the question, what is the selective advantage for the parasite of virulent effector proteins that kill the host within a few days, thus blocking the transmission cycle.

#### *1.4.2.4. IRG diversity in wild mice and resistance to virulent T. gondii strains*

Infections with virulent *T. gondii* strains are paradoxical, because virulence leads to lethality while avirulence will result in persistence and transmission in nature<sup>247</sup>. While laboratory mouse strains have a low genetic diversity in their IRG system, wild-derived inbred mouse strains on the contrary displayed considerable polymorphism in IRG proteins<sup>21</sup>, where almost every one carries a unique IRG genotype (Figure 1.7). IRGs vary in their level of diversity, for example, Irgm1, Irgd and Irgb10 are largely identical among mouse strains, whereas other IRGs such as Irgb6 and Irgb2-b1 are on the contrary exceptionally polymorphic. The polymorphic complexity found in many IRGs from wild-derived and wild mice reaches the level of the major histocompatibility complex (MHC)<sup>21</sup>. In fact, genetic diversity in the IRG system seems to be the key to protect mice against the activity of the virulent and highly polymorphic ROP proteins. The wild derived inbred mouse strain CIM (*Mus musculus castaneus*) from South India is able to counteract the virulence factors from Eurasian type I virulent strains, which allows parasite's encystment, and therefore potential transmission. More recently, resistance to the type I virulent strain RH has also been reported in

the PWK/PhJ strain, that belongs to the *Mus musculus musculus* subspecies<sup>248</sup>. In the case of CIM mice, initial in vitro experiments with CIM cells displayed a remarkable resistance to both virulent and avirulent strains as well as an efficient loading of Irga6 and Irgb6 to the PVM in both conditions<sup>21,55,202</sup>. Additionally, IFN $\gamma$ -stimulated CIM cells experience a rapid necrotic cell death upon infection with virulent and avirulent *T. gondii* strains, which controls the spread of the infection<sup>21</sup>.

Several different Eurasian type I *T. gondii* strains such as RH, BK and GT-1 that are virulent in laboratory strain mice, are avirulent in CIM mice. In breeding experiments between susceptible C57BL/6 mice and resistant CIM mice resistance to virulent *T. gondii* strains was mapped to highly polymorphic IRG genes located on chromosome 11<sup>21</sup>. This locus contains, among other polymorphic IRGs, the Irgb2-b1 gene.



**Figure 1.7.** Polymorphism at the protein level in IRG genes (Chromosome 11 and Chromosome 18) of inbred mouse strains. Each color block represents one IRG open reading frame. Numbers represent amino acid substitutions/indels relative to the C57BL/6 allele. The colors of the blocks indicate their phylogenetic relationship. 'ψ' indicates probable pseudogenes. Image taken from reference<sup>21</sup>.

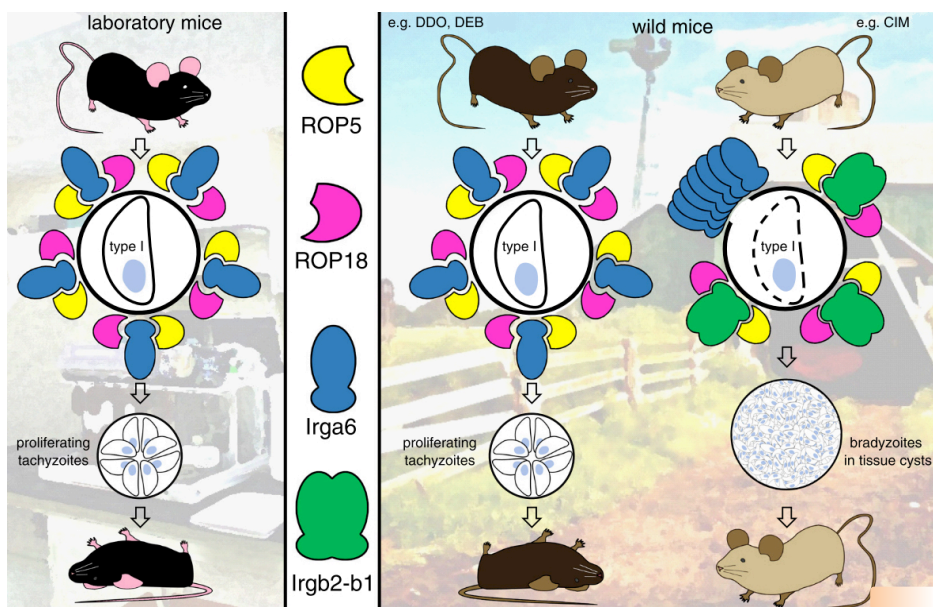
*T. gondii* ROP5 protein binds to the IRGA6 protein in the region around helix 3 (Arg-159)<sup>142</sup>. In the Irgb2 protein, that forms the N-terminal half of the Irgb2-b1 tandem protein, it is possible to find signs of recent divergent

selection in substitutions located in both the ROP5 binding site and in the switch loop I (the putative H4 and  $\alpha$ D structural domains), where the threonine residues targeted for phosphorylation by ROP17 and ROP18 are located. ROP5 protein and Irgb2-b1 interacting surfaces strongly suggests that there is indeed selection pressure from probably a long-term coevolutionary arms race<sup>21,242</sup>

Expression levels of Irgb2-b1 vary with the mouse strain, for example cells from C57BL/6 mice and most other laboratory strains, barely express this protein, whereas cells from some wild-derived mouse strains strongly express it. This difference segregates with the Irgb2-b1 allele and thus appears to depend exclusively on the local promoter and enhancers. This difference is also reflected in an striking difference in Irgb2-b1 accumulation on the PVM of IFN $\gamma$ -stimulated CIM cells (high) versus C57BL/6 cells (low), both infected with type I virulent *T. gondii* strains<sup>21,55,249</sup>. More importantly, in experiments in IFN $\gamma$ -stimulated C57BL/6 cells, expression of transiently transfected Irgb2-b1<sub>CIM</sub> protected endogenous Irga6<sub>C57BL/6</sub> from ROP5/ROP18/GRA7-mediated phosphorylation by the virulent strain RH<sup>21</sup>. Irgb2-b1<sub>CIM</sub> is itself phosphorylated by virulent ROP5/ROP18 complexes while Irga6 is not. Thus the Irgb2-b1<sub>CIM</sub> probably acts as a decoy, diverting the parasite active kinase complex from effector IRG proteins, allowing control of the tachyzoite growth phase<sup>21</sup> (Figure 1.8.). For this reason, the group of tandem IRG proteins is named as “Decoys” as distinct from the other single-unit GKS proteins, the “effectors” and the GMS proteins, the “regulators”.

A recent study has revealed the subtleties of the molecular interaction between ROP5 and Irgb2-b1. Murillo-León and collaborators<sup>250</sup> found that efficient binding of Irgb2-b1<sub>CIM</sub> occurs only to the RH strain-derived ROP5B isoform, but not to the ROP5A and ROP5C. This data contributed to confirm the critical role that ROP5 plays for the heightened virulence of *T. gondii*

type I strains against laboratory mice and it also suggest that other ROP pseudokinases may not play a significant role in virulence/avirulence behavior in mice against *T. gondii*. Resistance found in the wild-derived mouse strains CIM and PWK/PhJ most likely reflects the complexity associated to the interactions between natural allelic forms of the ROP proteins and the *Irgb2-b1* proteins in the wild. Thus, they shape not only the local population structure of the parasite, but also of the mouse host in human habited and wild environments<sup>58,59</sup>. CIM mice, for example, are susceptible to the infection with virulent *T. gondii* strains from South America such as VAND and AS28. The mismatch between *Irgb2-b1*<sub>CIM</sub> and ROP5 isoforms from these atypical strains represents a failure to control infection. Compatibility between a Eurasian mouse such as the CIM and type I *T. gondii* strains (frequently isolated in Asia) strongly suggest that *T. gondii* virulence and mouse resistance follow some form of local allele-matching evolutionary dynamics<sup>21,55,58,250</sup>. It is nevertheless notable that *Irgb2-b1* from PWK is also resistant to virulent type I strains despite substantial sequence differences from *Irgb2-b1*<sub>CIM</sub>, and despite the fact that virulent type I strains are not especially abundant in Eastern Europe and Northern Asia. Thus it follows that the interactions between virulence effectors and decoy IRG proteins allows a degree of redundancy.

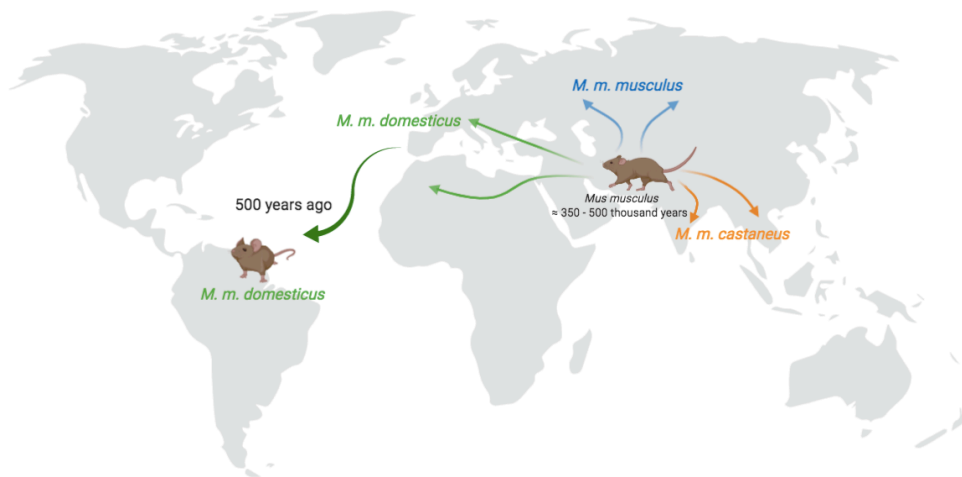


**Figure 1.8.** Susceptibility and resistance of mouse genotypes against type I virulent strains of *T. gondii*. A typical laboratory inbred strain such as C57BL/6 (left panel) expresses effector IRG proteins such as Irga6 that are phosphorylated and inactivated by a kinase complex of ROP5 and ROP18 secreted by the parasite. Immunity against *T. gondii* fails and the mice die within a few days of infection. Such susceptible IRG alleles are segregating in wild mouse populations with other IRG alleles that confer resistance to the virulent kinase complex (right panel). In particular, the highly polymorphic ‘tandem’ IRG protein, Irgb2-b1, of such resistant strains (e.g. CIM) acts as a decoy, diverting the parasite active kinase complex (ROP5/ROP18/GRA7) from effector IRG proteins (e.g. Irga6 and Irgb6) and allowing adequate immunity to develop to control the tachyzoite growth phase. In a mouse with such a genotype, “virulent” *T. gondii* strains become avirulent, form cysts, and can be transmitted. Other wild strains (such as DDO from Denmark and DEB from Spain) carry susceptible Irgb2-b1 haplotypes and will be as vulnerable as C57BL/6. Most likely, ROP kinases that are highly virulent for mice arise under pressure from evolutionarily relevant intermediate hosts that express IRG proteins capable of yielding sterile immunity to a primary infection. Image taken from reference <sup>58</sup>.

## 1.5. The house mouse world radiation

Globalization has turned the house mouse (*Mus musculus*) into perhaps the most successful invasive mammal in the world<sup>251</sup>. Originating in the Indo-Pakistan subcontinent and neighboring present-day Afghanistan and Iran<sup>252–</sup>

<sup>255</sup>, house mice comprise three main subspecies with different current global distributions: *M. m. domesticus*, *M. m. musculus* and *M. m. castaneus* (Figure 1.9.). Their closest rodent relatives are the non-human commensals mound building mouse, *Mus spicilegus*, and the Algerian mouse, *Mus spretus*<sup>256–258</sup>. The three subspecies of *M. musculus* started to diverge during the Pleistocene climatic oscillations (~350–500 thousand years ago (KYA))<sup>259–261</sup>. This split among the three subspecies occurred within a relatively short period of time. However, genetic and genomic data indicates that *M. m. castaneus* and *M. m. musculus* are more closely related to each other than either is to *M. m. domesticus*<sup>262,263</sup>. There is also evidence of multiple hybridization events among the subspecies in zones of secondary contact in Europe and Asia<sup>255,261</sup>.



**Figure 1.9.** Worldwide distribution of *Mus musculus* subspecies. The three subspecies of *M. musculus* started to diverge ~350–500 thousand years ago and the split among the three subspecies occurred within a short period of time. Current localization of the three subspecies is shown in green (*M. m. domesticus*), in blue (*M. m. musculus*) and in orange (*M. m. castaneus*). Dark green arrows indicate inferred routes of historical migrations and recent movements in association with humans (e.g. human colonization of the American continent 500 years ago). Image adapted from references<sup>55,258,264–266</sup> and made in Biorender.

Unique among other species in the genus *Mus*, house mice are human commensals and therefore primarily live in close proximity to humans. Although feral populations exist, house mice are commonly found in residential, agricultural, and commercial structures. Commensalism probably evolved independently in each of the three house mouse subspecies<sup>255,258</sup>. If correct, this implies that subspecific divergence preceded human association. Bioarchaeological evidence and genetic studies on modern populations agree on the origin of the *M. m. domesticus* commensal behavior associated with the Neolithic transition (10,000 - 12,000 years ago) in the broad region of the ancient Fertile Crescent, in the Middle East when mice entered the domestic space of humans<sup>267-269</sup>. Whether human sedentism and/or the rise of the agrarian societies were the key driving factors that led to this behavioral shift is still debated<sup>251</sup>. *M. musculus* phylogeography reflects human colonization and settlement patterns. In association with humans house mice have been spread around the world and have adapted to a wide range of environments. Mice can be found from sea level to over 4000 m in elevation, from tropical to subarctic locations, and in both dry and wet environments<sup>258,270</sup>.

#### **1.5.1. Expansion of the *M. m. domesticus* subspecies**

Among the three subspecies, *M. m. domesticus* is now the most common and widespread one, it currently occupies Western Europe, North and South America, Africa, Australasia and many oceanic islands worldwide<sup>255</sup>. Initial dispersion of *M. m. domesticus* occurred from the Middle East towards Europe, in a period of time that followed the Neolithic diaspora stemming from Southwest Asia. Zooarchaeological evidence collected from the Tigris-Euphrates Basin up to the eastern Mediterranean coasts suggests an increase in house mouse populations approximately during the Late Pleistocene<sup>267</sup>. In the area of Levant, sedentism of the human Natufian communities most likely created the ecological niche that promoted the



commensal behaviour of *M. m. domesticus* populations from 14,500 BP, and potentially the commensal pathway for another important species in the house mouse history, the domestic cat<sup>251</sup>. House mice were later involuntarily transported by humans mostly on ships and initially spread to new territories in Eurasia and north Africa<sup>271</sup>. One of the first direct evidences of house mice transportation as maritime stowaways was obtained on the shipwreck “Uluburun” from the late Bronze Age in Turkey. The mandible of a small mammal was recovered on the remains of the vessel and identified as a house mouse<sup>272</sup>. Later, with the improvement in ocean navigation techniques, humans expanded their territory and along with them the house mice, by exploiting human and livestock food supplies<sup>273</sup>.

The beginning of the 15<sup>th</sup> century marked a new historical era known as the “Age of Discovery”. Portugal and Spain were the leading Western European colonial empires to establish the first global trade network, followed by the Dutch, the French and the British<sup>274,275</sup>. Mice most likely were transported in those transatlantic trips to the Americas. In general, house mice in North and South American territories are descended primarily from *M. m. domesticus*, introduced into the New World by several waves of European explorers<sup>276</sup>. For example, Portuguese colonizers were probably the first to transport house mice (*M. m. domesticus*) on trading ships to the now-Brazilian territory in the 1500’s and continued to do so due to the constant commercial trade between Brazil and Portugal<sup>277</sup>.

### **1.5.2. Appearance of the laboratory and wild-derived mouse strains.**

House mice have been widely exploited as an excellent mammalian model for studying a wide variety of traits and diseases<sup>258</sup>. Human interest in the species started with the trade of mice with distinct coat colors and behaviors in ancient China, Japan, and Europe<sup>278</sup>. The first mice for experimentation

were used as early as the 17th and 18th centuries. During the 19th century zoologists in Europe bred fancy mice to study various characteristics, but their results could not be interpreted correctly until 1900 when Mendel's laws were rediscovered<sup>279</sup>. Since then, researchers saw the advantages of working with a mammal that could be housed in a small area, bred quickly, and that displayed many easily scored, variable traits<sup>258</sup>. Current classical laboratory inbred mouse strains are genetic mosaics of the three main subspecies, although they are primarily *M. m. domesticus* in origin<sup>258,280</sup>. However, these strains are derived from a limited set of founders and thus contain only a small subset of the genetic variation that is seen in nature<sup>280</sup>. Therefore, as an alternative in the search for animal models that better resemble the genetic diversity seen in nature, researchers since the 1970's have created new inbred lines using wild mice. These wild-derived inbred strains have much higher levels of genetic diversity, which is extremely useful, among other topics, in the study of host-pathogen interactions<sup>280,281</sup>.

## **1.6. Thesis aims.**

Recent studies on resistance to virulent *Toxoplasma gondii* strains in wild-derived Eurasian mice such as the CIM strain have shown the importance *in vitro* and *in vivo* of the Irgb2-b1 tandem protein<sup>21</sup>. Interestingly, CIM mice are only resistant to virulent strains such as type I strains (commonly found in Asia), but not to the highly virulent and genetically diverse South American *T. gondii* strains<sup>55</sup>. This strongly suggested that *T. gondii* virulence and mouse resistance might follow some form of allele-matching evolutionary dynamics, which is associated with local interactions. Given the relatively short period of time since the arrival of the first old-world *M. musculus domesticus* to the Americas (only 500 years ago) and the diverse genetic structure of the South American *T. gondii* strains, we hypothesized that the IRG alleles from South American house mice have evolved to

reflect co-adaptation to local *T. gondii* strains and therefore resistance. We then investigated:

1. IRG protein diversity in wild *M. m. domesticus* populations in both Europe and South America. Our focus was in Portuguese and Brazilian mouse populations.
2. The impact of *Irgb2-b1* alleles found in Brazilian mice in the resistance *in vitro* and *in vivo* to virulent South American *T. gondii* strains.
3. Genetic diversity of Brazilian *T. gondii* strains and isoform characterization in the *ROP18* and *ROP5* genes.

## Bibliography

1. Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat. Genet.* **32**, 569–577 (2002).
2. Retel, C., Kowallik, V., Huang, W., Werner, B. & Künzel, S. The feedback between selection and demography shapes genomic diversity during coevolution. *Sci. Adv.* **5**, (2019).
3. Thrall, P. H., Barrett, L. G., Dodds, P. N. & Burdon, J. J. Epidemiological and Evolutionary Outcomes in Gene-for-Gene and Matching Allele Models. *Front. Plant Sci.* **6**, 1–12 (2016).
4. Mode, C. J. A Mathematical Model for the Co-Evolution of Obligate Parasites and Their Hosts. *Soc. Study Evol.* **12**, 158–165 (1958).
5. Haldane, J. B. . Disease and evolution. *Ric. Sci. Suppl.* **19**, (1949).
6. Luijckx, P., Fienberg, H., Duneau, D. & Ebert, D. A matching-allele model explains host resistance to parasites. *Curr. Biol.* **23**, 1085–1088 (2013).
7. Flor, H. H. Genes for resistance to rust in Victory Flax. *Agron. J.* **43**, 527–531 (1951).
8. Kaloshian, I. Gene-for-gene disease resistance: bridging insect pest and pathogen defense. *J Chem Ecol.* **30**, 2419–2438 (2004).
9. Futuyma, D. J. in *Encyclopedia of Insects* (eds. Resh, V. & Cardé, R.) 175–179 (Elsevier, 2009).
10. Agrawal, A. & Lively, C. M. Infection genetics: Gene-for-gene versus matching-alleles models and all points in between. *Evol. Ecol. Res.* **4**, 79–90 (2002).
11. Hamilton, W. D., Axelrodtt, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites ( A Review ). *Proc. Natl. Acad. Sci. U. S. A.* **87**, 3566–3573 (1990).
12. Howard, R. S. & Lively, C. . Parasitism, mutation accumulation and the maintenance of sex. *Nature* **367**, 554–557 (1994).
13. Kawecki, T. J. Red Queen Meets Santa Rosalia : Arms Races and the Evolution of Host Specialization. *Am. Nat.* **152**, 635–651 (1998).
14. Gandon, S., Capowiez, Y., Dubois, Y., Michalakis, Y. & Olivieri, I. Local adaptation and gene-for-gene coevolution in a metapopulation model. *Proc. R. Soc.* **263**, 1003–1009 (1996).
15. Dybdahl, M. F. & Storfer, A. Parasite local adaptation : Red Queen versus Suicide King. *TRENDS Ecol. Evol.* **18**, 523–530 (2003).
16. Lively, C. M. Adaptation by a Parasitic Trematode to Local Populations of Its Snail Host. *Evolution (N. Y.)* **43**, 1663–1671 (1989).
17. Ballabeni, A. P. & Ward, P. I. Local Adaptation of the Tremadote Diplostomum phoxini to the European Minnow Phoxinus phoxinus , its Second Intermediate Host.

- Br. Ecol. Soc.* **7**, 84–90 (1993).
18. Landis, S. H., Kalbe, M., Reusch, T. B. H. & Roth, O. Consistent Pattern of Local Adaptation during an Experimental Heat Wave in a Pipefish-Trematode Host-Parasite System. *PLoS One* **7**, (2012).
  19. Refardt, D. & Ebert, D. Inference of parasite local adaptation using two different fitness components. *Eur. Soc. Evol. Biol.* **20**, 921–929 (2007).
  20. Routtu, J. & Ebert, D. Genetic architecture of resistance in *Daphnia* hosts against two species of host-specific parasites. *Heredity (Edinb.)*. **114**, 241–248 (2015).
  21. Lilue, J., Müller, U. B., Steinfeldt, T. & Howard, J. C. Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. *Elife* **2013**, 1–21 (2013).
  22. Fumagalli, M. *et al.* Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* **7**, (2011).
  23. Lederberg, J. J. B. S. Haldane (1949) on Infectious Disease and Evolution. *Genetics* (1999).
  24. Soares, M., Teixeira, L. & Moita, L. F. Disease tolerance and immunity in host protection against infection. *Nat. Rev. Immunol.* **17**, 83–96 (2017).
  25. Glass, E. J. The molecular pathways underlying host resistance and tolerance to pathogens. *Front. Genet.* **3**, 1–12 (2012).
  26. Caldwell, R. M., Schafer, J. F., Compton, L. E. & Patterson, F. L. Tolerance to Cereal Leaf Rusts. *Science (80-. )*. **128**, 714–715 (1958).
  27. Soares, M. P., Gozzelino, R. & Weis, S. Tissue damage control in disease tolerance. *Trends Immunol.* **35**, 483–494 (2014).
  28. Ferguson, W., Dvora, S., Fikes, R. W., Stone, A. C. & Boissinot, S. Long-term balancing selection at the antiviral gene *oas1* in central African chimpanzees. *Mol. Biol. Evol.* **29**, 1093–1103 (2012).
  29. Gorer, P. A., Lyman, S., Snell, G. D. & Haldane, J. B. . Studies on the genetic and antigenic basis of tumour transplantation Linkage between a histocompatibility gene and “fused” in mice. *Proc. R. Soc. B.* 499–505 (1948).
  30. Snell, G. & Higgins, G. Alleles at the histocompatibility-2 locus in the mouse as determined by tumor transplantation. *Genetics* **36**, 306–310 (1951).
  31. Radwan, J., Babik, W., Kaufman, J., Lenz, T. L. & Winternitz, J. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends Genet.* **36**, 298–311 (2020).
  32. Kaufman, J. Unfinished Business: Evolution of the MHC and the Adaptive Immune System of Jawed Vertebrates. *Annu. Rev. Immunol.* **36**, 383–409 (2018).
  33. Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility

- complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).
34. Bjorkman, P. J. *et al.* The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* **329**, 512–518 (1987).
  35. Blackwell, J. M., Jamieson, S. E. & Burgner, D. HLA and Infectious Diseases. *Clin. Microbiol. Rev.* **22**, 370–385 (2009).
  36. Haller, O., Staeheli, P., Schwemmler, M. & Kochs, G. Mx GTPases : dynamin-like antiviral machines of innate immunity. *Trends Microbiol.* **23**, 154–163 (2015).
  37. Haller, O., Arnheiter, H., Pavlovic, J. & Staeheli, P. The Discovery of the Antiviral Resistance Gene Mx : A Story of Great Ideas , Great Failures , and Some Success. *Annu. Rev. Virol.* **5**, 33–51 (2018).
  38. Haller, O., Acklin, M. & Staeheli, P. Influenza Virus Resistance of Wild Mice: Wild-Type and Mutant Mx Alleles Occur at Comparable Frequencies. *J. Interferon Res.* **7**, 647–656 (1987).
  39. Staeheli, P., Grob, R., Meier, E., Sutcliffe, J. G. & Haller, O. Influenza Virus-Susceptible Mice Carry Mx Genes with a Large Deletion or a Nonsense Mutation. *Mol. Cell. Biol.* **8**, 4518–4523 (1988).
  40. Haller, O., Staeheli, P. & Kochs, G. Interferon-induced Mx proteins in antiviral host defense. *Biochimie* **89**, 812–818 (2007).
  41. Crameri, M. *et al.* MxB is an interferon-induced restriction factor of human herpesviruses. *Nat. Commun.* **9**, 1–16 (2018).
  42. Schilling, M. *et al.* Human MxB Protein Is a Pan-herpesvirus Restriction Factor. *J. Virol.* **92**, 1–11 (2018).
  43. Nicolle, C. & Manceux, L. Sur une infection a corps de Leishman (ou organismes voisins) du gondi. *C R Acad Sci Hebd Seances Acad Sci D* **147**, 736. (1908).
  44. Splendore, A. Un nuovo protozoa parassita de'conigli. incontrato nelle lesioni anatomiche d'une malattia che ricorda in molti punti il Kala-azar dell'uomo. Nota preliminare pel. *Rev Soc Sci Sao Paulo* **3**, 109–112 (1908).
  45. Dubey, J. P. in *Toxoplasma gondii. The Model Apicomplexan: Perspectives and Methods* (eds. Kim, K. & Weiss, L. M.) (Elsevier, 2007).
  46. Douzery, E. J. P., Snell, E. A., Baptiste, E., Delsuc, F. & Philippe, H. The timing of eukaryotic evolution : Does a relaxed molecular clock reconcile proteins and fossils ? *Pnas* **101**, 15386–15391 (2004).
  47. Arisue, N. & Hashimoto, T. Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitol. Int.* **64**, 254–259 (2015).
  48. Wiesner, J., Reichenberg, A., Heinrich, S., Schlitzer, M. & Jomaa, H. The Plastid-Like Organelle of Apicomplexan Parasites as Drug Target. *Curr. Pharm. Des.* **14**, 855–871 (2008).
  49. Fichera, M. E. & Roos, D. S. A plastid organelle as a drug target in apicomplexan

- parasites. *Nature* **390**, 407–409 (1997).
50. Zhu, G., Marchewka, M. J. & Keithly, J. S. Cryptosporidium parvum appears to lack a plastid genome. *Microbiology* 315–321 (2000).
  51. Lorenzi, H. *et al.* Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. *Nat. Commun.* **7**, 10147 (2016).
  52. Kemp, L. E., Yamamoto, M. & Soldati-Favre, D. Subversion of host cellular functions by the apicomplexan parasites. *FEMS Microbiol. Rev.* **37**, 607–631 (2013).
  53. Hutchison, W. M. Experimental transmission of *Toxoplasma gondii*. *Nature* **206**, 961–962 (1965).
  54. Dubey, J. P., Miller, N. & Frenkel, J. K. Characterization of the New Fecal Form of *Toxoplasma gondii*. *J. Parasitol.* **56**, 447–456 (1970).
  55. Müller, U. B. Polymorphism in the IRG resistance system determines virulence of *Toxoplasma gondii* in mice. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2015).
  56. Martorelli Di Genova, B., Wilson, S. K., Dubey, J. P. & Knoll, L. J. Intestinal delta-6-desaturase activity determines host range for *Toxoplasma* sexual reproduction. *PLoS Biol.* **17**, 1–19 (2019).
  57. Boothroyd, J. . *et al.* *Forward and reverse genetics in the study of the obligate, intracellular parasite Toxoplasma gondii. Microbial Gene Techniques, Part B: Molecular Microbiology Techniques* (Elsevier, 1995).
  58. Müller, U. B. & Howard, J. C. The impact of *Toxoplasma gondii* on the mammalian genome. *Curr. Opin. Microbiol.* **32**, 19–25 (2016).
  59. Shwab, E. K. *et al.* Human impact on the diversity and virulence of the ubiquitous zoonotic parasite *Toxoplasma gondii*. *Proc. Natl. Acad. Sci.* **115**, 1–8 (2018).
  60. Johnson, W. E. *et al.* The Late Miocene Radiation of Modern Felidae: A Genetic Assessment. *Science* (80-. ). **127**, 1–6 (2006).
  61. Torrey, E. F. & Yolken, R. H. *Toxoplasma* oocysts as a public health problem. *Trends Parasitol.* **29**, 380–384 (2013).
  62. Shapiro, K. *et al.* Food and Waterborne Parasitology Environmental transmission of *Toxoplasma gondii* : Oocysts in water , soil and food. *Food Waterborne Parasitol.* **12**, e00049 (2019).
  63. Shapiro, K. *et al.* Type X strains of *Toxoplasma gondii* are virulent for southern sea otters ( *Enhydra lutris nereis* ) and present in felids from nearby watersheds. *Proc. R. Soc. B* **286**, (2019).
  64. Aramini, J. J., Stephen, C. & Dubey, J. P. *Toxoplasma gondii* in Vancouver Island Cougars ( *Felis concolor vancouverensis* ): Serology and Oocyst Shedding. *J. Parasitol.* **84**, 438–440 (1998).

65. de Moura, L. *et al.* Waterborne Toxoplasmosis, Brazil, from Field to Gene. *Emerg. Infect. Dis.* **12**, 326–329 (2006).
66. Bahia-oliveira, L. M. G. *et al.* Highly Endemic , Waterborne Toxoplasmosis in North Rio de Janeiro State , Brazil. *Emerg. Infect. Dis.* **9**, 55–62 (2003).
67. Ferreira, F. *et al.* The effect of water source and soil supplementation on parasite contamination in organic vegetable gardens. *Brazilian J. Vet. Parasitol.* **27**, 327–337 (2018).
68. Conrad, P. A. *et al.* Transmission of Toxoplasma : Clues from the study of sea otters as sentinels of Toxoplasma gondii flow into the marine environment. *Int. J. Parasitol.* **35**, 1155–1168 (2005).
69. Turner, D. . & Bateson, P. *The Domestic Cat, The Biology of its Behaviour.* (Cambridge University Press, 2014).
70. Vigne, J., Guilaine, J., Debue, K., Haye, L. & Gerard, P. Early Taming of the Cat in Cyprus. *Science (80-. ).* **304**, 259 (2004).
71. Dabritz, H. a. & Conrad, P. a. Cats and toxoplasma: Implications for public health. *Zoonoses Public Health* **57**, 34–52 (2010).
72. Galal, L. *et al.* Combining spatial analysis and host population genetics to gain insights into the mode of transmission of a pathogen: The example of Toxoplasma gondii in mice. *Infect. Genet. Evol.* 104142 (2019). doi:10.1016/j.meegid.2019.104142
73. Murphy, R. G. *et al.* The urban house mouse (Mus domesticus) as a reservoir of infection for the human parasite Toxoplasma gondii: An unrecognised public health issue? *Int. J. Environ. Health Res.* **18**, 177–185 (2008).
74. Smith, D. & Frenkel, J. K. Prevalence of antibodies to Toxoplasma gondii in wild mammals of Missouri and east central Kansas: biologic and ecologic considerations of transmission. *J. Wildl. Dis.* **31**, 15–21 (1995).
75. Galal, L. *et al.* Diversity of Toxoplasma gondii strains shaped by commensal communities of small mammals. *Int. J. Parasitol.* **49**, 267–275 (2019).
76. Gotteland, C. *et al.* Species or local environment, what determines the infection of rodents by Toxoplasma gondii? *Parasitology* **141**, 259–268 (2014).
77. Poulsen, A. *et al.* Prevalence and potential impact of Toxoplasma gondii on the endangered Amargosa vole (Microtus californicus scirpensis), California, USA. *J. Wildl. Dis.* **53**, 1–11 (2017).
78. Bajnok, J. *et al.* Prevalence of Toxoplasma gondii in localized populations of Apodemus sylvaticus is linked to population genotype not to population location. *Parasitology* **142**, 680–690 (2015).
79. Vilares, a. *et al.* Isolation and molecular characterization of Toxoplasma gondii isolated from pigeons and stray cats in Lisbon, Portugal. *Vet. Parasitol.* **205**, 506–511



- (2014).
80. Tenter, A. M., Heckeroth, A. R. & Weiss, L. M. *Toxoplasma gondii*: from animals to humans. *Int J Parasitol.* **30**, 1217–1258 (2000).
  81. Dubey, J. P. *Toxoplasmosis of Animals and Humans*. (CRC Press - Taylor and Francis Group, LLC, 2010).
  82. Berna, L. *et al.* Reevaluation of the *Toxoplasma gondii* and *Neospora caninum* genomes reveals misassembly, karyotype differences and chromosomal rearrangements. *bioRxiv Preprint*, (2020).
  83. Su, C. *et al.* Recent Expansion of *Toxoplasma* Through Enhanced Oral Transmission. *Science* (80-. ). **299**, 414–416 (2003).
  84. Darde, M. L., Bouteille, B. & Pestre-Alexandre, M. Isoenzymic characterization of seven strains of *Toxoplasma gondii* by isoelectrofocusing in polyacrylamide gels. *Am. J. Trop. Med. Hyg.* **39**, 551–558 (1988).
  85. Dardé, M. L., Bouteille, B. & Pestre-Alexandre, M. Isoenzyme analysis of 35 *Toxoplasma gondii* isolates and the biological and epidemiological implications. *J. Parasitol.* **78**, 786–794 (1992).
  86. Sibley, L. D. & Boothroyd, J. . Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature* **359**, 82–85 (1992).
  87. Darde, M. L., Ajzenberg, D. & Smith, J. in *Toxoplasma gondii. The Model Apicomplexan: Perspectives and Methods* (eds. Weiss, L. M. & Kim, K.) (Elsevier, 2007).
  88. Lehmann, T., Marcet, P. L., Graham, D. H., Dahl, E. R. & Dubey, J. P. Globalization and the population structure of *Toxoplasma gondii*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11423–11428 (2006).
  89. Khan, A. *et al.* Recent transcontinental sweep of *Toxoplasma gondii* driven by a single monomorphic chromosome. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14872–14877 (2007).
  90. Khan, A. *et al.* Genetic analyses of atypical *Toxoplasma gondii* strains reveal a fourth clonal lineage in North America. *Int. J. Parasitol.* **41**, 645–655 (2011).
  91. Su, C. *et al.* Globally diverse *Toxoplasma gondii* isolates comprise six major clades originating from a small number of distinct ancestral lineages. *Proc. Natl. Acad. Sci.* **109**, 5844–5849 (2012).
  92. Galal, L., Hamidović, A., Dardé, M. L. & Mercier, M. Diversity of *Toxoplasma gondii* strains at the global level and its determinants. *Food Waterborne Parasitol.* e00052 (2019). doi:10.1016/j.fawpar.2019.e00052
  93. Su, C., Zhang, X. & Dubey, J. P. Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP markers: A high resolution and simple method for identification of parasites. *Int. J. Parasitol.* **36**, 841–848 (2006).

94. Ajzenberg, D., Collinet, F., Mercier, A., Vignoles, P. & Darde, M.-L. Genotyping of *Toxoplasma gondii* Isolates with 15 Microsatellite Markers in a Single Multiplex PCR Assay. *J. Clin. Microbiol.* **48**, 4641–4645 (2010).
95. Su, C., Shwab, E. K., Zhou, P., Zhu, X. Q. & Dubey, J. P. Moving towards an integrated approach to molecular detection and identification of *Toxoplasma gondii*. *Parasitology* **137**, (2010).
96. Chaichan, P. *et al.* Geographical distribution of *Toxoplasma gondii* genotypes in Asia: A link with neighboring continents. *Infect. Genet. Evol.* **53**, 227–238 (2017).
97. Mercier, A. *et al.* Additional Haplogroups of *Toxoplasma gondii* out of Africa: Population Structure and Mouse-Virulence of Strains from Gabon. *PLoS Negl. Trop. Dis.* **4**, (2010).
98. Galal, L. *et al.* *Toxoplasma* and Africa: One Parasite, Two Opposite Population Structures. *Cell Press Rev.* **34**, (2018).
99. Bertranpetit, E. *et al.* Phylogeography of *Toxoplasma gondii* points to a South American origin. *Infect. Genet. Evol.* **48**, 150–155 (2017).
100. Ajzenberg, D. *et al.* Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. *Int. J. Parasitol.* **34**, 1185–1196 (2004).
101. Fux, B. *et al.* *Toxoplasma gondii* Strains Defective in Oral Transmission Are Also Defective in Developmental Stage Differentiation. *Infect. Immun.* **75**, 2580–2590 (2007).
102. Jensen, K. D. C. *et al.* *Toxoplasma gondii* Superinfection and Virulence during Secondary Infection Correlate with the Exact ROP5 / ROP18 Allelic Combination. *MBio* **6**, 1–15 (2015).
103. Behnke, M. S. *et al.* Rhoptry Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of *Toxoplasma gondii*. *PLOS Genet.* **11**, e1005434 (2015).
104. Gómez-Marín, J. E. *et al.* Toxoplasmosis in military personnel involved in jungle operations. *Acta Trop.* **122**, 46–51 (2012).
105. Alvarez, C. *et al.* Striking Divergence in *Toxoplasma* ROP16 Nucleotide Sequences From Human and Meat Samples. *J. Infect. Dis.* **211**, 2006–2013 (2015).
106. Gilbert, R. E. *et al.* Ocular Sequelae of Congenital Toxoplasmosis in Brazil Compared with Europe. *PLoS Negl. Trop. Dis.* **2**, 8–14 (2008).
107. De-la-Torre, A. *et al.* Severe South American Ocular Toxoplasmosis Is Associated with Decreased Ifn- $\gamma$  / IL-17a and Increased IL-6 / IL-13 Intraocular Levels. *PLoS Negl. Trop. Dis.* **7**, (2013).
108. Hakimi, M.-A., Olias, P. & Sibley, L. D. *Toxoplasma* Effectors Targeting Host Signaling and Transcription. *Clin. Microbiol. Rev.* **30**, 615–645 (2017).
109. Khan, A. *et al.* NextGen sequencing reveals short double crossovers contribute

- disproportionately to genetic diversity in *Toxoplasma gondii*. *BMC Genomics* **15**, 1168 (2014).
110. Khan, A. *et al.* A Monomorphic Haplotype of Chromosome 1a Is Associated with Widespread Success in Clonal and Nonclonal Populations of *Toxoplasma gondii*. *MBio* **2**, 1–10 (2011).
  111. Khan, A. *et al.* Global selective sweep of a highly inbred genome of the cattle parasite *Neospora caninum*. *PNAS* **116**, 22764–22773 (2019).
  112. Rastogi, S., Cygan, A. M. & Boothroyd, J. C. Translocation of effector proteins into host cells by *Toxoplasma gondii*. *Curr. Opin. Microbiol.* **52**, 130–138 (2019).
  113. Dubremetz, J. F. & Lebrun, M. Virulence factors of *Toxoplasma gondii*. *Microbes Infect.* **14**, 1403–1410 (2012).
  114. Morisaki, J. H., Heuser, J. E. & Sibley, L. D. Invasion of *Toxoplasma gondii* occurs by active penetration of the host cell. *J. Cell Sci.* **108**, 2457–2464 (1995).
  115. Boothroyd, J. C. & Dubremetz, J.-F. Kiss and spit: the dual roles of *Toxoplasma* rhoptries. *Nat. Rev. Microbiol.* **6**, 79 (2008).
  116. Besteiro, S., Dubremetz, J. F. & Lebrun, M. The moving junction of apicomplexan parasites: A key structure for invasion. *Cell. Microbiol.* **13**, 797–805 (2011).
  117. Nichols, B. A. & O'Connor, G. R. Penetration of mouse peritoneal macrophages by the protozoon *Toxoplasma gondii*. New evidence for active invasion and phagocytosis. *Lab. Investig.* **44**, 324–335 (1981).
  118. Mordue, B. D. G., Desai, N., Dustin, M. & Sibley, L. D. Invasion by *Toxoplasma gondii* Establishes a Moving Junction That Selectively Excludes Host Cell Plasma Membrane Proteins on the Basis of Their Membrane Anchoring. *J. Exp. Med.* **190**, 1783–1792 (1999).
  119. Charron, A. J. & Sibley, L. D. Molecular Partitioning during Host Cell Penetration by *Toxoplasma gondii*. *Traffic* **5**, 855–867 (2004).
  120. Dubremetz, J. F. Rhoptries are major players in *Toxoplasma gondii* invasion and host cell interaction. *Cell. Microbiol.* **9**, 841–848 (2007).
  121. Saeij, J. P. J. *et al.* *Toxoplasma* co-opts host gene expression by injection of a polymorphic kinase homologue. *Nature* **445**, 2–5 (2007).
  122. Saeij, J. P. J. *et al.* Polymorphic Secreted Kinases Are Key Virulence Factors in Toxoplasmosis. *Science* (80-. ). **161**, 1780–1784 (2006).
  123. Reese, M. L., Zeiner, G. M., Saeij, J. P. J., Boothroyd, J. C. & Boyle, J. P. Polymorphic family of injected pseudokinases is paramount in *Toxoplasma* virulence. *Pnas* **108**, 9625–9630 (2011).
  124. Behnke, M. S. *et al.* Virulence differences in *Toxoplasma* mediated by amplification of a family of polymorphic pseudokinases. *Pnas* **108**, 9631–9636 (2011).
  125. Lodoen, M. B., Gerke, C. & Boothroyd, J. C. A highly sensitive FRET-based

- approach reveals secretion of the actin-binding protein toxofilin during *Toxoplasma gondii* infection. *Cell. Microbiol.* **12**, 55–66 (2010).
126. Hermanns, T., Müller, U. B., Könen-Waisman, S., Howard, J. C. & Steinfeldt, T. The *Toxoplasma gondii* rhoptry protein ROP18 is an Irga6-specific kinase and regulated by the dense granule protein GRA7. *Cell. Microbiol.* **18**, 244–259 (2016).
  127. Gold, D. A. *et al.* The *Toxoplasma* Dense Granule Proteins GRA17 and GRA23 Mediate the Movement of Small Molecules between the Host and the Parasitophorous Vacuole. *Cell Host Microbe* **17**, 642–652 (2015).
  128. Rosowski, E. E. *et al.* Strain-specific activation of the NF- $\kappa$ B pathway by GRA15 , a novel *Toxoplasma gondii* dense granule protein. *J. Exp. Med.* **208**, 195–212 (2011).
  129. Sangaré, L. O. *et al.* *Toxoplasma* GRA15 Activates the NF- $\kappa$ B Pathway through Interactions with TNF Receptor-Associated Factors. *MBio* **10**, 1–13 (2019).
  130. Ma, J. S. *et al.* Selective and strain-specific NFAT4 activation by the *Toxoplasma gondii* polymorphic dense granule protein GRA6. *J. Exp. Med.* **211**, 2013–2032 (2014).
  131. Bougdour, A. *et al.* Article Host Cell Subversion by *Toxoplasma* GRA16 , an Exported Dense Granule Protein that Targets the Host Cell Nucleus and Alters Gene Expression. *Cell Host Microbe* **13**, 489–500 (2013).
  132. Braun, L. *et al.* A *Toxoplasma* dense granule protein , GRA24 , modulates the early immune response to infection by promoting a direct and sustained host p38 MAPK activation. *J. Exp. Med.* **210**, 2071–2086 (2013).
  133. Gay, G. *et al.* *Toxoplasma gondii* TgIST co-opts host chromatin repressors dampening STAT1-dependent gene regulation and IFN- $\gamma$  – mediated host defenses. *J. Exp. Med.* **213**, 1779–1798 (2016).
  134. Peixoto, L. *et al.* Integrative Genomic Approaches Highlight a Family of Parasite-Specific Kinases that Regulate Host Responses. *Cell Host Microbe* **8**, 208–218 (2010).
  135. English, E. D., Adomako-Ankomah, Y. & Boyle, J. P. Secreted effectors in *Toxoplasma gondii* and related species: determinants of host range and pathogenesis? *Parasite Immunol.* **37**, 127–140 (2015).
  136. Sidik, S. M. *et al.* A Genome-Wide CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes. *Cell* **166**, 1423–1435 (2016).
  137. Sangaré, L. O. *et al.* In Vivo CRISPR Screen Identifies TgWIP as a *Toxoplasma* Modulator of Dendritic Cell Migration. *Cell Host Microbe* **26**, 478–492.e8 (2019).
  138. Taylor, S. *et al.* A Secreted Serine-Threonine Kinase Determines Virulence in the Eukaryotic Pathogen *Toxoplasma gondii*. *Science (80-. )*. **314**, 1776–1781 (2006).
  139. Khan, A., Taylor, S., Ajioka, J. W., Rosenthal, B. M. & Sibley, L. D. Selection at a Single Locus Leads to Widespread Expansion of *Toxoplasma gondii* Lineages That

- Are Virulent in Mice. *PLoS Genet.* **5**, 1–14 (2009).
140. Hajj, H. El *et al.* The ROP2 family of *Toxoplasma gondii* rhoptry proteins : Proteomic and genomic characterization and molecular modeling. *Proteomics* **6**, 5773–5784 (2006).
  141. Niedelman, W. *et al.* The Rhoptry Proteins ROP18 and ROP5 Mediate *Toxoplasma gondii* Evasion of the Murine , But Not the Human, Interferon-Gamma Response. *PLoS Pathog.* **8**, e1002784 (2012).
  142. Reese, M. L., Shah, N. & Boothroyd, J. C. The *Toxoplasma* Pseudokinase ROP5 Is an Allosteric Inhibitor of the Immunity-related GTPases. *J. Biol. Chem.* **289**, 27849–27858 (2014).
  143. Ong, Y., Reese, M. L. & Boothroyd, J. C. *Toxoplasma* Rhoptry Protein 16 ( ROP16 ) Subverts Host Function by Direct Tyrosine Phosphorylation of STAT6. *J. Biol. Chem.* **285**, 28731–28740 (2010).
  144. Butcher, B. A. *et al.* *Toxoplasma gondii* Rhoptry Kinase ROP16 Activates STAT3 and STAT6 Resulting in Cytokine Inhibition and Arginase-1-Dependent Growth Control. *PLoS Pathog.* **7**, e1002236 (2011).
  145. Yamamoto, M. *et al.* A single polymorphic amino acid on *Toxoplasma gondii* kinase ROP16 determines the direct and strain-specific activation of Stat3. *J. Exp. Med.* **206**, 2747–2760 (2009).
  146. Dubey, J. P. Comparative Infectivity of *Toxoplasma gondii* Bradyzoites in Rats and Mice. *J. Parasitol.* **84**, 1279–1282 (1998).
  147. Takeuchi, O. & Akira, S. Pattern Recognition Receptors and Inflammation. *Cell* **140**, 805–820 (2010).
  148. Melo, M. B., Jensen, K. D. C. & Saeij, J. P. J. *Toxoplasma gondii* effectors are master regulators of the inflammatory response. *Trends Parasitol.* **27**, 487–495 (2011).
  149. Aliberti, J. *et al.* Molecular mimicry of a CCR5 binding-domain in the microbial activation of dendritic cells. *Nat. Immunol.* **4**, 485–490 (2003).
  150. Gazzinelli, R. T., Mendonça-Neto, R., Lilue, J., Howard, J. & Sher, A. Innate resistance against *Toxoplasma gondii*: An evolutionary tale of mice, cats, and men. *Cell Host Microbe* **15**, 132–138 (2014).
  151. Scanga, C. A. *et al.* Cutting Edge: MyD88 Is Required for Resistance to *Toxoplasma gondii* Infection and Regulates Parasite-Induced IL-12 Production by Dendritic Cells. *J. Immunol.* **168**, 5997–6001 (2002).
  152. Béla, S. R. *et al.* Impaired innate immunity in mice deficient in interleukin-1 receptor-associated kinase 4 leads to defective type 1 T cell responses, B cell expansion, and enhanced susceptibility to infection with *Toxoplasma gondii*. *Infect. Immun.* **80**, 4298–4308 (2012).

153. Robben, P. M. *et al.* Production of IL-12 by Macrophages Infected with *Toxoplasma gondii* Depends on the Parasite Genotype. *J. Immunol.* **172**, 3686–3694 (2004).
154. Bliss, S. K., Butcher, B. A. & Denkers, E. Y. Rapid Recruitment of Neutrophils Containing Prestored IL-12 During Microbial Infection. *J. Immunol.* **165**, 4515–4521 (2000).
155. Sturge, C. R. *et al.* TLR-independent neutrophil-derived IFN-  $\gamma$  is important for host resistance to intracellular pathogens. *Pnas* **110**, 10711–10716 (2013).
156. Bierly, A. L. *et al.* Dendritic Cells Expressing Plasmacytoid Marker PDCA-1 Are Trojan Horses during *Toxoplasma gondii* Infection. *J. Immunol.* **181**, 8485–8491 (2008).
157. Pepper, M. *et al.* Plasmacytoid Dendritic Cells Are Activated by *Toxoplasma gondii* to Present Antigen and Produce Cytokines. *J. Immunol.* **180**, 6229–6236 (2008).
158. Sousa, C. R. e *et al.* In Vivo Microbial Stimulation Induces Rapid CD40 Ligand-independent Production of Interleukin 12 by Dendritic Cells and their Redistribution to T Cell Areas. *J. Exp. Med.* **186**, 1819–1829 (1997).
159. Mashayekhi, M. *et al.* CD8 a+ Dendritic Cells Are the Critical Source of Interleukin-12 that Controls Acute Infection by *Toxoplasma gondii* Tachyzoites. *Immunity* **35**, 249–259 (2011).
160. Yarovinsky, F. Innate immunity to *Toxoplasma gondii* infection. *Nat. Rev. Immunol.* **14**, 109–21 (2014).
161. Andrade, W. A. *et al.* Combined action of nucleic acid-sensing Toll-like receptors and TLR11/TLR12 heterodimers imparts resistance to *Toxoplasma gondii* in mice. *Cell Host Microbe* **13**, 42–53 (2013).
162. Yarovinsky, F. *et al.* TLR11 activation of dendritic cells by a protozoan profilin-like protein. *Science* **308**, 1626–1629 (2005).
163. Koblansky, A. A. *et al.* Recognition of profilin by Toll-like receptor 12 is critical for host resistance to *Toxoplasma gondii*. *Immunity* **38**, 119–130 (2013).
164. Roach, J. C. *et al.* The evolution of vertebrate Toll-like receptors. *Pnas* **102**, 9577–9582 (2005).
165. Morger, J. *et al.* Naturally occurring Toll-like receptor 11 (TLR11) and Toll-like receptor 12 (TLR12) polymorphisms are not associated with *Toxoplasma gondii* infection in wild wood mice. *Infect. Genet. Evol.* **26**, 180–184 (2014).
166. Gazzinelli, R. T., Hieny, S., Wynn, T. A., Wolf, S. & Sher, A. Interleukin 12 is required for the T-lymphocyte-independent induction of interferon gamma by an intracellular parasite and induces resistance in T-cell-deficient hosts. *Proc. Natl. Acad. Sci.* **90**, 6115–6119 (1993).
167. Hunter, C. A., Subauste, C. S., Van Cleave, V. H. & Remington, J. S. Production of gamma interferon by natural killer cells from *Toxoplasma gondii*-infected SCID mice:

- regulation by interleukin-10, interleukin-12, and tumor necrosis factor alpha. *Infect. Immun.* **62**, 2818–2824 (1994).
168. Wilson, D. C., Matthews, S. & Yap, G. S. IL-12 Signaling Drives CD8<sup>+</sup> T Cell IFN- $\gamma$  Production and Differentiation of KLRG1<sup>+</sup> Effector Subpopulations during *Toxoplasma gondii* Infection. *J. Immunol.* **180**, 5935–5945 (2008).
  169. Gazzinelli, R. T. *et al.* Parasite-induced IL-12 stimulates early IFN-gamma synthesis and resistance during acute infection with *Toxoplasma gondii*. *J. Immunol.* **153**, 2533–2543 (1994).
  170. Schindler, C. & Darnell, J. E. Transcriptional responses to polypeptide ligands: The JAK-STAT Pathway. *Annu. Rev. Biochem.* **64**, 621–652 (1995).
  171. Darnell, J. E., Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* (80-. ). **264**, 1415–1421 (1994).
  172. Luder, C. G. K., Algner, M., Lang, C., Bleicher, N. & Grob, U. Reduced expression of the inducible nitric oxide synthase after infection with *Toxoplasma gondii* facilitates parasite replication in activated murine macrophages. *Int. J. Parasitol.* **33**, 833–844 (2003).
  173. Stutz, A., Golenbock, D. T. & Latz, E. Inflammasomes: too big to miss. *J. Clin. Invest.* **119**, 3502–3511 (2009).
  174. Schroder, K. & Tschopp, J. The Inflammasomes. *Cell* **140**, 821–832 (2010).
  175. Sergent, V. *et al.* Innate refractoriness of the Lewis rat to toxoplasmosis is a dominant trait that is intrinsic to bone marrow-derived cells. *Infect. Immun.* **73**, 6990–6997 (2005).
  176. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annu Rev Immunol* **32**, 513–545 (2014).
  177. Rusinova, I. *et al.* INTERFEROME v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res.* **41**, D1040–D1046 (2012).
  178. Kim, B., Shenoy, A. R., Kumar, P., Bradfield, C. J. & Macmicking, J. D. IFN-Inducible GTPases in Host Cell Defense. *Cell Host Microbe* **12**, 432–444 (2012).
  179. Horisberger, M. A., Staeheli, P. & Haller, O. Interferon induces a unique protein in mouse cells bearing a gene for resistance to influenza virus. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1910–1914 (1983).
  180. Cheng, Y. S., Colonno, R. J. & Yin, F. H. Interferon induction of fibroblast proteins with guanylate binding activity. *J. Biol. Chem.* **258**, 7746–7750 (1983).
  181. Boehm, U. *et al.* Two families of GTPases dominate the complex cellular response to IFN-gamma. *J. Immunol.* **161**, 6715–6723 (1998).
  182. Carlow, D. A., Marth, J., Clark-Lewis, I. & Teh, H. S. Isolation of a gene encoding a developmentally regulated T cell-specific protein with a guanine nucleotide

- triphosphate-binding motif. *J. Immunol.* **154**, 1724–1734 (1995).
183. Gilly, M. & Wall, R. The IRG-47 gene is IFN-gamma induced in B cells and encodes a protein with GTP-binding motifs. *J. Immunol.* **148**, 3275–3281 (1992).
  184. Lafuse, W. P., Brown, D., Castle, L. & Zwilling, B. S. Cloning and characterization of a novel cDNA that is IFN- $\gamma$ -induced in mouse peritoneal macrophages and encodes a putative GTP-binding protein. *J. Leukoc. Biol.* **57**, 477–483 (1995).
  185. Sorace, J. M., Johnson, R. J., Howard, D. L. & Drysdale, B. E. Identification of an endotoxin and IFN-inducible cDNA: possible identification of a novel protein family. *J. Leukoc. Biol.* **58**, 477–484 (1995).
  186. Taylor, G. A. *et al.* Identification of a novel GTPase, the inducibly expressed GTPase, that accumulates in response to interferon gamma. *J. Biol. Chem.* **271**, 20399–20405 (1996).
  187. Hunn, J. P., Feng, C. G., Sher, A. & Howard, J. C. The immunity-related GTPases in mammals: A fast-evolving cell-autonomous resistance system against intracellular pathogens. *Mamm. Genome* **22**, 43–54 (2011).
  188. Li, G., Zhang, J., Sun, Y., Wang, H. & Wang, Y. The Evolutionarily Dynamic IFN-Inducible GTPase Proteins Play Conserved Immune Functions in Vertebrates and Cephalochordates. *Mol. Biol. Evol.* **26**, 1619–1630 (2009).
  189. Bekpen, C. *et al.* The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. *Genome Biol.* **6**, R92 (2005).
  190. Olszewski, M. A., Gray, J. & Vestal, D. J. In Silico Genomic Analysis of the Human and Murine Guanylate-Binding Protein (GBP) Gene Clusters. *J. Interf. Cytokine Res.* **26**, 328–352 (2006).
  191. Shenoy, A. R. *et al.* Emerging themes in IFN-gamma-induced macrophage immunity by the p47 and p65 GTPase families. *Immunobiology* **212**, 771–784 (2007).
  192. Kresse, A. *et al.* Analyses of murine GBP homology clusters based on in silico, in vitro and in vivo studies. *BMC Genomics* **9**, 158 (2008).
  193. Hunn, J. P. *et al.* Regulatory interactions between IRG resistance GTPases in the cellular response to *Toxoplasma gondii*. *EMBO J.* **27**, 2495–2509 (2008).
  194. Howard, J. C., Hunn, J. P. & Steinfeldt, T. The IRG protein-based resistance mechanism in mice and its relation to virulence in *Toxoplasma gondii*. *Curr. Opin. Microbiol.* **14**, 414–421 (2011).
  195. Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat. Genet.* **39**, 830–832 (2007).
  196. Degrandi, D. *et al.* Extensive Characterization of IFN-Induced GTPases mGBP1 to mGBP10 Involved in Host Defense. *J. Immunol.* **179**, 7729–7740 (2007).



197. Praefcke, G. J. K. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat. Rev. Mol. Cell Biol.* **5**, 133–147 (2004).
198. Uthaiiah, R. C., Praefcke, G. J. K., Howard, J. C. & Herrmann, C. IIGP1, an interferon-gamma-inducible 47-kDa GTPase of the mouse, showing cooperative enzymatic activity and GTP-dependent multimerization. *J. Biol. Chem.* **278**, 29336–29343 (2003).
199. Papic, N., Hunn, J. P., Pawlowski, N., Zerrahn, J. & Howard, J. C. Inactive and active states of the interferon-inducible resistance GTPase, Irga6, in vivo. *J. Biol. Chem.* **283**, 32143–51 (2008).
200. Pawlowski, N. *et al.* The activation mechanism of Irga6, an interferon-inducible GTPase contributing to mouse resistance against *Toxoplasma gondii*. *BMC Biol.* **9**, 7 (2011).
201. Uthaiiah, R. C., Praefcke, G. J. K., Howard, J. C. & Herrmann, C. IIGP1, an interferon-g-inducible 47-kDa GTPase of the mouse, showing cooperative enzymatic activity and GTP-dependent multimerization. *J. Biol. Chem.* **278**, 29336–29343 (2003).
202. Lilue, J. Haplotypic polymorphism of the IRG protein family mediates resistance of mice against virulent strains of *Toxoplasma gondii*. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2012).
203. Ghosh, A., Uthaiiah, R., Howard, J., Herrmann, C. & Wolf, E. Crystal structure of IIGP1: A paradigm for interferon-inducible p47 resistance GTPases. *Mol. Cell* **15**, 727–739 (2004).
204. Pitossi, F. *et al.* A functional GTP-binding motif is necessary for antiviral activity of Mx proteins. *J. Virol.* **67**, 6726–6732 (1993).
205. Sigal, I. S. *et al.* Mutant ras-encoded proteins with altered nucleotide binding exert dominant biological effects. *Proc. Natl. Acad. Sci.* **83**, 952–956 (1986).
206. Martens, S. *et al.* Mechanisms regulating the positioning of mouse p47 resistance GTPases LRG-47 and IIGP1 on cellular membranes: retargeting to plasma membrane induced by phagocytosis. *J. Immunol.* **173**, 2594–2606 (2004).
207. Udenwobe, D. I. *et al.* Myristoylation: An Important Protein Modification in the Immune Response. *Front. Immunol.* **8**, 751 (2017).
208. Zerrahn, J., Schaible, U. E., Brinkmann, V., Gühlich, U. & Kaufmann, S. H. E. The IFN-inducible Golgi- and endoplasmic reticulum-associated 47-kDa GTPase IIGP is transiently expressed during listeriosis. *J. Immunol.* **168**, 3428–3436 (2002).
209. Tiwari, S., Choi, H.-P., Matsuzawa, T., Pypaert, M. & MacMicking, J. D. Targeting of the GTPase Irgm1 to the phagosomal membrane via PtdIns(3,4)P(2) and PtdIns(3,4,5)P(3) promotes immunity to mycobacteria. *Nat. Immunol.* **10**, 907–917 (2009).

210. Zhao, Y. *et al.* Virulent *Toxoplasma gondii* Evade Immunity-Related GTPase (IRG)-Mediated Parasite Vacuole Disruption Within Primed Macrophages. *J Immunol* **182**, 3775–3781 (2010).
211. Haldar, A. K. *et al.* IRG and GBP Host Resistance Factors Target Aberrant, “Non-self” Vacuoles Characterized by the Missing of “Self” IRGM Proteins. *PLoS Pathog.* **9**, (2013).
212. Springer, H. M., Schramm, M., Taylor, G. a & Howard, J. C. Irgm1 (LRG-47), a regulator of cell-autonomous immunity, does not localize to mycobacterial or listerial phagosomes in IFN- $\gamma$ -induced mouse cells. *J. Immunol.* **191**, 1765–74 (2013).
213. Collazo, C. M. *et al.* Inactivation of LRG-47 and IRG-47 reveals a family of interferon gamma-inducible genes with essential, pathogen-specific roles in resistance to infection. *J. Exp. Med.* **194**, 181–188 (2001).
214. Feng, C. G. *et al.* Mice deficient in LRG-47 display increased susceptibility to mycobacterial infection associated with the induction of lymphopenia. *J. Immunol.* **172**, 1163–1168 (2004).
215. Henry, S. C. *et al.* Impaired Macrophage Function Underscores Susceptibility to Salmonella in Mice Lacking Irgm1 (LRG-47). *J. Immunol.* **179**, 6963–6972 (2007).
216. Henry, S. C. *et al.* Balance of Irgm protein activities determines IFN-gamma-induced host defense. *J. Leukoc. Biol.* **85**, 877–885 (2009).
217. MacMicking, J. D., Taylor, G. A. & McKinney, J. D. Immune control of tuberculosis by IFN-gamma-inducible LRG-47. *Science* **302**, 654–659 (2003).
218. Santiago, H. C. *et al.* Mice Deficient in LRG-47 Display Enhanced Susceptibility to Trypanosoma cruzi Infection Associated with Defective Hemopoiesis and Intracellular Control of Parasite Growth. *J. Immunol.* **175**, 8165–8172 (2005).
219. Taylor, G. a., Feng, C. G. & Sher, A. Control of IFN- $\gamma$ -mediated host resistance to intracellular pathogens by immunity-related GTPases (p47 GTPases). *Microbes Infect.* **9**, 1644–1651 (2007).
220. Gutierrez, M. G. *et al.* Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell* **119**, 753–766 (2004).
221. Hunn, J. P. & Howard, J. C. The Mouse Resistance Protein Irgm1 (LRG-47): A Regulator or an Effector of Pathogen Defense? *PLOS Pathog.* **6**, 1–5 (2010).
222. Taylor, G. a *et al.* Pathogen-specific loss of host resistance in mice lacking the IFN-gamma-inducible gene IGTP. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 751–755 (2000).
223. Man, S. M. *et al.* IRGB10 Liberates Bacterial Ligands for Sensing by the AIM2 and Caspase-11-NLRP3 Inflammasomes. *Cell* **167**, 382–396.e17 (2016).
224. Coers, J. *et al.* Chlamydia muridarum Evades Growth Restriction by the IFN- $\gamma$ -Inducible Host Resistance Factor Irgb10. *J. Immunol.* **180**, 6237–6245 (2008).

225. Miyairi, I. *et al.* The p47 GTPases ligp2 and Irgb10 Regulate Innate Immunity and Inflammation to Murine Chlamydia psittaci Infection. *J. Immunol.* **179**, 1814–1824 (2007).
226. Lee, Y. *et al.* Initial phospholipid-dependent Irgb6 targeting to Toxoplasma gondii vacuoles mediates host defense. *Life Sci. Alliance* **3**, 1–16 (2019).
227. Nelson, D. E. *et al.* Chlamydial IFN-gamma immune evasion is linked to host infection tropism. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10658–10663 (2005).
228. da Fonseca Ferreira-da-Silva, M., Springer-Frauenhoff, H. M., Bohne, W. & Howard, J. C. Identification of the Microsporidian Encephalitozoon cuniculi as a New Target of the IFN $\gamma$ -Inducible IRG Resistance System. *PLoS Pathog.* **10**, e1004449 (2014).
229. Correia, A. *et al.* Predominant role of interferon- $\gamma$  in the host protective effect of CD8<sup>+</sup> T cells against Neospora caninum infection. *Sci. Rep.* **5**, 14913 (2015).
230. Liesenfeld, O. *et al.* The IFN- $\gamma$ -inducible GTPase, Irga6, protects mice against Toxoplasma gondii but not against Plasmodium berghei and some other intracellular pathogens. *PLoS One* **6**, e20568–e20568 (2011).
231. Martens, S. *et al.* Disruption of Toxoplasma gondii parasitophorous vacuoles by the mouse p47-resistance GTPases. *PLoS Pathog.* **1**, 0187–0201 (2005).
232. Khaminets, A. *et al.* Coordinated loading of IRG resistance GTPases on to the Toxoplasma gondii parasitophorous vacuole. *Cell. Microbiol.* **12**, 939–961 (2010).
233. Ling, Y. M. *et al.* Vacuolar and plasma membrane stripping and autophagic elimination of Toxoplasma gondii in primed effector macrophages. *J. Exp. Med.* **203**, 2063–2071 (2006).
234. Melzer, T., Duffy, A., Weiss, L. M. & Halonen, S. K. The gamma interferon (IFN-gamma)-inducible GTP-binding protein IGTP is necessary for toxoplasma vacuolar disruption and induces parasite egression in IFN-gamma-stimulated astrocytes. *Infect. Immun.* **76**, 4883–4894 (2008).
235. Zhao, Y. O. *et al.* Toxoplasma gondii and the immunity-related GTPase (IRG) resistance system in mice - A review. *Mem. Inst. Oswaldo Cruz* **104**, 234–240 (2009).
236. Zhao, Y. O., Khaminets, A., Hunn, J. P. & Howard, J. C. Disruption of the Toxoplasma gondii parasitophorous vacuole by IFN $\gamma$ -inducible immunity-related GTPases (IRG proteins) triggers necrotic cell death. *PLoS Pathog.* **5**, (2009).
237. Zhao, Z. *et al.* Autophagosome-Independent Essential Function for the Autophagy Protein Atg5 in Cellular Immunity to Intracellular Pathogens. *Cell Host Microbe* **4**, 458–469 (2008).
238. Reese, M. L. & Boothroyd, J. C. A helical membrane-binding domain targets the Toxoplasma ROP2 family to the parasitophorous vacuole. *Traffic* **10**, 1458–1470 (2009).

239. Fentress, S. J. *et al.* Phosphorylation of immunity-related GTPases by a *Toxoplasma gondii* secreted kinase promotes macrophage survival and virulence. *Cell Host Microbe* **8**, 484–495 (2010).
240. Steinfeldt, T. *et al.* Phosphorylation of mouse immunity-related gtpase (IRG) resistance proteins is an evasion strategy for virulent *Toxoplasma gondii*. *PLoS Biol.* **8**, (2010).
241. Boyle, J. P., Saeij, J. P. J., Harada, S. Y., Ajioka, J. W. & Boothroyd, J. C. Expression quantitative trait locus mapping of *toxoplasma* genes reveals multiple mechanisms for strain-specific differences in gene expression. *Eukaryot. Cell* **7**, 1403–1414 (2008).
242. Fleckenstein, M. C. *et al.* A *toxoplasma gondii* pseudokinase inhibits host irg resistance proteins. *PLoS Biol.* **10**, 14 (2012).
243. Behnke, M. S. *et al.* The polymorphic pseudokinase ROP5 controls virulence in *Toxoplasma gondii* by regulating the active kinase ROP18. *PLoS Pathog.* **8**, e1002992 (2012).
244. Alaganan, A., Fentress, S. J., Tang, K., Wang, Q. & Sibley, L. D. *Toxoplasma* GRA7 effector increases turnover of immunity-related GTPases and contributes to acute virulence in the mouse. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1126–1131 (2014).
245. Etheridge, R. D. *et al.* The *Toxoplasma* pseudokinase ROP5 forms complexes with ROP18 and ROP17 kinases that synergize to control acute virulence in mice. *Cell Host Microbe* **15**, 537–550 (2014).
246. Hunn, J. P., Feng, C. G. & Howard, J. C. The immunity-related GTPases in mammals: a fast-evolving cell- autonomous resistance system against intracellular pathogens. *Mamm Genome* **22**, 43–54 (2011).
247. Zhao, Y. & Yap, G. S. *Toxoplasma*'s Arms Race with the Host Interferon Response: A Ménage à Trois of ROPs. *Cell Host Microbe* **15**, 517–518 (2014).
248. Hassan, M. A., Olijnik, A.-A., Frickel, E.-M. & Saeij, J. P. Clonal and atypical *Toxoplasma* strain differences in virulence vary with mouse sub-species. *Int. J. Parasitol.* **49**, 63–70 (2019).
249. Müller, U. B. Evolutionary and functional studies of the Immunity-related GTPases in mice. University of Cologne. Department of Genetics. Cologne, Germany. *Diploma thesis* (2009).
250. Murillo-León, M. *et al.* Molecular mechanism for the control of virulent *Toxoplasma gondii* infections in wild-derived mice. *Nat. Commun.* **10**, 1–15 (2019).
251. Cucchi, T. *et al.* Tracking the Near Eastern origins and European dispersal of the western house mouse. *Sci. Rep.* **10**, 1–12 (2020).
252. Darvish, J., Orth, A. & Bonhomme, F. Genetic transition in the house mouse , *Mus musculus* of Eastern Iranian Plateau. *Folia Zool.* **55**, 349–357 (2006).

253. Hamid, H. S., Darvish, J., Rastegar-pouyani, E. & Mahmoudi, A. Subspecies differentiation of the house mouse *Mus musculus* Linnaeus , 1758 in the center and east of the Iranian plateau and Afghanistan. *Mammalia* **81**, 147–168 (2017).
254. Hardouin, E. A. *et al.* Eurasian house mouse (*Mus musculus* L.) differentiation at microsatellite loci identifies the Iranian plateau as a phylogeographic hotspot. *BMC Evol. Biol.* **15**, 26 (2015).
255. Boursot, P., Auffray, J. C., Britton-Davidian, J. & Bonhomme, F. The Evolution of House Mice. *Annu. Rev. Ecol. Syst.* **24**, 119–152 (1993).
256. Chevret, P., Veyrunes, F. & Britton-Davidian, J. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biol. J. Linn. Soc.* **84**, 417–427 (2005).
257. Tucker, P. K., Sandstedt, S. A. & Lundrigan, B. L. Phylogenetic relationships in the subgenus *Mus* (genus *Mus*, family Muridae, subfamily Murinae): examining gene trees and species trees. *Biol. J. Linn. Soc.* **84**, 653–662 (2005).
258. Phifer-Rixey, M. & Nachman, M. W. Insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife* **4**, 1–13 (2015).
259. Geraldès, A. *et al.* Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol. Ecol.* **17**, 5349–5363 (2008).
260. Geraldès, A., Basset, P., Smith, K. L. & Nachman, M. W. Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol. Ecol.* **20**, 4722–4736 (2011).
261. Duvaux, L., Belkhir, K., Boulesteix, M. & Boursot, P. Isolation and gene flow: inferring the speciation history of European house mice. *Mol. Ecol.* **20**, 5248–5264 (2011).
262. White, M. A., Ané, C., Dewey, C. N., Larget, B. R. & Payseur, B. A. Fine-Scale Phylogenetic Discordance across the House Mouse Genome. *PLOS Genet.* **5**, 1–12 (2009).
263. Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
264. Din, W. *et al.* Origin and radiation of the house mouse: clues from nuclear genes. *J. Evol. Biol.* **539**, 519–539 (1996).
265. Didion, J. P. & De Villena, F. P. M. Deconstructing *Mus gemischus*: Advances in understanding ancestry, structure, and variation in the genome of the laboratory mouse. *Mamm. Genome* **24**, 1–20 (2013).
266. Guénet, J. L. & Bonhomme, F. Wild mice: An ever-increasing contribution to a popular mammalian model. *Trends Genet.* **19**, 24–31 (2003).
267. Cucchi, T., Vigne, J. & Auffray, J. First occurrence of the house mouse ( *Mus musculus domesticus* Schwarz & Schwarz , 1943 ) in the Western Mediterranean : a zooarchaeological revision of subfossil occurrences. *Biol. J. Linn. Soc.* **84**, 429–445

- (2005).
268. Rajabi-Maham, H., Orth, A. & Bonhomme, F. Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. *Mol. Ecol.* **17**, 627–641 (2008).
  269. Bonhomme, F. & Searle, J. B. *House mouse phylogeography*. (Cambridge University Press, 2012).
  270. Jones, E. P. *et al.* Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evol. Biol.* **12**, 35 (2012).
  271. Pocock, M. J. O., Hauffe, H. C. & Searle, J. B. Dispersal in house mice. *Biol. J. Linn. Soc.* **84**, 565–583 (2005).
  272. Cucchi, T. Uluburun shipwreck stowaway house mouse: molar shape analysis and indirect clues about the vessel's last journey. *J. Archaeol. Sci.* **35**, 2953–2959 (2008).
  273. Gabriel, S. I. Colonisation history of a ubiquitous invasive, the house mouse (*Mus musculus domesticus*): a global, continental and insular perspective. (University of Lisbon, 2012).
  274. Greene, J. & Morgan, P. *Atlantic History: A Critical Appraisal*. (Oxford University Press, 2008).
  275. Russell-Wood, A. *The Portuguese Empire, 1415-1808: a World on the Move*. (JHUP in association with The Calouste Gulbenkian Foundation, 1992).
  276. Gardner, M., Kozak, C. & O'Brien, S. The Lake Casitas wild mouse: evolving genetic resistance to retroviral disease. *Trends Genet.* **7**, 22–27 (1991).
  277. Stephens, H. M. *Portugal - A História de uma Nação*. (Alma dos Livros, 2017).
  278. Keeler, C. E. *The Laboratory Mouse: its Origin, Heredity and Culture*. *Nature* **128**, 431 (Harvard University Press, 1931).
  279. Nishioka, Y. The origin of common laboratory mice. *Genome* **38**, 1–7 (1995).
  280. Yang, H. *et al.* Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* **43**, 648–655 (2011).
  281. Ideraabdullah, F. Y. *et al.* Genetic and haplotype diversity among wild-derived mouse inbred strains. *Genome Res.* **14**, 1880–1887 (2004).

## Chapter 2

---

### **2. Diversity of the IRG system in South American and European *Mus musculus***

Catalina Alvarez<sup>1</sup>, Urs Benedikt Müller<sup>2</sup>, Jingtao Lilue<sup>1</sup>, Luis Teixeira<sup>1</sup> and Jonathan Howard<sup>1, 2</sup>

<sup>1</sup> Fundação Calouste Gulbenkian, Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal. <sup>2</sup> Institute for Genetics, University of Cologne, 50674 Cologne, Germany.

## **Author contributions**

CA - Study concept and design; Mouse sampling; Cell culture; RNA extraction; PCR amplification; Phylogenetic analysis, Transcriptome analysis; Gene expression analysis; Statistical analysis; Drafting and editing the manuscript.

UBM - Mouse sampling; Phylogenetic analysis.

JL - Transcriptome analysis.

LT - Study concept and design; Supervision of execution.

JCH - Study concept and design; Supervision of execution; Drafting and editing the manuscript.



## 2.1. Abstract

The Immunity-related GTPases (IRGs) are IFN $\gamma$ -inducible genes essential for resistance against intracellular pathogens in the house mouse. In some wild-derived house mice, such as the Eurasian mice CIM strain, polymorphic IRG genes have been linked to the control of virulent strains of the intracellular parasite *Toxoplasma gondii*. The house mouse (*Mus musculus domesticus*) arrived in the Americas on the first European vessels in the 16th century, but they had never encountered infections with the highly virulent *T. gondii* strains present in South America. Here we increased the knowledge on the IRG genetic diversity in wild-caught and wild-derived Brazilian and Portuguese mice. Also, we reported the high prevalence of the *Irgb2*-b1<sub>PKW</sub> allele in the Brazilian mouse population. Nonetheless, we found that this high homogeneity of the *Irgb2*-b1 locus in the Brazilian mice is most likely associated with a soft sweep as a result of local adaptation from standing genetic variation to highly virulent *T. gondii* strains. Finally, analyses on the IRGB2<sub>PKW</sub> protein subunit revealed distinct polymorphisms in the H4 and  $\alpha$ D structural domains in comparison to the IRGB2BL6 and IRGB2<sub>CIM</sub>. In conclusion, our data showed a possible selection process in Brazilian mice on standing variation from European populations, possibly associated with fitness advantage (resistance) in Brazilian mice against highly virulent *T. gondii* strains

## 2.2. Introduction

Host-parasite interactions drive co-evolutionary processes in nature. However, these processes rely on dynamic equilibria created by the presence in a population of multiple host resistance alleles confronting multiple parasite virulence alleles<sup>1</sup>. These dynamic systems may be permissive for the persistence of both the host and the parasite populations but confrontations of incompatible alleles affect the equilibrium at the expense of individual interactions that could be fatal for either the parasite or the host<sup>2</sup>.

For *Toxoplasma gondii*, coevolution with its intermediate hosts demands enough virulence to survive the host's immune attack and encyst, but not so much that the host dies of the infection<sup>3</sup>. Immunity-related GTPases (IRGs) are IFN $\gamma$ -inducible genes of an essential cell-autonomous resistance system in the domestic mouse, *Mus musculus*, which is complicit in establishing this delicate balance<sup>4</sup>. The mouse IRG system has an extraordinary genetic complexity<sup>2,5-7</sup>, where these proteins appear to be as diverse in sequence as classical MHC proteins system, with extensive copy number variation and frequent gene conversion events<sup>2,8</sup>.

IRG proteins fall into five subfamilies, IRGA, IRGB, IRGC, IRGD, and IRGM. The three members of the IRGM (also called GMS) subfamily, Irgm1, Irgm2, and Irgm3, are negative regulators of the IRGA, IRGB, and IRGD effector (GKS) subfamilies that directly cause damage to the parasitophorous vacuole membrane of *T. gondii*. There are about 23 genes in the C57BL/6 genome in two adjacent clusters in chromosome 11 and a third one in chromosome 18<sup>2,5,7</sup>. Four of these genes are transcribed as adjacent pairs resulting in expression of proteins carrying two IRG domains, the so-called tandem IRG proteins (Irgb2-b1, Irgb3-b5, Irgb4-b5 and Irgb9-b8)<sup>2</sup>.

Wild mice and wild-derived inbred mice display a remarkable diversity in their IRG proteins in comparison to laboratory mice<sup>2</sup>. The wild-derived CIM mouse strain was the first *Mus musculus* reported to be fully resistant against infection with highly virulent type I *T. gondii* strains<sup>2</sup>. Recently, our lab has also shown that wild *M. m. castaneus* caught in South India display a similar repertoire of IRG proteins to the ones described for the CIM wild-derived strain<sup>6</sup>. This finding favors the hypothesis that wild mice might face infections with highly virulent strains that impose selection pressure favoring certain IRG alleles. Additionally to *M. m. castaneus*, previous studies<sup>9</sup> and unpublished data from our lab have shown that wild-derived *M. m. musculus* (PWK and PWD strains) can also resist infection against highly virulent type I *T. gondii* strains. However, there is still very little information about IRG proteins in wild *M. m. domesticus*.

The *M. m. domesticus* subspecies is the most prevalent house mouse in Europe, America, and Africa. This subspecies displays a highly genetically structured population that has experienced large events of expansion around the world in the past few thousand generations, but also possible bottlenecks and local extinctions, probably followed by anthropogenic mediated recolonizations<sup>10,11</sup>. Portugal is the westernmost country of continental Europe and represents an important geographic and historical location for the current distribution of *M. m. domesticus*<sup>6,12</sup>. Studies in fossil records found that this subspecies arrived in the Western Mediterranean zone approximately 3000 years ago<sup>12</sup>. A second wave of colonization first occurred in Italy and in the first millennium BC mice arrived at the Iberian Peninsula, specifically to Northeastern Spain and Portugal<sup>13</sup>.

*M. m. domesticus* arrived in South America in the 16<sup>th</sup> century probably in the ships of the first European settlers<sup>14,15,16</sup>. The area that corresponds nowadays to Brazil was conquered by Portuguese and maintained under their control until the 19<sup>th</sup> century. *M. m. domesticus* have arrived in the

Brazilian territory from those first Portuguese boats. Therefore, Portuguese mice constitute an interesting control mouse population because they are probably the direct ancestors of many of the *M. m. domesticus* found nowadays in Brazil.

The enormous genetic diversity of *T. gondii* found in South America suggests it may be the “birthplace” of the parasite<sup>17</sup>. The interaction between South American strains of *T. gondii* and *M. musculus* in the wild is only 500 years old and virtually all South American *T. gondii* strains tested are virulent for laboratory mice<sup>18,19</sup>. However, specific host-pathogen interactions between South American *M. musculus* and *T. gondii* strains are unknown.

In this study, we decided to increase the current knowledge of IRG protein diversity in wild *M. m. domesticus* populations in both Portugal and South America. We aimed to know whether the alleles of IRG genes in South American *M. musculus* reflect co-adaptation to local *T. gondii* strains.

## **2.3. Materials and methods**

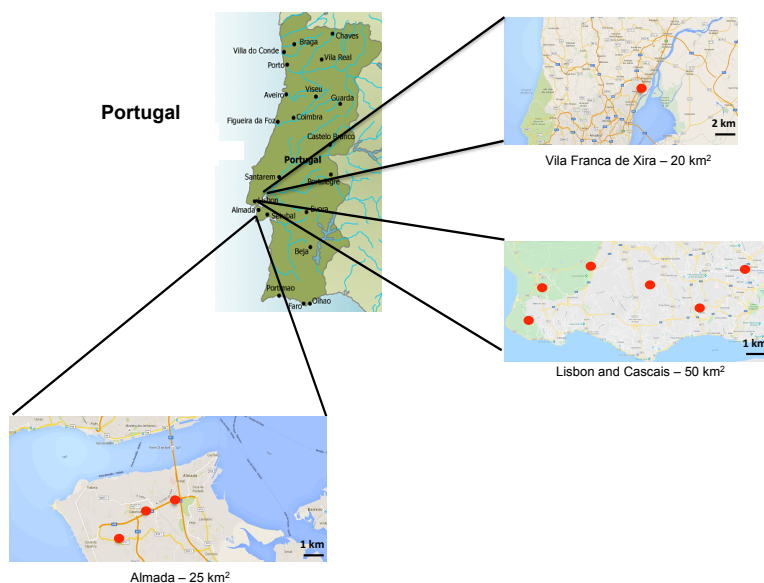
### **2.3.1. Ethical and collection permits**

In this study, we captured wild *Mus musculus* in Portugal under two licenses allocated by the Instituto de Conservação da Natureza e das florestas: Licence 653/2015/capt (métodos não selectivos) and Licence 97/2016/capt (métodos não selectivos). Both the national animal welfare authorities, and the institutional ethics committee in IGC approved all the procedures done in live mice (Instituto Gulbenkian de Ciencia Ethics Committee licence #003.2015 for 5 years (4-2015 to 4-2020) and Direção Geral de Alimentação e Veterinária (DGAV #0421/000/000/2016).

### **2.3.2. *Mus musculus* sample collection**

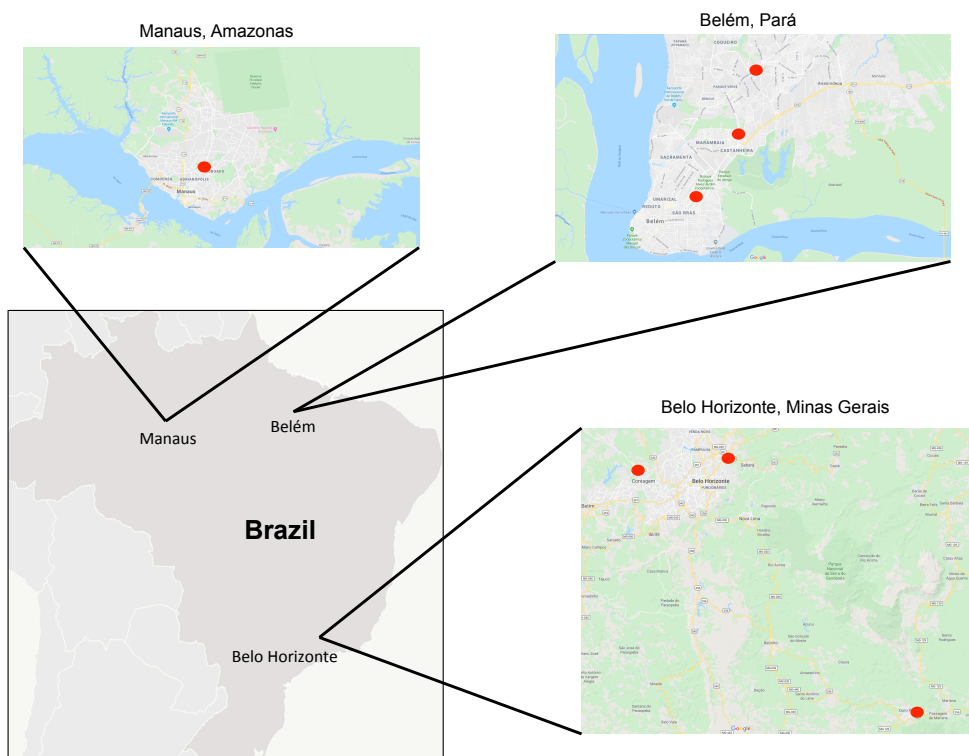
Previous studies from the Howard Lab have shown a high diversity of IRG proteins in wild-derived *Mus musculus* from different geographic locations<sup>2</sup>. Our study followed the rationale of mouse sample collection described by Dr. Urs Benedikt Müller<sup>6</sup>, who pioneered the search for IRG proteins in wild-caught *M. m. domesticus* in Europe and South America<sup>6</sup>.

For this project, the IRG genetic diversity was analyzed in animals collected from several places in Portugal and Brazil (**Supplemental Table 2.1**). To increase the knowledge on IRG protein diversity in wild-caught mouse populations, our study expanded the number of collection spots in both countries. Western European *M. m. domesticus* were sampled in two locations in the greater Lisbon area. The first one was located in the Northeast of Lisbon (Vila Franca de Xira municipality), which covered an area of 20 km<sup>2</sup>. The second spot was located in the south margin of Lisbon (Almada municipality), separated from Lisbon city by the Tagus River. In this location an area of 25 km<sup>2</sup> was screened (**Figure 2.1**).



**Figure 2.1.** Mouse sampling sites in Portugal. Red dots in the map mark the GPS coordinates of sample sites, as described in **Table 2.1** and **Supplemental Table 2.1**. Maps created with Google Maps 2019. Modified from Müller, 2015.

A previous study<sup>6</sup> sampled mice in the southeast and north of Brazil (States of Minas Gerais, and Pará). They aimed to cover a range of *M. m. domesticus* that might have come from Portuguese mouse founders. In the current study, we expanded the geographical area screened to the northwestern part of Brazil, in the city of Manaus, State of Amazonas (**Figure 2.2**). In this case the animals we used were not caught in the field. Instead we obtained animals from an already established colony of wild-derived mice whose founders were caught in Manaus and brought to the Nachman Laboratory at University of California Berkeley (**Table 2.1**).



**Figure 2.2.** Mouse sampling sites in Brazil. Red dots in the map mark the GPS coordinates of sample sites, as described in **Table 2.1** and **Supplemental Table 2.1**. Maps created with Google Maps 2019. Modified from (Müller, 2015).

**Table 2.1.** Mice sampled for this study.

Name	Sex	Sample location	GPS Coordinates
ALM-1	Male	Almada, Portugal	38°38'05.7"N 9°12'49.4"W
ALM-2	Female	Almada, Portugal	38°39'05.3"N 9°11'36.6"W
ALM-3	Female	Almada, Portugal	38°39'05.3"N 9°11'36.6"W
ALM-5	Male	Almada, Portugal	38°39'05.3"N 9°11'36.6"W
ALM-6	Male	Almada, Portugal	38°39'06.7"N 9°10'40.6"W
ALM-7	Female	Almada, Portugal	38°39'06.7"N 9°10'40.6"W
ALM-8	Female	Almada, Portugal	38°39'06.7"N 9°10'40.6"W
ALM-9	Female	Almada, Portugal	38°39'06.7"N 9°10'40.6"W
ALM-11	Male	Almada, Portugal	38°39'06.7"N 9°10'40.6"W
MANB	Female	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-2	Female	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-3	Male	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-4	Male	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-5	Female	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-6	Male	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-7	Female	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
MANA	Female	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-10	Male	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
BR-12	Female	Manaus, Amazonas, Brazil	UC Berkeley, California, USA
VFX-1	Male	Vila Franca da Xira, Portugal	38°58'35.2"N 8°59'00.2"W
VFX-2	Female	Vila Franca da Xira, Portugal	38°58'35.2"N 8°59'00.2"W
TC-1	Male	Tucson, Arizona, USA	UC Berkeley, California, USA
ED-2	Male	Edmonton, Canada	UC Berkeley, California, USA

### 2.3.3. Preparation and immortalization of tissue culture lines from wild-caught and wild-derived mice

Fibroblasts such as mouse embryonic fibroblast (MEFs) are widely used for *in vitro* experiments with *T. gondii*. However, they are not easy to obtain from mice recently derived from the wild. We therefore modified a method published by Antony et al., 1989, to isolate fibroblasts from the diaphragm of single adult wild-caught mice, we termed them Diaphragm-Derived Cells or DDCs<sup>2,21</sup>. It is important to appreciate that, like MEFs, DDCs are derived from a complex cell population, including macrophages, fibroblasts and other mesothelial cell types. DDCs might show population-specific

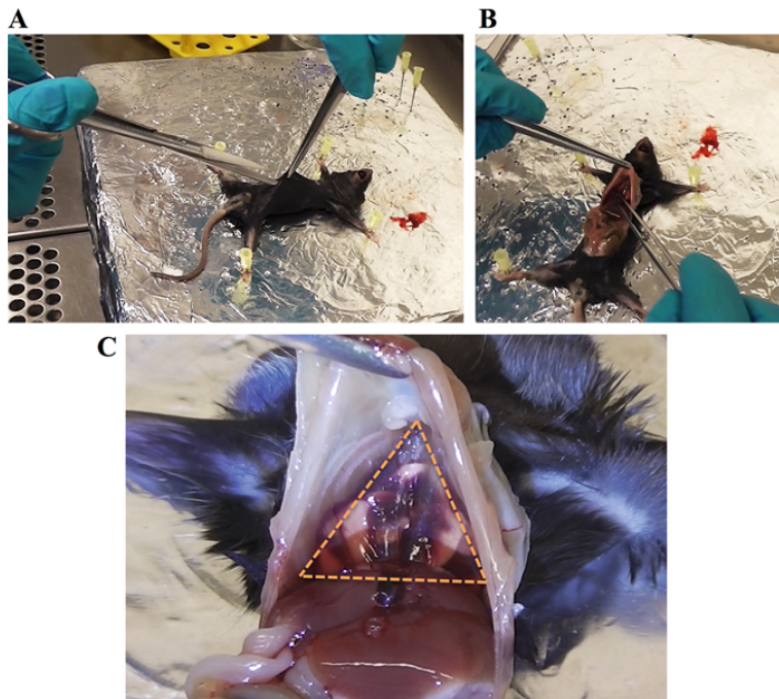
differences, due to the differential isolation of different cell subsets from the original tissue, as well as a differentially efficient transfection during the immortalization process<sup>21</sup>.

Collection of wild mice in the field was done using a modified Sherman trap (wooden made with a metallic grid on top), baited with small pieces of apple covered in chocolate. Traps were placed in each collection spot by the end of the day and those with live animals were collected next day early in the morning. Live animals were brought to IGC where they were anaesthetized, measured, weighed and sexed and finally killed by cervical dislocation. The abdomen was then sprayed with 70% Ethanol (EtOH) and opened, and the diaphragm removed (**Figure 2.3**). The diaphragm was washed briefly with sterile 1X phosphate-buffered saline (PBS) to remove erythrocytes and using fine scissors, cut into small pieces, as small as possible. Collagenase/Dispase 1mg/ml (25ul of stock solution (100mg/ml) in 2.5 ml of Dulbecco's modified Eagle's medium (DMEM) growth medium but without Heat-inactivated fetal bovine serum (HI-FBS), filtered with 0.2 mm syringe filter before adding to cells) was added to the diaphragm fragments. All the Collagenase/Dispase along with the diaphragm fragments was taken up and transferred to a sterile microcentrifuge tube. Fragments of diaphragm in Collagenase/Dispase for were incubated for 1h at 37°C using a ThermoMixer®. The tube was shaken at 900 rpm and every 15 min removed from the mixer and manually flipped a couple of times and returned to the mixer. The tube was spun down at low speed (100 g – 800rpm) for 10-15 sec and supernatant transferred to a fresh sterile microfuge tube. Sterile 1x PBS was added to the remaining diaphragm debris, shaken vigorously and spun again at low speed (100 g – 800rpm) for 10-15 sec. The new supernatant was transferred to the tube with the previous supernatant. Sterile 1X Trypsin / Ethylenediaminetetraacetic acid (EDTA) solution was added to diaphragm debris and incubate for 1h at 37°C in the ThermoMixer®, shaking at 750 rpm for 1 hour. Every 15 min the tube was



flipped by hand a couple of times and put back on the mixer. In the meantime, the tube containing the supernatants was spun down from the first incubation for 6 min at (500 g - 1800rpm). Supernatant was taken off carefully to avoid disturbing the cellular pellet. The pellet was re-suspended in residual supernatant and plated out. Fresh complete DMEM growth medium with 10% HI-FBS was added. To the suspension from the previous step, DMEM growth medium was added and plated out as a whole (diaphragm-debris + Trypsin / EDTA + DMEM growth medium). Fresh DMEM growth medium was added and cells were seeded in 6 well-plate cell (cultured treated) and immediately after placed in a cell culture incubator at 37°C (10% CO<sub>2</sub>). Medium from each well was changed once a day for 3 to 4 consecutive days initially to remove debris as far as possible. Approximately 3 to 4 days after plating small cell colonies start to grow up. Cells might take up to two weeks to reach 60-80% confluence. At this point, cells were either ready to be frozen or to be immortalized by transfection of an oncogenic plasmid. For the transfections, manufacturer's instructions for the ScreenFect®A Transfection Reagent kit <sup>22</sup> were followed up. Our laboratory has established that ScreenFect®A tranfection reagent is markedly more efficient in this setting than other transfection protocols that we have used in the past. Transfection rates for these primary cell cultures may be as high as 25% (Dr. Claudia Campos, personal communication). For this immortalization protocol we used the pSV3-neo plasmid expressing the immortalizing SV40 large T-antigen which alters the tumor suppressor proteins pRb and p53 <sup>23</sup>. After the transfection process, cells were placed in 6 well-plate and immediately after in a cell culture incubator at 37°C (10% CO<sub>2</sub>) and medium replaced after 48 hours. Cells were passaged to cell culture flasks (T25) and passaged at least 5 times more to ensure all cells were indeed immortalized. Established cell lines were maintained in Growth medium (DMEM supplemented with 10 % HI-FBS, 1X NEAA, 1X Penicillin-Streptomycin, 1mM Sodium Pyruvate and 2mM L-Glutamine) or were frozen

in DMEM+ 10% DMSO. This protocol was modified from Alvarez et al., 2020.



**Figure 2.3.** Mouse diaphragm dissection. A) Opening of the peritoneal cavity. B) Visualization of the diaphragm membrane. C) Triangular shaped-diaphragm membrane. Modified from Alvarez et al., 2020.

#### **2.3.4. Isolation of RNA from immortalized DDCs for Transcriptome analysis using Illumina sequencing**

Immortalized DDCs were stimulated with 200 U/ml (40 ng/ml) of IFN $\gamma$  for 24h. Cells were detached with a cell scraper and the cell pellet washed twice with PBS 1X. Total RNA was isolated using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Before quantification, total RNA was subjected to a DNase Treatment using the DNA-free<sup>TM</sup> kit (Ambion<sup>®</sup>) following the manufacturer's protocol. Library construction and whole transcriptome sequencing from total RNA were done at the Beijing

Genomics Institute (BGI, Hong Kong) using an Illumina-HiSeq2500 machine.

### 2.3.5. Isolation of genomic DNA from mouse tissues

Mouse DNA from wild-caught mice was isolated from approximately 25 mg of mouse liver. Total DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's protocol for animal tissues (Spin-Column protocol). In the first step of the protocol, an overnight incubation with Proteinase K (600 mAU/ml manufacturer solution) was added in order to increase tissue digestion and therefore DNA yield. DNA was eluted in 100 µl of Buffer AE (10 mM Tris·Cl, 0.5 mM EDTA, pH 9.0) and stored at -80°C.

### 2.3.6. Amplification of IRG genes

A 953 bp fragment of the *lrgd* gene was amplified using the primers *lrgd\_17B\_f* and *lrgd\_17B\_b* (The sequences of the primers used in this thesis can be found in the **key resources table** section). A 1400 bp fragment of the *lrgm1* gene was amplified using the primers *lrgm1\_655\_f* and *lrgm1\_655\_b*. The high fidelity DNA polymerase Herculase II (Agilent) was used for both genes with the following touchdown PCR protocol: 95°C 2 min, 10x (95°C 20 sec, 65°C 20 sec – cool down of 1°C/cycle until 55°C, 72°C 45 sec), 25x (95°C 20 sec, 55°C 20 sec, 72°C 45 sec), 72°C 5 min, 8°C o.n. MacroGen (Spain) sequenced the amplified fragments using Sanger sequencing technology.

The coding region of the gene *lrgb1* (1377 bp) was amplified using the primers B1FCA and B1RCA. For the gene *lrgb2* (1322 bp) the primers used were *lrgb2\_66F\_f* and *lrgb2\_66F\_b*. The high fidelity DNA polymerase Herculase II (Agilent) was used for both genes with the following touchdown

PCR protocol: 95°C 2 min, 10x (95°C 20 sec, 65°C 20 sec – cool down of 1°C/cycle until 55°C, 72°C 45 sec), 25x (95°C 20 sec, 55°C 20 sec, 72°C 45 sec), 72°C 5 min, 8°C o.n. MacroGen (Spain) sequenced the amplified fragments using Sanger sequencing technology.

A 9.3 kb fragment of the *lrgb2-b1* gene that covers the two exonic regions and the intronic region in between them was amplified. The primers used were *lrgb2\_66F\_f* and *purple20b2b1R*. The TaKaRa LA PCR Kit Ver. 2 for Long Range PCRs was used with the following thermal protocol: 30x (98°C 10 sec, 68°C 30 sec, 68°C 15 min), 8°C o.n.

All DNA fragments were mixed 5:1 with DNA Loading Buffer prior to agarose gel electrophoresis. DNA was visualized by the use of Midori Green Advance DNA Stain (1:30000) and ultraviolet (UV) light on TAE 1% agarose gels for fragments lower than 1.5 kb and 0.7% for fragments higher than 1.5 kb. The size of DNA fragments was compared to the GRS Ladder 100bp (100 bp – 1500 bp) (GRiSP) or with  $\lambda$ -Hind III digested MW Marker (2 kb – 23 kb) for bigger fragments.

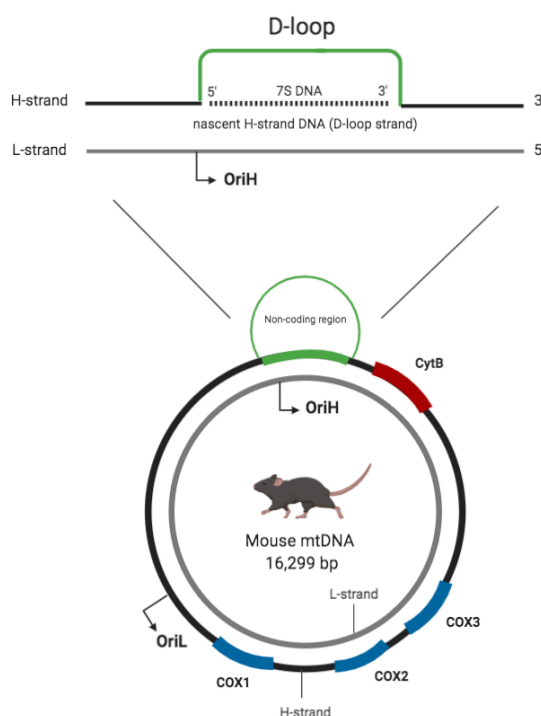
### **2.3.7. Nanopore sequencing *lrgb2-b1* gene**

In order to characterize polymorphic sites present in non-coding regions of the *lrgb2-b1* gene, the 9.3 kb fragment containing both coding exons of *lrgb2-b1* and the included intron was amplified as described above. The long-read Oxford Nanopore® Sequencing Technology was used to sequence the amplified fragment at the Genomics Unit of the Instituto Gulbenkian de Ciencia. The equipment used was a MinION™ Mk1B with flow cells version R9.4.1. Samples were previously barcoded with the PCR Barcoding Expansion Kit from Nanopore® using the primers Tailed-PrimerF-*lrgb2b1* and Tailed-PrimerR-*lrgb2b1*.

### **2.3.8. PCR and sequencing of mitochondrial DNA (D-loop)**

Sequences in the Non-coding region (D-loop) of the mtDNA (**Figure 2.4**) are widely used in phylogeographic approaches. Characteristics such as maternal inheritance, fast mutation rates, high copy number per cell and lack of recombination make this region a great candidate for recent ancestry and evolutionary genetics analysis<sup>24</sup>.

D-loop fragments from mitochondrial DNA were amplified using the DreamTaq™ Green DNA Polymerase (Fermentas GmbH). The sequencing PCR protocol used was as follows: 96 °C 5 min, 25x (96 °C 30 sec, 60 °C 15 sec, 72 °C 1 min), 72 °C 10 min, 8°C o.n. Macrogen (Spain) sequenced the amplified fragments using Sanger sequencing technology. This protocol is a modified version from Müller, 2015.



**Figure 2.4.** Map of the mouse mitochondrial genome. The Non-coding region (green) contains an origin of replication site for the H-strand (black bold line) called OriH. From this point onwards, the nascent H-strand (7S DNA - dashed lines) binds to the parental L-strand

(grey line) and displaces the parental H-strand to form the displacement loop (D-loop – green bubble). Adapted from (Shadel and Clayton, 1997) and (Müller, 2015)

### **2.3.9. Phylogenetic analysis of mitochondrial DNA (D-loop)**

The 867 bp sequences of 86 unique *M. m. domesticus* D-loop haplotypes were aligned using MUSCLE algorithm in the software Geneious® 9.1.5 (Biomatters, Auckland, New Zealand). Published D-loop sequences included in the alignment were adopted from several studies<sup>6,25–27</sup>. An unrooted phylogenetic tree was generated in Geneious® using the Jukes-Cantor genetic distance model and a neighbor-joining building method with no outgroup and bootstrap resampling method (5000 replicates).

### **2.3.10. Transcriptome analysis and IRG genes genotyping**

100 bp paired-end Illumina-reads from each of the mouse cell lines analyzed were mapped against the C57BL/6 mouse genome (Genome Reference Consortium Mouse Build 38 - GRCm38) using the RNA-seq aligner STAR (Alexander Dobin, Cold Spring Harbor Laboratory). The following parameters for the mapping process were established for all samples: runThreadN 8, alignIntronMax 500000 and sjdbOverhang 99.

Mapped BAM files from each mouse sample were trimmed in three regions, Chromosome 11 (positions 48860000-49100000), Chromosome 11 (58180000-58230000) and Chromosome 18 (60200000-60460000). Trimmed sequences were locally mapped to the ORFs of all known IRG gene alleles using Geneious® 9.1.5 (Biomatters, Auckland, New Zealand).

For each of the IRGs genes, an artificial chromosome representing the ORFs of all known sequences arranged 5' to 3' with 20 “N” gaps in between single ORFs was created. Variations in the mapping stringency (Maximum Mismatches Per Read) allowed us to find new alleles.

Color blocks were created according to the methodology described in Lilue et al., 2013. Briefly, each color block represents one IRG ORF. The numbers on each color block represent amino acid substitutions relative to the corresponding IRG allele present in the C57BL/6 mouse strain. The colors of the blocks indicate phylogenetic relationship. Taxa with similar colors are closer than those with contrasting colors<sup>6</sup>. Note, however, that the phylogenetic proximities are best represented by grades of color when multiple alleles are analyzed. For a di-allelic system, the color differential is the same whether the two alleles differ by 1 or 50 amino acids.

### **2.3.11. Individual analysis of SNPs in IRG genes alleles**

Sanger-sequenced fragments from the *Irgm1* and *Irgd* genes were aligned using the software MEGA 6.0<sup>28</sup>. A manual screening of polymorphic sites was performed, and each position recorded in order to create a consecutive sequence of polymorphic sites.

The 9.3 kb region amplified for *Irgb2-b1* gene was mapped against the Mouse Chromosome 11 using the Minimap2 software (Broad Institute, USA). A SNP calling analysis using samtools-1.9 (mpileup option) and bcftools 1.9 (call option) was performed to obtain a list with all polymorphic sites and their corresponding position.

A sequence containing 25 SNPs in linear order from the genomic genes *Irgm1*, *Irgb2-b1* and *Irgd* was created to cluster mouse genomic samples based on their sequence pattern with respect to these SNPs.

### **2.3.12. Expression levels of IRG genes (Transcriptome-based analysis)**

As part of the transcriptome data analysis, the total count of gene reads for each of the genes of the mouse genome was obtained using the STAR software (quantMode GeneCounts option). Since transcriptomes analyzed were sequenced with a different coverage, data normalization was done by taking the size of the mapped transcriptome with the lowest coverage (ALM1 – 1.62 GB) and dividing the total size of the other transcriptomes by 1.62. This will give a correction factor per sample used for each of the genes analyzed (**Supplemental Table 2.2**). Data visualization was done with the software Geneious® 9.1.5 (Biomatters, Auckland, New Zealand).

## **2.4. Results**

### **2.4.1. Polymorphisms in the IRG genes in wild *Mus musculus***

#### *2.4.1.1. IRG alleles on chromosome 11 in Portuguese mice*

The resistance to highly virulent *T. gondii* strains in wild-derived mice such as the CIM mouse strain is associated to the polymorphic tandem protein IRGB2-B1, located on chromosome 11<sup>2</sup>. We investigated whether IRG proteins located in the chromosome 11 of wild-caught mice in different geographical locations are also involved in the resistance to local *T. gondii* strains.

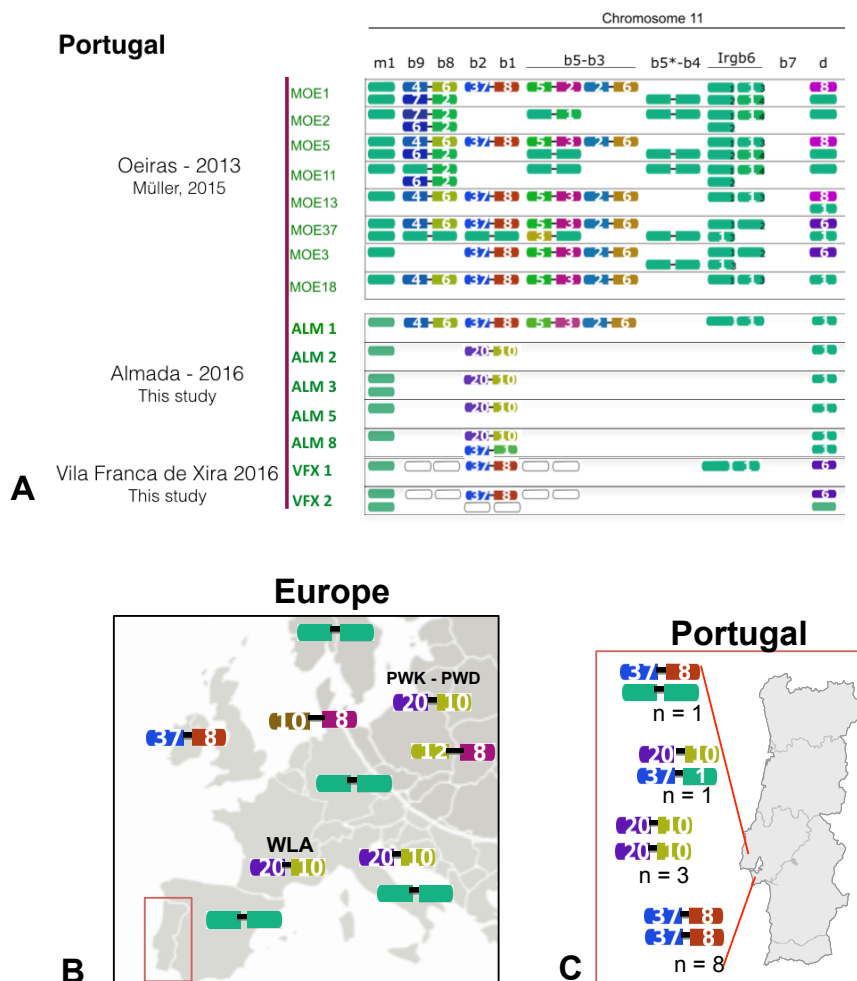
We started the genetic analysis of wild-derived and wild-caught mice in Europe, with a special focus on the Portuguese territory. Müller, 2015 reported that all heterozygous *M. m. domesticus* sampled in the great Lisbon area (Oeiras) (**Supplemental Figure 2.1**) display an IRG haplotype that in chromosome 11 is almost identical to that carried by the wild-derived *M. m. domesticus* from the Scottish Isles (mouse strains SIN and SIT) (**Figure 2.5, panel A**). In this study, we confirmed the prevalence of a Scottish-like IRG haplotype in chromosome 11 of mice sampled in other locations in the great Lisbon area (Vila Franca de Xira and Almada) (**Figure 2.5, panel A**).



These mice, however, display a lower prevalence of the Scottish-like IRG haplotype than the ones screened in Oeiras. We also found in some of these new Portuguese mouse populations an IRG haplotype similar to the one found in the *M. m. musculus* PWD and PWK strains and *M. m. domesticus* SPOS and WLA strains (**Figure 2.5, panel B**). Surprisingly, the animals that are carriers of this IRG haplotype (PWD/PWK-like) were only found in the south margin of the Tagus River (Almada).

As previously mentioned, certain allelic forms of the IRGB2-B1 tandem protein are directly involved in resistance against highly virulent *T. gondii* strains<sup>2,6,29</sup>. Portuguese mice display an overall low diversity in their IRG genotypes, including tandem genes such as the *Irgb2-b1* gene. We observed 4 allelic forms for the IRGB2-B1 protein in a population of 11 homozygous and 2 heterozygous mice (26 chromosomes in total). Compared to the overall diversity (10 IRGB2-B1 allelic forms) found in the European mouse population, Portuguese mice carry only a subset of this diversity. The IRGB2-B1<sub>CIM</sub> allelic form, associated with the resistance to highly virulent *T. gondii* strains was not found in any of the European mice analyzed (**Figure 2.5, panel B**).

Most of the Portuguese mice are carriers of the IRGB2-B1<sub>SIN/SIT</sub> allelic form (17 of 26 chromosomes) (**Figure 2.5, panel C**). A lower number of them are carriers of another allelic form, the IRGB2-B1<sub>PWK</sub> (7 of 26 chromosomes) (**Figure 2.5, panel C**). Unlike IRGB2-B1<sub>SIN/SIT</sub>, the *Irgb2-b1*<sub>PWK</sub> allelic form has a high expression level as well as it is reported to control efficiently virulent type I *T. gondii* strains<sup>6</sup>. In the Portuguese mouse population, one animal was found to be a carrier of a recombinant allelic form between IRGB2-B1<sub>PWK</sub> and a new allelic form that is a mixture of the IRGB2 from the SIN/SIT strain and the IRGB1 from the C57BL/6 strain (ALM8) (**Figure 2.5, panel A**).



**Figure 2.5.** Polymorphism of IRG proteins on chromosome 11 in Portuguese mice. Transcriptomes obtained from IFN $\gamma$ -stimulated DDCs were used to determine IRG proteins alleles in wild mice caught in Portugal. Protocols to determine colors and number of amino acid changes in relation to the C57BL/6 mouse (green color blocks) were according to Lilue et al., 2013 and Müller, 2015. (for details, see Material and Methods). Open blocks indicate homologues that were not sufficiently covered by Illumina reads mapped in that region. Heterozygous animals are depicted as two framed rows. A) IRG alleles from Portuguese *Mus musculus* caught in three different locations. First block displays data from mice caught in Oeiras by a previous study<sup>6</sup>. Second and third blocks display IRGs from mice caught in this study in Almada and Vila Franca de Xira, respectively. B) Irgb2-b1 alleles in European homozygous *Mus musculus*, data extracted from Lilue et al., 2013. C) Irgb2-b1 alleles in Portuguese heterozygous wild-caught *Mus musculus*. Total number of animals for each allele is placed below each color block.

#### 2.4.1.2. *IRG alleles in Brazilian and other North and South American mice.*

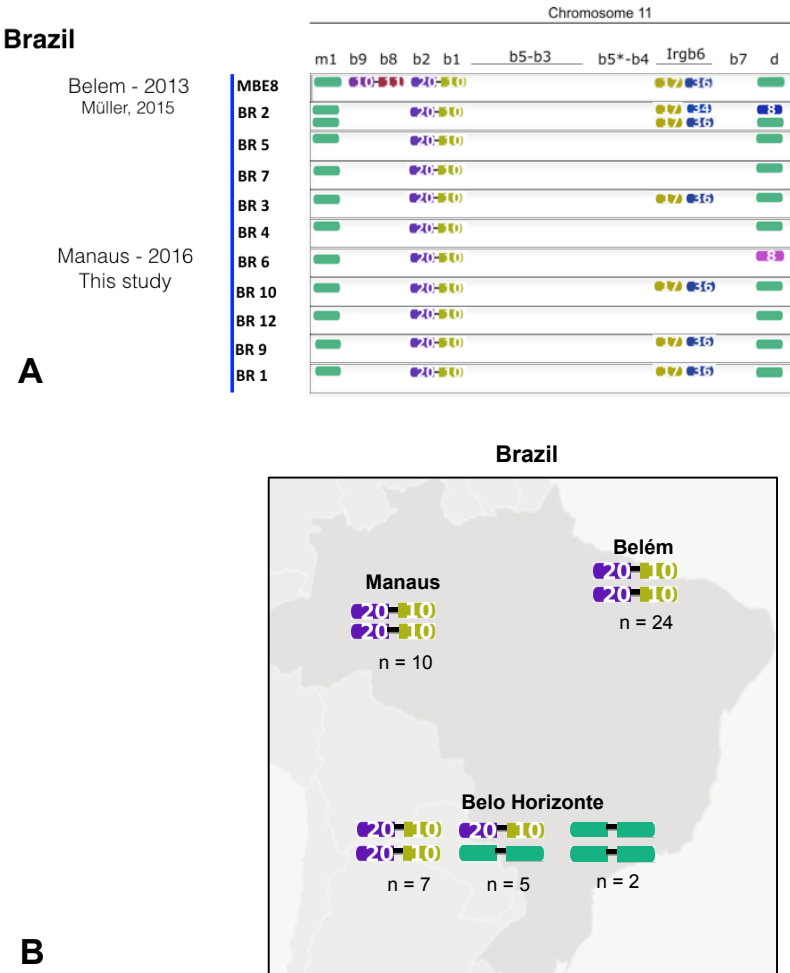
The recent colonization of the American continent by the originally Eurasian species *Mus musculus* (the domestic mouse) represents a potential opportunity to study local adaptation and/or co-evolution resulting from new contacts with originally South American parasitic species such as *T. gondii*.

Müller, 2015 reported that in the transcriptome from cells of the wild-caught mouse MBE8 in Belem (Brazil) it was possible to find for the large cluster of chromosome 11 an IRG haplotype that is very similar to the one reported for several wild-derived *M. m. musculus* strains. However, MBE8 cells have in the small cluster of the same chromosome, an IRG haplotype similar to the one found in *M. m. domesticus*. Since the number of Brazilian mice analyzed by previous studies was limited, we added new information from a wild-derived colony of mice from Manaus (Brazil).

The 10 mouse strains from Manaus have an almost identical IRG haplotype in the large cluster of chromosome 11 that the one reported for the MBE8 mouse (**Figure 2.6, panel A**). These IRGs are identical to those found in *M. m. musculus* mouse strains such as PWD and PWK. However, we observed non-coding nucleotide differences in highly conserved IRGs like IRGM1, as well as different allelic forms for other IRGs, like IRGD<sup>6</sup>.

Originally found by Dr. Urs Benedikt Müller, Brazilian mice display an extremely high prevalence of the IRGB2-B1<sub>PWK</sub> allelic form. To confirm this finding and since the initial analysis was done by the PCR amplification of only the *Irgb2* exon in wild-caught Brazilian mice (Belem and Belo Horizonte), we decided to also amplify the *Irgb1* exon from these populations to confirm the prevalence of the full *Irgb2-b1<sub>PWK</sub>* allele. Extending Muller's observation we found that all the 24 mice caught in

Belém and half of the 14 mice caught in Belo Horizonte are homozygous for the full *Irgb2-b1<sub>PWK</sub>* allele. We also found 5 mice from Belo Horizonte that are heterozygous for the *Irgb2-b1<sub>PWK</sub>* allele. Added to the finding that all Manaus mice are also homozygous for the same *Irgb2-b1<sub>PWK</sub>* allele, we can conclude that the *Irgb2-b1<sub>PWK</sub>* allele has a frequency of 90% (87 of 96 chromosomes) in Brazilian mice (**Figure 2.6, panel B**). These differences in allele frequency of the *Irgb2-b1<sub>PWK</sub>* allele in the Brazilian population (0.906) in comparison to the European population (0.291) are statistically significant (Fisher's exact test, two-tailed, \*\*\*\*p<0.0001).



**Figure 2.6.** Polymorphism of IRG proteins on Chromosome 11 in Brazilian mice. A) IRG alleles from Brazilian *Mus musculus*. First block displays data from a mouse caught in Belém

by a previous study<sup>6</sup>. Second block displays IRGs from 10 wild-derived inbred mouse lines from mice caught in Manaus by the Nachman Laboratory (UC Berkeley). B) Number of alleles for the *Irgb2-b1* gene in wild-derived and wild-caught *Mus musculus* from Brazil. Total number of animals for each allele is placed below each color block.

To test whether the *Irgb2-b1*<sub>PWK</sub> allele was also present in other American mice, we analyzed the transcriptome of IFN $\gamma$  induced DDCs from wild-derived mice from Peru and North America (Table 2.1). We found that the Peruvian wild-derived mouse strain PERA is a carrier of the *Irgb2-b1*<sub>BL6</sub>. Another mouse strain (PERC) derived from a wild mouse caught near to the PERA founder<sup>30</sup> carries a recombinant *Irgb2-b1* gene composed of an *Irgb2*-encoding sequence already reported in the German *M. m. domesticus* SCHEST and an *Irgb1*-encoding sequence found in Scottish (SIN and SIT) and Portuguese (MOE) mice. The North American wild-derived (and apparently not fully inbred) strain TC-1 is a heterozygous carrier of the Peruvian *Irgb2-b1*<sub>PERC</sub> allele and the Portuguese *Irgb2-b1*<sub>ALM8</sub>. Finally, the ED-2 mouse strain from Canada is a carrier of the *Irgb2-b1*<sub>WMP</sub>; this allele is a recombinant between *Irgb2*<sub>PWK</sub> and *Irgb1*<sub>BL6</sub>. Altogether, these results suggest that the high prevalence of the *Irgb2-b1*<sub>PWK</sub> allele is a unique characteristic of the Brazilian mouse population. This finding raised the question of whether the phenomenon seen in Brazilian mice is a result of a population bottleneck and consequent founder effect or positive selection pressure imposed on the *Irgb2-b1*<sub>PWK</sub> allele.

#### **2.4.2. Genetic diversity within the carriers of the *Irgb2-b1*<sub>PWK</sub> allele**

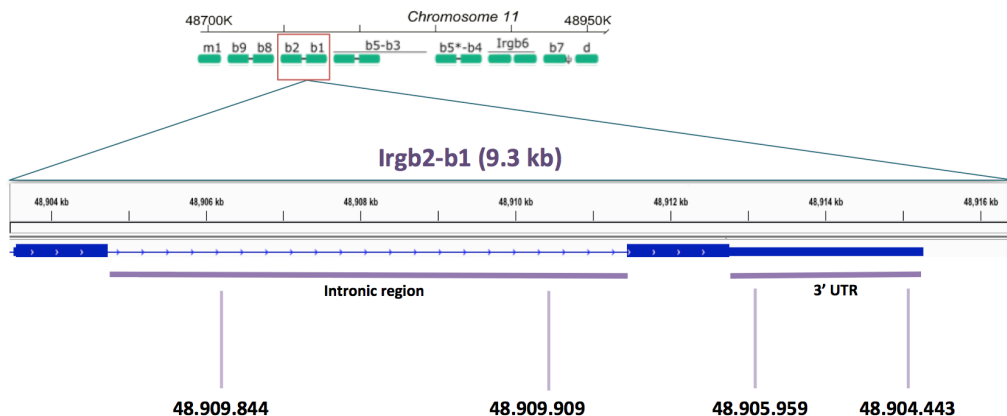
We asked whether the homogeneity of the *Irgb2-b1*<sub>PWK</sub> allele in the Brazilian *M. m. domesticus* population is the result of selection in Brazil on standing variation imported by the original immigrant population from Europe, or of a bottleneck event. Diversity in non-coding nucleotides could suggest complexity of the original immigrant population, arguing for selection at the

protein level rather than a founder effect. To address this question we looked for the nucleotide fine structure of haplotypes present in the carriers of the *Irgb2-b1<sub>PWK</sub>* allele in Brazil.

#### *2.4.2.1. SNPs in the Irgb2-b1<sub>PWK</sub> allele.*

We amplified a 9.3 kb region of the *Irgb2-b1* gene from 58 homozygous and heterozygous *M. m. musculus* and *M. m. domesticus* carriers of the *Irgb2-b1<sub>PWK</sub>* allele. This region contains two exonic regions (*Irgb2* and *Irgb1*) and one non-coding region (the intron in between the two exons). Transcriptomic data, where available, provided 3'-UTR sequences.

Long Nanopore reads confirmed a 100% nucleotide identity in the coding regions among all *Irgb2-b1<sub>PWK</sub>* allele carriers analyzed. However a total of four SNPs were found outside the coding regions, two in the intron in between the two exons (genomic positions 48.909.844 and 48.909.909) and two in the 3' UTR (genomic positions 48.905.959 and 48.904.443) (**Figure 2.7**). Altogether, 5 haplotypes representing different combinations of these SNPs were found for the *Irgb2-b1* gene. Two of them were exclusively found in Brazilian mice, two exclusively in European mice and a fifth one shared between Brazilian and European animals. These results thus document the presence of genetic diversity among the *Irgb2-b1<sub>PWK</sub>* alleles.



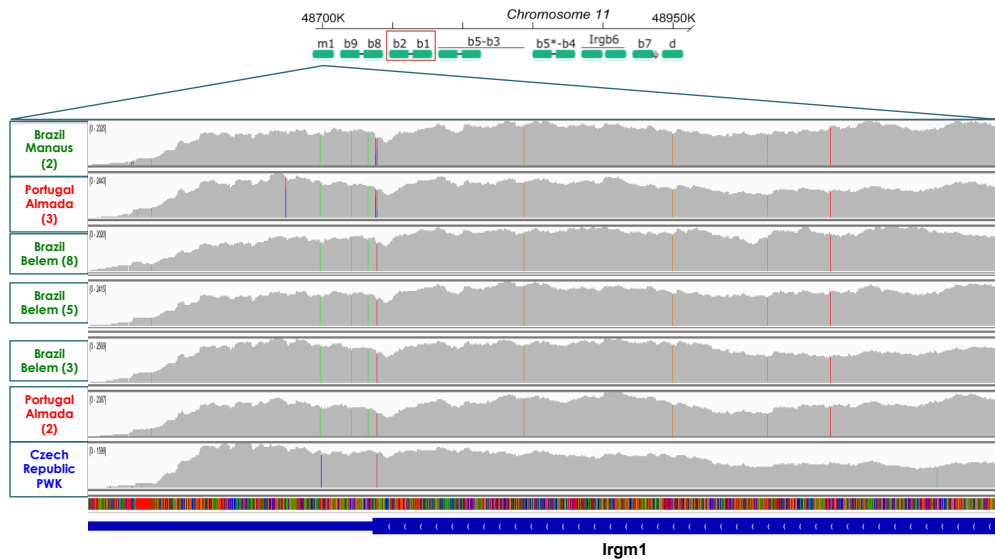
**Figure 2.7.** Schematic representation of the *Irgb2-b1* gene. A region of 9.3 kb was sequenced using Nanopore technology. Nanopore reads contained both coding regions (exons, thick blue line) and one non-coding region (the intron in between the two exons, thin blue line) ( $\approx 17500$  Nanopore reads depth). Analysis of the 3' UTR region (thin blue line) was done based on Transcriptomic data ( $\approx 2000$  Illumina reads depth). Genomic positions of four SNPs found in non-coding regions of the *Irgb2-b1*<sub>PWK</sub> allele are stated at the bottom of the image. Figure modified from IGV v. 2.3.72 in reference <sup>31</sup>.

#### 2.4.2.2. Variations in the *Irgm1* and *Irgd* genes among the *Irgb2-b1*<sub>PWK</sub> allele carriers.

To identify additional signs of genetic diversity among the carriers of the *Irgb2-b1*<sub>PWK</sub> allele, we analyzed the sequences of two IRG genes adjacent to the *Irgb2-b1* gene, the *Irgm1* and *Irgd* genes.

*Irgm1* is located in the large cluster of the chromosome 11, approximately 44 kb distant from the *Irgb2-b1* gene. This gene is highly conserved among several species and coding polymorphism in mice is limited to a single conservative exchange found in *M. m. castaneus* strains<sup>2</sup>. All mice analyzed have the same protein sequence for IRGM1. However, a total of 17 non-coding SNPs were found in the *Irgm1* sequences analyzed (**Figure 2.8**). Ten of them are located in the 3' UTR region and 7 in the coding region, generating a total of 10 different haplotypes. One haplotype was found only

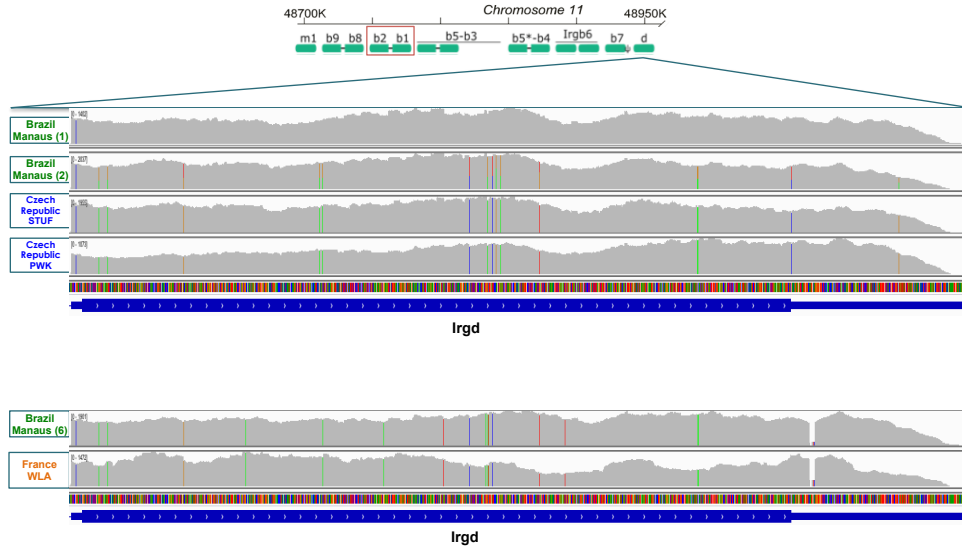
in Brazilian mice and two were shared between Brazilian and European mice.



**Figure 2.8.** Examples of the alleles found in the *Irgm1* gene in mice that are carriers of the *Irgb2-b1PWK* allele. Illumina reads were mapped against the *Irgm1* gene (Mouse reference genome: mm10). IGV browser view depicts depth of coverage for each mouse sample. Colored bars in the depth of coverage represent the relative proportion of non-reference reads across sites (SNPs), which vary depending on the site. A two-color bar means there is more than one nucleotide in that position (heterozygosity). Blue line in the bottom depicts the 3' UTR region (thin line) and the exonic region (thick line).

The first-ever reported p47 GTPase was *Irgd*<sup>32</sup>. This gene is also located in the large cluster of chromosome 11 and separated from the *Irgb2-b1* gene by approximately 190 kb. It is polymorphic in wild mice populations and is usually found as a single copy gene<sup>8</sup>. We found 23 SNPs in the *Irgd* gene from our wild mice (Figure 2.9) 17 of them in coding regions and 6 in non-coding regions (3' UTR). These SNPs represent a total of 5 haplotypes, where only 3 of them are shared between Brazilian and European mice.





**Figure 2.9.** Examples of the alleles found in the *Irgd* gene in mice that are carriers of the *Irgb2-b1<sub>PWK</sub>* allele. Illumina reads were mapped against the *Irgd* gene (Mouse reference genome: mm10). IGV browser view depicts depth of coverage for each mouse sample. Colored bars in the depth of coverage represent the relative proportion of non-reference reads across sites (SNPs), which vary depending on the site. A two-color bar means there is more than one nucleotide in that position (heterozygosity). Blue line in the bottom depicts the 3' UTR region (thin line) and the exonic region (thick line).

#### 2.4.2.3. Haplotypes in the carriers of the *Irgb2-b1<sub>PWK</sub>* allele based on SNPs in the *Irgm1*, *Irgb2-b1* and *Irgd* genes.

A 25-nucleotide sequence was created with all the SNPs found in linear order in the coding region of *Irgm1*, in the intronic region of *Irgb2-b1* and the coding region of *Irgd* (Table 2. 2). The analysis of this data showed that the 58 mouse genomic samples clustered into 23 haplotypes based on sequence patterns. Out of those 23 haplotypes, 16 are exclusively found in Brazilian mice. The most frequent of them, Haplotype 1, clustered a significant number of Brazilian mice from Belém and Manaus, whereas other Brazilian haplotypes diversely clustered animals from Belo Horizonte, Belem, and Manaus. From the other 7 haplotypes, one is shared between

European and Brazilian mice, 3 clustered only Portuguese mice and the 3 other did it exclusively with East European mice.

In conclusion, Brazilian *M. m. domesticus* have an exceptionally high prevalence of the Eurasian *Irgb2-b1<sub>PWK</sub>* allele. However, the finding of multiple haplotypes among members of this population makes this phenomenon unlikely to be due to a founder effect. Our data so far rather support selection on standing variation.

**Table 2. 2.** Haplotypes in the *Irgb2-b1*<sub>PWK</sub> allele carriers based on SNPs in the *Irgm1*, *Irgb2-b1* and *Irgd* genes. Genomic positions for each of the 25 SNPs analyzed in the coding and non-coding regions of the *Irgm1*, *Irgb2-b1* and *Irgd* genes are specified in the second row of the table. Subsequent rows correspond to the names of the mouse lines analyzed (Homozygous and Heterozygous animals). Reference allele belongs to the C57BL/6 mouse strain (green). Dash means identity with the reference sequence (C57BL/6).

	Irgm1							Irgb2-b1		Irgd																
Position	48.865. 807	48.86 5.810	48.86 6.092	48.86 6.377	48.86 6.559	48.86 6.680	48.86 6.884	48.90 9.844	48.90 9.909	49.09 5.589	49.09 5.700	49.09 5.831	49.09 5.837	49.09 5.945	49.09 6.052	49.09 6.098	49.09 6.127	49.09 6.131	49.09 6.132	49.09 6.133	49.09 6.139	49.09 6.146	49.09 6.154	49.09 6.223	49.09 6.268	
C57BL/6	C	C	T	C	A	C	G	G	C	T	G	G	G	G	A	T	G	G	C	G	T	A	G	G	C	
BEL1	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL2	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL3	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL4	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL5	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL6	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL7	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
MBE8	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL9	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL10	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL11	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL12	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL13	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL15	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL16	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL17	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL18	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL19	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL20	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL22	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL23	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL24	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
MANB	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BR3	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BR4	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BR5	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BR7	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
MANA	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BR10	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BR12	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
BEL14-1	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	T	
BEL14-2	-	T	G	G	G	T	-	C	A	G	A	-	A	A	T	-	A	T	-	C	-	-	-	-	-	
BH1-1	T	-	G	G	-	T	-	-	-	G	-	-	A	-	-	C	-	A	-	-	C	G	A	T	-	
BH2-1	T	-	G	G	-	T	-	-	-	G	-	-	A	-	-	C	-	A	-	-	C	G	A	T	-	
BH3-1	T	-	G	G	-	T	-	-	-	G	-	-	A	-	-	C	-	A	-	-	C	G	A	T	-	
BH5-1	T	-	G	G	-	T	-	-	-	G	-	-	A	-	-	C	-	A	-	-	C	G	A	T	-	

BH1-2	T	-	G	G	-	T	-	C	-	G	-	A	A	-	T	C	T	A	T	-	C	G	A	T	-
BH3-2	T	-	G	G	-	T	-	C	-	G	-	A	A	-	T	C	T	A	T	-	C	G	A	T	-
BH5-2	T	-	G	G	-	T	-	C	-	G	-	A	A	-	T	C	-	A	T	-	C	G	A	T	-
BH6-1	T	-	G	G	-	T	-	-	-	-	-	-	-	-	-	C	-	-	-	-	C	-	-	T	-
BH6-2	-	T	G	G	G	T	-	C	A	G	-	-	A	-	T	-	-	A	T	-	C	-	-	T	-
BH8	T	-	G	G	-	T	-	C	-	G	-	-	A	-	T	C	A	A	T	-	C	-	-	T	T
BH9-1	T	-	G	G	-	T	-	C	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH9-2	T	-	G	G	-	T	-	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH11-1	T	-	G	G	-	T	-	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH12-1	T	-	G	G	-	T	-	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH10-1	-	T	G	G	G	T	-	C	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH11-2	-	T	G	G	G	T	-	C	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH12-2	-	T	G	G	G	T	-	C	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH10-2	-	T	G	G	G	T	-	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH13	-	T	G	G	G	T	-	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BH14	-	T	G	G	G	T	-	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BR2-1	T	-	G	G	-	T	-	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BR6	-	T	G	G	G	T	-	C	A	G	A	-	A	A	T	C	A	A	T	-	C	-	-	T	T
SPOS	-	T	G	G	G	T	-	C	A	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
BR2-2	-	T	G	G	G	T	-	C	A	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
WMP	-	T	G	G	G	T	-	C	A	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
ALM2	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	A	-	A	-	-	-	-	-
ALM3-1	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	A	-	A	-	-	-	-	-
ALM5	-	T	G	G	G	T	-	C	A	-	-	-	-	-	-	-	-	A	-	A	-	-	-	-	-
ALM3-2	T	-	G	G	-	T	-	C	A	-	-	-	-	-	-	-	-	A	-	A	-	-	-	-	-
ALM8-2	T	-	G	G	-	T	-	C	A	-	-	-	-	-	-	-	-	A	-	A	-	-	-	-	-
ALM8-1	T	-	G	G	-	T	-	C	A	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-
STUF	-	T	-	-	-	-	A	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
PWK	-	T	-	-	-	-	A	-	-	G	-	A	A	-	-	C	-	A	-	-	C	G	A	T	-
DJO	-	-	-	-	-	-	A	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
WLA	-	T	G	G	G	T	-	-	A	G	A	-	A	A	-	C	A	A	T	-	C	-	-	T	T
BH4	T	-	G	G	-	T	-	-	-	G	-	-	A	-	T	C	A	A	T	-	C	-	-	T	T
BH7	T	-	G	G	-	T	-	-	-	G	-	-	A	-	T	C	A	A	T	-	C	-	-	T	T
BH2-2	T	-	G	G	-	T	-	C	-	G	-	A	A	-	T	C	A	A	T	-	C	G	A	T	-

### 2.4.3. Maternal phylogeny of wild and wild-derived *Mus musculus*

Reconstructions of Mitochondrial DNA sequences are a rather indirect and limited way to analyze phylogenies (due to its small size in comparison to genomic DNA) but could be useful to look for signatures of bottlenecks or founder effects in mammals. To know whether Brazilian mouse populations display a reduction in matrilineal ancestry diversity, possibly associated with a bottleneck effect, we expanded the mitochondrial phylogenetic reconstruction done by Müller, 2015. In that previous study, 867 bp D-loop sequences of mouse samples from distinct parts of Eurasia were retrieved from the literature<sup>25–27</sup> and aligned with sequences from Brazilian (Belém and Belo Horizonte), Portuguese (Oeiras) and Asian (India) wild-caught mice (**Supplemental Table 2.1**). To add an even wider framework for our Brazilian mouse population, we added sequences from mice caught in Brazil (Manaus), Portugal (Almada and Vila Franca de Xira), Peru (PERA and PERC) and North America (Tucson, Edmonton, Florida, and New York). A new phylogenetic reconstruction was done in 136 sequences using the Jukes-Cantor genetic distance model and a neighbor-joining building method (outgroup included, *Mus spretus* - SMON) (**Supplemental Figure 2.1**).

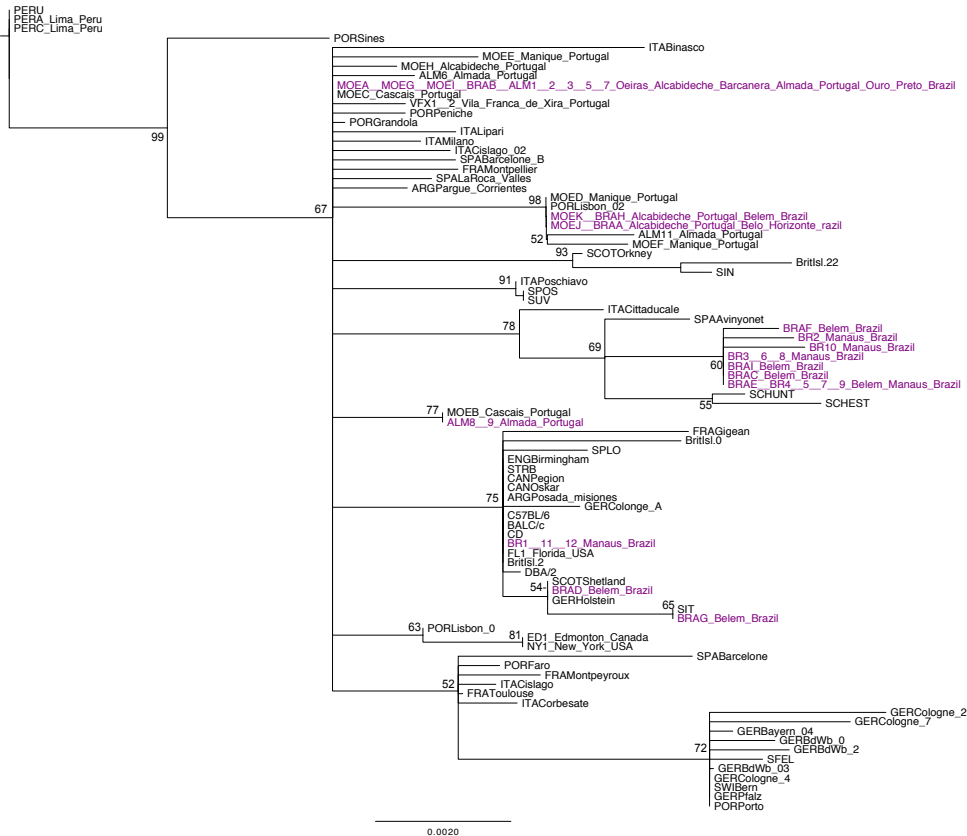
We confirmed the overall phylogenetic structure obtained by Müller, 2015, where the D-loop sequences analyzed grouped with known haplotypes from one of the three distinct branches for *M. m. domesticus*, *M. m. musculus* and *M. m. castaneus*. Additionally, the new sequences from Eurasian mice added in this study exclusively grouped with known haplotypes of mice from the same geographic origin. For example, Portuguese mice caught in Almada and Vila Franca de Xira clustered only with previously reported sequences from Portugal, such as MOE and POR.

On the other hand, we observed an interesting clustering in North American mice. We expected all American sequences to cluster with *M. m. domesticus* mice; however, that was not the case for the North American mouse TUC (Tucson, Arizona), which clustered in the *M. m. castaneus* branch, very closely to mice from Thailand and China.

#### *2.4.3.1. Brazilian Mus musculus domesticus ancestry*

*M. m. domesticus* successfully arrived and colonized the territory that nowadays corresponds to North and South America. Therefore, the phylogenetic reconstruction of the maternal lineages of this subspecies was vital for us to identify possible sources of genetic diversity reduction in our Brazilian mouse populations.

We ran a new phylogenetic reconstruction with only the D-loop sequences from *M. m. domesticus* samples without an outgroup, but also using the Jukes-Cantor genetic distance model and neighbor-joining building method. As expected, the overall topology of this tree (**Figure 2.10**) is slightly different from the one reconstructed for the three subspecies (**Supplemental Figure 2.1**) and it was possible to see better branch support for some of the important associations for us, such as the clustering of Portuguese and Brazilian samples.



**Figure 2.10.** Maternal phylogeny of wild and wild-derived *Mus musculus domesticus*. Unrooted phylogenetic tree based on a multiple sequence alignment of 86 unique 867 bp D-loop fragments from wild and wild-derived *M. m. domesticus*. Animals that are carriers of the *Irgb2-b1<sub>FWK</sub>* allele are labeled in purple. The Jukes-Cantor genetic distance model along with a neighbor-joining building method with no outgroup were computed in the software Geneious® 9.1.5 (Biomatters, Auckland, New Zealand). A bootstrap resampling method with 5000 replicates was run. Branch support is shown if >50. Scale=0.2 % differences. The D-loop sequences were either sequenced from genomic DNA or retrieved from published literature by Müller, 2015<sup>25–27,33</sup>. Müller, 2015 also established a haplotype nomenclature where mice caught in Portugal (Oeiras) are named as MOE, and mice caught in Brazil (Belo Horizonte and Belém) as BRA. Mice caught for this study in Portugal (Almada and Vila Franca de Xira) are named ALM and VFX, respectively and wild-derived mice from Brazil (Manaus) are named as BR. A Latin letter is placed after the name indicating the existence of multiple identical haplotypes in one haplogroup.

*M. m. domesticus* D-loop haplotypes displayed a high diversity on their maternal inter-subspecific structures, as it was already reported by Müller,

2015. Many of the new haplotypes introduced in this analysis, such as the ones from Portuguese mice in Almada and Vila Franca de Xira, clustered into three different branches. The first one includes a large haplogroup of Portuguese and Brazilian samples (MOEA=MOEG=MOEI= BRAB=ALM1-7; MOEC, ALM6, MOEH, and VFX1-2). This haplogroup alone contains 25% of the Brazilian mice (12 out of 47 animals) and 61% of Portuguese mice (34 out of 55). This a very diverse group of samples, considering that BRA-B was not restricted to one sample location in Brazil; animals in this haplotype were rather sampled in very distant places (about 2200 km apart) such as Ouro Preto (Minas Gerais) and Belém (Pará)<sup>6</sup>. The second haplogroup also contains samples from Portuguese and Brazilian mice (MOED, PORLisbon, MOEK=BRAH, MOEJ=BRAA, ALM11, and MOEF). Finally, the third haplogroup, relatively distant from the other branches that contain Portuguese samples, clustered exclusively animals from Portugal (MOEB and ALM8-9).

As previously mentioned, a significant percentage of Brazilian mice shared a close ancestry with Portuguese mice. Even though this is not a surprise, since Brazil was a colony of Portugal for almost 300 years and very likely multiples events of colonization by Portuguese mice occurred, a considerable number of Brazilian house mice also shared maternal lineage with other European *M. m. domesticus*.

One D-loop branch contains exclusively Brazilian mice (25% of them, 12 out of 47 animals), coming from Belém (Pará) and Manaus (Amazonas) (BRAAF, BR2, BR10, BR3-6-8, BRAI, BRAC, and BRAE=BR4-5-7-9). This appears at first glance to be surprising because the two cities are approximately 1500 km apart from each other, but they are connected by the Amazon River allowing easy transport and exchange. This branch of Brazilian mice clustered with different European haplotypes reported in Italian, Spanish



and German mice (ITACittaduale, SPAAvinyonet, SCHUNT, and SCHEST).

Finally, two more D-loop haplogroups containing Brazilian mice were identified: the first of them clustered a significant number of well-known laboratory mouse strains such as C57BL/6 and BALB/c with Brazilian mice from Manaus. The second haplogroup contained Brazilian mice from Belém and European mice from Scotland and Germany (SIT and BRAG; SCOTShetland, BRAD, and GERHolstein).

Overall, these results revealed the diverse maternal lineage present in the Brazilian mouse populations. Considering that even though these animals shared the characteristic of being carriers of the *Irgb2-b1*<sub>PWK</sub> allele, they have multiple maternal origins. This gives stronger support to the hypothesis of a selection on standing variation due to a fitness advantage like the resistance against highly virulent local *T. gondii* strains.

#### **2.4.4. *Irgb2-b1* expression in wild mice is allele dependent**

Previous studies have reported large differences in protein expression levels in the *Irgb2-b1* gene of laboratory mice versus wild-caught and wild-derived mice<sup>2,6,8,34</sup>. High levels of expression of the *Irgb2-b1* protein can be found in mice resistant to infection by virulent *T. gondii* strains, like the CIM and CAST mice.

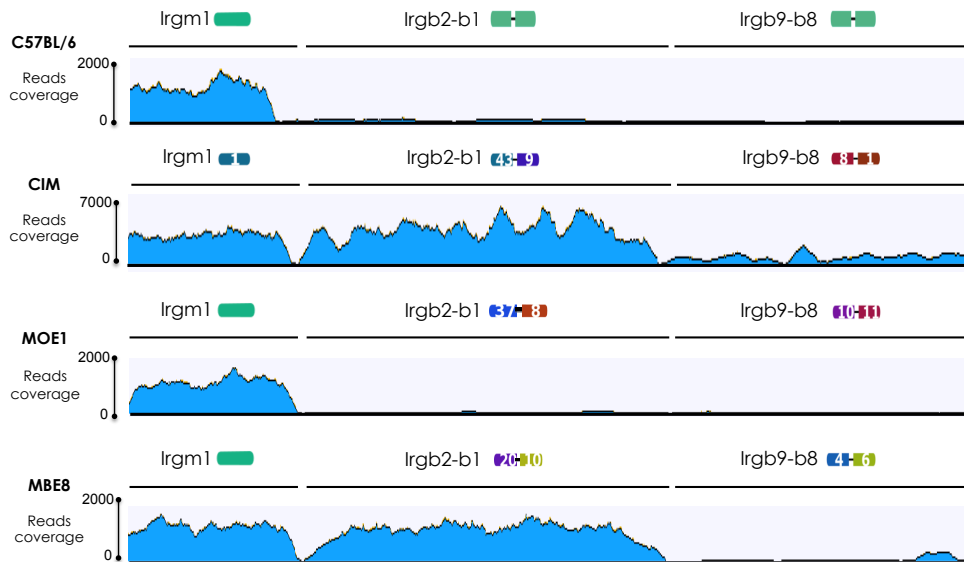
Lilue, 2012 found that even though there are no differences between the C57BL/6 and CIM or CAST mice in important domains of the *Irgb2-b1* ORF, or in regions in the promoter such as the Interferon Stimulated Response Element (ISRE, AGTTTCNNTTTCNY or TTTCRSTTT) or the Gamma activated sequences (GAS, TTCNNNGAA), differences in other regions can be found. Between the promoters of *Irgb2-b1*<sub>CIM</sub> and *Irgb2-b1*<sub>C57BL/6</sub>, there is

approximately a 10% sequence difference, however, genome recombination in the promoter of *Irgb2-b1*<sub>CAST</sub> similar to the ones observed in *Irgb6*<sub>C57BL/6</sub> enhances its expression level<sup>8</sup>.

As previously mentioned, almost 90% of the Brazilian mice analyzed for this study are carriers of the *Irgb2-b1*<sub>PWK</sub>. We compared the *Irgb2-b1* mRNA expression level of the Brazilian cell line MBE8 with the expression levels of other cell lines with different alleles for the *Irgb2-b1* gene. Transcriptomes from the C57BL/6, CIM, MOE1, and MBE8 were mapped against an artificial chromosome that contained the ORF from the *Irgm1*, *Irgb2-b1* and *Irgb9-b8* genes (Figure 2.11). We confirmed that the *Irgb2-b1*<sub>C57BL/6</sub> allele is associated with a very low endogenous expression, whereas the *Irgb2-b1*<sub>CIM</sub> is expressed at a very high level. However, this does not mean that all wild-derived/wild-caught mice always display a high expression level of the *Irgb2-b1* gene<sup>6</sup>. Expression levels seem to be directly associated with the allele carried. For example, the Portuguese mouse (MOE1) is a carrier of the *Irgb2-b1*<sub>SIT/SIN</sub> that displays *Irgb2-b1* expression levels that are comparable to the ones found in laboratory mice. On the contrary, a Brazilian mouse (MBE8), carrier of the *Irgb2-b1*<sub>PWK</sub>, has expression levels that are closer to the ones found in the CIM strain.

We also looked at the expression of other IRGs in these samples; we assessed the expression of *Irgm1* and another tandem IRG, the *Irgb9-b8* (Figure 2.11). We confirmed that unlike *Irgb2-b1*, the *Irgm1* always well-expressed irrespectively of the allele carried in this gene<sup>8</sup>. In other words, all mice analyzed have the same expression level in this highly conserved gene. However, for the *Irgb9-b8* we could not find any expression regardless of the mouse genetic background. Very low expression can be seen in the CIM mouse. It is not clear whether this is due to the higher coverage in this sample, reflecting low but real expression, or to misallocate reads due to the high sequence similarity among the tandem IRG genes.

This data confirms what has been previously reported by Lilue, 2012, where a reconstruction of *Irgb9-b8* was very difficult due to the reduced number of reads obtained.



**Figure 2.11.** Expression levels of the *Irgb2-b1* gene are allele dependent. Transcriptomes from IFN $\gamma$  induced cell lines from the C57BL/6, CIM, MOE1 and MBE8 mice were mapped against an artificial chromosome that contained the ORF from the *Irgb1*, *Irgb2-b1* and *Irgb9-b8* genes. Depth of coverage for each line is displayed in terms of number of reads. Note: The CIM transcriptome was sequenced with a coverage three times higher than the other mouse strains.

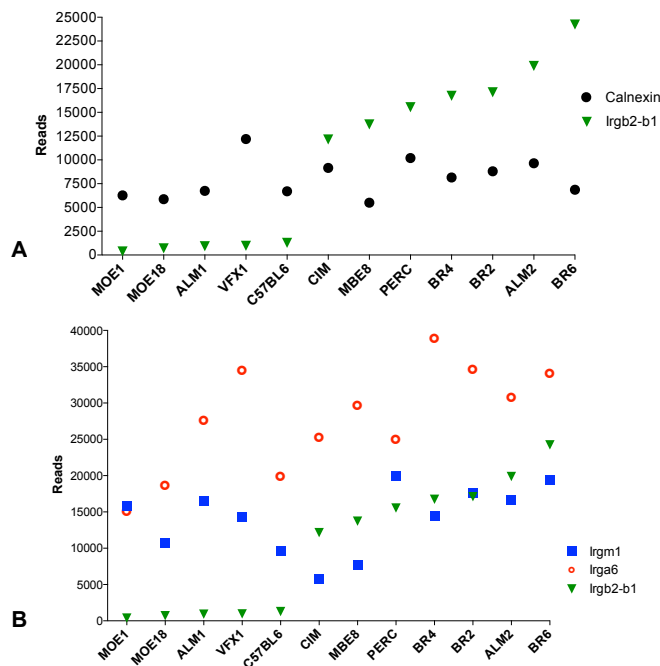
#### 2.4.5. *Irgb2-b1* expression levels are independent of other IRGs

To check whether the expression levels of the *Irgb2-b1* gene might also be influenced by the high or low expression of other IRG genes, we analyzed the transcriptome of multiple wild-caught and wild-derived mice. The C57BL/6 mouse (laboratory mouse) was used as our control.

An initial comparison between the expression levels of the *Irgb2-b1* gene and the housekeeping gene *Calnexin* (*Canx*) was done (**Figure 2.12, panel A**). We wanted to confirm whether the mouse strains analyzed had

comparable gene expression patterns in non-immune related genes. All mouse strains display comparable levels of Calnexin (from 6000 to 12000 reads). However, they are distinct in their *Irgb2-b1* expression, having samples with expression levels ranging from lower than 2500 reads to higher than 12000 reads.

Additionally, we compared the expression levels of the *Irgb2-b1* gene with the ones from two IRG genes that are crucial for the control of *T. gondii* strains, the *Irgm1* and the *Irga6* (Figure 2.12, panel B). Even though there is variation in the expression levels of both *Irgm1* and *Irga6* genes among the mice analyzed, we could not detect a direct correlation between the expression level of neither the regulator *Irgm1* nor the effector *Irga6*.



**Figure 2.12.** *Irgb2-b1* expression levels are not dependent on other IRG genes. Transcriptomes from multiple mouse cell lines were compared on their expression levels of A) the housekeeping gene Calnexin (Canx) and tandem *Irgb2-b1* gene and B) the regulator

Irgm1 gene, the effector Irga6 gene and tandem Irgb2-b1 gene. Mouse lines were organized from the lowest to the highest Irgb2-b1 gene expression.

#### 2.4.6. Polymorphisms in the IRGB2<sub>PWK</sub> protein subunit.

The putative H4 and  $\alpha$ D structural domains of the IRGB2<sub>CIM</sub> subunit are highly polymorphic and show recent divergent selection<sup>2</sup>. It has also been shown that interactions between these regions and virulence factors of the parasite such as the ROP5 pseudokinase are quite specific<sup>29</sup>. We wanted to know whether the IRGB2<sub>PWK</sub> subunit shows signs of possible selection imposed by the interaction with the ROP5 pseudokinases of virulent *T. gondii* strains. To address this question, we first aligned the polymorphic residues present in the C57BL/6, CIM and PWK strains (**Figure 2.13, panel A**). Residues from the reference mouse strains C57BL/6 were labeled in yellow, the ones from the CIM strain in orange and the ones uniquely found in the PWK strain in blue.

The IRGB2<sub>PWK</sub> protein subunit surprisingly displays an overall residue composition much closer to the IRGB2<sub>C57BL/6</sub> than to the IRGB2<sub>CIM</sub>: the  $\alpha$ D domain of IRGB2<sub>C57BL/6</sub> and IRGB2<sub>PWK</sub> is identical, whereas the IRGB2<sub>CIM</sub> has a remarkably different composition. This finding points to the H4 domain as responsible for the differences between the susceptible laboratory mouse C57BL/6 and the resistant PWK mouse. The H4 domain between the three strains has eight polymorphic sites. We found that between IRGB2<sub>C57BL/6</sub> and IRGB2<sub>PWK</sub> there are six polymorphic residues: from those, two corresponded to residues also found in the IRGB2<sub>CIM</sub> (positions 202 and 206) and four distinct from the IRGB2<sub>PWK</sub> (positions 198, 201, 213 and 214).

To know whether these residue differences represent significant changes in the protein characteristics of the IRGB2 subunit, we performed analyses of polarity and hydrophobicity. In terms of polarity (**Figure 2.13, panel B**), two residues (positions 198 and 213) displayed the biggest differences between

the IRGB2<sub>C57BL/6</sub> and IRGB2<sub>PWK</sub>. In position 198, IRGB2<sub>C57BL/6</sub> has the non-polar residue Isoleucine (I), whereas, in that same position, IRGB2<sub>PWK</sub> displays the polar (uncharged) residue Threonine (T).

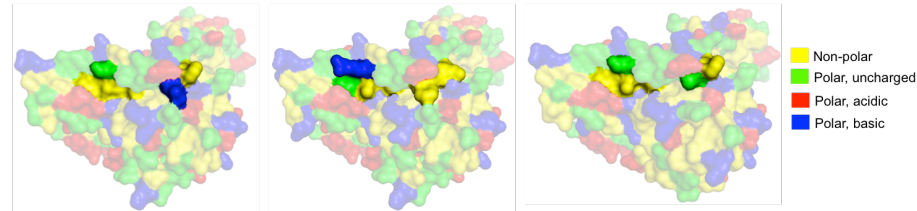
A similar situation happens in the 213 position, where the residue found in the Irgb2<sub>C57BL/6</sub> is a large polar (basic) Arginine and in the Irgb2<sub>PWK</sub> a large non-polar aromatic Tryptophan residue.

A hydrophobicity analysis of the Irgb2 subunit from C57BL/6, CIM and PWK strains was also done (**Figure 2.13, panel C**). We found differences in hydrophobicity for the six polymorphic residues located in the H4 domain (positions 198, 201, 202, 206, 213, 214). However, those differences were not remarkable. The most commonly seen events were changes from polar uncharged to polar basic residues. Only the residue located in position 213 showed the most striking differences, having for the C57BL/6 strain a hydrophilic Arginine (R) residue, an intermediate Cysteine (C) residue for the CIM strain and a hydrophobic Tryptophan (W) residue for the PWK strain.

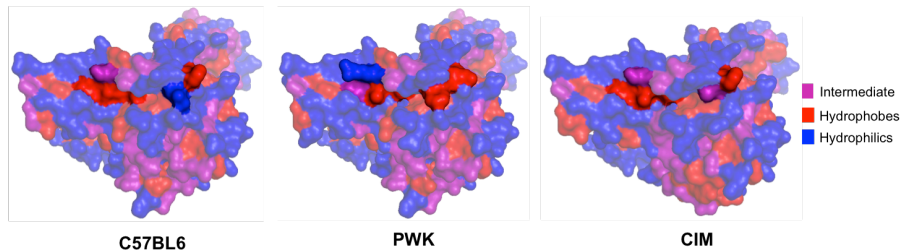
**A**

Residue	2	18	28	33	40	60	72	84	86	91	92	94	95	97	98	99	102	104	105	106	108	110	115	116	122	125	126	127	138	151	154	156	158	175	180	183	187	188	198	201	202	204	206	213	214	215	226	280	292	315	330	343	357	382
C57BL6	G	T	Q	S	D	V	A	V	D	A	A	S	T	V	V	H	T	R	T	P	T	T	V	T	S	S	T	A	E	I	K	S	I	R	S	E	L	C	I	S	I	I	L	R	F	Q	F	I	G	N	L	F	K	A
CIM	D	T	R	F	N	A	A	T	A	V	T	P	V	V	V	Y	I	K	K	S	P	A	A	I	A	F	H	H	K	G	K	G	V	R	L	K	L	V	I	S	L	N	I	C	F	H	V	M	G	D	V	L	N	V
PWK	G	K	Q	S	N	A	T	V	D	A	A	S	T	I	I	H	T	R	T	P	T	T	V	T	G	S	T	A	E	I	R	S	I	K	S	E	R	V	T	R	L	I	I	W	I	Q	F	I	E	D	L	F	K	V

**B**



**C**



**Figure 2.13.** Polymorphism for the IRGB2 protein subunit in three mouse lines. A) Polymorphic sites in the IRGB2 protein subunit for the C57BL/6, CIM and PWK mice. Residues from the reference strain C57BL/6 are labeled in yellow. Unique residues from the CIM strain are labeled in orange and from the PWK strain in blue. Putative interface regions with the *T. gondii* ROP5 protein ( $\alpha$ D and H4 domains) are delimited by bold black squares. B) 3D-structure of the Irgb2 subunit from C57BL/6, CIM and PWK. Protein modeling was done with the Phyre2 software. Residues in brighter colors corresponded to the six polymorphic sites located in the H4 domain among the three mouse strains. Protein surface was labeled based on the residue's polarity using the PyMol software. C) Protein surface was labeled based on the residue's hydrophobicity using the PyMol software

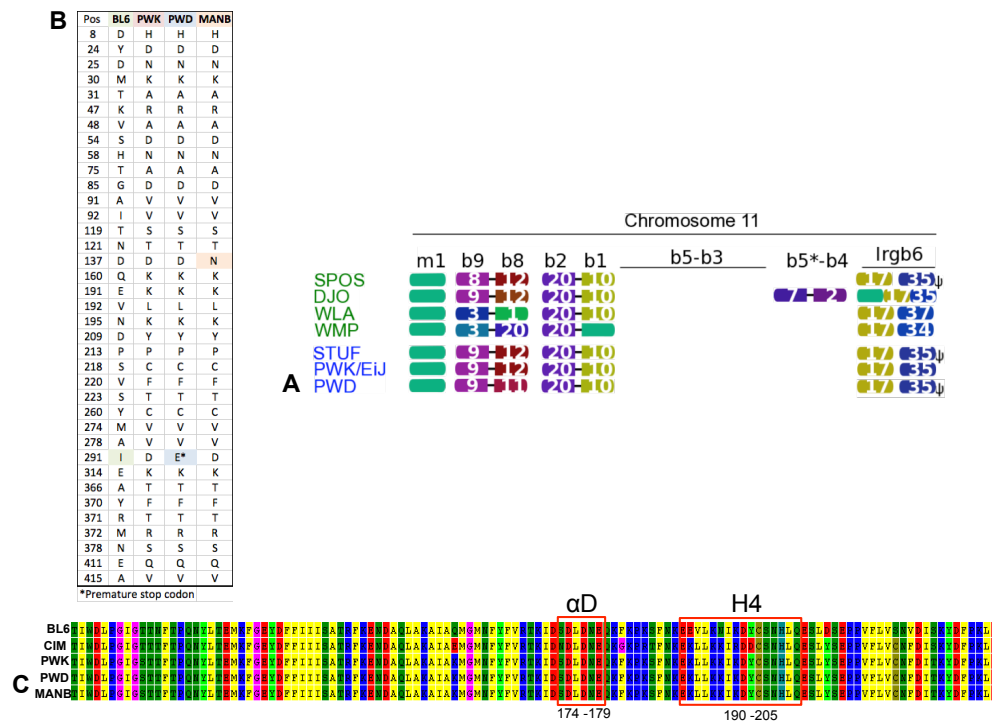
#### 2.4.7. Polymorphisms in the IRGB6 protein of mice carriers of the *Irgb2*<sub>PWK</sub> allele.

The IRGB6 protein is one of the most important IRG proteins as a pioneer in the loading of the PVM. Recently, Dr. Masahiro Yamamoto found that IRGB6-deficiency in mutant mice largely inhibits loading of the other effector IRG proteins, and the mouse is susceptible to infection with *T. gondii* *in vivo*<sup>35</sup>. We wanted to know if mice that are carriers of the *Irgb2*-b1<sub>PWK</sub> could also have differences in the alleles of the IRGB6 protein. A modified color block with the IRG proteins in the chromosome 11 (large cluster) from 7 representative mouse strains that are carriers of the *Irgb2*-b1<sub>PWK</sub> allele is shown (*M. m. domesticus* in green and *M. m. musculus* in blue) (**Figure 2.14, panel A**). As previously reported by Lilue, 2012, wild-derived mice normally have more than one copy of the *Irgb6* gene in each haploid genome. In this case, all animals have a conserved copy (Allele 17 - green), however, they differ in the composition and functional characteristics of the second copy. For example, WLA, WMP, and PWK strains have fully functional copies of the second *Irgb6* allele (34, 35 and 37 - blue), whereas SPOS, STUF, and PWD have truncated versions that are not functional (35 – blue).

A residue analysis of the second copy (blue allele) of the IRGB6 protein between the C57BL/6, PWK, PWD and the Brazilian MANB strain was done (**Figure 2.14, panel B**). A total of 37 polymorphic residues were identified, having the residue in position 291 as the one that shows the biggest differences between the strains analyzed. This residue is either an Isoleucine (I) like in C57BL/6, an Aspartic Acid (D) like in PWK and MANB or a STOP codon in PWD (responsible for the truncated version). Interestingly, the Brazilian mouse strain MANB has, in position 137, a polymorphic residue that is unique to this strain. Instead of an Aspartic Acid (D) as in C57BL/6, PWK, and PWD, the Brazilian mouse has an Asparagine (N).



Finally, we looked for differences in the putative H4 and  $\alpha$ D structural domains in the second copy (blue allele) of the IRGB6 protein (**Figure 2.14, panel C**). No differences were found neither in the  $\alpha$ D nor in the H4 domain of the *Irgb2*<sub>PWK</sub> carriers (PWK, PWD, and MANB). However, we found differences in other regions of the protein as previously mentioned. These results might go in line with what has been previously reported<sup>8</sup> for the IRGB6 protein, where residues are mostly in balancing selection, however, those with positive selection are not located in the same regions as in the IRGB2-B1 protein.



**Figure 2.14.** Polymorphisms in the IRGB6 protein. A) IRG alleles in chromosome 11 from wild-derived mouse lines carriers of the *Irgb2*-b1<sub>PWK</sub> allele (Modified from Müller, 2015). B) Polymorphic positions for the IRGB6 protein in the C57BL/6, PWK, PWD and MANB mouse strains. Residues uniquely found in the C57BL/6 are highlighted in green, in the PWD strain in blue and in the MANB strain in orange. Premature stop codons are marked with an

asterisk (\*). C) Alignment of the Irgb6 protein sequence from residue 110 to residue 230 in five different mouse strains. Residues from 174-179 correspond to the  $\alpha$ D helix. Residues from 190-205 correspond to the H4 helix.

## 2.5. Discussion

The Irgb2-b1 protein has a major role in the control of highly virulent *T. gondii* strains in wild-derived mice<sup>2</sup>. We have shown that Brazilian wild and wild-derived mouse populations have a remarkable homogeneity for the Irgb2-b1 gene, where a single allele called Irgb2-b1<sub>PWK</sub> is present in 90% of the mouse population analyzed (homozygous and heterozygous animals). *M. m. domesticus* colonized the current South American territory for the first time when the first human European settlers arrived in the 16<sup>th</sup> century<sup>16</sup> and it was presumably the first time house mice faced infections with highly virulent South American *T. gondii* strains. As mentioned by Charles Darwin, the first members of a species that arrive in a new territory might rapidly adapt to local conditions, this rapid evolutionary adaptation can help species counter stressful conditions, such as the exposure to new pathogens<sup>36</sup>. Furthermore, the first immigrants to arrive, adapting rapidly, have a selective advantage over subsequent, not yet adapted, immigrants. In fact, parasites and other pathogens are undoubtedly important drivers of adaptive evolution in their hosts<sup>37</sup>.

Genetic adaptation to pathogens can lead to allele homogeneity in natural populations when an advantageous allele has an increase in frequency<sup>38</sup>. Selection pressure imposed by pathogens usually acts on genes associated with immunological responses, such as in the case of human gene variants linked to malaria resistance (Hb, DARC, G6PD, CD36, GYPA, etc.), antiviral responses (type-III IFNs), NF- $\kappa$ B signaling (TLR10-TLR1-TLR6 cluster), MHC genes (class 1 and 2), etc<sup>38-42</sup>. Even though our data show a high allele frequency for the Irgb2-b1<sub>PWK</sub> allele in Brazilian mice, we also found genetic diversity in nearby genes, which together with data from non-coding

regions in the *Irgb2-b1* gene of *Irgb2-b1*<sub>PWK</sub> allele carriers indicates the presence of multiple haplotypes. Some of these haplotypes are for the time being unique in the limited Brazilian population tested, but others are shared between Brazilian and European mice. Further analyses of linkage disequilibrium signatures in regions around the *Irgb2-b1* gene are needed to confirm local recombination.

Our data also show that European mice are carriers of the *Irgb2-b1*<sub>PWK</sub> allele but the allele does not appear to be at high frequency. A remarkable difference in allele frequency between *M. m. domesticus* in Brazil versus their European counterparts is a possible indicator of genetic diversity reduction by demographic perturbations (e.g. bottlenecks, founder effect, etc.) or host-parasite co-evolutionary events (e.g. balancing selection, selective sweeps, etc.). Although our statistical data are not enough to fully support either one hypothesis or another, maternal phylogenetic data from mitochondrial sequences and the high genetic diversity present in nearby genes are consistent with a selective sweep, most likely a soft sweep as a result of local adaptation from standing genetic variation.

Approximately 400 to 600 generations of house mice have been evolving in the Americas since they arrived in association with European colonization (in the wild, mouse females breed seasonally with one or two litters per year if resources are available<sup>16,43</sup>), although this number of generations may seem small, South American house mice have spread rapidly and shown evidence of local adaptation<sup>44</sup>. This relatively short evolutionary time is compatible with a soft sweep from variation already present in the European immigrant population as being responsible for the high frequency of the *Irgb2-b1*<sub>PWK</sub> allele, not only because they occur faster than hard sweeps but also because they are commonly found in natural populations where they promote a rapid adaptation to a new territory<sup>45–47</sup>. Additionally, in contrast to a hard selective sweep where a single adaptive haplotype raises to a high

frequency within the population, in a soft selective sweep multiple adaptive haplotypes sweep through the populations simultaneously, which leads to persistence of genetic variation nearby the adaptive site<sup>47</sup>, which is also compatible to what was found in genes surrounding the *Irgb2-b1* locus in Brazilian mice.

We have also found that the *Irgb2-b1*<sub>PWK</sub> is one of the most highly expressed *Irgb2-b1* alleles and its expression is independent of other IRG gene expression. This data is consistent with previous studies showing high expression of the *Irgb2-b1*<sub>PWK</sub> allele<sup>2,6</sup>. Even though a direct correlation between expression levels of the *Irgb2-b1* protein with a high resistance level to virulent *T. gondii* has not been directly proven, we have seen low expression levels of the *Irgb2-b1* protein in mouse strains fully susceptible to virulent *T. gondii* strains (e.g. C57BL/6)<sup>2,6,29</sup>.

Finally, we found that polymorphisms in regions of the IRGB2 protein that directly interact with virulent ROP5 proteins of *T. gondii* such as the H4 and  $\alpha$ D structural domains are distinct in the IRGB2<sub>PWK</sub> subunit from those found in the IRGB2<sub>BL6</sub> (susceptible mouse strain) and IRGB2<sub>CIM</sub> (resistance mouse strain). This may be an important finding because when a particular host or parasite protein interacts, events such as selective sweep co-evolution can occur through alternating changes to functionally important segments of these proteins<sup>45</sup>. However, further studies in protein-protein interactions are needed to evaluate if for example binding affinity between virulent ROP5 protein isoforms and the IRGB2<sub>PWK</sub> subunit changes.

In conclusion, our results suggest selection on standing variation in Brazilian mice from European populations, possibly associated with fitness advantage (resistance) in Brazilian mice against highly virulent *T. gondii* strains. However, the complexity of these host-parasite interactions makes it difficult to precisely measure their evolutionary impact, therefore complementary

variability analysis at a whole-genome scale for both host and parasite are needed. Functional studies on the impact of the *Irgb2-b1<sub>PWK</sub>* allele for the control of virulent *T. gondii* strains are going to be developed in the coming chapters of this thesis.

## **2.6. Acknowledgments**

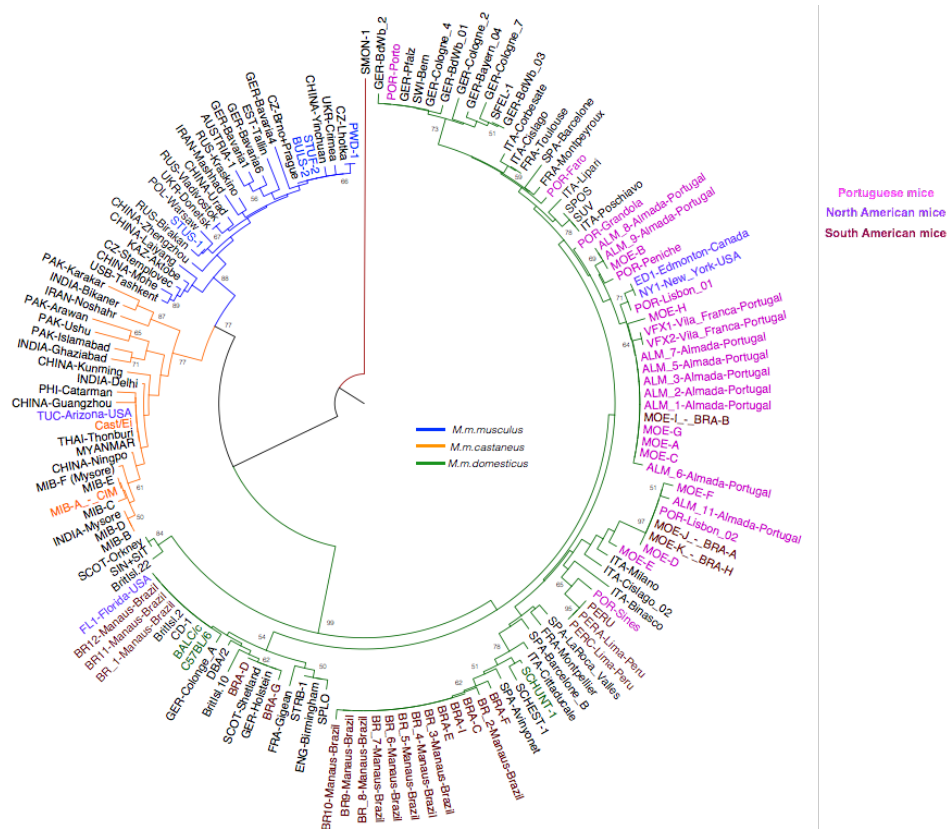
I would like to thank the current and former members of the Howard and Teixeira's laboratories who constantly gave me their feedback about this project. Special thanks to Dr. Michael Nachman and Dr. Ellen Robey from University of California Berkeley, who respectively provided us with the Manaus mouse samples and the laboratory conditions to work with them. I would also like to acknowledge the technical support provided by the IGC Animal Facility, supported by the research infrastructure Congento, project LISBOA-01-0145-FEDER-022170, and the Genomics facility. Joao Costa from the Genomics facility helped us with Nanopore sequencing and the corresponding reads analysis. This work was supported by central funds of the Instituto Gulbenkian de Ciência, by the Sonderforschungsbereiche 670 and 680 and Schwerpunkt 1399 of the Deutsche Forschungsgemeinschaft.

## 2.7. Supplementary material

**Supplemental Table 2.1.** Mice sampled in Brazil and Portugal by Urs Benedikt Muller and used for multiple analysis in this study. Modified from Müller, 2015.

Name	Sample location	GPS Coordinates	maternal DNA haplotype
BH-1	Belo Horizonte, Minas Gerais, Brazil	19°53'38.8"S 44°01'24.6"W	BRA-A=MOE-J, domesticus-like
BH-2	Belo Horizonte, Minas Gerais, Brazil	19°53'38.8"S 44°01'24.6"W	BRA-A=MOE-J, domesticus-like
BH-3	Belo Horizonte, Minas Gerais, Brazil	19°53'38.8"S 44°01'24.6"W	BRA-A=MOE-J, domesticus-like
BH-4	Belo Horizonte, Minas Gerais, Brazil	19°53'15.0"S 43°49'55.9"W	BRA-A=MOE-J, domesticus-like
BH-5	Belo Horizonte, Minas Gerais, Brazil	19°53'38.8"S 44°01'24.6"W	BRA-A=MOE-J, domesticus-like
BH-6	Sabará, Minas Gerais, Brazil	19°53'15.0"S 43°49'55.9"W	BRA-A=MOE-J, domesticus-like
BH-7	Sabará, Minas Gerais, Brazil	19°53'15.0"S 43°49'55.9"W	BRA-A=MOE-J, domesticus-like
BH-8	Sabará, Minas Gerais, Brazil	19°53'19.4"S 43°50'18.4"W	BRA-A=MOE-J, domesticus-like
BH-9	Sabará, Minas Gerais, Brazil	19°53'19.4"S 43°50'18.4"W	BRA-H=MOE-K, domesticus-like
BH-10	Ouro Preto, Minas Gerais, Brazil	20°22'44.9"S 43°30'40.9"W	BRA-B=MOE-I, domesticus-like
BH-11	Ouro Preto, Minas Gerais, Brazil	20°22'44.9"S 43°30'40.9"W	BRA-B=MOE-I, domesticus-like
BH-12	Ouro Preto, Minas Gerais, Brazil	20°22'44.9"S 43°30'40.9"W	BRA-B=MOE-I, domesticus-like
BH-13	Ouro Preto, Minas Gerais, Brazil	20°22'44.9"S 43°30'40.9"W	BRA-B=MOE-I, domesticus-like
BH-14	Ouro Preto, Minas Gerais, Brazil	20°22'44.9"S 43°30'40.9"W	BRA-B=MOE-I, domesticus-like
BEL-1	Belém, Pará, Brazil	1°25'42.0"S 48°27'09.3"W	BRA-B=MOE-I, domesticus-like
BEL-2	Belém, Pará, Brazil	1°25'42.0"S 48°27'09.3"W	BRA-B=MOE-I, domesticus-like
BEL-3	Belém, Pará, Brazil	1°25'42.0"S 48°27'09.3"W	BRA-C, domesticus-like
BEL-4	Belém, Pará, Brazil	1°25'42.0"S 48°27'09.3"W	BRA-D, domesticus-like
BEL-5	Belém, Pará, Brazil	1°25'42.0"S 48°27'09.3"W	BRA-D, domesticus-like
BEL-6	Belém, Pará, Brazil	1°25'42.0"S 48°27'09.3"W	BRA-B=MOE-I, domesticus-like
BEL-7	Belém, Pará, Brazil	1°25'42.0"S 48°27'09.3"W	BRA-B=MOE-I, domesticus-like
BEL-8	Belém, Pará, Brazil	1°23'43.6"S 48°27'05.8"W	BRA-B=MOE-I, domesticus-like
BEL-9	Belém, Pará, Brazil	1°23'43.6"S 48°27'05.8"W	BRA-B=MOE-I, domesticus-like
BEL-10	Belém, Pará, Brazil	1°23'43.6"S 48°27'05.8"W	BRA-I, domesticus-like
BEL-11	Belém, Pará, Brazil	1°23'43.6"S 48°27'05.8"W	BRA-D, domesticus-like
BEL-12	Belém, Pará, Brazil	1°23'43.6"S 48°27'05.8"W	BRA-B=MOE-I, domesticus-like
BEL-13	Belém, Pará, Brazil	1°23'43.6"S 48°27'05.8"W	BRA-E, domesticus-like
BEL-15	Belém, Pará, Brazil	1°23'43.6"S 48°27'05.8"W	BRA-C, domesticus-like
BEL-16	Belém, Pará, Brazil	1°26'47.4"S 48°28'07.9"W	BRA-F, domesticus-like
BEL-17	Belém, Pará, Brazil	1°26'47.4"S 48°28'07.9"W	BRA-F, domesticus-like
BEL-18	Belém, Pará, Brazil	1°26'47.4"S 48°28'07.9"W	BRA-F, domesticus-like
BEL-19	Belém, Pará, Brazil	1°26'47.4"S 48°28'07.9"W	BRA-F, domesticus-like
BEL-20	Belém, Pará, Brazil	1°26'47.4"S 48°28'07.9"W	BRA-F, domesticus-like
BEL-21	Belém, Pará, Brazil	1°26'47.4"S 48°28'07.9"W	BRA-F, domesticus-like
BEL-22	Belém, Pará, Brazil	1°26'41.9"S 48°28'06.2"W	BRA-F, domesticus-like
BEL-23	Belém, Pará, Brazil	1°26'41.9"S 48°28'06.2"W	BRA-F, domesticus-like
BEL-24	Belém, Pará, Brazil	1°26'41.9"S 48°28'06.2"W	BRA-F, domesticus-like
MOE-1	Oeiras, Portugal	38°42'13.7"N 9°19'09.3"W	MOE-A, domesticus-like
MOE-2	Oeiras, Portugal	38°42'13.7"N 9°19'09.3"W	MOE-A, domesticus-like

MOE-3	Oeiras, Portugal	38°42'13.7"N 9°19'09.3"W	MOE-A, domesticus-like
MOE-4	Oeiras, Portugal	38°42'13.7"N 9°19'09.3"W	MOE-A, domesticus-like
MOE-5	Oeiras, Portugal	38°42'13.7"N 9°19'09.3"W	MOE-A, domesticus-like
MOE-6	Cascais, Portugal	38°42'50.2"N 9°27'39.6"W	MOE-B, domesticus-like
MOE-7	Cascais, Portugal	38°42'50.2"N 9°27'39.6"W	MOE-B, domesticus-like
MOE-8	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-9	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-10	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-11	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-12	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-13	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-14	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-15	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-C, domesticus-like
MOE-16	Cascais, Portugal	38°43'17.4"N 9°27'38.8"W	MOE-I=BRA-B, domesticus-like
MOE-17	Manique, Portugal	38°43'21.8"N 9°21'53.1"W	MOE-D, domesticus-like
MOE-18	Manique, Portugal	38°43'21.8"N 9°21'53.1"W	MOE-E, domesticus-like
MOE-19	Manique, Portugal	38°43'21.8"N 9°21'53.1"W	MOE-F, domesticus-like
MOE-20	Manique, Portugal	38°43'21.8"N 9°21'53.1"W	MOE-F, domesticus-like
MOE-21	Manique, Portugal	38°43'21.8"N 9°21'53.1"W	MOE-J=BRA-A, domesticus-like
MOE-22	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-23	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-24	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-26	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-27	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-28	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-29	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-30	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-31	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-32	Barcarena, Portugal	38°43'33.1"N 9°16'53.0"W	MOE-I=BRA-B, domesticus-like
MOE-34	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-I=BRA-B, domesticus-like
MOE-36	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-G, domesticus-like
MOE-37	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-K=BRA-H, domesticus-like
MOE-38	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-H, domesticus-like
MOE-39	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-H, domesticus-like
MOE-40	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-H, domesticus-like
MOE-41	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-K=BRA-H, domesticus-like
MOE-42	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-I=BRA-B, domesticus-like
MOE-43	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-G, domesticus-like
MOE-44	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-G, domesticus-like
MOE-45	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-G, domesticus-like
MOE-46	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-H, domesticus-like
MOE-47	Alcabideche, Portugal	38°43'52.2"N 9°25'24.6"W	MOE-J=BRA-A, domesticus-like
MOE-50	Cascais, Portugal	38°42'50.2"N 9°27'39.6"W	MOE-K=BRA-H, domesticus-like



**Supplemental Figure 2.1.** Maternal phylogeny of wild and wild-derived *Mus musculus* species in the world. Unrooted phylogenetic tree based on a multiple sequence alignment of 136 unique 867 bp D-loop fragments from wild and wild-derived mice from the three *Mus musculus* subspecies. A Jukes-Cantor genetic distance model along with a neighbor-joining building method with outgroup (SMON-1) were computed in the software Geneious® 9.1.5 (Biomatters, Auckland, New Zealand). A bootstrap resampling method with 5000 replicates was run. Branch support is shown if >50. Scale=0.4 % differences. The D-loop sequences were either sequenced from genomic DNA obtained in this study and previous studies from the Howard lab or retrieved from published literature by Müller, 2015<sup>25–27,33</sup>. *Mus musculus* subspecies are indicated as *M. m. castaneus* (orange), *M. m. musculus* (blue) or *M. m. domesticus* (green). Portuguese mice are indicated in magenta, North American mice in purple and South American mice in brown.



**Supplemental Table 2.2.** Transcriptome data correction from different cell lines.

Cell line name	Mapped transcriptome size GB (.BAM)	Correction factor	Uncorrected data (number of reads)				Corrected data (number of reads)			
			Calnexin	Irgm1	Irga6	Irgb2-b1	Calnexin	Irgm1	Irga6	Irgb2-b1
C57BL/6	3,96	2,44	15404	22003	45408	2954	6302	9001	18576	1208
VFX1	1,66	1,02	12189	14305	34329	968	11895	13960	33502	945
ALM1	1,62	1,00	6744	16502	27454	922	6744	16502	27454	922
MOE1	3,67	2,27	13156	33346	31559	792	5807	14719	13931	350
MOE18	3,89	2,40	12928	23459	40756	1555	5384	9770	16973	648
ALM2	1,65	1,02	9644	16708	30637	19884	9469	16404	30080	19522
MBE8	3,59	2,22	10998	15460	59176	27478	4963	6976	26703	12400
BR2	1,62	1,00	8810	17580	34581	17107	8810	17580	34581	17107
BR4	1,73	1,07	8151	14480	38801	16737	7633	13559	36334	15673
BR6	1,72	1,06	6865	19442	34011	24240	6466	18312	32034	22831
PERC	1,77	1,09	10196	19961	24927	15537	9332	18269	22815	14220
CIM	16,58	10,23	87003	55320	238965	115423	8501	5405	23349	11278

## Bibliography

1. Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat. Genet.* **32**, 569–577 (2002).
2. Lilue, J., Müller, U. B., Steinfeldt, T. & Howard, J. C. Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. *Elife* **2013**, 1–21 (2013).
3. Pittman, K. J. & Knoll, L. J. Long-Term Relationships: the Complicated Interplay between the Host and the Developmental Stages of *Toxoplasma gondii* during Acute and Chronic Infections. *Microbiol. Mol. Biol. Rev.* **79**, 387–401 (2015).
4. Howard, J. C., Hunn, J. P. & Steinfeldt, T. The IRG protein-based resistance mechanism in mice and its relation to virulence in *Toxoplasma gondii*. *Curr. Opin. Microbiol.* **14**, 414–421 (2011).
5. Bekpen, C. *et al.* The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. *Genome Biol.* **6**, R92 (2005).
6. Müller, U. B. Polymorphism in the IRG resistance system determines virulence of *Toxoplasma gondii* in mice. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2015).
7. Müller, U. B. & Howard, J. C. The impact of *Toxoplasma gondii* on the mammalian genome. *Curr. Opin. Microbiol.* **32**, 19–25 (2016).
8. Lilue, J. Haplotypic polymorphism of the IRG protein family mediates resistance of mice against virulent strains of *Toxoplasma gondii*. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2012).
9. Hassan, M. A., Olijnik, A.-A., Frickel, E.-M. & Saeij, J. P. Clonal and atypical *Toxoplasma* strain differences in virulence vary with mouse sub-species. *Int. J. Parasitol.* **49**, 63–70 (2019).
10. Laurie, C. C. *et al.* Linkage disequilibrium in wild mice. *PLoS Genet.* **3**, 1487–1495 (2007).
11. Abolins, S. *et al.* The ecology of immune state in a wild mammal , *Mus musculus domesticus*. *PLoS Biol.* 1–24 (2018).
12. Cucchi, T., Vigne, J. & Auffray, J. First occurrence of the house mouse ( *Mus musculus domesticus* Schwarz & Schwarz , 1943 ) in the Western Mediterranean : a zooarchaeological revision of subfossil occurrences. *Biol. J. Linn. Soc.* **84**, 429–445 (2005).
13. Valenzuela-lamas, S., Baylac, M., Cucchi, T. & Vigne, J. House mouse dispersal in Iron Age Spain : a geometric morphometrics appraisal. *Biol. J. Linn. Soc.* 483–497

- (2011).
14. Bourso, P., Auffray, J. C., Britton-Davidian, J. & Bonhomme, F. The Evolution of House Mice. *Annu. Rev. Ecol. Syst.* **24**, 119–152 (1993).
  15. Gardner, M., Kozak, C. & O'Brien, S. The Lake Casitas wild mouse: evolving genetic resistance to retroviral disease. *Trends Genet.* **7**, (1991).
  16. Phifer-Rixey, M. & Nachman, M. W. Insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife* **4**, 1–13 (2015).
  17. Bertranpetit, E. *et al.* Phylogeography of *Toxoplasma gondii* points to a South American origin. *Infect. Genet. Evol.* **48**, 150–155 (2017).
  18. Behnke, M. S. *et al.* Rhopty Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of *Toxoplasma gondii*. *PLOS Genet.* **11**, e1005434 (2015).
  19. Jensen, K. D. C. *et al.* *Toxoplasma gondii* Superinfection and Virulence during Secondary Infection Correlate with the Exact ROP5 / ROP18 Allelic Combination. *MBio* **6**, 1–15 (2015).
  20. Antony, V. B., Owen, C. L. & Hadley, K. J. Pleural mesothelial cells stimulated by asbestos release chemotactic activity for neutrophils in vitro. *Am Rev Respir Dis* **139**, 199–206 (1989).
  21. Alvarez, C. *et al.* in *Toxoplasma gondii: Methods and Protocols, Methods in Molecular Biology* (ed. Tonkin, C. J.) 371–409 (Springer US, 2020). doi:10.1007/978-1-4939-9857-9\_20
  22. Incella. ScreenFect® A Transfection Reagent Protocol. 1–5 (2016).
  23. Southern, P. J. & Berg, P. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* **1** **4**, 327–341 (1982).
  24. Kivisild, T. Maternal ancestry and population history from whole mitochondrial genomes. *Investig. Genet.* 1–10 (2015). doi:10.1186/s13323-015-0022-2
  25. Bonhomme, F. *et al.* Genetic differentiation of the house mouse around the Mediterranean basin : matrilineal footprints of early and late colonization. *Proc. R. Soc.* 1034–1043 (2011). doi:10.1098/rspb.2010.1228
  26. Suzuki, H. *et al.* Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA. *Heredity (Edinb).* **111** **60**, 375–390 (2013).
  27. Goios, A., Pereira, L., Bogue, M., Macaulay, V. & Amorim, A. mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Res.* 293–298 (2007). doi:10.1101/gr.5941007.erations
  28. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6 : Molecular Evolutionary Genetics Analysis Version 6 . 0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

29. Murillo-León, M. *et al.* Molecular mechanism for the control of virulent *Toxoplasma gondii* infections in wild-derived mice. *Nat. Commun.* **10**, 1–15 (2019).
30. Wallace, M. E. An inherited agent of mutation with chromosome damage in wild mice. *J. Hered.* **76**, 271–278 (1985).
31. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
32. Bekpen, C. Evolutionary and functional studies of p47 GTPases involved in cell autonomous immunity. *PhD Thesis* Institute for Genetics, University of Cologne. (2006). at <<http://kups.ub.uni-koeln.de/volltexte/2006/1748/>>
33. Prager, E. M. *et al.* Mitochondria1 DNA sequence diversity and the colonization of Scandinavia by house mice from. *Biol. J. Linn. Soc.* 85–122 (1993).
34. Müller, U. B. Evolutionary and functional studies of the Immunity-related GTPases in mice. *Diploma thesis* (2009).
35. Lee, Y. *et al.* Initial phospholipid-dependent Irgb6 targeting to *Toxoplasma gondii* vacuoles mediates host defense. *Life Sci. Alliance* **3**, 1–16 (2019).
36. Santamaría, L. & Méndez, P. F. Evolution in biodiversity policy - current gaps and future needs. *Evol. Appl.* **5**, 202–218 (2012).
37. Ebel, E. R., Telis, N., Venkataram, S., Petrov, D. A. & Enard, D. High rate of adaptation of mammalian proteins that interact with *Plasmodium* and related parasites. *PLOS Genet.* **13**, 1–27 (2017).
38. Barreiro, L. B. & Quintana, L. Evolutionary and population ( epi ) genetics of immunity to infection. *Hum. Genet.* **139**, 723–732 (2020).
39. Grossman, S. R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703–713 (2013).
40. Fumagalli, M. *et al.* Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* **7**, (2011).
41. Fumagalli, M. & Sironi, M. Human genome variability, natural selection and infectious diseases. *Curr. Opin. Immunol.* **30**, 9–16 (2014).
42. Lenz, T. L. Adaptive value of novel MHC immune gene variants. *Proc. Natl. Acad. Sci.* **115**, 1414–1416 (2018).
43. Pocock, M. J. O., Searle, J. B. & White, P. C. L. Adaptations of animals to commensal habitats: population dynamics of house mice *Mus musculus domesticus* on farms. *J. Anim. Ecol.* **73**, 878–888 (2004).
44. Phifer-Rixey, M. *et al.* The genomic basis of environmental adaptation in house mice. *PLOS Genet.* **14**, 1–28 (2018).
45. Ebert, D. & Fields, P. D. Host–parasite co-evolution and its genomic signature. *Nat.*

*Rev. Genet.* (2020). doi:10.1038/s41576-020-0269-1

46. Anderson, T. J. C. *et al.* Population Parameters Underlying an Ongoing Soft Sweep in Southeast Asian Malaria Parasites. *Mol. Biol. Evol.* **34**, 131–144 (2017).
47. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent Selective Sweeps in North American *Drosophila melanogaster* Show Signatures of Soft Sweeps. *PLOS Genet.* **11**, 1–32 (2015).

## Chapter 3

---

### **3. IRG-mediated resistance in wild *Mus musculus* against South American *Toxoplasma gondii***

Catalina Alvarez<sup>1</sup>, Claudia Campos<sup>1</sup>, Urs Benedikt Müller<sup>2</sup>, Tobias Steinfeldt<sup>3</sup>, Luis Teixeira<sup>1</sup> and Jonathan C. Howard<sup>1, 2</sup>

<sup>1</sup> Fundação Calouste Gulbenkian, Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal, <sup>2</sup> Institute for Genetics, University of Cologne, 50674 Cologne, Germany, <sup>3</sup> Institute of Virology, Medical Center University of Freiburg, 79104 Freiburg, Germany.

## **Author contributions**

CA - Study concept and design; Mouse sampling; Cell culture; Protein quantification; Flow cytometry experiments; *In vivo* experiments; Statistical analysis; Drafting and editing the manuscript.

CC - Parasite propagation; cloning; *in vivo* experiments.

UBM - Mouse sampling; *in vivo* experiments.

TS - CRISPR/Cas9 gene deletion; Lentiviral transduction for the overexpression of *Irgb2-b1* allelic forms.

LT - Study concept and design; Supervision of execution.

JCH - Study concept and design; Supervision of execution; Drafting and editing the manuscript.

### 3.1. Abstract

Some strains of the ubiquitous intracellular parasite *Toxoplasma gondii* (e.g. RH and GT-1 strains) display a high virulence on its natural intermediate host, the house mouse. Virulent *T. gondii* strains secrete pseudokinases and kinases that inactivate the Immunity-Related GTPase (IRG) resistance system in laboratory mice. However, some wild-derived Eurasian mice have polymorphic IRG proteins that inhibit the replication of such virulent *T. gondii* strains. Our previous studies have shown a high prevalence of the *Irgb2*-b1<sub>PWK</sub> allele in wild-caught and wild-derived mice from Brazil and Europe. Here we tested the ability of the *Irgb2*-b1<sub>PWK</sub> allele to confer resistance against virulent *T. gondii* strains from South American. We found that expression in *Irgb2*-b1KO cells of the *Irgb2*-b1<sub>PWK</sub> protein rescues resistance to *T. gondii* strains such as RH strain, but also against highly virulent Brazilian strains. Our *in vivo* experiments demonstrated that European mice carriers of the *Irgb2*-b1<sub>PWK</sub> allele are fully resistant to type I strains. However, their resistance to virulent South American strains varies from intermediate to absent. We also showed that South American mice that carry the *Irgb2*-b1<sub>PWK</sub> allele are resistant to virulent RH strain, but only the Brazilian MANA strain is also resistant to the South American TgMmBr1 strain. Our findings provide insights into a possible local adaptation of Brazilian mice to resist highly virulent *T. gondii* strains.



### 3.2. Introduction

All natural populations are exposed to infections with parasitic pathogens; therefore host immune response has evolved as one of the first barriers to contain the attack of those pathogens. The immune system is constantly subjected to selection pressure imposed by virulence factors, strongly promoting co-evolutionary processes<sup>1</sup>. However, the interplay between ecological and genetic factors is crucial to determine whether a host and a parasite can adapt to each other<sup>2</sup>.

IRG proteins (immunity-related GTPases) are an early host immune defense mechanism against certain pathogens in mice. IRG proteins are encoded in the mouse genome on chromosomes 11 and 18<sup>3</sup> and are strongly activated by IFN $\gamma$ . About 20 single coding units are encoded in the inbred laboratory mouse strain C57BL/6, however, four of those genes are transcribed as adjacent pairs resulting in expression of proteins carrying two IRG domains, the so-called tandem IRG proteins (Irgb2-b1, Irgb3-b5, Irgb4-b5 and Irgb9-b8)<sup>4-6</sup>. Wild-caught and wild-derived mouse strains have a much higher genetic diversity in their IRG system<sup>5</sup> than laboratory mice.

*Toxoplasma gondii* is an obligate intracellular Apicomplexan parasite and extremely abundant worldwide. *T. gondii* might currently infect one-third of the human population. Differences in seroprevalence can change according to geographical locations and socioeconomic status<sup>7</sup>. This parasite is known for its extraordinarily broad host range. It reproduces sexually (definitive host) in all true cats (Felidae) and asexually in all blood-warm animals including birds (intermediate host). Intermediate hosts get naturally infected by inadvertently foraging environmentally resistant oocysts shed in the feces by the definitive hosts<sup>8</sup>.

IRG proteins accumulate onto the parasitophorous vacuole membrane

(PVM) of cells infected with avirulent *T. gondii* strains<sup>9,10</sup>. Once there, IRGs promote PVM disruption and consequently vacuole rupture. This resistance process ultimately leads to parasite death and the necrotic death of the infected host cell<sup>11</sup>. However, virulent *T. gondii* strains can inactivate IRG proteins using effectors such as the rhoptry-derived pseudokinase ROP5 and kinase ROP18<sup>12–15</sup>. IRG inactivation ultimately leads to mouse death within days. The evolution of virulence in *T. gondii* seems paradoxical because the death of the mouse before encystment interrupts the parasite life cycle. Nonetheless, specific IRG alleles from wild mice confer resistance against virulent parasites. *Irgb2-b1* gene in the CIM mice (wild-derived Indian mouse strain) protects against Type I virulent strains<sup>5,6</sup>. Recently, it was shown that CIM<sub>*Irgb2-b1*KO</sub> cells do not resist infections against the RH *T. gondii* strain<sup>6</sup>.

South American *T. gondii* strains are genetically highly diverse, and many of them display high virulence in laboratory mice<sup>16,17</sup>. CIM mice are fully susceptible to the infection with South American *T. gondii* strains, which is related to an *Irgb2-b1* allelic composition that is not capable to fully inhibit the ROP5/ROP18/GRA7 virulence kinase complex of some strains such as VAND<sup>6</sup>.

In this study, we tested the capacity of the *Irgb2-b1*<sub>PWK</sub> allelic form, which is highly prevalent in Brazilian *M. m. domesticus*, to confer resistance against South American *T. gondii* strains *in vitro* and *in vivo*.

### **3.3. Materials and methods**

#### **3.3.1. Ethical permits**

In the laboratory, animal housing and all the procedures done with live mice were in accordance with national regulations on animal experimentation and welfare. Both the national animal welfare authorities, and the institutional

Ethics Committee at the IGC approved these protocols (Instituto Gulbenkian de Ciência Ethics Committee licence #003.2015 for 5 years (4-2015 to 4-2020)).

### 3.3.2. Mice

All *in vivo* experiments were performed at the Instituto Gulbenkian de Ciência (Portugal). C57BL/6J mice between 8 to 12 weeks were obtained from animal SPF facility at IGC and brought immediately to the quarantine facility for experimental infections. Wild-derived mouse strains were obtained from different sources. Mouse strains used in this study and their origin are given in the **key resources section** of this thesis. All strains were maintained in inbreeding (brother-sister), housed in individual ventilated cages, kept on a 12/12 hour light/dark cycle and fed with pelleted food and water available *ad libitum*. The cage bedding was supplemented with Enviro-dri®, nestlets and igloos as environmental enrichment.

### 3.3.3. Cell lines

Diaphragm Derived Cells (DDCs)<sup>18</sup> isolated from wild-caught animals were used in several experimental protocols in this thesis (**Chapter II**). The names of the DDC cell line correspond in each case to the individual wild-derived or wild-caught mice.

Engineered cell lines such as T17 DDCs and the corresponding *Irgb2-b1* allelic form complemented lines were generated in collaboration with Dr. Tobias Steinfeldt from the University of Freiburg, Germany. The generation of these lines followed the procedures for CRISPR/Cas9 gene deletion and lentiviral transduction for the overexpression of *Irgb2-b1* allelic forms described in Murillo-León et al., 2019. Cell lines were maintained by serial transfers in new T75 flasks in medium at a dilution of 1:10 for a maximum of 10 transfers. Frozen cell lines were kept in tanks with liquid nitrogen (-

196°C) and a new cell line aliquot was thawed once cells reached the maximum number of transfers mentioned above.

### **3.3.4. Sodium dodecylsulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and Western blot**

#### ***3.3.4.1. Generation of IFN $\gamma$ -induced cell lysates for SDS-PAGE***

DDCs were counted and plated in duplicate in 6-well plates. 24 hours later, cells were induced with mouse IFN $\gamma$  (200 U/ml). After an induction period of 24 hours, cells were washed once with 1X PBS before being detached physically with cell scrapers. Cells were centrifuged, the supernatant discarded and the cell pellet lysed in Cell Lysis Buffer (140 mM NaCl, 20 mM Tris-HCl (pH 8.3), 5 mM MgCl<sub>2</sub>, 0.5% Nonidet P-40 (NP-40), supplemented with cOmplete<sup>TM</sup> protease inhibitor (Roche, Germany) (1 tablet per 10 ml). The solution was kept for 30 minutes on ice followed by centrifugation for 30 minutes at 4 °C (>14,000 xg). The pellet was discarded and the postnuclear supernatant transferred to a new tube. Supernatants were stored at -80°C before SDS-PAGE.

#### ***3.3.4.2. Estimation of protein concentration by BCA assay***

The Pierce<sup>TM</sup> BCA Protein Assay (Thermo Scientific<sup>TM</sup>) was used following the manufacturer's instructions to determine protein concentration in the samples analyzed. This is a detergent-compatible formulation based on bicinchoninic acid (BCA). The method combines the reduction of Cu<sup>+2</sup> to Cu<sup>+1</sup> by protein in an alkaline medium (the biuret reaction) with the detection of the cuprous cation (Cu<sup>+1</sup>) using a reagent containing BCA. The purple-colored reaction product of this assay is formed by the chelation of two molecules of BCA with one cuprous ion. This complex exhibits absorbance at 562nm that is nearly linear with increasing protein concentrations over a range of 20 to 2000 $\mu$ g/mL<sup>19</sup>.

All samples were normalized to the sample with the lowest concentration in the batch. Dilutions were done in 1X lysis buffer.

#### 3.3.4.3. SDS-PAGE

Protein samples were mixed 5:1 with Protein Sample Buffer 6X (0.25 M Tris-HCl (pH 6.8), 25% glycerol, 5%  $\beta$ -mercaptoethanol, 5% SDS, 0.125 % bromophenol blue) and boiled for 5 minutes at 95°C. Samples were centrifuged for 1 minute at 300 xg and cooled down prior to loading them on the gel. Protein samples were loaded onto gels with a 5% acrylamide stacking gel (H<sub>2</sub>O, 1.0 M Tris-HCl (pH 6.8), 30 % acrylamide mix, 10 % SDS, 10 % ammonium peroxydisulfate (APS), tetramethylethylenediamine (TEMED)) and a 10% acrylamide resolving gel (H<sub>2</sub>O, 1.5 M Tris-HCl (pH 8.8), 30% acrylamide mix, 10% SDS, 10% APS and TEMED). Gels were run at 120 volts, and once the samples entered the resolving gel, voltage was increased to 150 volts.

#### 3.3.4.4. Western blot (*BioRad Semi-dry transfer system*)

After SDS-PAGE proteins were transferred to a polyvinylidene difluoride (PVDF) membrane (Whatman, Germany) by electro-blotting in Towbin Transfer Buffer (25 mM Tris base, 152 mM glycine, 20% MeOH, in double-distilled water (ddH<sub>2</sub>O) for a final pH 8.3) for 1 hour and 30 minutes at RT, 25 volts max. and constant 0.32 A in a semi-dry transfer system (BioRad). Membranes were blocked with Blocking Buffer (0.1% Tween 20 in 1X PBS, 5% skim milk powder) for 30 minutes at room temperature (RT) or at 4°C over-night (o.n.). After blocking, membranes were washed twice for 5 minutes with Washing Buffer (PBS-T) (0.1 % Tween 20 in 1X PBS) and incubated in constant rocking with primary antibodies diluted in PBS-T (1h at RT or at 4 °C o.n.). Primary antibodies used were the 165 and the 954 antisera (sources and antibodies' concentrations can be found in the **key resources table** section) for IRGA6 protein and IRGB2-B1 protein

detection, respectively, and the anti-tubulin antibody for detection of Tubulin protein.

Incubated membranes were washed three times for 5 minutes with PBS-T. Primary antibodies were detected by incubation for 30 minutes at RT with the appropriate DyLight™ secondary antibodies and then washed two times for 5 minutes with PBS-T in constant rocking, followed by a wash in PBS and a last wash in milliQwater (de-ionized water). DyLight™ secondary antibodies were visualized by fluorescence using the Li-Cor Odyssey® scanner.

### **3.3.5. Propagation of *Toxoplasma gondii* strains**

All experiments involving live *T. gondii* were done in a BSL-2 room. *T. gondii* strains were maintained by serial passage in confluent monolayers of Hs27 cells (ATCC). When parasites were ready to be harvested, host cells were detached with a cell scraper and the solution passed through a 25G x 5/8" needle to disrupt host cells and release *T. gondii* tachyzoites. To ensure a satisfactory purification from host cell debris, two alternative methods were followed: 1) a differential centrifugation during 5 minutes at 100 ×g and the supernatant centrifuged for 15 minutes at 700 ×g leading to a toxoplasma suspension with few host cells or 2) by filtering the solution with a 3,0µm Nuclepore Membrane (Whatman, Germany) mounted on a 6ml syringe attached to a Swin-Lok Holder 25 mm (Whatman, Germany) (leading to a toxoplasma suspension with no host cells). Purified parasites, in a suspension of IMDM with 5%FBS, were counted using an Improved Neubauer chamber (Marienfeld®, Germany) and immediately used for *in vivo* infections in mice or *in vitro* infections in cells.

### **3.3.6. Immunofluorescence microscopy**

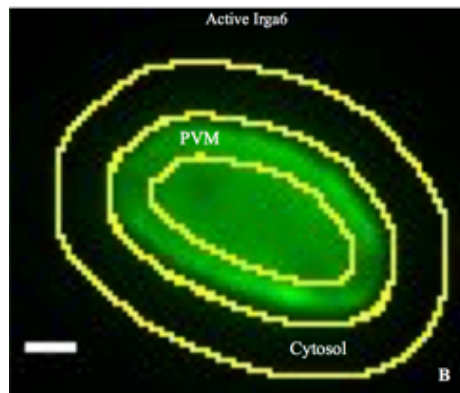
DDCs were counted and seeded on sterile 24-well plates with 13 mm glass coverslips in each well. 24 hours later, cells were induced with mouse IFN $\gamma$  (200 U/ml) for 24 hours. DDCs with approximately 80% confluence were infected with *T. gondii* tachyzoites at a multiplicity of infection (MOI) of 10 for 2 hours. Infected cells were washed three times with 1X PBS and subsequently fixed with 4% paraformaldehyde (PFA) for 30 minutes at RT. Fixed cells were washed three times with 1X PBS and permeabilized with 0.1% saponin in PBS for 10 minutes at RT in constant rocking. Later, coverslips were blocked with Blocking buffer (3% Bovine serum albumin (BSA), 0.1% saponin in PBS) for at least 30 minutes at RT (or at 4°C o.n). Fixed cells were incubated with primary antibodies diluted in Blocking buffer for 1 hour. Primary antibodies used were the 10E7 mouse monoclonal antibody, the 954 antiserum and the anti-GRA7 rat monoclonal antibody, to detect respectively IRGA6, IRGB2-B1 protein and intracellular parasites. After incubation, cells were washed twice with blocking buffer. Primary antibodies were detected by incubation for 30 minutes at RT with the appropriate secondary reagents diluted in blocking buffer. The following secondary antibodies were used: Alexa Fluor<sup>®</sup> 488 Donkey Anti-Mouse, Alexa Fluor<sup>®</sup> 555 Goat Anti-Rabbit, Alexa Fluor<sup>®</sup> 647 Goat Anti-Rat. DAPI (4',6-Diamidine-2'-phenylindole dihydrochloride) was used for nuclear staining. Cells were washed twice with blocking buffer and twice with 1x PBS for 10 minutes. Coverslips with stained cells were transferred to a slide with ProLong<sup>®</sup> Gold Antifade Reagent (Life Technologies). Images were taken at least 12 hours after mounting in a Zeiss Axio Observer Z1 fluorescence microscope, equipped with an AxioCam MRm camera and Zen 2011 software (ZEISS, Germany). Images of infected cells were captured automatically in all relevant channels at 40x or 63x magnifications<sup>18</sup>. All images from immunofluorescence experiments were stored and subsequently analyzed for the quantification of IRG loading

### **3.3.7. Quantification of IRG proteins loading**

Quantification of IRG loading onto the PVM has to be unbiased (done blindly) as well as assessed only for intracellular parasites. The GRA7 protein was used as a marker for selection of intracellular parasites because it is released by intracellular parasites shortly after invasion and localizes to the intravacuolar network and the PVM<sup>10</sup>. Only parasitophorous vacuoles (PVs) from parasites expressing GRA7 protein were scored for IRG protein loading and the quantifications were always performed blindly. **Note:** Blinding process involved that all prepared microscopic slides were blinded with tape and coded before imaging by a person non-involved with the experiment. Only after imaging and data analysis were done, real slide labels were revealed.

To automate the measurement of fluorescence in the PV membrane (PVM) and the host cytosol in the immediate vicinity of the PVM, an ImageJ macro created by Ana Lina Rodrigues in our lab was used<sup>18</sup>. The macro assumes images were loaded into ImageJ in composite mode (all relevant colours) after achromatic pixel-shift correction to optimise overlay. The user firstly defines the PV manually (by varying the “tolerance”), but later the macro automatically defines a ROI (Region of Interest) for the cytosol and the PVM (**Figure 3.1**). A table that contains different measurements from the PV, PVM, and cytosol is generated. From this table it is possible to obtain the fluorescence intensity (Mean, StDev, Min, Max and Modal fluorescence intensities; in arbitrary units) for each of the vacuoles analyzed<sup>18</sup>.





**Figure 3.1.** Quantification of IRG proteins in IFN $\gamma$ -stimulated mouse cells infected with *T. gondii*. To measure the intensity of IRG signal on *T. gondii* PVs, the Fiji 2.0.0-rc69/1,52i software along with a MACRO were used. The approach was designed to allow parasite segmentation from the cytosol and rapid analysis of numerous PVs. ROIs defining PV, PVM and cytosol are displayed. Adapted from reference <sup>18</sup>.

### 3.3.8. Quantification of host cell necrosis and rate of infection in response to *T. gondii* strains.

A Fluorescence-activated cell sorting (FACS)-based assay developed in our laboratory by Dr. Joana Loureiro was used for the quantification of host cell necrosis and rate of infection upon a challenge with several strains of *T. gondii*. This method involves three steps. First, incubation with a LIVE/DEAD<sup>®</sup> dye (Molecular Probes, USA) which is an indicator of plasma membrane permeability, second, cell fixation with paraformaldehyde and third, staining of a surface antigen from *T. gondii* in infected cells<sup>18</sup>.

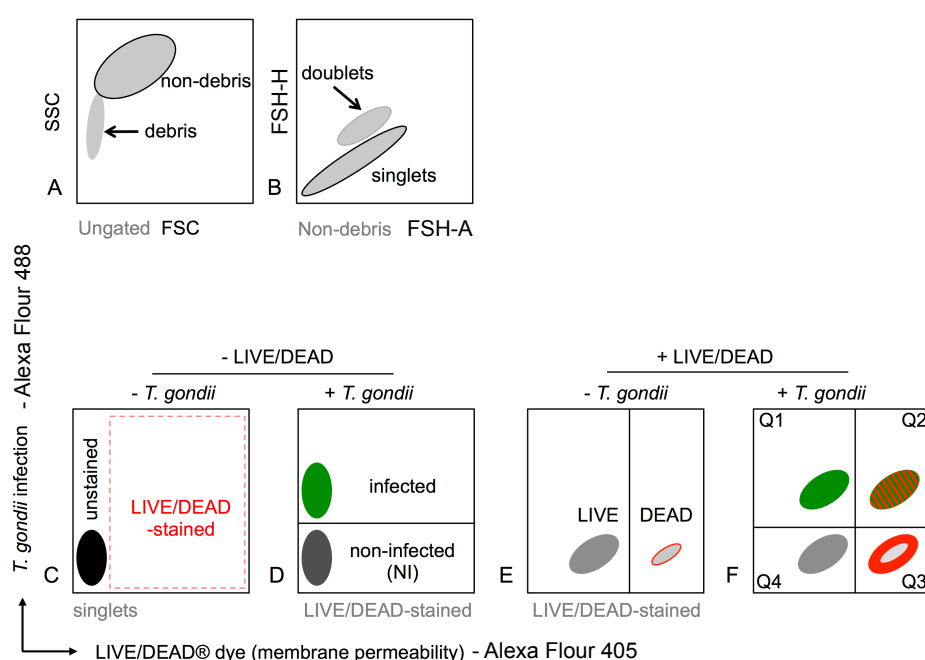
DDCs were counted and seeded on sterile 12-well plates; after 24 hours cells were induced with mouse IFN $\gamma$  (200 U/ml). After 24 hours of IFN $\gamma$  induction, host cells were infected with the desired *T. gondii* strain at MOI 30. Infection was stopped 8 hours later and supernatants containing necrotic cells were collected. Cells that remained adhered to the plate were

detached with Accutase (BioLegend) and transferred to the same Eppendorf tube with the corresponding supernatants. Cells were pelleted (5 minutes at 500xg at RT) and re-suspended in 1X PBS. Solutions were transferred to a V-bottom 96 well plate. Cells were re-pelleted and re-suspended in a solution of violet LIVE/DEAD® dye (Working Dilution concentration of 1:200 to 1:500 in 1X PBS) and incubated for 30 minutes at RT in the dark. After the incubation time, cells were washed twice with 1X PBS and re-suspended in 4% PFA. Fixation process was done for at least 20 minutes at RT or overnight at 4°C<sup>18</sup>. After this stage, all incubations were performed at 4°C with constant rocking.

Cells were washed in their 96-well plates with ice-cold 1X PBS, pelleted and permeabilized with ice-cold permeabilization/wash (P/W) buffer (BD Pharmingen™) for 20 minutes. To determine infection levels, cells were incubated for 30 min with the TP3 mouse monoclonal antibody directed against the *T. gondii* cell surface molecule, SAG1, diluted in P/W buffer (1:250). Subsequently, cells were washed twice with P/W buffer and incubated with the secondary antibody Alexa Fluor® 488 Donkey Anti-Mouse dissolved in B/W buffer (1:300) for 30 minutes. Later, cells were washed twice with P/W buffer, pelleted, re-suspend in FACS buffer (1X PBS + 1% BSA) and transferred to tubes for flow cytometry analysis<sup>18</sup>. Cells were analyzed in a BD Fortessa X-20 cytometer using the FACSDiVa software.

Gating strategy and data analysis were performed in the FlowJo™ 10.4.1 software according to the rationale proposed in reference <sup>18</sup>. Briefly, an initial gating screen based on side and forward scatter was applied in order to eliminate free parasites and small cellular debris from the analysis (**Figure 3.2**). Cells defined by these parameters were analyzed in two fluorescent channels, one giving the infected status, the other giving the live/dead status (**Figure 3.2, panel C-F**). After the appropriate gating, the cell population was divided into four subpopulations, one low and another high signal

populations for the SAG1 *T. gondii* antigen (*T. gondii* – Alexa Fluor 488+), as well as either low or high for live/dead. Each population located in a specific quadrant: quadrant Q4 contains the live, uninfected cell sub-population, Q1 live infected cells, and Q2 necrotic infected cells. With time after necrosis, the intracellular levels of parasite antigen decrease. As a result, quadrant Q3 contains necrotic cells that were infected but have lost the SAG1+ signal, as well as necrotic uninfected victims of manipulation from cell preparations and the staining protocol (**Figure 3.2, panel F**)<sup>18</sup>.



**Figure 3.2.** Gating strategy for detection by flow cytometry of host cell necrosis and rate of infection in *T. gondii* infected cells. A) Removal of small particles (debris) by using forward scatter (FSC) (linear scale) and side scatter (SSC) (linear scale) parameters. B) Cell doublets can be gated out using the forward scatter height (FSC-H) (linear scale) and forward scatter area (FSC-A) (log scale) parameters. C) Removal of un-stained/non-infected cells (black population) from properly stained cells with the violet LIVE/DEAD dye (dashed red line). All further analysis was done only in properly stained cells. D) Host cells necrosis and rate of infection in live cells were assessed using two-parameter density plots (Alexa Fluor 488 vs. Alexa Fluor 405). Shown are hypothetical dot plot density profiles of cells either uninfected (*T. gondii*-) (dark grey population) or infected with *T. gondii* (*T. gondii* – Alexa Fluor 488+)

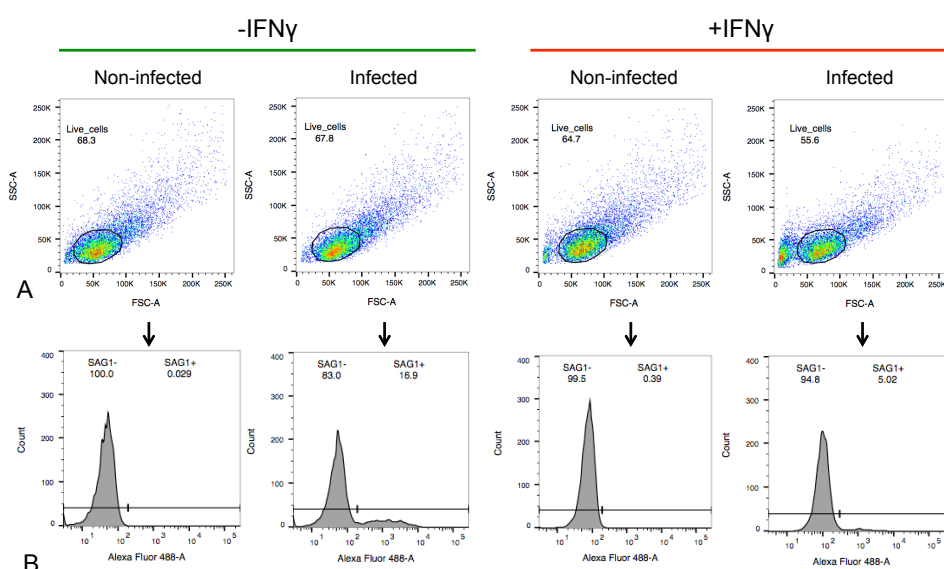
(green population) but not incubated with violet LIVE/DEAD® dye. E) This panel depicts hypothetical dot plots of cells incubated with the fixable violet LIVE/DEAD® dye (LIVE/DEAD – Alexa Fluor 405+) and are either uninfected (panel E) or infected (panel F). The light grey population represents the background of necrotic uninfected cells due to harvesting of adherent cells from culture dishes and manipulation during LIVE/DEAD® staining. Modified from reference <sup>18</sup>.

### 3.3.9. Quantification of inhibition of parasite replication

The capacity of host cells to inhibit the replication of *T. gondii* upon induction with IFN $\gamma$  was quantified by a flow cytometry based assay. This protocol is a modified version from the one described in reference <sup>6</sup>. DDCs were seeded in 12-well plates and induced with mouse IFN $\gamma$  (200U/ml) for 24 hours. Cells were subsequently infected with different *T. gondii* strains for 24 hours. Removing the medium stopped infection and cells were washed once with 1X PBS+3% Fetal Calf Serum (FCS) (PBS/FCS) in the 12-well plates. Remaining cells were detached with Accutase (BioLegend) and washed twice with PBS/FCS. Cells were fixed in 4% PFA o.n.<sup>6</sup>. All subsequent steps were performed at 4°C with constant rocking. Fixed cells were washed once with ice-cold 1X PBS, pelleted and permeabilized with ice-cold permeabilization/wash (P/W) buffer (BD Pharmingen™) for 20 minutes. As in the protocol for host cell necrosis and rate of infection, in order to determine infection levels, cells were incubated with the anti-SAG1 TP3 mouse monoclonal antibody dissolved in P/W buffer at a concentration of 1:250 for 30 minutes. After incubation, cells were washed twice with P/W buffer and incubated with the secondary antibody Alexa Fluor® 488 Donkey Anti-Mouse dissolved in P/W buffer (1:300) for 30 minutes. Later, cells were washed twice with P/W buffer, pelleted, re-suspended in PBS/FCS and transferred to tubes for flow cytometry analysis. Cells were assayed in a BD Fortessa X-20 cytometer using the FACSDiVa software<sup>18</sup>.

Further analyses were performed using FlowJo™ 10.4.1 software. Cells were initially gated based on side and forward scatter (SSC vs. FSC

channel). A small gate limiting the area with the highest density of cells in the plot helped to eliminate free parasites and small cellular debris (**Figure 3.3, panel A**). Subsequent analysis was performed only on the gated population. Cell populations were divided into two by their SAG1 marker (Alexa Fluor 488 +) intensity (**Figure 3.3, panel B**). Percentage inhibition of *T. gondii* replication was defined as follows:  $100 - (\text{mean number of IFN}\gamma\text{-induced cells positive for SAG1} / \text{mean number of non-induced cells positive for SAG1}) \times 100$ .



**Figure 3.3.** Gating strategy for the inhibition of parasite replication assay by flow cytometry. A) Representative data from an inhibition of parasite replication assay. Cells were gated using forward scatter (FSC) (linear scale) and side scatter (SSC) (linear scale) parameters to remove small particles (debris and free Toxoplasmas). All further analyses were done only on this gated population of cells. B) Cell populations were divided into two by their SAG1 marker (Alexa Fluor 488 +) intensity. Populations were established in a histogram of number of cells vs. Alexa Fluor 488 channel intensity.

### 3.3.10. Mouse survival assay

All procedures with live animals were performed in accordance with national regulations for animal welfare and experimentation. Mice were infected

intraperitoneally (i.p.) with *T. gondii* tachyzoites resuspended in 200 µl of RPMI (High glucose w/o L-glutamine w/o sodium pyruvate) using 25G hypodermic needles. All *T. gondii* strains used in this study for *in vivo* experiments are considered virulent in the laboratory mouse model (susceptible animals die between 5 to 10 days post-infection). Mice showing clear symptoms and distress were killed by cervical dislocation to reduce suffering. Criteria suggested for *T. gondii* acute infections in reference 20, were followed: 1) inability to reach food or water for more than 24 hours; 2) if the animal develops a condition that results in significant pain; 3) failure to respond to handling and 4) extreme lethargy and difficult breathing. After >25 days post infection, survivors were killed by carbon dioxide (CO<sub>2</sub>) overdose and a blood sample taken by cardiac puncture. Serum samples were tested for sero-conversion using the Toxocell Latex Kit (Biokit).

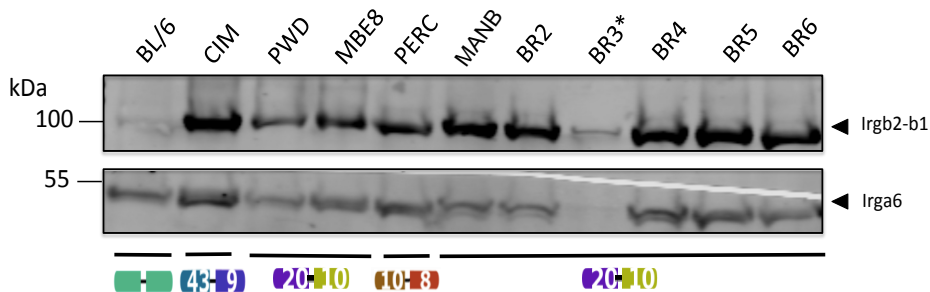
## 3.4. Results

### 3.4.1. IRG protein expression in Brazilian mouse cells

High levels of expression of the IRGB2-B1 protein seem to be associated with resistance to virulent *T. gondii* strains<sup>5,21</sup>. For example, the *Irgb2-b1*<sub>CIM</sub> allele has a much higher expression, both at the transcriptome level and protein level, than the *Irgb2-b1*<sub>C57BL/6</sub> allele. Therefore, we analyzed the protein expression levels of the Brazilian wild mice cells induced with IFN $\gamma$  to know whether their *Irgb2-b1* allele might be highly expressed and possibly associated with resistance (Figure 3.4.).

The IRGA6 protein was used in this western blot analysis as a control for IRG expression levels. Slightly higher expression was observed in the CIM mouse than in the other mouse strains, possibly associated with the sample loading process. Each of the induced Brazilian cell lines from Manaus (BR) and Belém (MBE8) (11 cell lines in total, only representative strains shown) that carry the *Irgb2-b1*<sub>PWK</sub> expressed the IRGB2-B1 protein in levels

comparable to the CIM mice. Moreover, we confirmed the very low expression of the IRGB2-B1 protein in laboratory mice (C57BL/6) reported by previous studies<sup>5,6,21</sup>. Our analysis also revealed that the new allelic form *Irgb2-b1*<sub>PERC</sub> from a Peruvian mouse also displays a strong expression of the IRGB2-B1 protein.



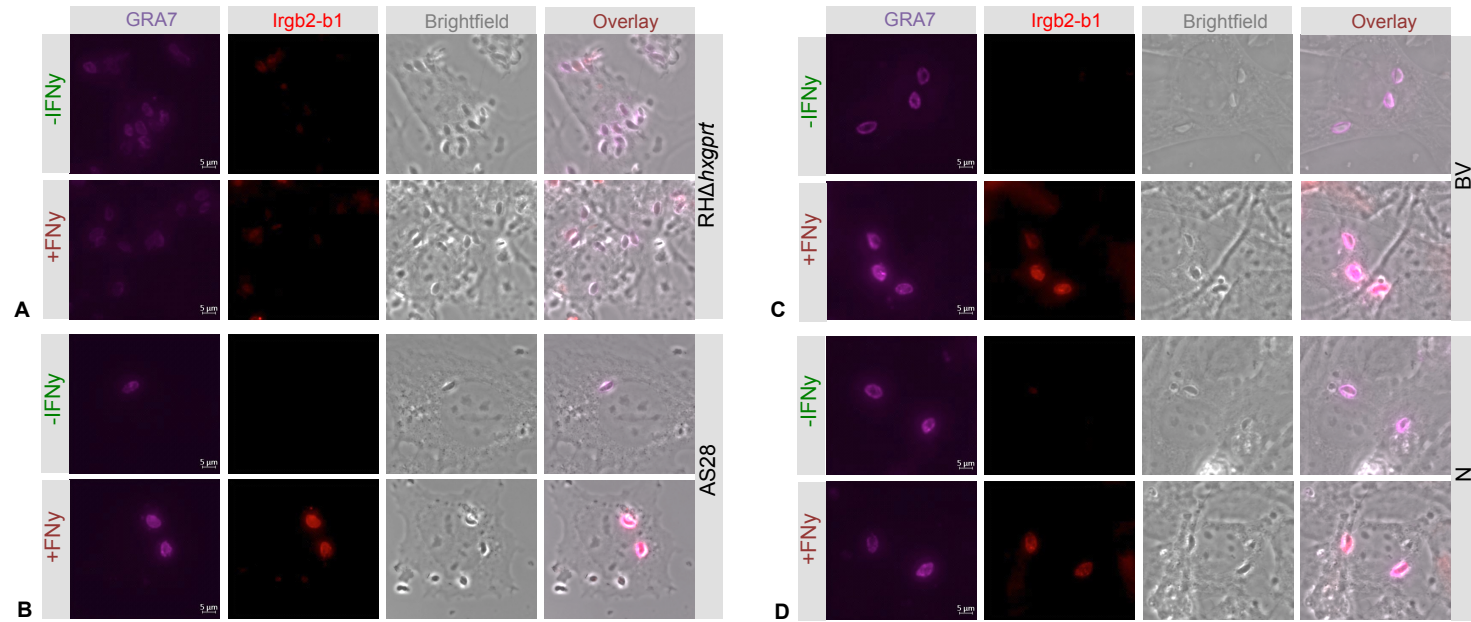
**Figure 3.4.** IRG proteins expression in Brazilian cell lines. IFN $\gamma$ -stimulated DDCs from different mouse cell lines were lysed. Lysates were blotted and membranes stained with the 165 antibody for the IRGA6 protein and the 954 antibody for the IRGB2-B1 protein. *Irgb2-b1* allelic forms are indicated at the lower part of the image (Color blocks). Sample BR3 was loaded with a lower protein concentration than the other samples (\*). The 47kDa protein IRGA6 was detected in the lysates of all DDCs, although detection in sample BR3 was lower due to lower protein loading as indicated above. 100 kDa IRG tandem protein (IRGB2-B1) was weakly detected in DDCs from the laboratory mouse strain C57BL/6. However, high expression of *Irgb2-b1* protein was detected in the cell lines from the wild-derived mouse strains CIM, PWD, PERC, the Brazilian MANB, BR2, BR4, BR5 and BR6, and in the wild-caught Brazilian MBE8.

### 3.4.2. Vacuolar intensities of IRGB2-B1 in Brazilian cells infected with different virulent *T. gondii* strains.

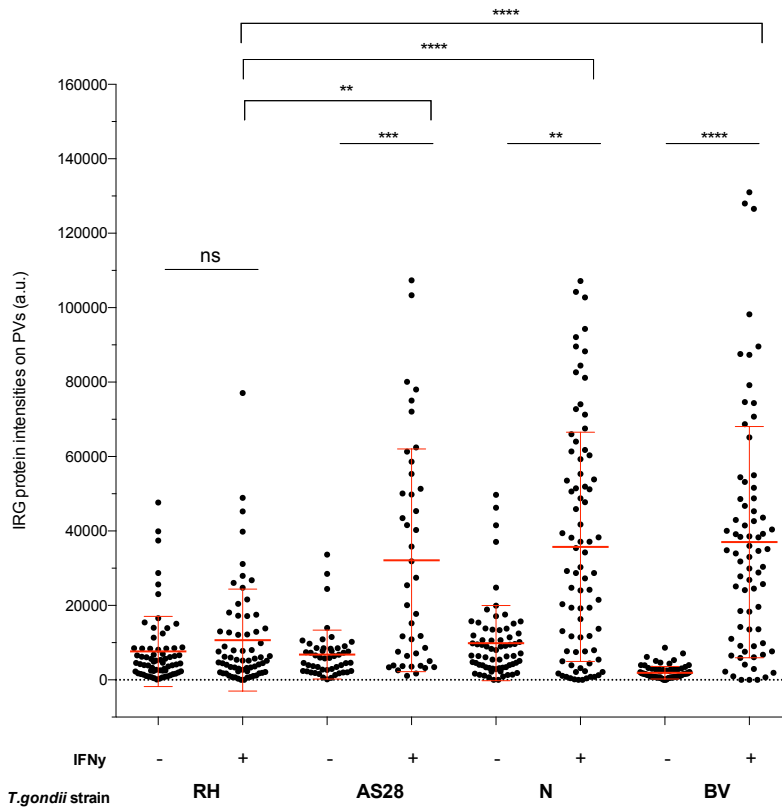
We used fluorescent microscopy to measure the vacuolar intensity of the IRGB2-B1 protein in MANB cells upon infection with local (Brazilian) and non-local virulent *T. gondii* strains (Figure 3.5.). When infected with the virulent laboratory-adapted RH $\Delta$ *etahxgprt* strain, induced cells show an increase in the loading intensity of the *Irgb1-b2* protein in relation to the non-induced ones (Figure 3.6). A similar situation occurred in cells infected with

the Brazilian strains BV, N, and AS28; however, differences between induced and non-induced cells are much more significant. An overall comparison also shows that induced cells infected with Brazilian strains have much higher IRGB2-B1 intensities than those infected with the *RHdeltahxgprt* strain.





**Figure 3.5.** Representative fluorescent images of *T. gondii* strains RH $\Delta$ tahxprt, AS28, BV and N derived vacuoles. MANB DDCs were non-stimulated or stimulated with 200U/ml of IFN $\gamma$  for 24 hours. Unstimulated and stimulated cells were infected with *T. gondii* strain A) RH $\Delta$ tahxprt, B) AS28 C) BV and D) N at a MOI of 10. After 2 hours of infection, cells were prepared for immunofluorescence analysis as described in Materials and Methods. GRA7 protein in purple (left hand panels), IRGB2-B1 protein in red (second panels from left to right), brightfield in grey (third panels from left to right) and overlay (right hand panels) are shown. All pictures were taken with the same exposure time. Scale bar is 5  $\mu$ m.



**Figure 3.6.** Higher IRGB2-B1 loading intensity in IFN $\gamma$  induced Brazilian cells infected with local *T. gondii* strains. IFN $\gamma$ -induced MANB DDCs (200 U/ml) were infected for 2 h with indicated *T. gondii* strains at a MOI of 10. Individual IRGB2-B1 positive vacuoles were identified with anti-IRGB2-B1-specific antiserum (954 antibody). Error bars indicate the mean and SD of one experiment with two technical replicates. Kruskal–Wallis test followed by Dunn's multiple comparisons were used to test differences between groups; \*\*\*\*p < 0.0001; \*p < 0.005; ns: no significant. Statistical analyses were done in the GraphPad Prism v.6 software. **Note:** This experiment was done once, it must be repeated before publication.

### 3.4.3. In vitro IRG-mediated control of Brazilian *T. gondii* strains by Irgb2-b1<sub>PWK</sub> carriers.

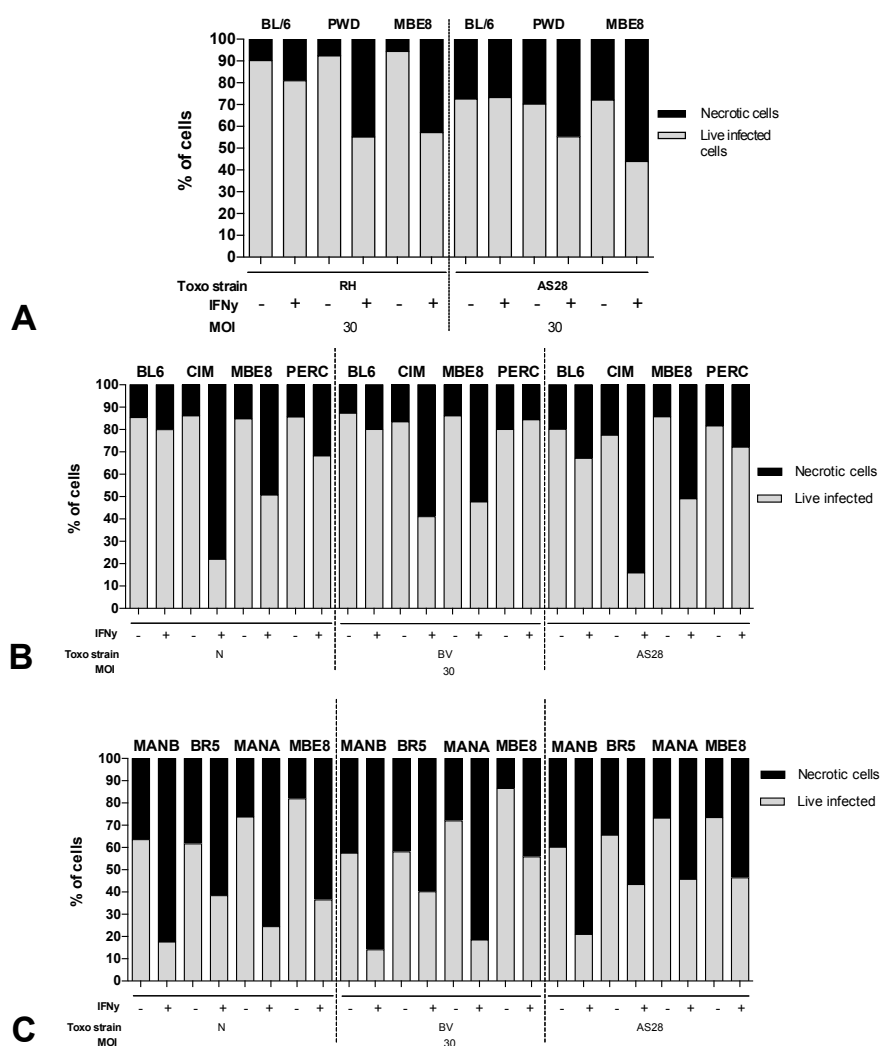
To test whether cells that are Irgb2-b1<sub>PWK</sub> carriers can efficiently control the infection with Brazilian *T. gondii* strains we used a flow cytometry-based assay to measure IRG-mediated host cell necrosis and rate of infection. We first compared the control of the highly virulent RH*delta*hxgprt strain in C57BL/6, PWD and MBE8 DDCs (the last two cell lines are carriers of the

Irgb2-b1<sub>PWK</sub> allelic form) (**Figure 3.7, panel A**). A non-significant control (reduced percentage of necrotic cells and a high percentage of infection) was observed in IFN $\gamma$ -induced C57BL/6 DDCs in comparison with non-induced cells. However, both PWD and MBE8 IFN $\gamma$ -induced DDCs displayed a significant increase in the percentage of necrotic cells, as well as a reduction of the percentage of infected cells in relation to non-induced cells. We then challenged the same panel of host DDCs against the virulent Brazilian *T. gondii* strain AS28 (**Figure 3.7, panel A**). Results indicated no IRG-mediated control in the C57BL/6 DDCs, whereas PWD and MBE8 cells are capable of controlling the AS28 strain. This is reflected by an increase in the percentage of necrotic cells in induced DDCs, as well as a reduction in infected live cells.

Considering the previous results that showed good control of the Brazilian AS28 strain by cells that are carriers of the Irgb2-b1<sub>PWK</sub> allelic form, we decided to analyze the death response of a new set of DDCs to the AS28 strain, but also to N and BV (**Figure 3.7, panel B**). This new set of DDCs included C57BL/6 and CIM as controls, the Irgb2-b1<sub>PWK</sub> carrier MBE8 and the Peruvian strain PERC. As previously reported by Müller, 2015, when IFN $\gamma$ -induced CIM cells are infected with Brazilian strains, they efficiently control them. On the contrary, IFN $\gamma$ -induced C57BL/6 cells show a very weak control of these strains. The MBE8 IFN $\gamma$ -induced cells display a very similar phenotype to that reported for CIM, where all Brazilian strains are well controlled. To our surprise, IFN $\gamma$ -induced PERC cells that have a strong expression of the IRGB2-B1 protein are inefficient in controlling Brazilian strains.

Finally, we analyzed if the ability of MBE8 cells to control infection by Brazilian *Toxoplasma* strains was a shared characteristic with other Irgb2-b1<sub>PWK</sub> carrier cells. To do this, we infected four Brazilian mouse cell lines (MANB, BR5, MANA and MBE8) with the Brazilian strains AS28, BV, and N

(Figure 3.7, panel C). We found that all IFN $\gamma$ -induced Brazilian DDCs tested could control the infection with their local *T. gondii* strains. However, there are some intrinsic differences between them. For example, the cell lines MANB and BR5 display a higher background in non-induced cells (higher initial percentage of necrosis) than the MANA and MBE8 cells. This has as a consequence that the first two cell lines might not show the same significant level of control as the other two. In conclusion, our results *in vitro* confirm that all Brazilian cell lines tested display high levels of control of local highly virulent *T. gondii* strains.



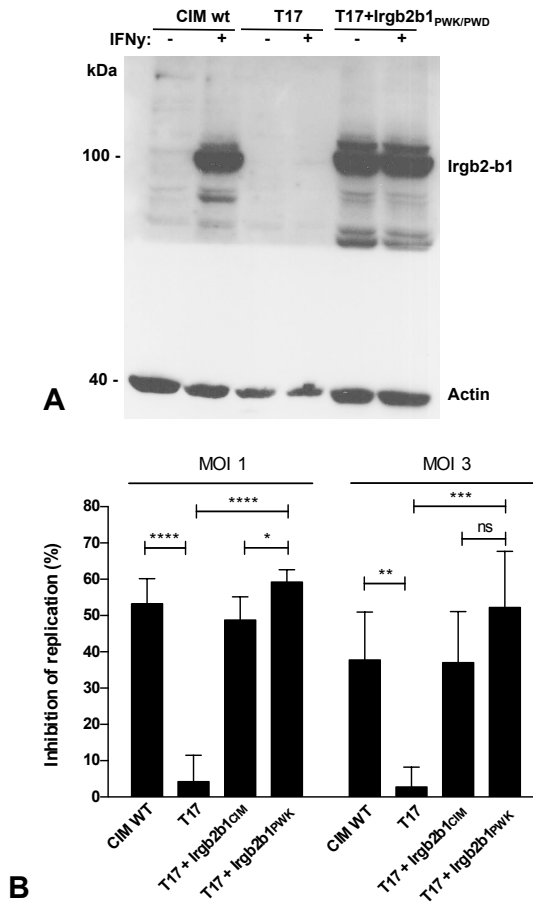
**Figure 3.7.** Brazilian *T. gondii* strains are controlled by *Irgb2-b1<sub>PWK</sub>* allelic form carriers. Assay of control of *T. gondii* by the detection of host cell necrosis and rate of infection using flow cytometry. DDCs non-induced and induced with IFN $\gamma$  (200 U/ml) were infected with the indicated *T. gondii* strain at a MOI of 30. Eight hours later, cells were harvested and incubated with the violet LIVE/DEAD<sup>®</sup> dye prior to fixation to detect live/dead cells. Fixed cells were stained with the TP3 antibody against the SAG1 *T. gondii* protein. Percentages of necrotic cells and live infected cells were obtained based on the gating strategy detailed in Material and Methods. A. C57BL/6, PWD and MBE8 DDCs were infected with the RH*deltahxgprt* and AS28 *T. gondii* strains. B. C57BL/6, CIM, MBE8 and PERC DDCs were infected with the N, BV and AS28 *T. gondii* strains. C. Brazilian DDCs MANB, BR5, MANA and MBE8 were infected with the N, BV and AS28 *T. gondii* strains.

#### **3.4.4. The *Irgb2-b1<sub>PWK</sub>* allelic form confers resistance against the Eurasian highly virulent RH-YFP strain *in vitro*.**

To test the relevance of the *Irgb2-b1<sub>PWK</sub>* allelic form for the resistance against the virulent RH-YFP strain, we used T17 DDCs (CIM<sub>*Irgb2-b1<sup>KO</sup>*</sub> DDCs)<sup>6</sup>. In collaboration with Dr. Tobias Steinfeldt, we observed a constitutive expression of the IRGB2-B1 protein independent of IFN $\gamma$  induction in T17 cells complemented with *Irgb2-b1<sub>PWK</sub>* by lentiviral transduction (T17+ *Irgb2-b1<sub>PWK</sub>*) (**Figure 3.8, panel A**). We also found a decreased expression of actin in T17 DDCs, however, this phenotype was only seen in this experiment, probably due to a under loading of the sample. Previous western blots in these cells did not show a reduction in actin expression<sup>6</sup>.

The ability of IFN $\gamma$ -induced CIM WT, T17 cells, T17+*Irgb2-b1<sub>CIM</sub>* and T17+*Irgb2-b1<sub>PWK</sub>* to control replication of RH-YFP was compared in infected cells by flow cytometry (**Figure 3.8, panel B**). As previously reported<sup>6</sup>, T17 cells have lost their ability to inhibit the replication of the *T. gondii* virulent strain RH-YFP shown by CIM wt cells. However, once the T17 cells are complemented with the *Irgb2-b1<sub>CIM</sub>* allelic form, RH-YFP infection control is rescued. A similar phenotype was observed in T17 cells complemented with

the Irgb2-b1<sub>PWK</sub>. Comparable results can be seen regardless of the MOI (1 or 3) used for the experiment.

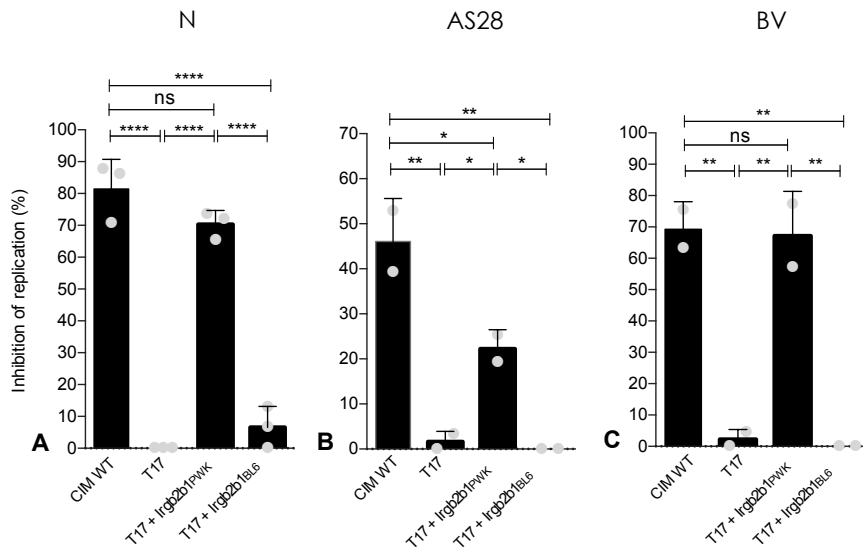


**Figure 3.8.** Irgb2-b1<sub>PWK</sub> is involved in the resistance against highly virulent *T. gondii* strains *in vitro*. A) Expression levels of the IRGB2-B1 protein in complemented T17 cells with the Irgb2-b1<sub>PWK</sub> allelic form. Unstimulated and stimulated (200 U/ml – 24 hours) DDCs from CIM w/t, Irgb2-b1<sub>CIM</sub> KO (T17) and T17+Irgb2-b1<sub>PWK</sub> were lysed. Lysates were blotted and membranes stained with the 954 antibody to detect the 100 kDa IRGB2-B1 protein and with the anti-tubulin antibody (loading control). IRGB2-B1 signal in complemented cells is visible in stimulated and unstimulated cells. Irgb2-b1<sub>CIM</sub> in CIM wt cells is only detectable upon stimulation with IFN $\gamma$ . B) CIM w/t, T17, T17+Irgb2-b1<sub>CIM</sub> and T17+Irgb2-b1<sub>PWK</sub> DDCs were non-induced or induced with 200 U/ml of IFN $\gamma$  for 24 h and infected with *T. gondii* RH-YFP at a MOI of 1 or 3. Intracellular parasite growth was determined by flow cytometry 24 hours post infection. Inhibition of replication was calculated as: 100 – (mean number of IFN $\gamma$ -induced

cells positive for SAG1/mean number of non-induced cells positive for SAG1)  $\times$  100 (see details in Materials and Methods). Error bars indicate the mean and SD of three independent experiments. Two-way ANOVA followed by a Tukey's multiple comparisons test were used to test differences between groups; \*\*\*\*p < 0.0001; ns: not significant. Statistical analyses were done in the GraphPad Prism v.6 software.

### **3.4.5. In vitro resistance to virulent Brazilian *T. gondii* strains is mediated by the *lrgb2-b1*<sub>PWK</sub> allelic form.**

The second chapter of this thesis showed the presence of two *lrgb2-b1* allelic forms in *M. m. domesticus* caught in different locations in Brazil: the *lrgb2-b1*<sub>C57BL/6</sub> and *lrgb2-b1*<sub>PWK</sub>. From these two, the *lrgb2-b1*<sub>PWK</sub> allelic form displays the highest frequency, close to 90%. The ability of T17 cells complemented with either the *lrgb2-b1*<sub>C57BL/6</sub> or the *lrgb2-b1*<sub>PWK</sub> allelic form to control virulent Brazilian *T. gondii* strains was compared in infected cells by flow cytometry (Figure 3.9). For this experiment, the three Brazilian strains N (Figure 3.9, panel A) AS28 (Figure 3.9, panel B) and BV (Figure 3.9, panel C) were used. As we observed for RH-YFP, CIM wt cells efficiently inhibited parasite replication for all strains but T17 cells were unable to do so. Complementation of T17 cells with *lrgb2-b1*<sub>PWK</sub> rescued resistance against the N and BV strains to wt levels and slightly lower levels for the AS28 strain. However, we could not restore any resistance of T17 to infection when the *lrgb2-b1*<sub>C57BL/6</sub> allelic form<sup>6</sup> was present and highly expressed, indicating a direct responsibility of the *lrgb2-b1*<sub>PWK</sub> in controlling Brazilian *T. gondii* strains (Figure 3.9).



**Figure 3.9.** *Irgb2-b1<sup>PWK</sup>* controls replication of Brazilian *T. gondii* strains. CIM w/t, T17, T17+*Irgb2-b1<sup>PWK</sup>* and T17+*Irgb2-b1<sup>BL6</sup>* DDCs were non-induced and induced with 200 U/ml of IFN $\gamma$  for 24 h and infected with the indicated *T. gondii* strain at a MOI of 1. Intracellular parasite growth was determined by flow cytometry 24 hours post infection. Percentages of inhibition of replication were calculated as described in Material and Methods. A) Induced and non-induced cells were infected with the Brazilian N strain. Error bars indicate the mean and SD of three independent experiments. B) Induced and non-induced cells were infected with the Brazilian AS28 strain. Error bars indicate the mean and SD of two independent experiments. C) Induced and non-induced cells were infected with the Brazilian BV strain. Error bars indicate the mean and SD of two independent experiments. Two-way ANOVA followed by a Tukey's multiple comparisons test were used to test differences between groups; \*\*\*\*p < 0.0001; \*\*p < 0.001; ns: not significant. Statistical analyses were done in the GraphPad Prism v.6 software.

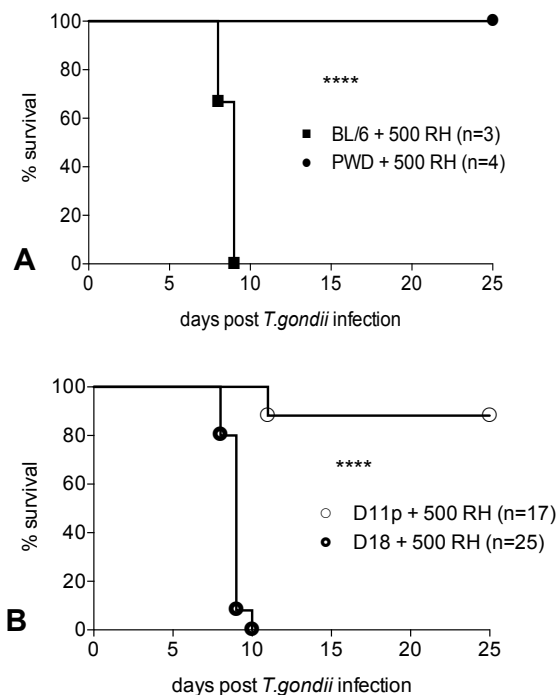
### 3.4.6. Mice carrying the *Irgb2-b1<sup>PWK</sup>* allelic form can control the growth of Eurasian *T. gondii* strains in vivo.

The *M. m. castaneus* strain CIM was until recently the only wild-derived mouse strain shown to be resistant to infection with highly virulent Eurasian *T. gondii* strains. We compared the virulence of the RH $\Delta$ ltahxgprt in laboratory C57BL/6 and wild-derived PWD mice (European carrier of the *Irgb2-b1<sup>PWK</sup>* allelic form) (Figure 3.10, panel A).



Unlike C57BL/6 mice that died between day 9 and 10 post-infection, PWD mice survived for the full length of the infection (25 days). No signs of acute infection were observed in the PWD mice. However, all mice displayed seroconversion for anti-*T. gondii* antibodies.

To know whether the IRG system is also involved in resistance against virulent Eurasian *T. gondii* strains in the PWD mouse, we used the consomic strains D11p (C57BL/6 mouse containing the IRG encoding region of Chromosome 11 from PWD) and D18 (C57BL/6 mouse with the Chromosome 18 from PWD containing the IRGA genes) (**Figure 3.10, panel B**). While D11p mice survived the infection against the RH*delta*hxgprt strain, all D18 mice succumbed within less than 10 days post-infection (Dr. Urs Benedikt Müller, unpublished data). These results suggest that the region responsible for the resistance against virulent laboratory *T. gondii* strains in PWD mice is located in Chromosome 11. The lack of any apparent contribution from the IRGA genes on chromosome 18 is consistent with similar results from the CIM mouse resistant strain. Most likely it is the IRGB2-B1 protein, encoded in this region, as confirmed by the restoration of *in vitro* resistance to T17 by expression of the IRGB2B1<sub>PWK</sub> allelic form. Data in reference <sup>20</sup> have independently shown similar results to the ones presented in this section.



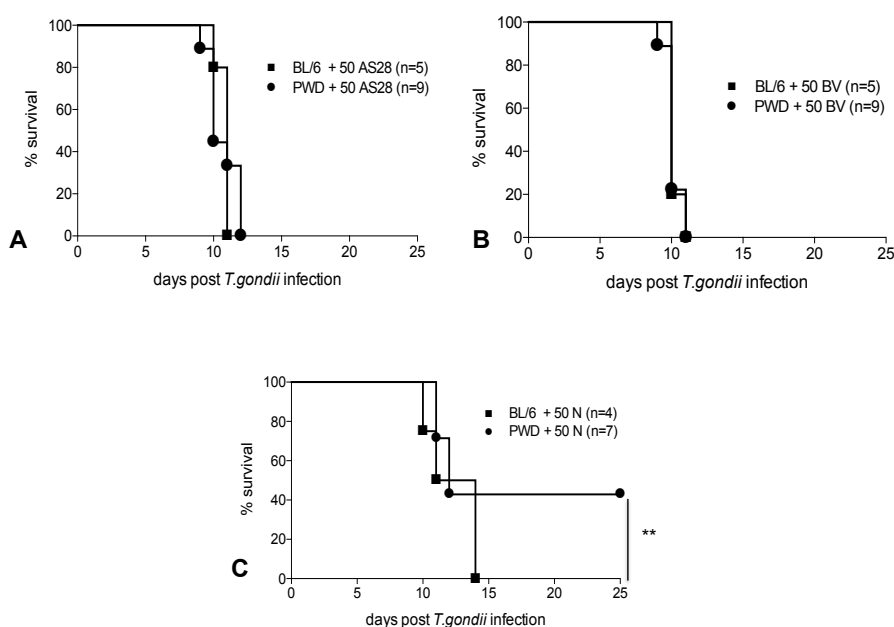
**Figure 3.10.** PWD mice are highly resistance to type I virulent *T. gondii* strains. A) PWD and C57BL/6 mice were intraperitoneally (i.p) infected with 500 tachyzoites of the type I virulent *T. gondii* strain RHdelta $hxgprt$ . Mouse survival was monitored for 25 days. B) Consomic strains D11p and D18<sup>22</sup> were intraperitoneally (i.p) infected with 500 tachyzoites of the type I virulent *T. gondii* strain RHdelta $hxgprt$ . Mouse survival was monitored for 25 days (Unpublished data from experiments performed by Dr. Urs Benedikt Müller). All survivors were verified for *T. gondii* infection based on the presence of anti-*T. gondii*-specific antibodies and DNA from their brains was extracted to check with a PCR-RFLP for the corresponding pattern of the *T. gondii* strain injected. Only seropositive and/or PCR positive survivors are displayed in the graph. n=number of mice infected. Survival curves were compared by using the Log-rank (Mantel-Cox) test. \*\*\*\* p < 0.0001. Statistical analyses were done in the GraphPad Prism v.6 software.

### 3.4.7. European carriers of the *Irgb2-b1*<sub>PWK</sub> allelic form display an intermediate resistance against infection with some Brazilian *T. gondii* strains.

The most common South American *T. gondii* strains are genetically very diverse and highly virulent in laboratory mice, with a similar phenotype to the one seen with Type I strains<sup>16</sup>. The *M. m. musculus* PWD is a mouse strain

like CIM capable of surviving infection with Type I virulent *T. gondii* strains<sup>5,6,20</sup>. However, it is not known whether PWD can survive infection with South American *T. gondii* strains. Since the IRGB2-B1 haplotypes found on Chromosome 11 of the PWD mouse are identical to those found in Brazilian mice, we used the PWD mouse to test its ability to resist the infection with Brazilian *T. gondii* strains. PWD mice were infected with three different Brazilian strains, N, BV and AS28 (Figure 3.11).

PWD mice succumbed to the infection with both AS28 (Figure 3.11, panel A) and BV strain (Figure 3.11, panel B) at a rate of death similar to the one seen in C57BL/6 susceptible mice. However, we observed an intermediate survival for the infection with the N strain, where 3 out of 7 animals survived.



**Figure 3.11.** PWD mice are susceptible to the Brazilian virulent *T. gondii* strains AS28 and BV, but display an intermediate resistance against the N strain. A) PWD and C57BL/6 mice were intraperitoneally (i.p) infected with 50 tachyzoites of the atypical virulent *T. gondii* strain AS28. B) BV and C) N. Mouse survival was monitored for 25 days. All survivors were verified for *T. gondii* infection based on the presence of anti-*T. gondii*-specific antibodies and DNA from their brains was extracted to check with a PCR-RFLP for the corresponding pattern of

the *T. gondii* strain injected. Only seropositive and/or PCR positive survivors are displayed in the graph. n=number of mice infected. Survival curves were compared by using the Log-rank (Mantel-Cox) test. \*\*p < 0.004.

### **3.4.8. Resistance to South American *T. gondii* strains relies on host genetic background and local adaptation.**

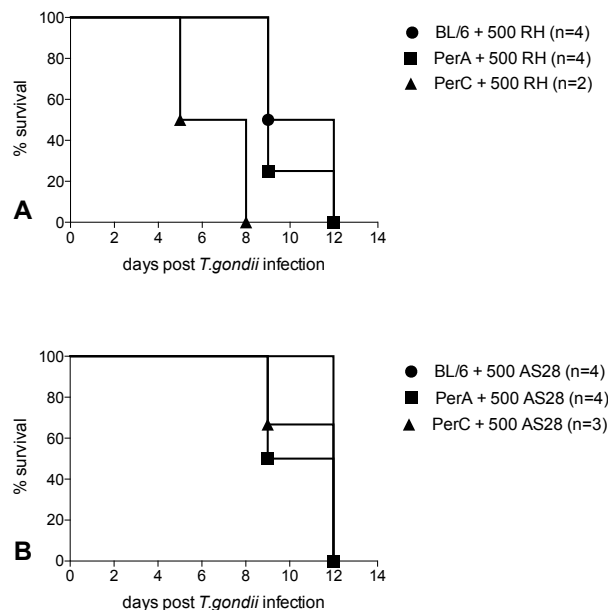
#### **3.4.8.1. Peruvian mice are susceptible to the infections with Eurasian Type I and Brazilian *T. gondii* strains.**

There are only a few wild-derived mouse strains from South American *Mus musculus domesticus*. These animals represent a perfect opportunity to test resistance against infection with sympatric and allopatric *T. gondii* strains in the South American house mouse. We used two well-known wild-derived mouse strains from South America, the Peruvian PerA and PerC mice, to test their ability to survive the infection with the virulent Type I RHdeltahxgprt strain and the Brazilian AS28 strain (Figure 3.12).

PerA mice were as susceptible as the C57BL/6 mice to infection with both RHdeltahxgprt and the AS28 strains (Figure 3.12, panel A). This is an expected result considering that the IRGB2-B1 from the PerA strain is the Irgb2-b1<sub>C57BL/6</sub> allelic form, associated with susceptibility to virulent strains. PerC mice carry a strongly expressed Irgb2-b1 allele with comparable levels of expression to those reported for the CIM and PWD mice<sup>5,6,21,23</sup> (Figure 3.4). Nevertheless, these were also fully susceptible to infection with the two virulent strains used (Figure 3.12, panel B).

Altogether, the fact that mice from localities West of the Andes Mountains are susceptible to *T. gondii* strains from East of the Andes Mountains suggest that adaptation to sympatric *T. gondii* strains is probably important for resistance to Brazilian isolates. Local interactions that might lead to resistance in South American mice might be very restricted in their geographic location. Therefore, close proximity on location between the host

and the parasites, as well as, the proper host genetic background define resistance to South American *T. gondii* strain, not as previously suggested, mouse subspecies<sup>20</sup>.



**Figure 3.12.** Peruvian mice are susceptible to virulent *T. gondii* strains from Brazil. A) C57BL/6, PERA and PERC mice were intraperitoneally (i.p) infected with 500 tachyzoites of the virulent *T. gondii* strain RH*deltahxgprt*. Mouse survival was monitored for 12 days. B) C57BL/6, PERA and PERC mice were intraperitoneally (i.p) infected with 500 tachyzoites of the Brazilian virulent *T. gondii* strain AS28. Mouse survival was monitored for 12 days. n=number of mice infected.

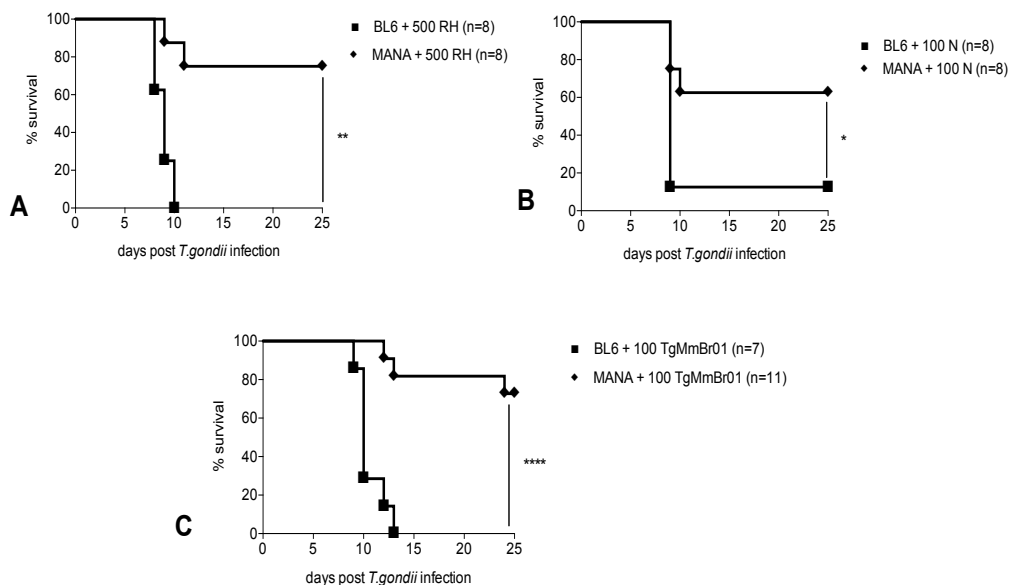
### 3.4.8.2. Brazilian mice display different levels of resistance against virulent Type I and sympatric *T. gondii* strains

Co-adaptation between host and parasite is dependent upon geographical proximity, and favours the common host genotypes<sup>24</sup>. To test the ability of Brazilian mice to resist infection with highly virulent Brazilian *T. gondii* strains, we used wild-derived *M. m. domesticus* from Manaus, Brazil (MANA) and challenged them with the typical Type I virulent strain RH*deltahxgprt* and with two different Brazilian strains. Brazilian strains used

were the N strain, isolated in São Paulo, Brazil in 1952 and the TgMmBr01 isolated by Dr. Urs Benedikt Müller from the Howard Lab in Belém, Brazil in 2014 (Unpublished data).

MANA mice displayed resistance to the infection with the highly virulent RH*deltahxgprt* strain, where 6 out of 8 animals survived the infection (**Figure 3.13, panel A**). We could detect DNA from the RH*deltahxgprt* strain in brains from infected animals (a clear sign of a chronic stage of the disease) (data not shown).

MANA mice also displayed intermediate resistance against the Brazilian N strain, with 5 out of 8 animals still surviving after of infection (**Figure 3.13, panel B**). All of these seroconverted, and parasite DNA was detected in their brains by PCR. MANA mice also showed significant resistance (8 survivors from 11 mice) against TgMmBr01, a strain recently isolated from a wild mouse in Belém, Brazil (**Figure 3.13, panel C**).



**Figure 3.13.** Brazilian mice display different levels of resistance against virulent Type I and local *T. gondii* strains. A) C57BL/6 and MANA mice were intraperitoneally (i.p) infected with

500 tachyzoites of the virulent *T. gondii* strain RH~~delta~~hxgprt. B) 100 tachyzoites of the atypical Brazilian *T. gondii* strain N and C. 100 tachyzoites of the Brazilian *T. gondii* strain TgMmBr01. Mouse survival was monitored for 25 days. All survivors were verified for *T. gondii* infection based on the presence of anti-*T. gondii*-specific antibodies and DNA from their brains was extracted to check with a PCR-RFLP for the corresponding pattern of the *T. gondii* strain injected. Only seropositive and/or PCR positive survivors are displayed in the graph. n=number of mice infected. Survival curves were compared by using the Log-rank (Mantel-Cox) test. \*p 0.03; \*\*p 0.001; \*\*\*\*p < 0.0001.

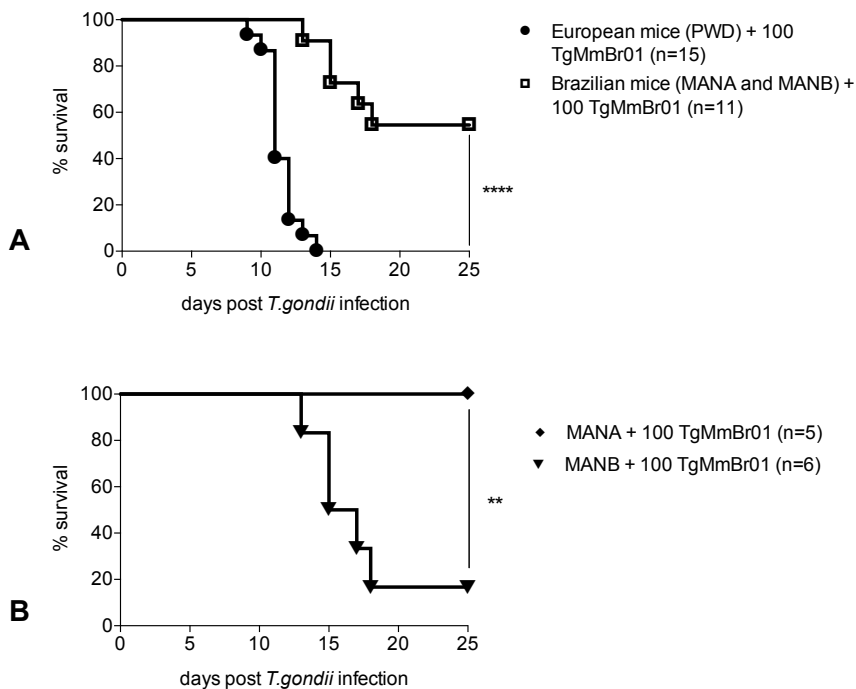
### **3.4.9. European and Brazilian *Irgb2-b1<sub>PWK</sub>* allelic form carriers resist differently to the infection against highly virulent Brazilian *T. gondii* strains.**

Our experiments *in vitro* have shown that the *Irgb2-b1<sub>PWK</sub>* allelic form is responsible for the resistance against highly virulent Eurasian and South American *T. gondii* strains in European and South American mouse cell lines. However, it is not clear yet whether in infections *in vivo*, the *Irgb2-b1<sub>PWK</sub>* allelic form has a role in the control of Brazilian *T. gondii* strains.

We infected European PWD and Brazilian MANA and MANB mice, which are all carriers of the *Irgb2-b1<sub>PWK</sub>* allelic form; with the Brazilian *T. gondii* strain TgMmBr01 (**Figure 3.14, panel A**). Their ability to control the infection is very different. Unlike the Brazilian mice that displayed an intermediate resistance (5 out of 11 mice died), all PWD animals (15 animals) succumbed to the infection within less than 15 days. This result not only shows the importance of adaptation to sympatric pathogens that ultimately leads to resistance, but also questions the role of the *Irgb2-b1<sub>PWK</sub>* allelic form as the single allele responsible for the resistance against Brazilian *T. gondii* strains.

Finally, to further investigate the importance of the *Irgb2-b1<sub>PWK</sub>* allelic form as the single allele responsible for the resistance against Brazilian *T. gondii* strains *in vivo*. We compared the resistance against TgMmBr01 in MANA and MANB mice (**Figure 3.14, panel B**). Even though, these two strains are

both Brazilian carriers of the *Irgb2-b1*<sub>PWK</sub> allelic form, as well as sharing IRG haplotypes for the Chromosome 11, we observed a significant variation for the resistance against TgMmBr01. All 5 MANA mice infected survived the full length of the infection (25 days), whereas only 1 out of 6 MANB mice survived. Altogether, these results (Figure 3.14) indicate that the *Irgb2-b1*<sub>PWK</sub> allele may be necessary but not sufficient to control Brazilian *T. gondii* strains.



**Figure 3.14.** Brazilian mice have a higher resistance to Brazilian *T. gondii* strains than European PWD mice. A) PWD and MANA and B) MANA and MANB mice were intraperitoneally (i.p) infected with 100 tachyzoites of the Brazilian virulent *T. gondii* strain TgMmBr01. Mouse survival was monitored for 25 days. All survivors were verified for *T. gondii* infection based on the presence of anti-*T. gondii*-specific antibodies and DNA from their brains was extracted to check with a PCR-RFLP for the corresponding pattern of the *T. gondii* strain injected. Only seropositive and/or PCR positive survivors are displayed in the graph. n=number of mice infected. Survival curves were compared by using the Log-rank (Mantel-Cox) test. p\*\* 0.008, p\*\*\*\* < 0.0001.

### 3.5. Discussion



IRG protein accumulation at the PVM of infected IFN $\gamma$ -induced mouse cells is crucial to restrict *T. gondii* growth. This loading process promotes PVM rupture, parasite death, and subsequent necrotic death of the host cell<sup>9,11</sup>. *T. gondii* virulent strains can overcome the IRG system by the secretion of virulence factors that block IRG function. The highly polymorphic tandem IRG protein Irgb2-b1<sub>CIM</sub> was the first allelic form of this protein to be reported to protect CIM mice from *T. gondii* virulent strain infections. Using an integrative approach of *in vitro* and *in vivo* techniques, we have found a new IRG haplotype (Irgb2-b1<sub>PWK</sub>) in wild and wild-derived *M. m. domesticus* and *M. m. musculus* that is strikingly resistant to highly virulent *T. gondii* strains from different geographic locations. We confirmed the high protein expression of IRG tandem proteins for the Irgb2-b1<sub>PWK</sub> allelic form in IFN $\gamma$ -induced cells. So far, no laboratory mouse strains have been reported to express a significant amount of any of the IRG tandem proteins, only some wild-derived mouse strains have shown high levels of expression (e.g. Irgb2-b1<sub>CIM</sub>, Irgb2-b1<sub>BULS</sub>, Irgb2-b1<sub>STUF</sub>, etc.)<sup>5,21</sup>. Interestingly, these wild-derived strains belong to all the three mouse subspecies, which shows that high expression of IRG tandem proteins is not determined by the mouse subspecies, as previous studies suggested<sup>20</sup>, but it is rather associated with a specific mouse strain Irgb2-b1 allele (polymorphisms). This conclusion was further confirmed by our finding that *M. m. domesticus* from Brazil carrying the Irgb2-b1<sub>PWK</sub> allele have similar levels of expression as those found in European *M. m. domesticus*, and *M. m. musculus* carriers of the same allele.

*T. gondii* virulence has been directly correlated with IRG protein inactivation<sup>9</sup>. The Irgb2-b1 protein works as a decoy for the *T. gondii* ROP5 protein, a virulence factor that participates in inactivation of IRG proteins. We found that the Irgb2-b1<sub>PWK</sub> allele loaded more efficiently to PVMs of Brazilian cells infected with local *T. gondii* strains than to those infected with

the laboratory strain RH. Even though there is still no direct evidence that a higher PVM loading of the *Irgb2-b1* protein represents a better control of virulent *T. gondii* strains, reduction of p(T108)*Irga6*<sub>CIM</sub> has been observed when *Irgb2-b1*<sub>CIM</sub> loads efficiently on PVMs from RH infected CIM cells<sup>5,6</sup>. Moreover, we found that when Brazilian mouse cell lines, carriers of the *Irgb2-b1*<sub>PWK</sub> allelic form, are challenged with local *T. gondii* strains they display good control of the infection, namely a reduction of infected cells as well as an increase in necrotic death. Our findings give further support to the role of *Irgb2-b1* protein as a decoy for virulent *T. gondii* ROP kinases, as well as to a possible increase in *Irgb2-b1* loading when local interactions are involved (e.g. Brazilian mouse cells vs. Brazilian *T. gondii* strains). Recent studies on the *Irgb2-b1*<sub>CIM</sub> protein have found that expression of this protein form on CIM<sub>*Irgb2-b1*KO</sub> cells rescues the resistant phenotype against virulent type I *T. gondii* strains<sup>6</sup>. We found that expression of the *Irgb2-b1*<sub>PWK</sub> protein form in previously susceptible CIM<sub>*Irgb2-b1*KO</sub> cells is also sufficient to fully confer resistance, not only to the type I RH strain but also to highly virulent Brazilian strains. Consistently, these data demonstrate that the expression of the proper *Irgb2-b1* allele is key to confer resistance against *T. gondii* virulent strains.

The IRG system is essential for mouse survival following infection with all strains of *T. gondii*. Many of the critical steps needed for resistance in mice against *T. gondii* happen within 2 to 3 hours after the pathogen has entered a cell induced with IFN $\gamma$ . Therefore, *in vitro* approaches have been very important to measure many of the parameters associated with *T. gondii* control, such as parasite inhibition and host necrotic cell death. However, so far none of these approaches can be considered fully reliable to extrapolate resistance *in vivo*. This has been already observed in CIM mice, in which IFN $\gamma$  stimulated cells showed resistance against South American *T. gondii*, but mice were fully susceptible to infection<sup>6,21</sup>. We found that European mice which are carriers of the *Irgb2-b1*<sub>PWK</sub> allele can resist type I virulent strains,

which is coherent with our *in vitro* data. However, their resistance to virulent South American *T. gondii* strains is either intermediate or absent according to the strain used, which contrasts with the *in vitro* data that would predict full resistance to these strains. Once again, multiple factors can directly impact on differences observed in resistance levels *in vivo* versus *in vitro*. First, cells used for this study (DDCs) are derived from a complex cell population and given the procedures needed for their isolation and immortalization may show differences between preparations. Even though our previous studies indicate that cell heterogeneity does not affect *T. gondii* infection, probably other factors such as capacity to the cells to react to IFN $\gamma$ , adherence to the plates, transfection efficiency, etc. can influence the outcome of *in vitro* experiments<sup>18</sup>. Second, in an ideal scenario, *in vivo* experiments must be the closest to infections in nature, which is not always possible. We do believe that the infection route (e.g. oral, intraperitoneal, etc.), initial dosage, and cellular immune responses on the early stage of infection can impact disease progression *in vivo*. In this context, it is important to recall that natural infection of intermediate hosts is usually with sporocysts, not tachyzoites.

We also reported the first wild-derived mouse strain able to resist infections with virulent South American *T. gondii* strains. This is an important finding if we consider that no mice (neither laboratory, wild-derived nor wild strains) so far has been reported to be fully resistant to any virulent South American *T. gondii* strain<sup>6,20</sup>. Even though the Brazilian wild-derived *M. m. domesticus* strains used in this study were planned as models for natural infection of wild mice in Brazil, given their similarities in genetic background with wild-caught mice (Chapter II, this thesis), we are aware that the use of wild-derived mice instead of wild-caught mice has limitations. Although wild-derived strains exhibit a much higher genetic diversity than classical laboratory mouse strains<sup>25,26</sup>, they have a lower genetic diversity than wild-caught mice, which can interfere with the relevance of these mice as mimics

for natural infections with *T. gondii*. Additionally, it would be interesting to do experiments in wild-caught mice with recently isolated *T. gondii* strains and through a more “natural” route of infection (e.g. oral gavage with oocysts).

We have shown that resistance to South American *T. gondii* strains relies on host genetic background and local adaptation. Therefore, variations in resistance to South America *T. gondii* strains might be expected, as in the case of the two wild-derived Brazilian mice used for this study. The ManA and ManB mouse strains are both homozygous carriers for the *Irgb2-b1*<sub>PWK</sub> allele; nonetheless, ManB is susceptible whereas ManA is fully resistant to infection with the virulent strain TgMmBr01. We have seen that resistance to virulent Eurasian and South American *T. gondii* strains in *Irgb2-b1*<sub>PWK</sub> Brazilian mouse carriers remains even when the resistant allele is present in the heterozygous condition (data not shown). Therefore, differences between the two wild-derived Brazilian mouse strains strongly suggest that besides IRG tandem proteins there are other factors in the resistance against local strains that we still need to identify. Further studies must be done on the genetic basis of these differences, for example, by classical genetic backcrosses between resistant and susceptible mice to identify a hypothetical new gene(s) inside or outside the IRG resistance system (e.g. *Irgb6* or genes on Chromosome 18). Also, it is unknown if the *Irgb2-b1*<sub>PWK</sub> protein interacts with specific ROP5 isoforms (e.g. ROP5B) in the same way as the *Irgb2-b1*<sub>CIM</sub> does and therefore resistance to different parasites will rely on the expression of certain ROP5 proteins. Finally, the hypothesis of collaborative action between the *Irgb2-b1* protein and other proteins must be also considered in the future. Altogether, our data suggest a strong adaptation to local *T. gondii* strains in Brazilian mice, possibly associated with selection pressure imposed by the parasites.

### **3.6. Acknowledgments**

I would like to acknowledge the work of previous and current members of the laboratory who contributed to the development of the study, especially to Dr. Joana Loureiro who developed the flow cytometry-based method for the quantification of host cell necrosis and rate of infection, Ana Lina Rodrigues who created the ImageJ macro and Dilşan Özcanoğlu who performed some of the fluorescence microscopy experiments. Many thanks to Dr. Michael Nachman and Dr. Ellen Robey from University of California Berkeley, who respectively provided the Manaus mouse samples and the conditions to work with them. IGC facilities services were also crucial for the development of this study, in particular the Animal Facility, supported by the research infrastructure Congento, project LISBOA-01-0145-FEDER-022170, the Advanced Imaging Unit, supported by the project PPBI-POCI-01-0145-FEDER-022122 and the Flow Cytometry Unit, supported by the project LISBOA-01-0145-FEDER-007654. This work was supported by central funds of the Instituto Gulbenkian de Ciência, by the Sonderforschungsbereiche 670 and 680 and Schwerpunkt 1399 of the Deutsche Forschungsgemeinschaft.

## Bibliography

1. Keogh, C. L., Sanderson, M. E. & Byers, J. E. Local adaptation to parasite selective pressure: comparing three congeneric co-occurring hosts. *Oecologia* **180**, 137–147 (2016).
2. Thrall, P. H., Barrett, L. G., Dodds, P. N. & Burdon, J. J. Epidemiological and Evolutionary Outcomes in Gene-for-Gene and Matching Allele Models. *Front. Plant Sci.* **6**, 1–12 (2016).
3. Bekpen, C. Evolutionary and functional studies of p47 GTPases involved in cell autonomous immunity. *PhD Thesis* Institute for Genetics, University of Cologne. (2006). at <<http://kups.ub.uni-koeln.de/volltexte/2006/1748/>>
4. Bekpen, C. *et al.* The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. *Genome Biol.* **6**, R92 (2005).
5. Lilue, J., Müller, U. B., Steinfeldt, T. & Howard, J. C. Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. *Elife* **2013**, 1–21 (2013).
6. Murillo-León, M. *et al.* Molecular mechanism for the control of virulent *Toxoplasma gondii* infections in wild-derived mice. *Nat. Commun.* **10**, 1–15 (2019).
7. Montoya, J. G. & Liesenfeld, O. Toxoplasmosis. *Lancet* **363**, 1965–1976 (2004).
8. Dubey, J. P. in *Toxoplasma gondii. The Model Apicomplexan: Perspectives and Methods* (eds. Kim, K. & Weiss, L. M.) (Elsevier, 2007).
9. Khaminets, A. *et al.* Coordinated loading of IRG resistance GTPases on to the *Toxoplasma gondii* parasitophorous vacuole. *Cell. Microbiol.* **12**, 939–961 (2010).
10. Martens, S. *et al.* Disruption of *Toxoplasma gondii* parasitophorous vacuoles by the mouse p47-resistance GTPases. *PLoS Pathog.* **1**, 0187–0201 (2005).
11. Zhao, Y. O., Khaminets, A., Hunn, J. P. & Howard, J. C. Disruption of the *Toxoplasma gondii* parasitophorous vacuole by IFN $\gamma$ -inducible immunity-related GTPases (IRG proteins) triggers necrotic cell death. *PLoS Pathog.* **5**, (2009).
12. Reese, M. L., Zeiner, G. M., Saeij, J. P. J., Boothroyd, J. C. & Boyle, J. P. Polymorphic family of injected pseudokinases is paramount in *Toxoplasma* virulence. *Pnas* **108**, 9625–9630 (2011).
13. Saeij, J. P. J. *et al.* Polymorphic Secreted Kinases Are Key Virulence Factors in Toxoplasmosis. *Science (80-. )*. **161**, 1780–1784 (2006).
14. Fentress, S. J. *et al.* Phosphorylation of immunity-related GTPases by a *Toxoplasma gondii* secreted kinase promotes macrophage survival and virulence. *Cell Host Microbe* **8**, 484–495 (2010).
15. Taylor, S. *et al.* A Secreted Serine-Threonine Kinase Determines Virulence in the

- Eukaryotic Pathogen *Toxoplasma gondii*. *Science* (80-. ). **314**, 1776–1781 (2006).
16. Behnke, M. S. *et al.* Rhoptry Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of *Toxoplasma gondii*. *PLOS Genet.* **11**, e1005434 (2015).
17. Jensen, K. D. C. *et al.* *Toxoplasma gondii* Superinfection and Virulence during Secondary Infection Correlate with the Exact ROP5 / ROP18 Allelic Combination. *MBio* **6**, 1–15 (2015).
18. Alvarez, C. *et al.* in *Toxoplasma gondii: Methods and Protocols, Methods in Molecular Biology* (ed. Tonkin, C. J.) 371–409 (Springer US, 2020). doi:10.1007/978-1-4939-9857-9\_20
19. Pierce Biotechnology. Pierce <sup>TM</sup> BCA Protein Assay Kit. *Thermo Scientific* (2000). at <www.thermoscientific.com/pierce>
20. Hassan, M. A., Olijnik, A.-A., Frickel, E.-M. & Saeij, J. P. Clonal and atypical *Toxoplasma* strain differences in virulence vary with mouse sub-species. *Int. J. Parasitol.* **49**, 63–70 (2019).
21. Müller, U. B. Polymorphism in the IRG resistance system determines virulence of *Toxoplasma gondii* in mice. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2015).
22. Gregorová, S. *et al.* Mouse consomic strains: Exploiting genetic domesticus subspecies. *Genome Res.* **18**, 509–515 (2008).
23. Lilue, J. Haplotypic polymorphism of the IRG protein family mediates resistance of mice against virulent strains of *Toxoplasma gondii*. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2012).
24. Lively, C. M. & Dybdahl, M. F. Parasite adaptation to locally common host genotypes. *Nature* **405**, 679–681 (2000).
25. Poltorak, A., Apalko, S. & Sherbak, S. Wild-derived mice: from genetic diversity to variation in immune responses. *Mamm. Genome* **29**, 577–584 (2018).
26. Ideraabdullah, F. Y. *et al.* Genetic and haplotype diversity among wild-derived mouse inbred strains. *Genome Res.* **14**, 1880–1887 (2004).

## Chapter 4

---

### **4. Genetic characterization of South American *Toxoplasma gondii* strains**

Catalina Alvarez<sup>1</sup>, Claudia Campos<sup>1</sup> and Jonathan Howard<sup>1, 2</sup>

<sup>1</sup> Fundação Calouste Gulbenkian, Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal. <sup>2</sup> Institute for Genetics, University of Cologne, 50674 Cologne, Germany.



**Author contributions**

CA - Study concept and design; Mouse tissue collection; Genome analysis; SNP calling analysis; Phylogenetic analysis; Drafting and editing the manuscript.

CC - Study concept and design; Parasite propagation; DNA purification and extraction; PCR-RFLP experiments.

JCH - Study concept and design; Supervision of execution; Drafting and editing the manuscript.

## 4.1. Abstract

South America is considered as the “birthplace” of the parasite *Toxoplasma gondii*. Parasite isolates from this geographic region are genetically diverse and display an atypical population structure, usually without dominant clones. Most of the South American *T. gondii* strains are highly virulent in laboratory mice. Virulent strains have evolved to resist the host's innate immune response by the secretion of virulence factors such as the pseudokinase ROP5 and the kinase ROP18. ROP5 and ROP18 are highly polymorphic proteins considered as the two major virulence factors in *T. gondii* isolates from North and South America. Using a whole-genome analysis approach in the three Brazilian *T. gondii* strains N, BV, and AS28, we found these strains share high levels of genetic similarity. Also, we found that the ROP18 gene and ROP5 isoforms found in them are identical. Moreover, analysis of the ROP5 isoforms identified six different types, all of them previously reported for other virulent strains. Finally, we developed a PCR-RFLP based protocol for the genotyping of the Brazilian strains used for this study that can be potentially used for other *T. gondii* strains in which the standard protocols for genotyping are not sufficient.

## 4.2. Introduction

*Toxoplasma gondii* is an obligate intracellular protozoan parasite that belongs to the Apicomplexa phylum, as do other parasites like *Plasmodium* and *Cryptosporidium*. However, in contrast to other parasites, *T. gondii* has an extraordinarily broad host range, as well as the capacity to infect any nucleated vertebrate cell<sup>1</sup>. *T. gondii* is also distributed world-wide, however, there are regions that display geographically distinct *T. gondii* population structures. Europe and North America show a dominant clonal genotype structure, where almost all isolates belong to four clonal lineages, types I, II, III and 12<sup>2</sup>. East Asia and North Africa have types II and III isolates and regional clonal lineages (e.g. Chinese 1 in China or Africa 1 and Africa 3 in Africa)<sup>3-6</sup>. Central and South America on the other hand have a striking genetic diversity in *T. gondii* isolates<sup>7-10</sup>, there are at least 150 distinct genotypes classified within haplogroups 4 to 16. It is thought that this geographical region is the “birthplace” of the parasite, due to the absence of signs of a recent genetic bottleneck, as well as an atypical population structure without dominant clones<sup>11</sup>.

A majority of South American *T. gondii* strains are reported to be virulent in laboratory mice<sup>12,13</sup>. Recent data showed that 261 out of 427 isolates from South and Central America were highly virulent to laboratory mice<sup>9,10</sup>. On the contrary, only 14 out of 193 isolates from North America displayed a virulent phenotype<sup>9,10</sup>. Acute primary toxoplasmosis in humans is also common in South America. It is still unclear to what extent host-parasite genetics, host immune response, and rate of exposure to the parasite may contribute to the high virulence of South American strains. Nevertheless, differences in the genetic complexity of *T. gondii* strains certainly play an important role<sup>11</sup>.

*T. gondii* genotyping can be done by the analysis of multilocus makers using techniques such as PCR-restriction fragment length polymorphism (PCR-RFLP) or microsatellite (MS) analysis. The standard PCR-RFLP genotyping for *T. gondii* is based on the analysis of 10 markers distributed along 8 chromosomes and one marker from the Apicoplast<sup>6,9,14,15</sup>. The combination between PCR-RFLP patterns and microsatellites is translated into a specific number for each genotype according to the codification adopted by the ToxoDB database (<http://www.toxodb.org/toxo/>)<sup>14,16,17</sup>.

The effector *T. gondii* proteins ROP5 and ROP18 have been identified as the most important virulent factors of the parasite for mice. The combination of ROP18 and ROP5 allele types is highly predictive of *T. gondii* virulence across a broad range of global *T. gondii* isolates<sup>18</sup>. Genetic crosses between type I and type III or between type II and type III strains identified ROP18<sup>19,20</sup>, whereas crosses between type II and III or type I and II strains lead to the identification of ROP5<sup>21,22</sup>. Together these proteins counteract the effect of Immunity Related GTPases (IRGs), which are proteins from the innate immune response of the host<sup>23–25</sup>.

Multiple non-identical tandem copies that vary according to the *T. gondii* genetic background encode the pseudokinase ROP5. Among the Eurasian strains, type I strains have approximately 6 copies, type II strains approximately 10 copies, and type III approximately 4 copies of ROP5<sup>22</sup>. These multiple copies can be classified into three different major isoforms, named ROP5A, ROP5B, and ROP5C. Isoforms display characteristic differences between the virulent (type I and type III) and avirulent (type II) loci. Genetic variations in ROP5 and ROP18 account for 50% to 90% of the virulence *in vivo* in laboratory mouse strains<sup>21,22,25</sup>.

The ROP18 protein is a Serine/Threonine kinase secreted into the host cytosol. It is a major virulence factor and responsible for the differences

between highly virulent type I and avirulent type II strains<sup>20</sup> and between type II and type III<sup>19</sup>. Type III strains are less virulent in mice compared with types I and II, since they do not express ROP18 due to a 2.1kb insertion in the ROP18 promoter<sup>19</sup>. A single copy gene encodes the ROP18 protein, which displays high levels of genetic diversity and signs of strong positive selection<sup>26</sup>. The combination of ROP18 and ROP5 alleles does not only determine strain virulence in North American/European archetypal lineages, but also does it in the diverse isolates from South/central America and Asia<sup>18</sup>.

In conclusion, type I strains are highly virulent due to their ROP5 and ROP18 alleles, while type II strains are less virulent due to avirulent ROP5 alleles and an avirulent ROP18 allele. Finally, type III strains are non-virulent because of a lack of ROP18 expression, even though they have virulent ROP5 alleles.

In this study, we present the full genomic sequences of three mouse-virulent South American *T. gondii* strains (AS28, BV and N) and analyse the genetic diversity found in these previously poorly genetic characterized strains. Whole genome analyses were done to identify SNPs and find genetic markers that could uniquely identify these three strains using a PCR-RFLP approach. Finally, an analysis of the sequences from the major virulence factors from *T. gondii*, the ROP18 and ROP5 genes, were completed to identify possible factors associated with phenotypic differences in *in vivo* experiments with these strains.

## **4.3. Materials and methods**

### **4.3.1. Cell culture maintenance and propagation of *Toxoplasma gondii* strains**

The origin and characteristics of all *T. gondii* strains used for this study are described in the key sources section of this thesis. *T. gondii* strains were maintained by serial passage in confluent monolayers of human Hs27 foreskin fibroblasts (ATCC). Due to adaptation to laboratory conditions or recent isolation, *T. gondii* strain growth rate *in vitro*<sup>27</sup> changes (time between infection and host cell lysis is variable); therefore adjustments in passaging time were made based on the specific *T. gondii* strain growth rate. When parasites were ready to be harvested, host cells were detached with a cell scraper and the suspension passed through a 25G x 5/8" needle to lyse host cells and free *T. gondii* tachyzoites. The parasite/host cell suspension was placed in a 15 ml Falcon tube for further procedures.

#### **4.3.2. Separation and purification of *Toxoplasma gondii* tachyzoites**

*T. gondii* tachyzoites suspensions usually contain a large number of host cells and debris. Proper purification of *T. gondii* tachyzoites may be important for the success of subsequent experiments. The human host cell genome with 3,100 Mbp is approximately 50x larger than the *T. gondii* genome at 65 Mbp. Thus even minor contamination with host cell gDNA can interfere with molecular biology procedures such as PCR and genome sequencing<sup>27,28</sup>.

Wu et al., 2012 described a purification protocol using an  $\alpha$ - cellulose matrix (CF-11 - Sigma-Aldrich®). We modified this protocol to increase the yield of purified *T. gondii* tachyzoites from human host cells, while simultaneously decreasing the contamination of human genomic DNA.

CF-11 was hydrated overnight at 4°C in 1X PBS. We prepared syringe barrels with a thin cotton wool sinter and layered it with 2 cm of hydrated CF-11. The column was washed twice with 1X PBS and allowed to drain until approximately 1 mm of liquid was left on top of the CF-11 column. The parasite/host cell suspension in a 15ml falcon tube was spun down for 3

minutes at 30 xg to remove large debris. Supernatant was collected and pellet discarded. A re-centrifugation of the supernatant was performed at 700xg for 15 minutes at RT. During this time, the CF-11 column was washed twice with 6ml of DMEM 0% FBS and allowed to drain. From the centrifugation process, supernatant was discarded and the tachyzoite pellet was re-suspended in 4ml of DMEM with 0% FBS. A new collection tube with 120 µl of FBS was prepared and placed underneath the column. A total of 4ml of the re-suspended tachyzoites suspension was filtered through the CF-11 column and collected in the falcon tube with FBS. Subsequently, the column was washed with 2ml of DMEM 0% FBS (without FBS) to increase the yield of tachyzoite recovery. The filtered suspension was spun down at 700 xg for 15 minutes. The remaining pellet was re-suspended in 1ml of PBS+2% FBS and transferred to an eppendorf tube. **Note:** before re-suspension, the pipette tip was coated by pipetting twice the solution of PBS+2%FBS. Re-suspended pellet was centrifuged at 1000 xg for 15 minutes and the supernatant discarded.

To remove residual DNA from the medium, the purified tachyzoite pellet was re-suspended in 200 µl of DNAase solution (DNA-free™ Kit - Ambion®) (179 µl 1X PBS + 20 µl of 10x DNase buffer + 2 µl of TurboDNase) and digested for 30 minutes at 37°C. Later, 800 µl of 1X PBS was added. DNase activity was stopped by adding 10 µl of EDTA 0.5 M and subsequently incubating at 75°C for 10 minutes. The solution was centrifuged at max speed for 3 minutes and pellet washed twice with 1X PBS. Supernatant was discarded and the pellet, containing the purified *T. gondii* tachyzoites, was frozen at -20°C until further analysis.

#### **4.3.3. Genomic DNA isolation**

Total DNA from purified tachyzoites was extracted using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's protocol for cultured

cells (Spin-Column protocol). To increase DNA yield, before the precipitation step with EtOH, samples were incubated at 56°C for 10 minutes. DNA was eluted in 100 µl of Buffer EB (10 mM Tris·Cl, pH 9.0) and stored at -80°C.

#### **4.3.4. Validation of *Toxoplasma gondii* tachyzoites purification**

Two different PCR protocols for the amplification of human specific genes were used to assess the presence of DNA of the human host cells and the efficiency of our modified protocol for the purification of *T. gondii* tachyzoites from human host cells.

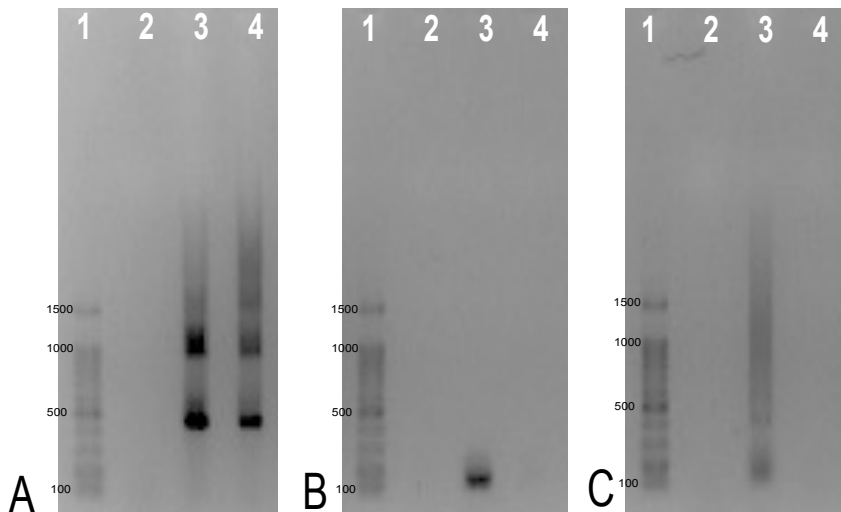
Amplification of human mitochondrial DNA has often been used to determine contamination with human DNA in samples from other organisms. In this case, primers L16124F and H16227R reported by reference <sup>29</sup>, were used for the amplification of a fragment of the human mtDNA in DNA from purified *T. gondii* tachyzoites.

Alu elements are short interspersed elements (SINEs) of approximately 300 nucleotides in length and primate-specific. There are more than 1 million Alu copies in the human genome<sup>30</sup>. Therefore, the amplification of Alu copies is an accurate approach to determine the presence of small amounts of human DNA in samples from other organism. One set of primers, 101 F and 206 R, were used for the amplification of Alu elements from the human 7SL RNA gene<sup>30</sup>.

To assess integrity of *T. gondii* DNA, we used primers Tox-5 and Toxo-8 were used. This set of primers targets the repetitive 529bp region of the *T. gondii* genome. The DreamTaq™ Green DNA Polymerase (Fermentas GmbH) was used for the three PCR protocols described above. The PCR protocol was as follows: 96 °C 5 min, 25x (96 °C 30 sec, 59 °C 15 sec, 72 °C 1 min), 72 °C 10 min, 8°C o.n.



The modified CF-11 cellulose protocol allowed the purification of *T. gondii* genomic DNA, while removing traces of human host cells and human DNA (Figure 4.1, panel A-C).



**Figure 4.1.** Validation of *T. gondii* tachyzoite purification by CF-11 cellulose column. A) Amplification of the 529 bp fragment from *T. gondii* using the primers Tox-5 and Tox-8. Positions: 1. 100bp DNA ladder; 2. Negative control (No DNA); 3. Non-purified Tachyzoites/host cells solution; 4. Purified Tachyzoites/host cells solution after CF-11 cellulose/DNase protocol. Expected bands: 400bp and 900bp. B) Amplification of Human mtDNA using the primers L16124F and H16227R. Positions: 1. 100bp DNA ladder; 2. Negative control (No DNA); 3. Non-purified Tachyzoites/host cells solution; 4. Purified Tachyzoites/host cells solution after CF-11 cellulose/DNase protocol. Expected band: 200bp. C) Amplification of Alu elements from the human 7SL RNA gene using the primers 101 F and 206 R. Positions: 1. 100bp DNA ladder; 2. Negative control (No DNA); 3. Non-purified Tachyzoites/host cells solution; 4. Purified Tachyzoites/host cells solution after CF-11 cellulose/DNase protocol I. Expected band: 150bp.

#### 4.3.5. PCR-RFLP *Toxoplasma gondii* genotyping

There are few genotypes in *T. gondii* strains isolated from humans and animal sources in Europe and North America. This contrasts with the coexistence of hundreds of genotypes in strains isolated in South

America<sup>9,14</sup>. Proper genotyping of *T. gondii* strains is crucial to relate *in vivo* or *in vitro* phenotypes with genetic characteristics.

A well-known multiplex multilocus nested PCR-RFLP (Mn-PCR-RFLP) typing strategy was used for the distinction of three South American strains (AS28, BV and N). A total of 12 genetic markers have been reported to provide high resolution for *T. gondii* genotyping<sup>31</sup>. In this study two markers (Alt. SAG2 and L358) were analyzed in the South American strains. Selection of these markers was based on a personal communication by Dr. Chunlei Su (University of Tennessee, USA), who suggested those as the best markers for the distinction of South American *T. gondii* strains. Multiplex multilocus nested PCR and subsequent RFLP analysis were done following the protocol described by previous publications<sup>31</sup>

#### **4.3.6. *Toxoplasma gondii* whole genome sequencing**

Whole genome sequencing using Illumina technology was done in three genetically uncharacterized Brazilian *T. gondii* strains (AS28, BV and N) and the laboratory adapted ME49 strain (Control). Library construction and whole genome sequencing from total DNA were done at the Genomics unit of the Instituto Gulbenkian de Ciência (Portugal). The Illumina NextSeq 500 machine was used with an average coverage of 200X.

#### **4.3.7. Genome and SNP calling analyses**

150 bp paired-end Illumina-reads were mapped against the reference genome ME49 (ToxoDB-version 44) using the Burrows-Wheeler Aligner (BWA – Heng Li, Broad Institute). The algorithm applied was BWA-MEM, which is a fast and highly accurate mapper for long Illumina reads. It does not take into account CNV. To obtain a list with all polymorphic sites and their corresponding position, mapped BAM files were subjected to a SNP

calling analysis using the softwares samtools-1.9 (mpileup option) and bcftools 1.9 (call option).

#### **4.3.8. PCR-RFLP design for Brazilian *Toxoplasma gondii* strains**

SNP calling analysis results were used as an input for the identification of differential sites between AS28, BV and N. Regions containing these differential SNPs were identified and primers designed for their corresponding PCR amplification. Selection of the regions to be amplified, as well as the most suitable enzymes for the design of a PCR-RFLP panel to distinguish between the three strains was done with the Lasergene Molecular Biology software, option seqbuilder Pro (DNASTAR Inc). Four SNPs located in the genes TG\_293820, TG\_294270, TG\_321410 and TG\_264740 were chosen.

The DreamTaq™ Green DNA Polymerase (Fermentas GmbH) was used for all PCR reactions. The PCR protocol was as follows: 96 °C 2 minutes, 35x (96 °C 30 sec, 58 °C 1 minute, 72 °C 1 minute), 72 °C 5 minute, 8°C o.n.

The amplification of the TG\_293820 gene was done with the primers 293820\_N\_Fwd and 293820\_N\_Rev and the enzymatic digestion for the PCR-RFLP with the enzyme HinfI at 37°C for 1 hour. For the TG\_294270 and TG\_321410 genes, primers 294270\_BV\_Fwd and 294270\_BV\_Rev and 321410\_AS\_N\_Fwd and 321410\_AS\_N\_Rev were respectively used. Enzymatic digestions for both genes were done with the enzyme BsaJI at 60°C for 1 hour. Finally, amplification of the TG\_264740 gene was done with the primers 264740\_AS\_Fwd and 264740\_AS\_Rev and the enzymatic digestion for the PCR-RFLP with the enzyme AclI at 37°C for 1 hour.

#### **4.3.9. ROP5 and ROP18 gene sequence analysis**

Several studies based on gene knockout and forward genetics agree that *T. gondii* rhoptry protein 5 and 18 (ROP5 and ROP18) are the most potent mouse virulence factors identified<sup>32</sup>.

It has been shown as well that unlike among other known virulence genes, the ROP5 locus has been considerably expanded. This single-exon effector gene displays multiple tandem copies (isoforms) with different copy number and different alleles among different *T. gondii* strains. Genetic characterization of these different ROP5 isoforms is crucial to understand in vivo and in vitro virulence patterns in *T. gondii* strains<sup>22</sup>.

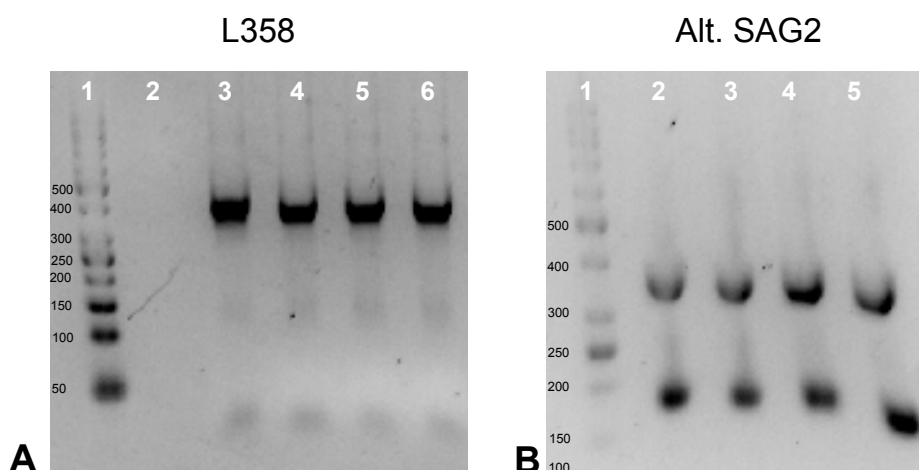
A homology-based approach using an artificial chromosome representing the ORFs of 6 ROP5 isoforms from avirulent (ME49 strain – ToxoDB database) and 10 isoforms from virulent (RH – Dr. Jon Boyle by Nanopore and PacBio reads)<sup>33</sup> strains arranged 5' to 3' with 20 “N” gaps in between single ORFs was created. Geneious® 9.1.5 (Biomatters, New Zealand) software was used to map a trimmed fragment (10kb) from the Chromosome XII that contains the ROP5 isoforms against the artificial chromosome created.

ROP18 is a single-exon secreted serine-threonine kinase of which both expression level and allelic differences contribute to variation in acute virulence in infection of mice with *T. gondii* strains<sup>20,32</sup>. Sequences from the ROP18 gene were extracted from the mapped BAM file from each of the *T. gondii* strains analyzed. Allelic differences were defined based on the sequences reported for Type I, II and III *T. gondii* strains.

## **4.4. Results**

### **4.4.1. Commonly used PCR-RFLP markers for *T. gondii* genotyping are not suitable for the distinction of some Brazilian strains.**

In the previous chapter of this thesis, multiple experiments *in vivo* and *in vitro* were done with the Brazilian *T. gondii* strains AS28, BV and N. However, since infections were done simultaneously with more than one of these strains, the need for constant confirmation of the genotypes was clear. We used two markers for genotyping, Alt. SAG2 and L358, reported to be suitable for the discrimination of South American strains (Figure 4.2.) To our surprise, none of the markers used displayed a differential pattern that could allow us to distinguish between the three strains. Type I (ToxoDB #10) RH strain was used as a control, and all three Brazilian strains presented the same band pattern as RH.



**Figure 4.2.** Multiplex multilocus nested PCR-RFLP (Mn-PCR-RFLP) analysis of *Toxoplasma gondii* samples using two gene markers. A) L358 marker. Positions: 1. 50bp DNA ladder; 2. Negative control (No DNA); 3. AS28 strain; 4. BV strain; 5. N strain; 6. RH strain. Enzymes HaeIII+NlaIII (double digest). Type I strain expected band size ( $\approx 350$  bp). B) Alt. SAG2 marker. Positions: 1. 50bp DNA ladder; 2. AS28 strain; 3. BV strain; 4. N strain; 5. RH strain. Enzymes HinfI+TaqI. Type I expected bands size ( $\approx 175$ bp and  $\approx 375$  bp).

#### **4.4.2. Brazilian *T. gondii* strains AS28, BV and N can be differentiated by their distinct genetic backgrounds.**

The genomes of AS28, BV, and N have not been previously sequenced. Having identified differences in the disease phenotypes obtained upon infection with these strains, we expected to find genomic sequence variation that could be used to differentiate the strains.

Mapping against the reference strain ME49 of the three Brazilian *T. gondii* strains showed the following percentages of reads mapped, for the BV strain 99.36% (46,096,232 reads), for the N strain 99.31% (102,288,319 reads) and for the AS28 strain 99.20% (84,299,295 reads). After paired-end Illumina-reads mapping, a SNP calling analysis was done in the three genomes. We found a total of 57 SNPs between the three strains (Table 4.1). They were evenly distributed along the 14 chromosomes and approximately 60 Mb analyzed. 22 of the SNPs map to annotated genes and encoded proteins with a wide variety of functions, based on predicted conserved domains (ToxoDB v. 46, Nov 2019). This set of SNPs contained genes coding for proteins associated with ubiquitination pathways (ubiquitin transferases and fusion degradation proteins), kinases, RNA polymerases, etc. From the remaining SNPs, 24 are located in genes that encoded for hypothetical proteins and 11 more were located in non-coding regions (Table 4.1).

**Table 4.1.** SNPs between AS28, BV, N and ME49 *Toxoplasma gondii* strains.

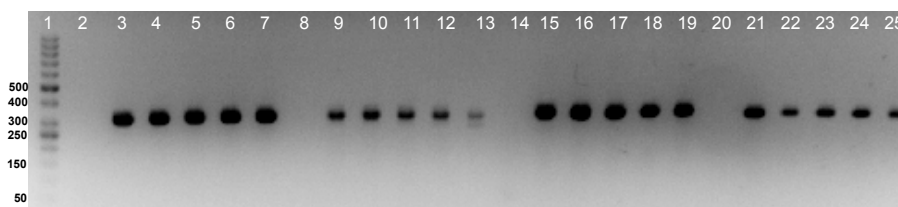
\*Positions used for the PCR-RFLP analysis design.

Chromosome	Gene name	Gene description	Position	<i>Toxoplasma gondii</i> strain			
				AS28	BV	N	ME49
ChrIa	TGME49_293820	Calpain family cysteine protease domain-containing protein	579316*	C	C	G	C
ChrIa	TGME49_294270	Histone arginine methyltransferase PRMT4/CARM1	766344*	C	G	C	C
ChrIa	TGME49_294400	Hypothetical protein	889587	G	Indel	G	A
Chrlb	TGME49_321410	Hypothetical protein	1830220*	G	C	G	C
ChrlI	TGME49_221700	Hypothetical protein	581782	A	G	A	A
ChrlI	TGME49_221700	Hypothetical protein	581821	A	G	A	A
ChrIII	TGME49_254370	Guanylyl cyclase	1657010	C	G	G	G
ChrlX	TGME49_264740	Hypothetical protein	2075570*	C	A	A	A
ChrlX	TGME49_264650	Phosphoacetylglucosamine mutase	2148655	G	G	A	G

ChrIX	TGME49_289370	Hypothetical protein	3240461	T	C	T	C
ChrIX	Non-coding region	-	5408869	A	G	A	G
ChrV	Non-coding region	-	1478445	T	G	G	G
ChrV	TGME49_285700	Ubiquitin fusion degradation protein UFD1AP	2282757	C	A	A	A
ChrV	TGME49_285680	Dihydrolipoamide acyltransferase, putative	2292377	G	C	G	C
ChrV	Non-coding region	-	2587920	G	A	A	A
ChrV	Non-coding region	-	101497	A	G	G	G
ChrV	Non-coding region	-	1340991	T	A	A	A
ChrVI	TGME49_240770	Cytochrome b5 family heme/steroid binding domain-containing protein	1439651	A	G	A	G
ChrVI	TGME49_288440	NEK kinase	235174	T	C	C	C
ChrVI	Non-coding region	-	2157385	C	G	G	G
ChrVI	TGME49_243430	OTU family cysteine protease	2627969	C	A	C	A
ChrVI	TGME49_238390	Hypothetical protein	186979	C	C	T	C
ChrVIIa	TGME49_280660	HECT-domain (ubiquitin-transferase) domain-containing protein	226764	G	G	A	G
ChrVIIa	TGME49_204120	Hypothetical protein	1975548	G	C	G	C
ChrVIIa	TGME49_203050	AP2 domain transcription factor AP2VIIa-6	2830357	T	C	T	C
ChrVIIa	TGME49_202120	Hypothetical protein	3597345	T	T	G	T
ChrVIIa	TGME49_281980	Phosphatidate cytidyltransferase	4303894	A	T	A	T
ChrVIIa	TGME49_201390	Hypothetical protein	3909373	A	T	A	T
ChrVIIb	Non-coding region	-	476394	G	C	G	G
ChrVIIb	TGME49_261690	Hypothetical protein	1577017	G	T	G	G
ChrVIIb	TGME49_260570	Hypothetical protein	2154371	A	G	A	A
ChrVIIb	TGME49_257070	Hypothetical protein	4244915	G	A	G	G
ChrVIII	TGME49_230570	Transporter, major facilitator family protein	863430	T	T	C	T
ChrVIII	TGME49_274130	TBC domain-containing protein	3007472	C	T	T	T
ChrVIII	Non-coding region	-	3605665	C	T	C	T
ChrVIII	TGME49_270150	Hypothetical protein	5383515	T	G	T	G
ChrX	TGME49_224870	Hypothetical protein	2760520	T	C	T	C
ChrX	TGME49_214545	Hypothetical protein	6350780	T	T	A	T
ChrX	Non-coding region	-	3929676	C	A	A	A
ChrX	TGME49_207080	Histone lysine acetyltransferase MYST-B	7379305	C	A	C	C
ChrXI	TGME49_311220	Hypothetical protein	1875788	T	C	C	C
ChrXI	TGME49_310910	WD domain, G-beta repeat-containing protein	1667896	C	G	C	C
ChrXI	TGME49_311360	Protein kinase G AGC kinase family member PKG	2001193	A	A	G	A
ChrXI	Non-coding region	-	3146013	T	G	T	T
ChrXI	TGME49_313770	Hypothetical protein	3684860	G	G	A	G
ChrXI	TGME49_315560	ATP-binding cassette G family transporter ABCG77	4809945	T	C	C	C
ChrXI	TGME49_315980	EREBP-4 family protein	5116644	T	G	T	G
ChrXI	TGME49_217040	Hypothetical protein	5711338	G	A	A	A
ChrXI	TGME49_215910	Hypothetical protein	6532330	T	G	T	T
ChrXII	TGME49_218880	SF-assemblin, putative	1240466	G	G	A	A
ChrXII	TGME49_217580	DNA-directed RNA polymerase III RPC6	1926464	T	G	T	T
ChrXII	TGME49_247600	Hypothetical protein	3438189	C	G	G	G
ChrXII	TGME49_249370	Translation initiation factor SU11, putative	4499335	G	A	G	A
ChrXII	TGME49_249450	Hypothetical protein	4545836	A	G	A	G
ChrXII	TGME49_251890	Hypothetical protein	5647166	T	C	C	C
ChrXII	Non-coding region	-	6168791	C	G	C	C
ChrXII	TGME49_247600	Hypothetical protein	3438192	G	A	A	A

#### 4.4.3. A new PCR-RFLP assay allows the genotyping of the Brazilian AS28, BV, and N strains.

As previously mentioned, we were not able to genotype the three Brazilian strains by using previously reported markers for PCR-RFLP in *T. gondii*<sup>31</sup>. Therefore, we decided to design primers to amplify four differential SNPs between the three strains (Table 4.1). The first gene selected was the TGME49\_293820, where a non-synonymous SNP found in position 579316 (ChrIa) allows the distinction of the N strain. We also chose the TGME49\_294270 gene; the non-synonymous SNP found in this position (766344 - ChrIa) allows the distinction of the BV strain. Finally, two more genes, the TGME49\_321410 in position 1830220 (Chrlb), which allows the distinction of the N and AS28 strains, and the TGME49\_264740 in position 2075570 (ChrIX) for the distinction of the AS28 strain were selected.

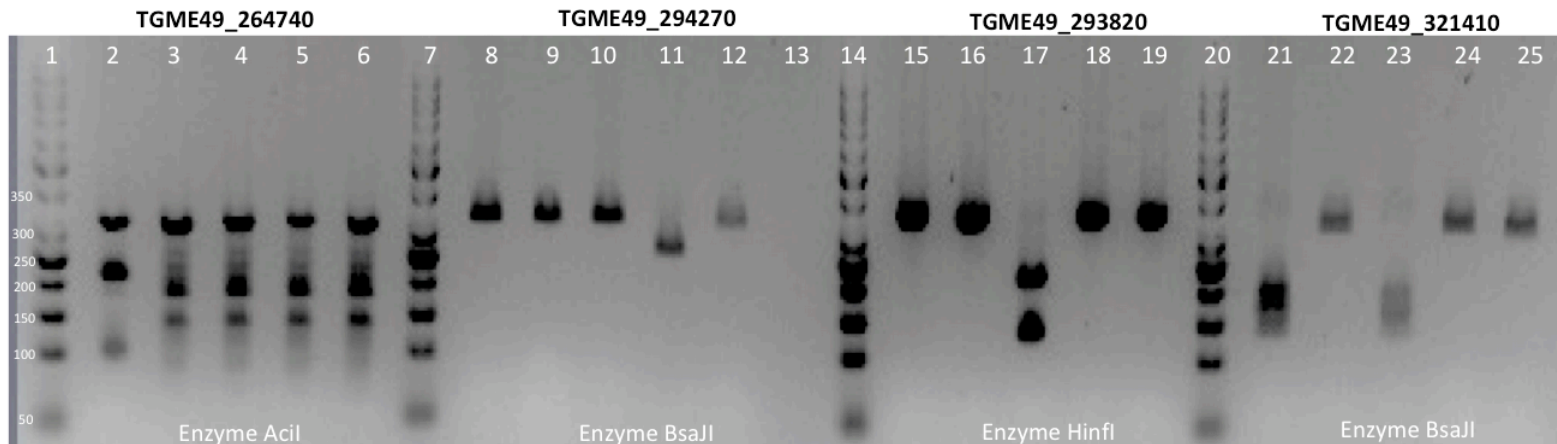


**Figure 4.3.** PCR amplification of *Toxoplasma gondii* samples using four gene markers. Positions 2 – 7: TGME49\_264740 gene. 1. 50bp DNA ladder; 2. Negative control (No DNA); 3. AS28 strain; 4. TgMmBr01 strain; 5. N strain; 6. BV strain; 7. ME49 strain. Expected amplicon: 325 bp. Positions 8 – 13: TGME49\_294270 gene. 8. Negative control (No DNA); 9. AS28 strain; 10. TgMmBr01 strain; 11. N strain; 12. BV strain; 13. ME49 strain. Expected amplicon: 321 bp. Positions 14 – 19: TGME49\_293820 gene. 14. Negative control (No DNA); 15. AS28 strain; 16. TgMmBr01 strain; 17. N strain; 18. BV strain; 19. ME49 strain. Expected amplicon: 351 bp. Positions 20 – 25: TGME49\_321410 gene. 20. Negative control (No DNA); 21. AS28 strain; 22. TgMmBr01 strain; 23. N strain; 24. BV strain; 25. ME49 strain. Expected amplicon: 331 bp.

PCR amplification for the four genes selected was done (Figure 4.3). All of them gave us amplicons between 300 to 400 bp, as expected. PCR products were later subjected to enzymatic digestion (Figure 4.4). Gene



TGME49\_264740 upon digestion with the enzyme *Acil* displayed a two-band pattern for the AS28 strain (120, 230 and an undigested band of 321 bp). For the gene TGME49\_294270, restriction with *Bsa*JI led to a two-band pattern (a faint band of 65 bp and clear band of 256 bp) for the BV strain. Restriction with the *Hinfl* enzyme in the TGME49\_293820 gene results in two bands (130 and 220 bp) for the N strain. We used again the enzyme *Bsa*JI to obtain a two-band pattern (150 and 180 bp) for the strains AS28 and N.

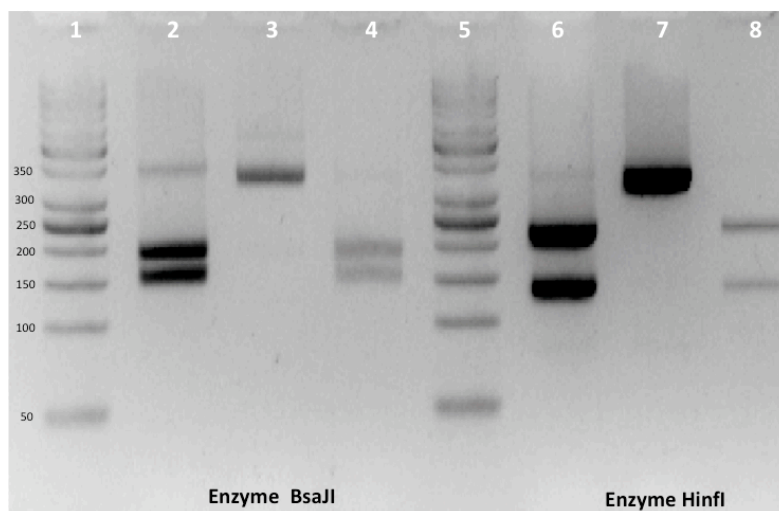


**Figure 4.4.** PCR-RFLP of *Toxoplasma gondii* samples using four gene markers. Positions 1 – 6: TGME49\_264740 marker to distinguish AS28 strain. 1. 50bp DNA ladder; 2. AS28 strain; 3. TgMmBr01 strain; 4. N strain; 5. BV strain; 6. ME49 strain. Enzyme Acil. Positions 8 – 13: TGME49\_294270 marker to distinguish BV strain. 7. 50bp DNA ladder; 8. AS28 strain; 9. TgMmBr01 strain; 10. N strain; 11. BV strain; 12. ME49 strain; 13. Negative control (No DNA). Enzyme BsaJI Positions 14 – 19: TGME49\_293820 marker to distinguish N strain. 14. 50bp DNA ladder; 15. AS28 strain; 16. TgMmBr01 strain; 17. N strain; 18. BV strain; 19. ME49 strain. Enzyme HinfI Positions 20 – 25: TGME49\_321410 marker to distinguish AS28 and N strains. 20. 50bp DNA ladder; 21. AS28 strain; 22. TgMmBr01 strain; 23. N strain; 24. BV strain; 25. ME49 strain. Enzyme BsaJI.

#### 4.4.4. New PCR-RFLP assay allows the distinction of Brazilian *T. gondii* strains using DNA extracted from tissue samples.

We tested the new markers for the PCR-RFLP identification of Brazilian *T. gondii* strains in tissues isolated from infected mice. We used isolated DNA from the brain of a MANA mouse chronically infected with the N strain to test the BsaII TGME49\_321410 marker. As previously mentioned, this marker allows the distinction of the AS28 and N strains from other virulent strains. Upon amplification and enzymatic treatment, DNA from the infected brain displayed a two-band pattern that matched with the control DNA (isolated from cell-cultured tachyzoites) from the N strain (**Figure 4.5**). We also used as a control DNA from the RH strain, which unlike the N strain, displayed a single band pattern.

An additional marker, the HinfI TGME49\_293820 marker was also tested. Isolated DNA from a mouse infected with the N strain displayed a two-band pattern that corresponds to the expected bands for the N strains. This two-band pattern can also be observed in the DNA control for the N strain (**Figure 4.5**).



**Figure 4.5.** PCR-RFLP of *Toxoplasma gondii* samples using two gene markers to distinguish N strain in infected mouse brains. Positions 1 – 4: TGME49\_321410 marker to distinguish AS28 and N strains. 1. 50bp DNA ladder; 2. N strain; 3. RH strain; 4. mouse brain

infected with N strain; Enzyme BsaJI. Positions 5 – 8: TGME49\_293820 marker to distinguish N strain. 1. 50bp DNA ladder. 2. N strain; 3. RH strain; 4. mouse brain infected with N strain; Enzyme HinfI.

#### **4.4.5. Genetic diversity in virulence factors of the Brazilian strains AS28, BV, and N.**

##### ***4.4.5.1. ROP5 isoforms in Brazilian strains display characteristics of highly virulent genetic backgrounds.***

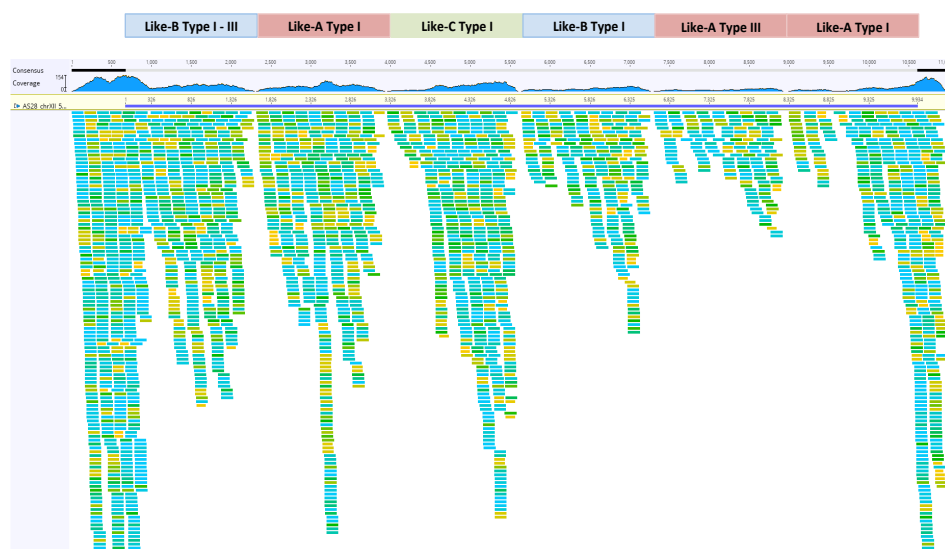
The Brazilian strains AS28, BV, and N have been associated with a virulent phenotype in laboratory and wild-derived mice, as well as with a strong IFN $\gamma$  mediated control in *in vitro* assays (Chapter III – This thesis). Nonetheless, nothing is known about the genetic diversity of virulence genes in these strains.

The ROP5 protein is one of the most important virulence factors of *T. gondii*. Isoforms of this gene are classified into three distinctive categories, A, B and C. The Steinfeldt and Howard labs have recently shown that the ROP5B isoform in the RH strain strongly binds to the IRGB2-B1 protein from the CIM mouse strain<sup>34</sup>. This binding affinity defines in the CIM mouse its resistance against infection with highly virulent Type I strains, like RH.

The full genome sequencing of the three Brazilians strains AS28, BV, and N allowed us to extract their ROP5 isoform sequences. The RH genome sequenced and annotated by Dr. Jon Boyle using long reads (Nanopore and PacBio technology) was used as a reference for ROP5 isoforms due to its high level of curation for polymorphic genes such as ROP5 in relation to other *T. gondii* strains reported only sequenced by Illumina<sup>33</sup>. Initial analysis in the AS28 strain showed 6 ROP5 isoforms (**Figure 4.6**). These isoforms display characteristics of highly virulent *T. gondii* strains, such as those found in Type I but also in Type III and South American strains. We specifically found three A-like (**Supplemental Figure 4.1; panels B, E and F**),

two B-like (**Supplemental Figure 4.1; panels A and D**) and one C-like (**Supplemental Figure 4.1, panel C**) isoforms. However, we do not exclude the possibility that more isoforms could be present in this strain.

Surprisingly, after the same analysis performed in the BV and N strains, we found that all three strains are carriers of the same ROP5 isoforms. Sequences with 100% of identity were found in AS28, BV, and N. These isoforms are not the same as the ones present in the RH reference genome.

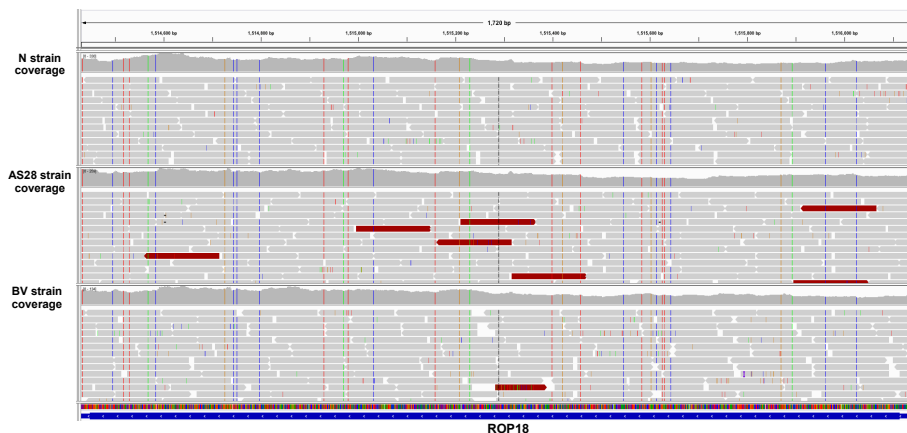


**Figure 4.6.** Isoforms of the ROP5 gene in Brazilian *Toxoplasma gondii* strains. An homology-based approach using an artificial chromosome representing the ORFs of several ROP5 isoforms from avirulent (ME49) and virulent (RH) strains was used to characterize ROP5 isoforms. Depth of coverage of sequence reads for the ROP5 gene mapped to the six isoforms found in the AS28 strain genome. Isoforms found in the AS28, BV and N strain were identical and they are depicted as color blocks in the top of the figure. Localization of the different ROP5 isoforms does not represent their real genomic organization. Figure modified from Geneious® 9.1.5 (Biomatters, New Zealand) software.

#### 4.4.5.2. The AS28, BV, and N strains are carriers of identical ROP18 sequences.

Allelic differences in the ROP18 gene are directly associated with virulence in different *T. gondii* strains. In particular, among South American strains

there is a high level of polymorphisms in this gene. We extracted the ROP18 sequences from the AS28, BV, and N strains. Our findings show that, similarly to the ROP5 gene, these three strains share the same sequence in their ROP18 gene (Figure 4.7).



**Figure 4.7.** ROP18 gene in the AS28, BV and N *Toxoplasma gondii* strains. Depth of coverage of sequence reads for the AS28, BV and N strain mapped against the reference genome ME49 (ToxoDB-version 44). Grey horizontal bars represent paired end reads, which mates are in the same chromosome to the one shown in the image. Red horizontal bars represent paired end reads, which mates are in another chromosome to the one shown in the image.

paired end reads that are coded by the chromosome on which their mates can be found

This sequence displays the polymorphic sites associated with highly virulent *T. gondii* strains. After a BLAST analysis, the ROP18 sequence found in AS28, BV and N strains has an identity higher than 99% for strains such as the Type I GT1 and the South American strains CAST, GUYKOE, TgCatBR18, TgCatBr5, etc. Also, a phylogenetic analysis done in ROP18 protein sequences from multiple strains showed that the ROP18 protein sequence found in this study, clusters with strains that are highly virulent for laboratory mice (Figure 4.8).



**Figure 4.8.** Phylogenetic analysis of ROP18 sequences from multiple *T. gondii* strains. Unrooted phylogenetic tree from ROP18 protein sequences generated in MEGA 6 using Jukes-Cantor genetic distance model and neighbor-joining building method with no outgroup and bootstrap resampling method (500 replicates). Only bootstrap values above 50 are shown.

## 4.5. Discussion

South American *T. gondii* strains show a much greater genetic diversity than their North American, European, Asian, and African counterparts, where the parasite's population structure is dominated by a few prevalent clonal strains<sup>35–37</sup>. They are also characterized by their lack of signs of a recent genetic bottleneck and it is thought that the most recent common ancestor of modern *T. gondii* populations emerged 1.5 Ma ago in the Amazonian

forest<sup>6,9,11,17</sup>. We have done full genome analyses of three Brazilian *T. gondii* strains used for our *in vivo* and *in vitro* experiments on the resistance to local strains in Brazilian mice (Chapter III, this thesis), but which were previously poorly genetically characterized<sup>38,39</sup>.

The three strains analyzed in this study are characterized by being atypical (i.e not types I, II or III) and highly virulent in laboratory mice. They were isolated from different sources (mouse, sheep, and rabbit), geographic locations (São Paulo and Minas Gerais states), and timing of isolation (1952, 1969, and 1975)<sup>38</sup>. Therefore, we expected a high diversity at the genome level. However, we found that these three strains display high levels of genetic similarity all along the genomes. Even though it is known that *T. gondii* strains isolated from different geographic areas might show a high degree of genetic similarities<sup>9</sup>, this was an unexpected finding considering the particularly high diversity found among other Brazilian strains, where the presence of multiple genotypes in Brazilian wildlife and domestic animals in various regions of the country is common<sup>36,40</sup>. Nonetheless, they displayed differences in genes involved in ubiquitination pathways (ubiquitin transferases and fusion degradation proteins), kinases, RNA polymerases, etc. They all came from the same laboratory in Brazil, which does not remove the possibility of a mix up of these strains. Further analyses are necessary on the population structure of these strains and others in the same geographic area to identify how regionally or locally they are diversifying.

The ROP5 and ROP18 genes are the major murine virulence factors in South American strains<sup>12</sup>. We characterized both ROP5 and ROP18 genes in the AS28, BV, and N strains. For the ROP5 gene, we found that these three strains shared the same isoform types and their corresponding sequences. The characterization of ROP5 isoforms is crucial because they directly impact *T. gondii*'s virulence<sup>22</sup>. In this case, all isoforms characterized



(three A-like, two B-like, and one C-like) phylogenetically clustered with type I, III, and some haplogroups commonly associated with virulent strains (including strains from “exotic” haplogroups such as the CPHT and P89). This means that all ROP5 isoforms found in these three Brazilian strains have been already reported in others strains.

Recent studies have shown the predominant responsibility of the ROP5B isoform from the RH strain for the virulence in laboratory mice of Eurasian *T. gondii* strains. Also, it is been shown that Irgb2-b1<sub>CIM</sub> protein form is the direct antagonist of ROP5B in CIM mice<sup>34,41</sup>. Nevertheless, ROP5 isoforms from the South American strain VAND are not targeted by Irgb2-b1<sub>CIM</sub>. The three South American strains analyzed for this study are also virulent in CIM mice<sup>42</sup>, and future studies must be done on the ROP5B isoforms from these strains to know whether they are targeted by Irgb2-b1<sub>CIM</sub>. Most importantly, functional analysis (e.g. Yeast-two-hybrid experiments) with the ROP5 isoforms identified in this study should be done to reveal their contribution in the resistance to local strains observed in Brazilian wild-derived mice (Irgb2-b1<sub>PWK</sub>) (Chapter III, this thesis). Nonetheless, we are aware that more isoforms may be present in these strains, but we could not identify them given the limitations for the characterization of isoforms when using short-read sequencing technologies (e.g. Illumina)<sup>33,43</sup>. Finally, it is known that complementation of *T. gondii* ROP5KOs with certain isoforms can rescue their virulent phenotypes. However, ROP5 isoforms can be differentially expressed and that determines the capacity of a *T. gondii* strain to evade the IRG system<sup>43</sup>. Therefore, it would be interesting to do studies on the expression level of the ROP5 isoforms characterized on a cell infection context, as well as the identification of possible variants in the upstream promoter region that could affect their expressions.

For the ROP18 gene, we identified the same allele in all three strains (AS28, BV, and N). This allele has the typical characteristics of a virulent ROP18

allele (there is not presence of a 2.1kb insertion in the ROP18 promoter) and phylogenetically clusters with *T. gondii* strains that are carriers of ROP18 alleles 1 and 4 (strongly associated with virulent South American isolates<sup>18,44,45</sup>). Although the ROP18 allele type alone is not a strong predictor of virulence in mice, the combined ROP18/ROP5 allele type is a strong predictor. Additionally, to look at both allele types can reveal the segregation of different alleles between the two loci and their possible influence on murine virulence by preventing a proper function of the IRG system<sup>18,25</sup>. Our results showed that the Brazilian strains AS28, BV, N strains have a virulent combination of ROP18 and ROP5 alleles which are in agreement with previous studies that report a high proportion of this combination in South American *T. gondii* strains<sup>18</sup>.

Current genetic markers for *T. gondii* genotyping might not be sufficient to accurately identify some South American *T. gondii* strains<sup>9</sup>. We were not able to distinguish the three Brazilian strains necessary for several of the *in vitro* and *in vivo* experiments done for this thesis using conventional genotyping markers. Therefore, we developed a PCR-RFLP based protocol for the genotyping of these strains. An SNP calling analysis revealed that several SNPs among these strains could be potentially used to genotype them accurately. Our approach was sufficient to distinguish each strain accurately, as well as to be sensitive enough even in samples from infected animals (mouse brain). We propose that more studies should be done to evaluate the potential of these new markers for the genotyping of other *T. gondii* strains.

## **4.6. Acknowledgments**

I would like to acknowledge previous and current members of the Howard and Teixeira laboratories for their constant feedback. Thanks to the IGC facilities that contributed to the development of this study, in particular the

Animal Facility, supported by the research infrastructure Congento, project LISBOA-01-0145-FEDER-022170, Genomics facility and to the head of the Bioinformatics facility, Dr. Jingtao Lilue. I would like to thank as well Dr. Ricardo W. A. Vitor from Universidade Federal de Minas Gerais in Brazil who provided us with the original isolates of the N, BV and AS28 strains. This work was supported by central funds of the Instituto Gulbenkian de Ciência, by the Sonderforschungsbereiche 670 and 680 and Schwerpunkt 1399 of the Deutsche Forschungsgemeinschaft.

## 4.7. Supplementary material

### ROP5 sequences

>AS28\_chrXII\_575591\_577431\_ROP5\_1\_like-B\_type\_I\_III

ATGGCGACGGATGCCAGGAGACTAGCCACCGGCCCTTGCTCTGTTAACGTGCCTGGTATGGCGGGCAGGCGCATTTACGCCATCGCCT  
CCAACCTCCAGAGCCAATGAGCTGACGTCGGGCATTCCCGGTGAGGCTCCGGAGAACACTGCGGAGGAGCGAGGAATCGGAGACAT  
CTCAGATTACCCCTCATTCACTGGTGCAGATGCGGCCGATATTAGCGGTGGCACAGTTCCCGAAGGGCCAGGATCGCCTGCCAGCAC  
ATCTGGGTCTGGGGGTGCTTCAACAGATTATTTAGCAGATTTCTGTCAGGTAAGTGATCCACACAAGGCACGGGAGTCGCTGACGAA  
CCTCAGCAGGGGCCCGACCGTTACTTCGGGAGAGACTTGCTCAGCACTTCCGCAGGCTGAGGGTCTTGTTGGGCGCCTTATGCCG  
AGGTGGGGCTTTGGATTTCGTCGCCATGTAGGCAGGTGGTGGCCAAGGAGGTGGCCGACCCGCTGTTTCTTGATGGAACCGGGG  
GATGAGTTCACTTCGTACCTGGTTAGAGCGAAGAAAGACCGGATTGGTTTCTACCGCCGGGAATGGTGGAAGGATACGAGGTGGAG  
GTTAAAGCTGTGACGGCAGCTATTTGGCCCCAGAATACTGCCAGAGAAGTGCGCTCGCTGTTGGACAGAAGAAAAAGGACACTGAGAG  
TGGTTGGGGCTTTTGAGCAAGTGTTTCGATCTGTCTTGATCTGGCGCAAGATGTGGAGACTCAAGAGCATATAGCTGCAGAAGTCTTC  
ACTCTGACCGCGCAGAGTCGAGGTTGGACCTTCAGCGTGTGCATGACAACTTGTTGTCTCAGCGCGGTTTGGTCTGAGAGCCCTG  
TGCTGTCAAGGGACAGGAGGCGTCTTTGCTTCCCTACGATGCTGTGACAGTTCCGACTCAACCCCCCTTCGAGGAAGTGAATCCAGG  
ACAGAGCAATTATCCAGTAGCGAACTATTTCTCTCATGCCCCGCCAATGATGAGCCTTGAGACGCTCCATAGAATAGTCGATCAG  
AAGAATTCTCAGGGCGACATCGGCCGAGTAGTACGTATGGTCTTGACTGCGGAACTCATAAGAAATTCAGCAAAACATTAGATCCGC  
GGACTTGTGCACGGACGTATCACATCAGAAAACCTTTTATTATGCCCGATGGCCGCTGATGCTGGGGGATGTATCCGCGTTGAGGA  
AGGTCGGAACCCGAGGACCGGTGTCGAGTGTTCCAGTCACCTATGCCCCACGCGAGTCTTGTGCGAACACAGAGACGGCGGCATTTA  
CACACGCACTCGATGCGTGGCAATTAGGCCTCGTTATACATCGAATTTGGTGCCCTCATCTTGCCCTTCGAGCTGGTGACACCTCGAATG  
AAGAGATCAGGGAGGAGGCCAAGCATGCGGGTTCAGGACTAGACGTTCTGTCACTGAAGCGTGACGCCAATGCCCGACGCTGTA  
GAGATGCTGGTGAGGCATTTCTCAATTTCAACACTCGAAAACGCTGCTTCCCTGACAGCCATGGGGACGCTGAATTCGCCCAGC  
TACAGCTCGACATATCCACCGCCTGTGCATCAAAATAA

>AS28\_chrXII\_571224\_572927\_ROP5\_2\_like-A\_type\_I

ATGGCGACGGATGCCAGGAGACTAGCCACCGGCCCTTGCTCTGTTAACGTGCCTGGTATGGCGGGCAGGCGCATTTACGCCATCGCCT  
CCAACCTCCAGAGCCAATGAGCTGACGTCGGGCATTCCCGGTGAGGCTCCGGAGAACACTGCGGAGGAGCGAGGAATCGGAGACAT  
CTCAGATTACCCCTCATTCACTGGTGCAGATGCGGCCGATATTAGCGGTGGCACAGTTCCCGAAGGGCCAGGATCGCCTGCCAGCAC  
ATCTGGGTCTGGGGGTGCTTCAACAGATTATTTAGCAGATTTCTGTCAGGTAAGTGATCCACACAAGGCACGGGAGTCGCTGACGAA  
CCTCAGCAGGGGCCCGACCGTCACTTCGGGAGAGACTTGCTCAGCACTTCCGCAGGCTGAGGGTCTTGTTGGGCGCCTTATGCCG  
AGGTGGGGCTTTGGATTTCGTCGCCATGTAGGCAGGTGGTGGCCAAGGAGGTGGCCGACCCGCTGTTTCTTGATGGAACCGGGG  
GATGAGTTCACTTCGTCTAGACTGAGTCAGACAGAATATCGGACTCGCTTTTACCGCCGGGAATGGTGGAAGGTACGAGGTGGAGG  
TTAAAGCTGTGACGGCAGCTATTTGGCCCCAGAATACTGCCAGAGAAGTGCGCTCGCTGTTGGACAGAAGAAAAAGGACACTGAGAG  
GGTTGGGGCTTTTGAGCAAGTGTTTCGATCTGTCTTGATCTGGCGCAAGATGTGGAGAGTCAAGAGCGCATGGCCATCGAAGTCTTC  
ACTCTGACGAGCGGGAACACTGTATCAGACCTTGTCTGTGTCATGACAACTTGTTGCTATGACGGGGTGGTTAGTGAAGCCCCC  
AGCAGGCACGGGACATCTGCAGACTCCTTCTCCACCGATGCTGTGACAGTTCCGACTCAACCCCCCTTCGAGGAAGTGAATCCAGG  
ACAGAGCAATTATCCAGTAGCGAACTATTTCTCTCATGCCCCGCCCTATAATGAGCCTTGAGCCGCTCCATAGAATAGTCGATCAG  
AAGAATTCTCAGGGCGACATCGGCCGAGTAGTACGTATGGTCTTGACTGCGGAACTCATAAGAAATTCAGCAAAACATTAGATCCGC  
GGACTTGTGCACGGACGTATCACATCAGAAAACCTTTTATTATGCCCGATGGCCGCTGATGCTGGGGGATGTATCCGCGTTGAGGA  
AGGTCGGAACCCGAGGACCGGTGTCGAGTGTTCCAGTCACCTATGCCCCACGCGAGTCTTGTGCGAACACAGAGACGGCGGCATTTA  
CACACGCACTCGATGCGTGGCAATTAGGCCTCGTTATACATCGAATTTGGTGCCCTCATCTTGCCCTTCGAGCTGGTGACACCTCGAATG  
AAGAGATCAGGGAGGAGGCCAAGCATGCGGGTTCAGGACTAGACGTTCTGTCACTGAAGCGTGACGCCAATGCCCGACGCTGTA  
GAGATGCTGGTGAGGCATTTCTCAATTTCAACACTCGAAAACGCTGCTTCCCTGACAGCCATGGGGACGCTGAATTCGCCCAGC  
TACAGCTCGACATATCCACCGCCTGTGCATCAAAATAA

>AS28\_ROP5\_3\_like-C\_Type\_I

ATGGCGACGAAGCTCGCTAGACTAGCCACGTGGCTTGCTTGGTAGGTTGCCTGTTGTGGCGGGCGGGGCGCAGTTTCAGCTCTCCGC  
ATAACTCCAGGACGAATGATCTGGCTTCGGGAACCCCGCATGTGGCTCGTGGGACACTGAGGCGCAGTACGAAGTGGAGACGATT  
CAGATTTTCCGCAGGCCGTGGCCGAAGAGGTGGCAGATATGAGCGCGGCAGAGTTCCCGAGTGCCAGCATCGTCTACCCACAT  
CTGCGTCCGAGGGGATTTTCAAGATTAGTTGCGAGACTTCGTCGGGGAAGAGGAACCGCAGATGGCGCAGGAGTTGCTGACGAAA  
CCCATCAGGAGCCGCGCCGCCACTTCGGAAGAGACTTGCTCAGCACTTCCGTAGGCTGAGGGGCTTCTTCGACGCCCTTACGCCGA  
GGTGGCTCTCCGCTCTCGGCCCGCGGGCGCAAGATGGTGGAGAGGGAGACAGAGACCGCTGCTGGACCCCTCGTTTATCGGGTTG  
GAAGCTGGAGATTCTTCATGCGCGACCTGCTGAAACGTGAAGAAGAGCTGATTGGATACTGTGCGGAAGAAGCGTTGAAAGAACCTG  
CAGCGATGGTTGAGGCTGTCACGGCAACTGTATGGCCGCAAAATGCTGAAACAACCGTGGATTTCGCTTTTGTGTCAGGAGAGCGGAA  
GTTGAAATTTGGTGAGCCTCTTCGAGTCGGTGACCGATCTGTCGATTTTATGAAGGATGTAGAGCGCCTGGAGGATTTGCTCTGA

AGGTCTTCACTATGGGTGCCGAGAATTCCCGATCAGAGCTGGAGCGGTTGCATGAAGCGACTTTTTCGGCAGCGAGGTTGCTTGGGG  
AGAGTCCAGAGGAGGCACGGGACAGACGCAGGCTTTTACTTCCCTCCGATGCTGTGGCAGTTCAGTCTCAGCCCCdTTTCGCTCAGCTG  
AGTCCAGGACAGAGCGACTATGCAGTCCGGAACCTACTTGCCTCATGCCCGCTGCGTCGGTGGATCTTGAATTGCTCTTTAGGACATT  
GGATTTCTGTATGTACTCAGGACTCAGAGAAGATTTTTTAGCGCTTACATACTAACGGCACAGCTGATCCGTCTGGACGCCAACCTGCA  
GAGCAAAGGACTTGTGCATGGACATTTACACCCGGATAACCTTTTATTATGCCCGATGGCCGCTGATGCTGGGGgATGATCCGTGT  
TGAGGAAGGTCCGAACCCGAGGACCGGCATCAAGCGTCCCGGTTACCTATGCGCCTCGCGAGTTCCTGAATGCAAGCACGGCAACAT  
TTACACAGCGCTGGATGCGTGGCAACTGGGTCTTAGCATATACCGGGTTTGGTGCCTATTCTTGCCTTTCGGACTCGTGACACCTGG  
GATCAAAGGTCATGGAAaAGACCAAGTCTACGAGTTCAGGGACTGACAGTCTGGCATTTCGGCTCATGTACACCTCTGCCTGACTTCG  
TGCAGACACTTATTGGACGGTTCTCAACTTCGATAGGCGTCGACGCTGCTTCCCTTGGAGGCCATGGAGACGCCAGAGTTCTCCCA  
GCTCCAAAACGAAATATCGAGCAGCCTATCAACAGGACAACCCATTGCTGCGCCCTCAGTCGCTTGA

>AS28\_ROP5\_4\_like-B\_Type\_I

ATGGCGACGAAGCTTGCTAGACTAGCCACGTGGCTTGTCTTGGTAGGTTGCCTCTTGTGGCGGGCGGGGCGAGTTCAACTCTCTCCCG  
CAAACCTCCAGGACGAATGATCTGGCTTCGGGAACCCCGCATGTGGCTCGTGGGGACACTGAGGCGCAGTCAGGAACCTGGAGACGATT  
CAGATTTTCCGAGCGCGTGGTCCGAAGAGGTGGCAGATATGAGCGCGCGCAGAGTTCCTCCGAGTGCCAGCATCGTCTACCAACCAT  
CTGCGTCCGAGGGGATTTTCAAGAGATTAGTTCGACAGCTTCGTCGGGAAGAGGAACCGCAGATGGCGCAGGAGTTGCTGACGAAA  
CCCATCAGGGGCCGCGCCGCCACTTCGGAAGAGACTTGCTCAGCACTTCGCTAGGCTGAGGGGCTTCTTCGGACGCCCTTACGCCGA  
GGTGGCTCTCCGTCTCGGCCGCCGGCGCAAAGATGGTGGAGAGGGAGACAGAGACCGCTGCTGGATCCTTCGTTTTCATGGGTTG  
GAAGCTGGAGATTCTTCATGCGCGACCTGCTGAAACGTGAAGAAGAGCTGATTGGATACTGTGCGGAAGAAGCGTTGAAAGAACCTG  
CAGCGATGGTTGAGGCTGTACGGCAACTGTATGGCCGCAAAATGCTGAAACAACCGTGGATTCACTTTTGAAGTCAAGGAGAGCGGAA  
GTTGAAATTGGTGGAGCCTCTTCGAGTCGGTGACCGATCTGTCGTAATTTTAGTAAGGGATGTAGAGCGCCTGGAGGATTTTCGCTCTGA  
AGGTCTTCACTATGGGTGCCGAGAATTCCCGATCAGAGCTGGAGCGGTTGCATGAAGCGACTTTTTCGGCAGCGAGGTTGCTTGGGG  
AGAGTCCAGAGGAGGCACGGGACAGACGCAGGCTTTTACTTCCCTCCGATGCTGTGGCAGTTCAGTCTCAGCCCCCTTTCGCTCAGCT  
GAGTCCAGGACAGGACGACTATGCAGTCGCGAACTATTTGCTTTCATGCCCGCTGCGTCGGTGGATCTTGAATTGCTCTTTAGCACAT  
TGGACTTCGTGTATGTATTAGGGGGGACGAAGGTATTTTAGCGCTTACATACTAACGGCACAGCTGATCCGTCTGGCAGCCAACT  
GCAGAGCAAAGGACTTGTGCATGGACATTTACACCCGGATAACCTTTTATTATGCCCGATGGCCGCTGATGCTGGGGGATGTATCC  
GCATTGTGGAAGGTCCGAACCCGAGGACCGGCATCAAGCGTCCCGGTTACCTATGCGCCTCGGGAGTTCCTGAATGCAAGCACGGCA  
ACATTTACACACGCGCTGGATGCGTGGCAACTGGGTCTTAGCATATACCGGGTTTGGTGCCTATTCTTGCCCTTCGGACTCGTGACACC  
TGGGATCAAAGGTCATGAAAAGGCCAAGTCTACGAGTTCAGGGACTGACAGTCTGGCATTTCGGCTCATGTACACCTCTGCCTGAC  
TTCGTGAAGACACTTATTGGACGGTTCCTCAACTTCGATAGGCGTCGACGCTGCTTCCCTTGGAGGCCATGGAGACGCCAGAGTTCC  
TCCAGCTCCAAAACGAAATATCGAGCAGCCTATCAACAGGACAACCTACTGCTGCGCCCTCAGTCGCTTGA

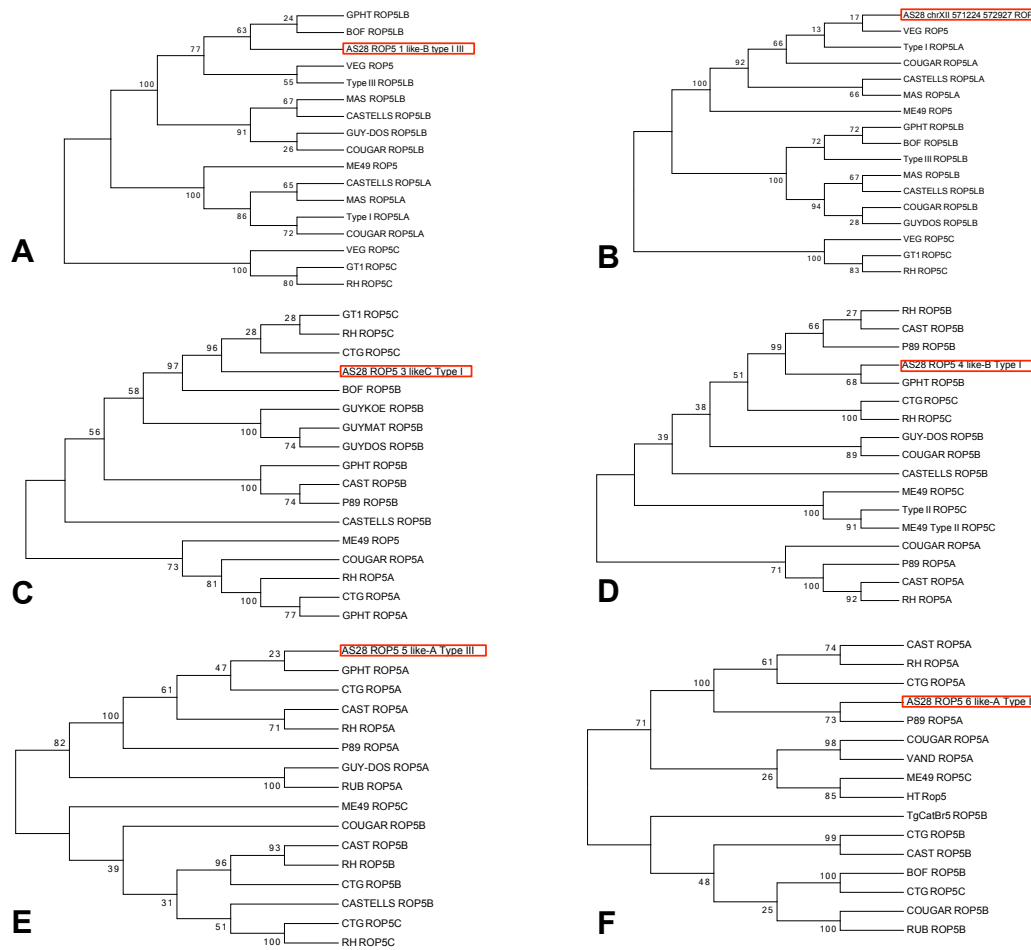
>AS28\_ROP5\_5\_like-A\_Type\_III

ATGGCGACGAAGCTCGCTAGACTAGCCACGTGGCTTGTCTTGGTAGGTTGCCTGTTGTGGCGGGCAGGGGCGAGTTCAGCTCTCTCCG  
CCAACTCCAGGACGAATGATCTGGCTTCGGGAACCCCGCATGTGGCTCGTGGGGACACTGAGGCGCAGTCAGGAACCTGGAGACGAT  
TCAGATTTTCCGAGCGCGTGGCCGAAGAGGTGGCAGATATGAGCGCGCGCAGAGTTCCTCCGAGTGCCAGCATCGTCTACCAACACA  
TCTGCATCCGAGGGGATTTTCAAGAGATTAGTTCGACAGCTTCGTCGGGAAGAGGAACCGCAGATGGCGCAGGATTCGTGACGAAA  
CCCATCAGGGGCCGCGCCGCCACTTCGGAAGAGACTTGCTCAGCACTTCGCTAGGCTGAGGGGCTTCTTCGGACGCCCTTACGCCGA  
GGTGGCTCTCCGTCTCGGCCGCCGGCGCAAAGATGGTGGAGAGGGAGACAGAGACCGCTGCTGGACCTTCGTTTTCATGGGTTG  
GAAGCTGGAGATTCTTCATGCGCGACCTGCTGAAACGTGAAAAGAGCTGATTGGATACTGTGCGGAAGAAGCGTTGAAAGAACCTG  
CAGCGATGGTTGAGGCTGTACGGCAACTGTATGGCCGCAAAATGCTGAAACAACCGTGGATTCACTTTTGAAGTCAAGGAGAGCGGAA  
GTTGAAATTGGTGCAGCCTCTTCGAGTCGGTGACCGATCTGTCGTAATTTTAGTAAGGGATGTAGAGCGCCTGGAGTATTTTCGCTCTGA  
AGGTCTTCACTATGGGTGCCGAGAATTCCCGATCAGAGCTGGAGCGGTTGCATGAAGCGACTTTTTCGGCAGCGAGGTTGCTTGGGG  
AGAGTCCAGAGGAGGCACGGGACAGACGCAGGCTTTTACTTCCCTCCGATGTTGTGGCAGTTCAGTCTCAGCCCCCTTTCGCTCAGCT  
GAGTCCAGGACAGAGCGACTATGCAGTCGCGAACTATTTGCTTCTCATGCCCGCTGCGTCGGTGGATCTTGAATTGCTCTTTAGCACAT  
TGAACCTTCGTGTATGTATTAGGGGGGGAGAAGGTATTTTAGCGCTCACATACTAACGGCACAGCTGATCCGTCTGGCAGCCAACT  
GCAGAGCAAAGGACTTGTGCATGGACGCTTACACCCGGATAACCTTTTCTATGCCGTATGGCCCCGTGATGCTGGGGGATGTATCC  
GCATTGTGGAAGGTCCGAACCCGAGGACCGGCATCAAGCGTCCCGGTTACCTATGCGCCTCGGGAGTTCCTGAATGCAAAACACGGCA  
ACATTTACACACGCGCTCAATGCGTGGCAACTGGGTCTTAGCATATACCGGGTTTGGTGCCTAGTCTTGCCCTTCGGACTCGTGACACC  
TGGGATCAAAGGTCATGAAAAGGCCAAGTCTACTAGTTCAGGGACTGACAGTCTGTCACTTCTCCCATGTGACCTGTGCCTGACT  
TCGTGGAGACACTGATTAGACGGTTCCTCAACTTCGATAGGCGGCGACGCTGCTTCCCTTGGAGGCCATGGAGACGCCAGAGTTCTC  
CCAGCTCCAAAACGAAATATCGAGGCGCCTATCAACAGGACAACCTACTGCTGCGCCCTCAGTCGCTTGA

>AS28\_ROP5\_6\_like-A\_Type\_I

ATGGCGACGAAGCTCGCTAGACTAGCCACGTGGCTTGTCTTGGTAGGTTGCCTGTTGTGGCGGGCAGGGGCGAGTTCAGCTCTCTCCG  
CCAACTCCAGGACGAATGATCTGGCTTCGGGAACCCCGCATGTGGCTCGTGGGGACACTGAGGCGCAGTCAGGAACCTGGAGACGAT

TCAGATTTTCCGCAGGCCGTGGCCGAAGAGGTGGCAGATATGAGCGCGGCAGAGTTCCCCGAGTGCCAGCATCGTCTACCACCACA  
TCTGCRITCCGAGGGGATTTTCAGAAGATTAGTTCGCAGACTTCGTGGGGGAAGAGGAACCGCAGATGGCGCAGGAGTTGCTGACGAA  
ACCCATCAGGAGCCGCGCCGCCACTTCGGAAGAGACTTGCTCAGCACTTCCGTAGGCTGAGGGGCTTCTTCGGACGCCCTTACGCCG  
AGGTGGCTCTCCGGTCTCGGCCCGCGGCGCAAAGATGGTGGAGAGGGAGACAGAGACCGCTGCTGGACCCCTTCGTTTCATGGGT  
GGAAGCTGGAGATTCTGTCATGCGCGACCTGCTGAAACGTGAAGAAGAGCTGATTGGATACTGTCGCGAAGAAGCGTTGAAAGAACCT  
GCAGCGATGGTTGAGGCTGTACGGCAACTGTATGGCCGCAAAATGCTGAAACAACCGTGGATTCACTTTTGAGTCAGGGAGAGCGGA  
AGTTGAAATTGGTGAGCCTCTTCGAGTCGGTGACCGATCTGTCGTATTTTAGTAAGGATGTAGAGCGCCTGGAGTATTTGCTCTG  
AAGGTCTTCACTATGGGTGCCGAGAATCCCGATCAGAGCTGGAGCGGTTGCATGAAGCGACTTTTTCGGCAGCGAGGTTGCTTGGG  
GAGAGTCCAGAGGAGGCACGGGACAGACGCAGGCTTTTACTTCCCTCCGATGTTGTGGCAGTTCAGTCTCAGCCCCCTTCGCTCAGC  
TGAGTCCAGGACAGAGCGACTATGCAGTCGCGAACTATTTGCTTCTCATGCCCGCTGCGTCGGTGGATCTTGAATTGCTCTTTAGCACA  
TTGAACTTCGTGTATGTATTAGGGGGGGAGAAGGTATTTAGCGCGTCACATACTAACGGCACAGCTGATCCGTCTGGCAGCCAACCT  
GCAGAGCAAAGGACTTGTGCATGGACGCTTACACCGGATAACCTTTTTCTTATGCCGTATGGCCCCGTGATGCTGGGGGATGTATCC  
GCATTGTGAAGGTCGGAACCCGAGGACCGGCATCAAGCGTCCCGTTACCTATGCGCCTCGGGAGTTCTTGAATGCAACACGGCA  
ACATTTACACACGCGCTCAATGCGTGGCAACTGGGTCTTAGCATATACCGGGTTTGGTGCCTAGTCTTGCCTTTCGGACTCGTGACACC  
TGGGATCAAAGGGTCATGGAAAAGGCCAAGTCTACTAGTTCAGGGACTGACAGTCTGTCACTTCTCCCATGTGCACCTGTGCCTGACT  
TCGTGGAGACACTGATTAGACGGTTCCTCAACTTCGATAGCGCGGACGCCTGCTTCCCTTGGAGGCCATGGAGACGCCAGAGTTCCT  
CCAGCTCCAAAACGAAATATCGAGGCGCCTATCAACAGGACAACCTACTGCTGCGCCCTCAGTCGCTTGA



**Supplemental Figure 4.1.** Phylogenetic analysis of ROP5 sequences from AS28 *T.gondii* strain. Phylogenetic trees based on multiple amino acid sequence alignments of ROP5 isoforms (A, B and C-like) from different *T.gondii* strains. The Jukes-Cantor genetic distance model along with a neighbor-joining building method with no outgroup were computed in the software MEGA 6. Phylogenetic tree for the A) AS28 ROP5-1, B) AS28 ROP5-2, C) AS28 ROP5-3, D) AS28 ROP5-4, E) AS28 ROP5-5, F) AS28 ROP5-6.

## ROP18 sequences

>ROP18\_N\_strain\_TGME49\_chrvIIa\_1514430-1516150

```
CCTTTTCATTTTGAGTCTTTTATTCTGTGTGGAGATGTTCTGCTGTTTGAAGTTTTGCGCGGCGCCTTTTACTACTGAATCCATCTCTTT
AAACGCTGCTGTCTCCAAGGCTTGATGCGGGAGCATCCGTTTCTGAGGATCTCGATTCAAGAGGCTTCGGATGAGGTCTTGAACCAAGC
TCAGGCGTGAAGGACAATCTCGGAAGTGAAGTAGTCCCAGATGCCGTCGGCCGGAGTTGGACGTTCTTGCACCAAGATGCAGTACA
AAGTTATACCGAGTTGCCACGCGTCAGTGGGGAATGTATAGGTGATGCTGTAGCCTGAAACGGTCGCTCCGGAGGCTCGTAACCGG
GAGTACCTATCGCGCGTCCAACCGAATTATTGATTCTATACGTTCCGAAGTCGCCGAGAAACAGGCGACCGTCTTTCAAGAGGAGGAAA
TTCGCGCGTTTGATATCCGTATGCACAATCCCTGAGCTTGAACATTGGCCACTAGTTTGATTGCCTGACTGGATAGGTACATTGCAACG
ACCAGGCCAAATTCACTTTTATTGACAGACGCATCTCCAAATACCCATGAAATGACTTTGCTCATGTCCGCTTCTGCCCGCATCATGAGA
AGAAAAAGTTTGGTACCCACCGAGTCGTCAAACCAATCAGGACCTCTGTGGATGCTGGCTGTCCCTCTAACATCACACATCACTAGG
AACCATGAATCTACAGCTTTCTGGGCATCCTTCGCGCTTGGCTAGACTAAAGTTCTGTAGCAGGACGACTCCCTCTGCAAGTCAA
GCATAGTCTCATCGTGGGCTCCTTTTCTGACATGAAACCTTAACAGCCAACCTTTCATTGCTCTCCACGCTGTAGCCTCATATACAG
TGGCGAATCCACCAGAGCCAAGGGGTGCACCTCTCACCAGCGTCCGGGTTTACCAGGTAGTCAGAAAAAAATCTTTTGGAAACGTC
CGGAGGCCAGGCGCGCAGATCAGAAATCTATCCAGTCTTACTATTTCTAGCGCTTGAGGCTGCCGATTGTTCTTGCAATGTGCG
CCAGGATGCGTTTGCCGATGACGCATCCAGAATCCGCTTCTCAAAAACCAATTCAGATCGCCGCTTTTGTCGCCGAAACTTCGCTGT
CGGTTACGCCCTTGAGGAAAGTACCTCCGGATTCCGCGTTTGAATAACGATTTGCTCTACGCCAAAGTTGAGCAAGCCTCTGCTGAC
GGACAGACGGGGTCTCGAGGAGCTATCAGCTGCGGAACCCGAGACTTCACCATCTCTTTCTCTCGTCTCAAGAGATTGAGAAAGCCTC
TGAAAAATCGTGCGCCGTTGGTTTGTTCGTCATTTACCGGTGTTGAGCCCTTCTCCAGCAATGAAACGCTCTCGAGTGTCTTCT
GTCGCTCTGCTAAATGTTGTAGTGGCCAACTGTAGCAAGCCACCGCTTGTACGCAACGTGCTGTTGCGAATCCAACTTGTGCGTTT
GGAGTCGAGTTTGAAGGACCAAGTGTGATTCCGGTCCCATCTGGACTTGAAGAGCAGGAAGACAAGCGCTACATTTAACCCCGCA
AGACAGGCTGTCTTCGGGAGAAGAGTCGCTAAACCCATTCCGACGACGGTACGCGTAAGAGGTGGCCGCTGTACCGAAAAACATCACA
ACTTTCACACAACTGGACTGGGGTGAGTACAA
```

>ROP18\_AS28\_strain\_TGME49\_chrvIIa\_1514430-1516150

```
CCTTTTCATTTTGAGTCTTTTATTCTGTGTGGAGATGTTCTGCTGTTTGAAGTTTTGCGCGGCGCCTTTTACTACTGAATCCATCTCTTT
AAACGCTGCTGTCTCCAAGGCTTGATGCGGGAGCATCCGTTTCTGAGGATCTCGATTCAAGAGGCTTCGGATGAGGTCTTGAACCAAGC
TCAGGCGTGAAGGACAATCTCGGAAGTGAAGTAGTCCCAGATGCCGTCGGCCGGAGTTGGACGTTCTTGCACCAAGATGCAGTACA
AAGTTATACCGAGTTGCCACGCGTCAGTGGGGAATGTATAGGTGATGCTGTAGCCTGAAACGGTCGCTCCGGAGGCTCGTAACCGG
GAGTACCTATCGCGCGTCCAACCGAATTATTGATTCTATACGTTCCGAAGTCGCCGAGAAACAGGCGACCGTCTTTCAAGAGGAGGAAA
TTCGCGCGTTTGATATCCGTATGCACAATCCCTGAGCTTGAACATTGGCCACTAGTTTGATTGCCTGACTGGATAGGTACATTGCAACG
ACCAGGCCAAATTCACTTTTATTGACAGACGCATCTCCAAATACCCATGAAATGACTTTGCTCATGTCCGCTTCTGCCCGCATCATGAGA
AGAAAAAGTTTGGTACCCACCGAGTCGTCAAACCAATCAGGACCTCTGTGGATGCTGGCTGTCCCTCTAACATCACACATCACTAGG
AACCATGAATCTACAGCTTTCTGGGCATCCTTCGCGCTTGGCTAGACTAAAGTTCTGTAGCAGGACGACTCCCTCTGCAAGTCAA
GCATAGTCTCATCGTGGGCTCCTTTTCTGACATGAAACCTTAACAGCCAACCTTTCATTGCTCTCCACGCTGTAGCCTCATATACAG
TGCGGAATCCACCAGAGCAAGGGGTGCACCTCTCACCAGCGTCCGGGTTTACCAGGTAGTCAGAAAAAAATCTTTTGGAAACGTC
CGGAGGCCACGCGCGCAGTCAGATCAGAAATCTATCCAGTCTTACTATTTTCTAGCGCTTGAGGCTGCCGATTGTTGTTGATGCTGCG
CCAGGATGCGTTTGCCGATGACGCATCCAGAATCCGCTTCTCAAAAACCAATTCAGATCGCCGCTTTTGTCGCCGCAAACTTCGCTGT
CGGTTACGCCCTTGAGGAAAGTACCTCCGGATTCCGCGTTTGAATAACGATTTGCTCTACGCCAAAGTTGAGCAAGCCTCTGCTGAC
GGACAGACGGGGTCTCGAGGAGCTATCAGCTGCGGAACCCGAGACTTCACCATCTCTTTCTCTCCGCTCAAGAGATTGAGAAAGCCTC
TGAAAAATCGTGCGCCGTTGGTTTGTTCGTCATTTACCGGTGTTGAGCCCTTCTCCAGCAATGAAACGCTCTCGAGTGTCTTCT
GTCGCTCTGCTAAATGTTGTAGTGGCCAACTGTAGCAAGCCACCGCTTGTACGCAACGTGCTGTTGCGAATCCAACTTGTGCGTTT
GGAGTCGAGTTTGAAGGACCAAGTGTGATTCCGGTCCCATCTGGACTTGGAAGAGCAGGAAGACAAGCGTACATTTTAACCCCGCA
AGACAGGCTGTCTTCGGGAGAAGAGTCGCTAAACCCATTCCGACGACGGTACGCGTAAGAGGTGGCCGCTGTACCGAAAAACATCACA
ACTTTCACACAACTGGACTGGGGTGAGTACAA
```

>ROP18\_BV\_strain\_TGME49\_chrvIIa\_1514430-1516150

```
CCTTTTCATTTTGAGTCTTTTATTCTGTGTGGAGATGTTCTGCTGTTTGAAGTTTTGCGCGGCGCCTTTTACTACTGAATCCATCTCTTT
AAACGCTGCTGTCTCCAAGGCTTGATGCGGGAGCATCCGTTTCTGAGGATCTCGATTCAAGAGGCTTCGGATGAGGTCTTGAACCAAGC
TCAGGCGTGAAGGACAATCTCGGAAGTGAAGTAGTCCCAGATGCCGTCGGCCGGAGTTGGACGTTCTTGCACCAAGATGCAGTACA
AAGTTATACCGAGTTGCCACGCGTCAGTGGGGAATGTATAGGTGATGCTGTAGCCTGAAACGGTCGCTCCGGAGGCTCGTAACCGG
GAGTACCTATCGCGCGTCCAACCGAATTATTGATTCTATACGTTCCGAAGTCGCCGAGAAACAGGCGACCGTCTTTCAAGAGGAGGAAA
TTCGCGCGTTTGATATCCGTATGCACAATCCCTGAGCTTGAACATTGGCCACTAGTTTGATTGCCTGACTGGATAGGTACATTGCAACG
ACCAGGCCAAATTCACTTTTATTGACAGACGCATCTCCAAATACCCATGAAATGACTTTGCTCATGTCCGCTTCTGCCCGCATCATGAGA
AGAAAAAGTTTGGTACCCACCGAGTCGTCAAACCAATCAGGACCTCTGTGGATGCTGGCTGTCCCTCTAACATCACACATCACTAGG
AACCATGAATCTACAGCTTTCTGGGCATCCTTCGCGCTTGGCTAGACTAAAGTTCTGTAGCAGGACGACTCCCTCTGCAAGTCAA
GCATAGTCTCATCGTGGGCTCCTTTTCTGACATGAAACCTTAACAGCCAACCTTTCATTGCTCTCCACGCTCTGTAGCCTCATATACAG
```



TGGCGAATCCACCAGAGCCAAGGGGTGCACCTCTCACCAGCGTCCGGGTTTCACCGGTAGTCACAGAAACAAATCTTTTGGAACGTC  
CGGAGGCCAGGCGGCGACAGTCAGAATTCTATCCAGTCGTTCACTATTTTCTAGCGCTTGAGGCTGCCCCGATTGTCTTGTCATGTGCG  
CCAGGATGCGTTTGCCGATGACGCATCCAGAATCCGCCTTCTCAAAAACCAATTCAGATCGCCGCTTTGTGCCCGCAAACCTTCGCTGT  
CGGTTACGCCCTTGAGGAAAGTACCTCCGGATTCCGCGTTTGAATAACGATTTGCTCTACGCCAAAGTTGAGCAAGCCTCTGTCGTAC  
GGACAGACGGGGTCTCGAGGAGCTATCAGCTGCGGAACCCGAGACTTCACCATCTCTTCTCTCCGTCTCAAGAGATTGAGAAGCCTC  
TGAAAAATCGTGCGCCGTTGGTTTGTCTTGCGCATTACCCGGTGTGAGCCCTTTCCTCCAGCAATGAAACGTCTCGAGTGCTTTCT  
GTCGCTCCTGCTAAATGTTTGTAGTGGCCAACTGTAGCAAGCCACCGCTTGTCAGCAACGTGCTGTTGCGAATCCAACTTGTGGTTT  
GGAGTCGAGTTTGAAGGACCAAGTGTGATTCCGGTCCCATCCTGGACTTGGAAGAGCAGGAAGACAAGCGCTACATTTAACCCCGCA  
AGACAGGCTGTCTTCGGGAGAAGAGTCGCTAAACCCATTCCGACGACGGTACGCGTAAGAGGTGGCCGCTGTACCGAAAACATCACA  
ACTTTCACACAACTGGACTGGGGTGAGTACAA

## Bibliography

1. Dubey, J. P. in *Toxoplasma gondii. The Model Apicomplexan: Perspectives and Methods* (eds. Kim, K. & Weiss, L. M.) (Elsevier, 2007).
2. Khan, A. *et al.* Genetic analyses of atypical *Toxoplasma gondii* strains reveal a fourth clonal lineage in North America. *Int. J. Parasitol.* **41**, 645–655 (2011).
3. Mercier, A. *et al.* Additional Haplogroups of *Toxoplasma gondii* out of Africa : Population Structure and Mouse-Virulence of Strains from Gabon. *PLoS Negl. Trop. Dis.* **4**, (2010).
4. Chaichan, P. *et al.* Geographical distribution of *Toxoplasma gondii* genotypes in Asia: A link with neighboring continents. *Infect. Genet. Evol.* **53**, 227–238 (2017).
5. Galal, L. *et al.* *Toxoplasma* and Africa : One Parasite , Two Opposite Population Structures. *Cell Press Rev.* **34**, (2018).
6. Galal, L., Hamidović, A., Dardé, M. L. & Mercier, M. Diversity of *Toxoplasma gondii* strains at the global level and its determinants. *Food Waterborne Parasitol.* e00052 (2019). doi:10.1016/j.fawpar.2019.e00052
7. Howe, D. K. & Sibley, L. D. *Toxoplasma gondii* Comprises Three Clonal Lineages : Correlation of Parasite Genotype with Human Disease. *J. Infect. Dis.* **172**, 1561–1566 (1995).
8. Darde, M. L., Ajzenberg, D. & Smith, J. in *Toxoplasma gondii. The Model Apicomplexan: Perspectives and Methods* (eds. Weiss, L. M. & Kim, K.) (Elsevier, 2007).
9. Shwab, E. K. *et al.* Geographical patterns of *Toxoplasma gondii* genetic diversity revealed by multilocus PCR-RFLP genotyping. *Parasitology* **141**, 453–61 (2014).
10. Shwab, E. K. *et al.* Human impact on the diversity and virulence of the ubiquitous zoonotic parasite *Toxoplasma gondii*. *Proc. Natl. Acad. Sci.* 1–8 (2018).
11. Bertranpetit, E. *et al.* Phylogeography of *Toxoplasma gondii* points to a South American origin. *Infect. Genet. Evol.* **48**, 150–155 (2017).
12. Behnke, M. S. *et al.* Rhoptry Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of *Toxoplasma gondii*. *PLOS Genet.* **11**, e1005434 (2015).
13. Jensen, K. D. C. *et al.* *Toxoplasma gondii* Superinfection and Virulence during Secondary Infection Correlate with the Exact ROP5 / ROP18 Allelic Combination. *MBio* **6**, 1–15 (2015).
14. Su, C., Zhang, X. & Dubey, J. P. Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP markers : A high resolution and simple method for identification of parasites. *Int. J. Parasitol.* **36**, 841–848 (2006).
15. Dubey, J. P. *et al.* Biologic and genetic comparison of *Toxoplasma gondii* isolates in

- free-range chickens from the northern Para state and the southern state Rio Grande do Sul , Brazil revealed highly diverse and distinct parasite populations. *Vet. Parasitol.* **143**, 182–188 (2007).
16. Ajzenberg, D., Collinet, F., Mercier, A., Vignoles, P. & Darde, M.-L. Genotyping of *Toxoplasma gondii* Isolates with 15 Microsatellite Markers in a Single Multiplex PCR Assay. *J. Clin. Microbiol.* **48**, 4641–4645 (2010).
  17. Lorenzi, H. *et al.* Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. *Nat. Commun.* **7**, 10147 (2016).
  18. Shwab, E. K. *et al.* The ROP18 and ROP5 gene allele types are highly predictive of virulence in mice across globally distributed strains of *Toxoplasma gondii*. *Int. J. Parasitol.* **46**, 141–146 (2016).
  19. Saeij, J. P. J. *et al.* Polymorphic Secreted Kinases Are Key Virulence Factors in Toxoplasmosis. *Science (80-. )*. **161**, 1780–1784 (2006).
  20. Taylor, S. *et al.* A Secreted Serine-Threonine Kinase Determines Virulence in the Eukaryotic Pathogen *Toxoplasma gondii*. *Science (80-. )*. **314**, 1776–1781 (2006).
  21. Behnke, M. S. *et al.* Virulence differences in *Toxoplasma* mediated by amplification of a family of polymorphic pseudokinases. *Pnas* **108**, 9631–9636 (2011).
  22. Reese, M. L., Zeiner, G. M., Saeij, J. P. J., Boothroyd, J. C. & Boyle, J. P. Polymorphic family of injected pseudokinases is paramount in *Toxoplasma* virulence. *Pnas* **108**, 9625–9630 (2011).
  23. Khaminets, A. *et al.* Coordinated loading of IRG resistance GTPases on to the *Toxoplasma gondii* parasitophorous vacuole. *Cell. Microbiol.* **12**, 939–961 (2010).
  24. Hunter, C. a. & Sibley, L. D. Modulation of innate immunity by *Toxoplasma gondii* virulence effectors. *Nat. Rev. Microbiol.* **10**, 766–778 (2012).
  25. Fleckenstein, M. C. *et al.* A *toxoplasma gondii* pseudokinase inhibits host irg resistance proteins. *PLoS Biol.* **10**, 14 (2012).
  26. Khan, A., Taylor, S., Ajioka, J. W., Rosenthal, B. M. & Sibley, L. D. Selection at a Single Locus Leads to Widespread Expansion of *Toxoplasma gondii* Lineages That Are Virulent in Mice. *PLoS Genet.* **5**, 1–14 (2009).
  27. Roos, D. S., Donald, R. G., Morrisette, N. S. & Moulton, a L. Molecular tools for genetic dissection of the protozoan parasite *Toxoplasma gondii*. *Methods Cell Biol* **45**, 27–63 (1994).
  28. Wu, L. *et al.* Separation and purification of *Toxoplasma gondii* tachyzoites from in vitro and in vivo culture systems. *Exp. Parasitol.* **130**, 91–94 (2012).
  29. Malmström, H., Storå, J., Dalén, L., Holmlund, G. & Götherström, A. Extensive Human DNA Contamination in Extracts from Ancient Dog Bones and Teeth. *Mol. Biol. Evol.* **22**, 2040–2047 (2005).

30. Funakoshi, K., Baghe, M., Zhou, M., Suzuki, R. & Abe, H. Highly sensitive and specific Alu- based quantification of human cells among rodent cells. *Sci. Rep.* **7**, 1–12 (2017).
31. Su, C., Shwab, E. K., Zhou, P., Zhu, X. Q. & Dubey, J. P. Moving towards an integrated approach to molecular detection and identification of *Toxoplasma gondii*. *Parasitology* **137**, (2010).
32. English, E. D., Adomako-Ankomah, Y. & Boyle, J. P. Secreted effectors in *Toxoplasma gondii* and related species: determinants of host range and pathogenesis? *Parasite Immunol.* **37**, 127–140 (2015).
33. Xia, J., Venkat, A., Le Roch, K., Ay, F. & Boyle, J. Third generation sequencing revises the molecular karyotype for *Toxoplasma gondii* and identifies emerging copy number variants in sexual recombinants. *Prepr. from bioRxiv* (2020). doi:10.1101/2020.03.10.985549
34. Murillo-León, M. *et al.* Molecular mechanism for the control of virulent *Toxoplasma gondii* infections in wild-derived mice. *Nat. Commun.* **10**, 1–15 (2019).
35. Ajzenberg, D. *et al.* Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. *Int. J. Parasitol.* **34**, 1185–1196 (2004).
36. Pena, H. F. J., Gennari, S. M., Dubey, J. P. & Su, C. Population structure and mouse-virulence of *Toxoplasma gondii* in Brazil. *Int. J. Parasitol.* **38**, 561–569 (2008).
37. Mercier, A. *et al.* Human impact on genetic diversity of *Toxoplasma gondii*: Example of the anthropized environment from French Guiana. *Infect. Genet. Evol.* **11**, 1378–1387 (2011).
38. Ferreira, A. M., Martins, M. S. & Vitor, R. W. A. Virulence for balb/c mice and antigenic diversity of eight *Toxoplasma gondii* strains isolated from animals and humans in Brazil. *Parasite* **8**, 99–105 (2001).
39. Ferreira, A. D. M., Vitor, R. W. A., Carneiro, A. C. A. V, Brandão, G. P. & Melo, M. N. Genetic variability of Brazilian *Toxoplasma gondii* strains detected by random amplified polymorphic DNA-polymerase chain reaction ( RAPD-PCR ) and simple sequence repeat anchored-PCR ( SSR-PCR ). *Infect. Genet. Evol.* **4**, 131–142 (2004).
40. Vitaliano, S. N. *et al.* Genetic characterization of *Toxoplasma gondii* from Brazilian wildlife revealed abundant new genotypes. *Int. J. Parasitol. Parasites Wildl.* **3**, 276–283 (2014).
41. Lilue, J., Müller, U. B., Steinfeldt, T. & Howard, J. C. Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. *Elife* **2013**, 1–21 (2013).
42. Müller, U. B. Polymorphism in the IRG resistance system determines virulence of *Toxoplasma gondii* in mice. University of Cologne. Institute for Genetics. Cologne,

- Germany. *PhD Thesis* (2015).
43. Niedelman, W. *et al.* The Rhoptry Proteins ROP18 and ROP5 Mediate *Toxoplasma gondii* Evasion of the Murine , But Not the Human, Interferon-Gamma Response. *PLoS Pathog.* **8**, e1002784 (2012).
  44. Rêgo, W. M. F. *et al.* Association of ROP18 and ROP5 was efficient as a marker of virulence in atypical isolates of *Toxoplasma gondii* obtained from pigs and goats in Piauí , Brazil. *Vet. Parasitol.* **247**, 19–25 (2017).
  45. Hamilton, C. M. *et al.* Comparative virulence of Caribbean, Brazilian and European isolates of *Toxoplasma gondii*. *Parasit. Vectors* **12**, 104 (2019).

## **Chapter 5**

---

### **5. General discussion**

Adaptation to local conditions is critical in the co-evolution of host-pathogen interactions. Pathogens constitute a constant threat for multicellular organisms, which have evolved in many cases an immune system good enough to allow them to combat these virulent microbes. Selection pressure has been a major player not only for the emergence of immune defense mechanisms in the host but also to allow the pathogen to counteract immune effector mechanisms long enough for efficient transmission to a new host<sup>1</sup>.

This thesis has focused on the dynamic ecological and evolutionary relationship between the house mouse, *Mus musculus* and its natural pathogen, the apicomplexan protozoan, *Toxoplasma gondii*. We argue that a functional balance between the potential virulence of the parasite and a mouse cell-autonomous resistance mechanism is maintained by selection operating between two polymorphic systems, the ROP kinase-based virulence system of the parasite, and the interferon-inducible Immunity-Related GTPases, IRG proteins, of the mouse, known to be essential resistance factors against *T. gondii*. In particular, we have studied the impact of colonization of South America on *Mus musculus*, accompanying the early European colonists, by confrontation with novel and highly virulent South American *T. gondii* strains that can defeat the IRG resistance mechanism.

Many *T. gondii* strains have been found to be highly virulent in mouse laboratory strains<sup>2</sup>. Yet, virulence in *T. gondii* when it infects mice is paradoxical from an evolutionary point of view. Death of the mouse before the parasite's encystment interferes with the possibility of the parasite to pass to a new intermediate (any warm-blooded animal) or a definitive host (a member of the Felidae family)<sup>3</sup>. Recent studies have shown considerable polymorphism in IRG proteins among several wild and wild-derived inbred mouse strains. Some wild-derived mice, like the CIM mouse strain from

India (*M. m. castaneus*)<sup>3</sup> and the PWK mouse strain from the Czech Republic (*M. m. musculus*)<sup>4,5</sup> display natural resistance to highly virulent type I *T. gondii* strains. In both cases resistance was linked to IRG genes located on Chromosome 11, and for the CIM strain, resistance was fully linked to the polymorphic tandem IRG gene *Irgb2-b1*. Transfection of the *Irgb2-b1*CIM allele is sufficient to confer resistance to type I *T. gondii* strains to C57BL/6 cells previously susceptible to the infection<sup>3</sup>. Increasing evidence suggests *T. gondii* virulence and mouse resistance follow a form of allele-matching evolutionary dynamics.

Some of the most virulent *T. gondii* strains come from South America, and these are virulent even in wild-derived Eurasian mice resistant to virulent type I Eurasian strains. There has so far been no analysis of the genetic diversity of IRG proteins in natural mouse populations in order to understand whether IRG alleles can be linked to resistance to local *T. gondii* strains. It is important to invest in a clearer understanding of the IRG genetic diversity in wild and wild-derived mouse populations and how this diversity changes due to possible selection pressure imposed by pathogens. We made use of field-collected mouse samples, wild-derived mouse strains, newly established cell lines, and local isolated *T. gondii* strains to investigate IRG protein diversity and possible adaptation to local *T. gondii* strains in South American mice. The following is a summary of the findings presented in this work and a discussion of several concepts emerging from this research.

## 5.1. Summary of the findings.

This work began in Chapter 2 with the transcriptome analysis of immortalized cell lines from wild-caught and wild-derived mice from Portugal and Brazil. We found that Portuguese mice have a diverse battery of *Irgb2-b1* alleles, including those found in mice from the Scottish Isles (*Irgb2-b1*<sub>SIN/SIT</sub>) at the highest prevalence and less frequently the *Irgb2-b1*<sub>PWK</sub> allele



from *M. m. musculus* mice. In contrast, Brazilian mice possessed only two different *Irgb2-b1* alleles, namely a strikingly high prevalence of the *Irgb2-b1<sub>PWK</sub>* allele, and a correspondingly low prevalence of *Irgb2-b1<sub>BL6</sub>*. We looked for signs of selection or of bottleneck events in the Brazilian mouse population that could explain the high prevalence of the *Irgb2-b1<sub>PWK</sub>* allele. Multiple polymorphic sites were found in non-coding regions in the *Irgb2-b1* gene of *Irgb2-b1<sub>PWK</sub>* allele carriers, many but not all of which were also found in European mice. This is most likely explained by a hitchhike of these SNPs through very close linkage to *Irgb2-b1* alleles in ancestral European chromosomes that we probably have not found yet. Additionally, analysis of closely-linked genes revealed the presence of a significant number of haplotypes that uniquely belong to the Brazilian population, as well as other shared between Brazilian and European mice. A maternal phylogeny analysis showed Brazilian mice have a diverse matrilineal ancestry with origins not only in Portugal but also Scotland, Germany, Spain, and Italy. Although, this situation might be explained as well by reverse gene flow from Europe. Finally, expression levels based on transcriptomic data of the different *Irgb2-b1* alleles confirmed that the *Irgb2-b1<sub>PWK</sub>* is a highly expressed allele and this level of expression is regulated in cis. Finally, we found that amino acid sequences in the H4 and  $\alpha$ D structural domains of *IRGB2<sub>PWK</sub>* subunit are distinct from those found in the *IRGB2<sub>BL6</sub>* and *IRGB2<sub>CIM</sub>*. No differences in the same domains were found for the *IRGB6* protein of *IRGB2<sub>PWK</sub>* carriers. Our results suggest selection on standing variation in Brazilian mice from European populations, possibly associated with fitness advantage (resistance) in Brazilian mice to highly virulent indigenous *T. gondii* strains.

In Chapter 3, we focused our efforts to test the ability of the *Irgb2-b1<sub>PWK</sub>* allele to confer resistance against infection with highly virulent South American *T. gondii* strains. We confirmed the high expression of the *Irgb2-b1<sub>PWK</sub>* protein in IFN $\gamma$ -induced cells from different geographic locations. We

also showed that *Irgb2-b1<sub>PWK</sub>* protein loaded more efficiently onto parasitophorous vacuole membranes (PVMs) of cells from Brazilian mice infected with local *T. gondii* strains than to cells infected with the laboratory strain RH. Furthermore, we found that expression of the *Irgb2-b1<sub>PWK</sub>* protein in *Irgb2b1*-deficient CIM cells is sufficient to confer full resistance not only to the RH strain but also to highly virulent Brazilian strains. These *in vitro* results strongly supported the role of the *Irgb2-b1<sub>PWK</sub>* allele in resistance to virulent *T. gondii* strains and we investigated whether this phenotype can also be seen *in vivo*. We found that European mice which are carriers of the *Irgb2-b1<sub>PWK</sub>* allele can resist European type I virulent strains but their resistance to virulent South American *T. gondii* strains is either intermediate or absent according to the strain used. We also demonstrated that some Brazilian mice carriers of the *Irgb2-b1<sub>PWK</sub>* allele are significantly more resistant against South American strains than their European counterparts. Interestingly, the two Brazilian mouse strains tested did not display the same resistance to South American strains, which might indicate there are yet unidentified factors in the resistance to these local strains in Brazilian mice. Altogether, our data suggest adaptation to local *T. gondii* strains in Brazilian mice, possibly associated with selection pressure imposed by those parasites.

Finally, in chapter 4 we did a whole-genome sequence analysis of the three Brazilian *T. gondii* strains used for this thesis. We found high levels of genetic similarity among them. Even though differences along the genome were commonly found, genes that encode the major parasite's virulence factors, *ROP18* and *ROP5*, are identical among these Brazilian highly virulent strains. We also identified isoform types for the *ROP5* gene and their phylogenetic relationship with other strains. All isoforms clustered phylogenetically with those reported for other virulent strains, consistent with the virulent characteristics of the strains analyzed. Given the difficulties encountered for the genotyping of the Brazilian strains used for this thesis,

we developed a PCR-RFLP based protocol for the genotyping of our strains. SNPs found among these strains from a whole-genome analysis were shown to be reliable for the distinction of these strains and may be valuable for the determination of other *T. gondii* strains in which usual protocols for genotyping are not sufficient.

## **5.2. High prevalence of the *Irgb2-b1*<sub>PWK</sub> in South American house mice and the rapid adaptation to local parasites.**

The encounter of a host with a new pathogen can result in intense selection pressure on both populations, which is represented in rapid evolutionary changes for the host and the pathogen. By analyzing the allele frequency of the *Irgb2-b1* allele in Brazilian and European populations of wild-caught and wild-derived house mice, we showed that the *Irgb2-b1*<sub>PWK</sub> allele has an extremely high prevalence in the Brazilian population. Multiple *Irgb2-b1* alleles can be found in European house mouse populations, and the *Irgb2-b1*<sub>PWK</sub> allele was found in low prevalence. Such a difference in *Irgb2-b1* allele prevalence between European and Brazilian mouse populations may follow from demographic perturbations or host-parasite co-evolutionary events. For the first explanation, a genetic bottleneck or a founder effect looks plausible since current Brazilian house mice have as ancestors the surely limited but equally surely unknown numbers of mice arriving with European settlers in the 15<sup>th</sup> century and subsequently. However, a species that colonizes a new territory is subject to novel and perhaps intense selection pressures including from novel parasites, as is certainly the case for mice arriving in Brazil, suggesting the plausibility of a host-parasite co-evolutionary event as the cause of genetic diversity reduction on loci functional in immune resistance.

Populations that are exposed to highly virulent pathogens usually experience changes in their genetic structure, for example, human African

populations exposed to severe malaria (*Plasmodium sp.*) infections exhibit mutations and high prevalence of malaria resistance associated alleles such as in the HbS Sickle cell, Hemoglobin-B (HBB) and Hemoglobin-C (HBC) genes and MHC class I and II genes, as result of natural selection driven by evolutionary pressure<sup>6-9</sup>. Another group of hemoglobin disorders, thalassemia ( $\alpha$  and  $\beta$ ), can also protect against malaria with an incidence of up to 30% among communities of West Africa<sup>10</sup>. Whilst our statistical data cannot provide full support to either demographic perturbations or host-parasite co-evolutionary events for the high prevalence of *Irgb2-b1*<sub>PWK</sub> in the Brazilian mice, data obtained from maternal phylogenetic analyses (D-loop region from mitochondrial DNA) and the high genetic diversity present in nearby genes suggested that rather than a bottleneck due to a colonization process, a selective sweep event has occurred in this mouse population. This is most likely a soft sweep result of local adaptation to pathogens from standing genetic variation, as multiple adaptive haplotypes sweep through the populations simultaneously, which leads to persistence of genetic variation near the adaptive site<sup>11</sup>. The variation we found in genes sequences both within and adjacent to the *Irgb2-b1* locus in Brazilian mice is fully compatible with this interpretation.

Standing variation can help to accelerate resistance emergence to new pathogens. This phenomenon has been typically reported in interactions that occurred in a relatively short evolutionary time. One of the best-documented field examples of this fast co-evolutionary process is the infection of the wild European rabbit (*Oryctolagus cuniculus*) with the myxoma virus (genus *Leporipoxvirus*, family *Poxviridae*). The first exposure of the European (native to the Iberian peninsula) rabbit to the virus occurred only seventy years ago when the myxoma virus (originally found in rabbits from North and South America) was released in Australia and subsequently in Europe (France, United Kingdom, Denmark, and Sweden) as a biological control for the exponentially growing rabbit population<sup>12</sup>. The originally

released strain of the myxoma virus displayed a 99.8% case fatality, which imposed an intense selection pressure on the rabbit population, resulting in a rapid arms-race against the rabbit immune system and the virus<sup>13,14</sup>. Almost immediately after the first outbreaks, the rapid development of resistance to myxomatosis was reported in Australia, France, and the UK due to both evolution of less virulent viral phenotypes and increased resistance in rabbit populations. Many of the genetic changes in the rabbits occurred in immunity-related genes (CD96, FCRL, CD200-R and IFN- $\alpha$ ) which all play a regulatory role in the innate immune response, resulting in an enhanced antiviral immunity. A cumulative effect of multiple alleles shifting in frequency across the genome was observed, as opposed to a few major-effect changes to the immune response, supporting a polygenic basis of resistance. Most likely a selection on standing genetic variation that was present over 800 years ago in continental Europe and was later carried with the rabbits that colonized the UK and Australia was the cause of the fast emergence of resistance<sup>14</sup>.

As the European wild rabbit had a short evolutionary time to adapt to a highly virulent new pathogen, so house mice arrived in South America for the first time only 500 years ago and probably their first encounter with several new pathogens including the highly virulent South American *T. gondii* strains occurred just then. Selection for resistance against infectious diseases can change rapidly in these natural populations, as new pathogens appear<sup>15</sup>. Wild mice can carry a substantial burden of infections and multiple microorganisms present in South America might have infected this new mammalian host. This includes a wide range of pathogens such as viruses, bacteria, ectoparasites, and intestinal nematode infections<sup>16</sup>. Therefore, selection pressure from one or more of these pathogens might have occurred in the IRG system. As previously mentioned, IRG proteins have been reported to participate in the cellular resistance against bacterial parasites of mice such as *C. trachomatis* and *C. psittaci*, *Francisella*

*novicida*, *Escherichia coli*, *Salmonella Typhimurium* as well the microsporidian fungus, *E. cuniculi*<sup>17–21</sup>. However, so far there is no evidence about the role of IRG tandem proteins (Irgb2-b1) in resistance against any other parasite except *T. gondii*.

House mice in Europe and Asia have been exposed to local *T. gondii* strains for a long time, but most of these strains are considered avirulent due to the ability of these mice to efficiently control the infection. Until now there is no report yet of house mice able to survive infection with virulent South American strains, which are proven to evade and sabotage host defenses<sup>22,23</sup>. As highlighted in previous studies, the existence of virulent *T. gondii* strains that proliferate unrestricted and kill a susceptible mouse is a Pyrrhic victory<sup>3,4,24</sup>. There is little chance for a dead mouse to pass a virulent *T. gondii* to either other intermediate or definitive host (for sexual replication)<sup>4</sup>. Therefore, we expected to find some South American house mice with a certain level of resistance against local virulent *T. gondii* strains. Our data indicated that the Irgb2-b1<sub>PWK</sub> allelic form, widely present in the Brazilian mouse population but to a lower extent in European populations, displays several properties associated with the capacity to confer resistance against virulent strains that are also found in the Irgb2-b1<sub>CIM</sub><sup>1,3</sup> allelic form. The Irgb2-b1<sub>PWK</sub> protein, like Irgb2-b1<sub>CIM</sub> has high expression levels. Low expression levels of Irgb2-b1 have thus far been associated with mouse strains fully susceptible to virulent *T. gondii* strains (e.g. C57BL/6)<sup>1,3,4</sup>.

*In vitro*-based approaches, especially those used to evaluate cell-autonomous IRG resistance, where parasite vacuole disruption and host cellular necrosis is strong, are useful to monitor the efficacy of a given Toxoplasma–host genetic combination<sup>25</sup>. A recent study has shown by gene deletion and complementation that polymorphic variation in Irgb2-b1 is responsible for the restriction of virulent *T. gondii* strains in CIM cells and the failure of restriction in C57BL/6 cells<sup>1</sup>. In agreement with this study, we

found that complementation of CIM<sub>Irgb2-b1<sup>KO</sup></sub> cells with Irgb2-b1<sub>PWK</sub> confers full resistance against type I strains and South American strains (AS28, BV, and N). Complementation with Irgb2-b1<sub>CIM</sub> confers a lower level of resistance and complementation with Irgb2-b1<sub>BL6</sub> confers no resistance at all. Additionally, an IRG based resistance phenotype (reduction of infected cells and increased necrosis) is also observed in cells isolated from mice that carry the Irgb2-b1<sub>PWK</sub> and challenged with multiple virulent *T. gondii* strains. Altogether, these data confirmed that polymorphic variations in the Irgb2-b1 gene are determinants of mouse resistance *in vitro* to virulent *T. gondii* strains.

As has been shown in this study and many others in the host-parasite interactions field, *in vitro*-based approaches are particularly useful for high throughput applications. However, *in vivo* and *ex vivo* studies are still crucial to understand the complex nature of parasite virulence and immune host's responses<sup>26</sup>. Compared to other parasites, *T. gondii* (tachyzoites) easily propagate in cell culture but their virulence effects might be different both qualitatively and quantitatively from what occurs during natural infection. Additionally, the infection route (e.g. oral, intraperitoneal, etc.), initial dosage, and cellular immune responses in the early stage of infection can impact disease progression *in vivo*. In this study, we have used different laboratory-reared wild-derived mouse strains (*M. m. musculus* and *M. m. domesticus*) as exemplars for natural infection with virulent *T. gondii* strains. However, we are aware that wild animals experience a range of complex symbiont exposures and environmental stressors that cannot be adequately replicated in captivity and could affect the outcome of these artificial infections<sup>27</sup>. Hassan et al., 2019 have suggested that levels of resistance against virulent *T. gondii* strains is correlated with mouse sub-species<sup>5</sup>. Our data, previous studies from our lab (Dr. Urs Benedikt Müller, unpublished data) have demonstrated that IRG genes in Chromosome 11 from *M. m.*

*musculus* (PWK and PWD strains) are responsible for the resistance *in vivo* to virulent *T. gondii* type I strains, independently of subspecies. We showed that rather than mouse subspecies, *Irgb2-b1* polymorphic variations account for resistance *in vivo* to virulent Type I strains and are necessary against some South American strains. The resistant PWK (a *M. m. musculus*) allele of *Irgb2-b1* from the Czech Republic, is present at a significant frequency in wild *M. m. domesticus* in Europe, and as we show, abundantly in *M. m. domesticus* in Brazil. Hassan et al., 2019 position could be considered tenable for the high resistance of the *M. m. castaneus* strains, CAST from Thailand and CIM from South India. However, this resistance is due exclusively to the distinctive *Irgb2-b1* allele common in *M. m. castaneus* strains, and confers complete resistance to type I virulent *T. gondii* strains. In an F2 cross between CIM and C57BL/6 mice, complete resistance is associated with the presence of the CIM allele in the homozygous state, while mice homozygous for the C57BL/6 allele of *Irgb2-b1* are precisely as susceptible as the C57BL/6 strain itself<sup>3</sup>. It is not known whether the resistant allele common in *M. m. castaneus* strains is also present in other wild mouse subspecies (it has not yet been found).

The outcome of the C57BL/6 F2 cross with CIM showed conclusively that there is no other independent resistance allele associated with the resistance of CIM. However, our data on the resistance of the two Manaus strains from Brazil, MANA and MANB (both homozygous carriers for the *Irgb2-b1*<sub>PWK</sub> allele), showed that these house mice are fully resistant to virulent RH, but the MANA strain is also highly if not completely resistant to infection with the virulent South American strain TgMmBr01, while the MANB strain is not. This result suggest that besides IRG tandem proteins there may be other factors in the resistance against South American *T. gondii* strains that are still waiting to be identified, as Hassan et al., 2019 have suggested<sup>5</sup>.



*M. m. domesticus* is the house mouse subspecies found in North and South America. The detection of an *M. m. musculus*-like *Irgb2-b1* allele in *M. m. domesticus* in Brazil might have different explanations. One is there is a mixing of the two species in the Brazilian territory. Although, it might be unlikely that *M. m. musculus* have first colonized the American continent probably accompanying the first human colonizers that crossed from Asia to America through the Bering strait more than 14,000 years ago<sup>28</sup>, the mixing of the two mouse subspecies might have occurred already in Europe. Available evidence indicates that even though *M. m. domesticus* and *M. m. musculus* have been separated for almost 300,000 years ago, which in numbers of generations and relative molecular divergence is similar to the split of chimpanzees and humans<sup>29</sup>, hybridization (gene flow) among these two subspecies happens in zones of secondary contact, like in Central Europe<sup>30,31</sup>. Moreover, several studies have shown that most wild-derived mouse strains have a strong level of introgression from other subspecies (29% from two subspecies and 56% from all three subspecies)<sup>32</sup>. More importantly, given the resistance-conferring properties of the *M. m. musculus*-like *Irgb2-b1*<sub>PWK</sub> allele and the short time for selection to bring forward new resistance alleles<sup>33</sup> since the first European house mice arrived in Brazil, it is overwhelmingly likely that standing genetic variation present in the immigrant European population contributed to accelerate the process of resistance emergence against local *T. gondii* strains.

### **5.3. Genotype–genotype interactions: *Irgb2-b1* alleles and ROP5/ROP18 isoforms.**

Both host and parasite genotypes are fundamental determinants of infection outcome. Genotype–genotype interactions that determine infectivity in host-parasite systems have been studied for a long time. One of the prime models for antagonistic co-evolution is the one between the water flea *Daphnia magna* infected with the castrating bacterial pathogen *Pasteuria ramose*<sup>31</sup>. This system follows a matching-allele model (MAM), that

assumes that an exact match between suites of host and parasite alleles is required for a successful immune response to occur<sup>32</sup>, for example, a parasite can only infect when its alleles match those of its host<sup>33</sup>. Recent studies have demonstrated that specific *Irgb2-b1* protein allelic forms have a crucial role in the resistance against virulent type I *T. gondii* strains by targeting allele-specific ROP5/ROP18 isoforms<sup>1,2,3,4</sup>, possibly indicating that this system follows a MAM. The direct interaction between the ROP5B protein from the virulent type I RH strain and *Irgb2-b1*<sub>CIM</sub> protein has been proven to be responsible for resistance to this otherwise virulent strain in *M. m. castaneus*<sup>1</sup>. However, none of the ROP5 isoforms from the South American strain VAND are targeted by *Irgb2-b1*<sub>CIM</sub>, making CIM mice fully susceptible to VAND<sup>1</sup>. These data suggest a high degree of molecular specificity in the interaction between *Irgb2-b1* alleles and ROP5 isoforms.

Resistance against virulent *T. gondii* strains depends on interaction of the *Irgb2-b1* protein of a mouse host with the ROP5 protein isoform of the infecting *T. gondii* strain. Nonetheless, it is not clear yet to what extent resistance against South American *T. gondii* strains can be solely attributed to *Irgb2-b1* and ROP5 isoforms interactions. The three South American *T. gondii* strains used for this study (AS28, BV, N) are virulent in CIM mice<sup>3</sup> but displayed different levels of virulence in PWD and MANA mice. Our whole-genome analysis revealed that these three strains carry identical ROP5 allele specific isoforms and a single ROP18 allele. Their genetic combination of ROP5/ROP18 alleles corresponds to that found in several virulent South American *T. gondii* strains<sup>35</sup>. However, we are aware that more ROP5 isoforms may be present in these strains, but we could not identify them given the limitations for the characterization of isoforms when using short-read sequencing technologies (e.g. Illumina)<sup>36,37</sup>. Moreover, differential expression levels of the ROP5 isoforms found in these strains could determine their interaction with the distinct *Irgb2-b1* protein forms, under the assumption that the ROP5 isoform that interacts with *Irgb2-b1* is

the one with the highest expression level, but we did not perform this analysis. Other studies have suggested that atypical *T. gondii* strains that are virulent in wild-derived mice such as *M. m. castaneus* and *M. m. musculus* might have either a unique ROP5 isoform that can evade *Irgb2-b1* targeting or have extra copies of ROP5 isoforms for binding both *Irgb2-b1* or other effector IRGs<sup>4</sup>. Nonetheless, these differences in resistance against South American *T. gondii* strains in wild-derived mice might involve common ROP5 isoforms that could be better targeted by *Irgb2-b1* alleles such as the *Irgb2-b1*<sub>PWK</sub>. For example, the atypical VAND and TgCATBr5 strains display a virulent and avirulent phenotype, respectively, in consomic C57BL/6/PWD Chr.11.1 mice<sup>4</sup>. In this context, we would expect at least one of the six ROP5 isoforms identified in the Brazilian strains from this study to show a high level of identity with ROP5 isoforms from TgCATBr5, but that is not the case.

As previously suggested by others studies<sup>4</sup>, the possibility that a different effector protein, not the ROP5/ROP18 complex, could be involved in differential resistance against South American *T. gondii* strains in wild-derived mouse strains (*M. m. musculus* and *M. m. domesticus*), cannot be excluded. Other IRG proteins might also play some role in this process. The effector protein IRGB6 (highly polymorphic in wild-derived mice<sup>2,21</sup>) could be another candidate. A very recent publication and unpublished data from our lab have demonstrated that *Irgb6*-deficient cells are defective in the recruitment of ubiquitin and other IFN $\gamma$ -inducible GTPase effectors onto the *T. gondii* PVM<sup>38</sup>. Additionally, *Irgb6*-deficient mice are unable to fully control infection with avirulent *T. gondii* strains, causing an increase in parasite burden in tissues, increased inflammatory responses, and subsequent death<sup>38</sup>. Perhaps, the ROP5/ROP18 complex from some South American *T. gondii* could interact with the *Irgb6* protein from wild-derived *M. m. musculus* and *M. m. domesticus* that are resistant to the infection with these strains. Our data showed the putative H4 and  $\alpha$ D structural domains (the IRG –

ROP5 interface) in the IRGB6 protein of the *Irgb2<sub>PWK</sub>* carriers PWK and PWD (*M. m. musculus*) and MANB (*M. m. domesticus*) displayed no differences. However, we found differences in other regions of the protein that might be relevant for the interaction with *T. gondii* virulence factors (residues mostly in balancing selection)<sup>21</sup>. This raises the question whether a single-domain IRG protein such as IRGB6 could autonomously inactivate phosphorylation by the virulence kinases and therefore replace or enhance *Irgb2-b1* decoy function.

## 5.4. Concluding remarks and future perspectives

Over the last two decades, many studies have shown the importance of the IRG proteins and their role in resistance against intracellular pathogens in mammals<sup>42–46</sup>. More recently, the role of the IRG system in laboratory mice against avirulent *T. gondii* strains has been widely studied<sup>47</sup>. However, virulent strains of *T. gondii* can evade the host immune system by secreting kinases and pseudokinases that inactivate IRG proteins<sup>48,49</sup>. Our lab has demonstrated the complexity and variability of IRG genes in wild mice that allows these rodents to survive virulent *T. gondii* strains. The work developed during this thesis exploits a natural experiment in host-pathogen interactions provided by the colonization of South America, rich in endemic *T. gondii* strains, by the Eurasian species, *Mus musculus*, brought in the 15<sup>th</sup> century aboard the ships of the European colonists. Our work shows high homogeneity of one resistance allele of Eurasian origin at the *Irgb2-b1* locus found in Brazilian *M. m. domesticus*. Evidence from mitochondrial sequencing showing extensive variation, and the presence of multiple SNP haplotypes in and surrounding the homogenized *Irgb2-b1* allele both argue against a population bottleneck to account for the homogenization of the resistant *Irgb2-b1* allele. We conclude that the IRG system has been subjected to strong selection pressure by the highly virulent *T. gondii* strains in South America, resulting in rapid adaptation to local *T. gondii* strains by

selection on standing polymorphic variation in the IRG system. We have also extended the current knowledge of the role of polymorphisms in the *Irgb2-b1* tandem protein to target the virulence factors of the ROP5/ROP18 complex.

Wild house mice can sustain infections with a wide range of pathogens<sup>16</sup>. Also, *T. gondii* can infect any warm-blood animal, which makes this a very diffuse system where multiple hosts and parasites can be involved in the shaping and evolution of the IRG system. In that context, one of the remaining questions we still have in this field is what is the real evolutionary cost of the IRG system. We strongly suspect that it has indeed a cost, since, for example, most birds and some reptiles have apparently lost their IRG genes. It is, however, unclear both why the infectious environment allows loss of the IRG system in some animal groups and its persistence at high complexity in others, and also what the cost of maintaining the IRG system is due to. However, the cost to maintain a single IRG allele in a population might be different and therefore events such as the high prevalence of the *Irgb2-b1*<sub>PWK</sub> in the Brazilian mouse population can be maintained. This situation of low cost to maintain certain alleles in immune-related genes is not new. One of the best-characterized examples is the MHC system and the emergence of alleles that might never be used against a pathogen, but are still present in the population without being counter selected. Therefore, further studies are still needed in different infection models between wild mice and *T. gondii* strains to better understand how local selection works and how it can affect the spread of IRG alleles that can confer resistance in a certain population. Globally, house mice are probably now the most important evolutionary intermediate host of *T. gondii* due to their large, worldwide population and their cohabitation with domestic cats. However in South America, until the 16<sup>th</sup> century free from both domestic cats and mice from Eurasia, *T. gondii* has presumably co-evolved for a much longer time with other South American rodents, large and small, which are common

prey species for South American felines (e.g. Jaguars) and have distinctive IRG proteins (Howard Lab, unpublished data). Since South America seems to have been the cradle of *T. gondii* evolution, it would be interesting to investigate the possible role of the IRG system in coevolving native rodents.

## Bibliography

1. Murillo-León, M. *et al.* Molecular mechanism for the control of virulent *Toxoplasma gondii* infections in wild-derived mice. *Nat. Commun.* **10**, 1–15 (2019).
2. Sibley, L. D. & Boothroyd, J. . Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature* **359**, 82–85 (1992).
3. Lilue, J., Müller, U. B., Steinfeldt, T. & Howard, J. C. Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. *Elife* **2013**, 1–21 (2013).
4. Müller, U. B. Polymorphism in the IRG resistance system determines virulence of *Toxoplasma gondii* in mice. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2015).
5. Hassan, M. A., Olijnik, A.-A., Frickel, E.-M. & Saeij, J. P. Clonal and atypical *Toxoplasma* strain differences in virulence vary with mouse sub-species. *Int. J. Parasitol.* **49**, 63–70 (2019).
6. Ohashi, J. *et al.* Extended Linkage Disequilibrium Surrounding the Hemoglobin E Variant Due to Malarial Selection. *Am. J. Hum. Genet.* **74**, 1198–1208 (2004).
7. Modiano, D. *et al.* Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature* **414**, 305–308 (2001).
8. Dominguez-Andres, J. & Netea, M. G. Impact of Historic Migrations and Evolutionary Processes on Human Immunity. *Cell Press Rev.* **40**, 1105–1119 (2019).
9. Hill, A. V. S. *et al.* Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595–600 (1991).
10. Williams, T. N. & Weatherall, D. J. World distribution, population genetics, and health burden of the hemoglobinopathies. *Cold Spring Harb. Perspect. Med.* **2**, a011692–a011692 (2012).
11. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent Selective Sweeps in North American *Drosophila melanogaster* Show Signatures of Soft Sweeps. *PLOS Genet.* **11**, 1–32 (2015).
12. Kerr, P. J. Myxomatosis in Australia and Europe: A model for emerging infectious diseases. *Antiviral Res.* **93**, 387–415 (2012).
13. Kerr, P. J. *et al.* Next step in the ongoing arms race between myxoma virus and wild rabbits in Australia is a novel disease phenotype. *Proc. Natl. Acad. Sci.* **114**, 9397–9402 (2017).
14. Alves, J. M. *et al.* Parallel adaptation of rabbit populations to myxoma virus. *Science*. **363**, 1319–1326 (2019).
15. Woolhouse, M. E. J., Haydon, D. T. & Antia, R. Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.* **20**, 238–244

- (2005).
16. Abolins, S. *et al.* The comparative immunology of wild and laboratory mice, *Mus musculus domesticus*. *Nat. Commun.* **8**, 14811 (2017).
  17. Miyairi, I. *et al.* The p47 GTPases Irgp2 and Irgb10 Regulate Innate Immunity and Inflammation to Murine *Chlamydia psittaci* Infection. *J. Immunol.* **179**, 1814–1824 (2007).
  18. da Fonseca Ferreira-da-Silva, M., Springer-Frauenhoff, H. M., Bohne, W. & Howard, J. C. Identification of the Microsporidian *Encephalitozoon cuniculi* as a New Target of the IFN $\gamma$ -Inducible IRG Resistance System. *PLoS Pathog.* **10**, e1004449 (2014).
  19. Eren, E. *et al.* Irgm2 and Gate-16 cooperatively dampen Gram-negative bacteria-induced caspase-11 response. *EMBO Rep.* **21**, e50829 (2020).
  20. Finethy, R. *et al.* Dynamin-related Irgm proteins modulate LPS-induced caspase-11 activation and septic shock. *EMBO Rep.* **21**, e50830 (2020).
  21. Linder, A. & Hornung, V. Irgm2 and Gate-16 put a break on caspase-11 activation. *EMBO Rep.* **21**, e51787 (2020).
  22. Jensen, K. D. C. *et al.* *Toxoplasma gondii* Superinfection and Virulence during Secondary Infection Correlate with the Exact ROP5 / ROP18 Allelic Combination. *MBio* **6**, 1–15 (2015).
  23. Behnke, M. S. *et al.* Rhoptry Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of *Toxoplasma gondii*. *PLOS Genet.* **11**, e1005434 (2015).
  24. Lilue, J. Haplotypic polymorphism of the IRG protein family mediates resistance of mice against virulent strains of *Toxoplasma gondii*. University of Cologne. Institute for Genetics. Cologne, Germany. *PhD Thesis* (2012).
  25. Alvarez, C. *et al.* in *Toxoplasma gondii: Methods and Protocols, Methods in Molecular Biology* (ed. Tonkin, C. J.) 371–409 (Springer US, 2020). doi:10.1007/978-1-4939-9857-9\_20
  26. Swann, J., Jamshidi, N., Lewis, N. E. & Winzeler, E. A. Systems analysis of host-parasite interactions. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **7**, 381–400 (2015).
  27. Jackson, J. a. Immunology in wild nonmodel rodents: an ecological context for studies of health and disease. *Parasite Immunol.* **37**, 220–232 (2015).
  28. Bourgeon, L., Burke, A. & Higham, T. Earliest Human Presence in North America Dated to the Last Glacial Maximum: New Radiocarbon Dates from Bluefish Caves, Canada. *PLoS One* **12**, e0169486 (2017).
  29. Staubach, F. *et al.* Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (*Mus musculus*). *PLoS Genet.* **8**, (2012).
  30. Phifer-Rixey, M. & Nachman, M. W. Insights into mammalian biology from the wild house mouse *Mus musculus*. *Elife* **4**, 1–13 (2015).



31. Duvaux, L., Belkhir, K., Boulesteix, M. & Boursot, P. Isolation and gene flow: inferring the speciation history of European house mice. *Mol. Ecol.* **20**, 5248–5264 (2011).
32. Yang, H. *et al.* Subspecific origin and haplotype diversity in the laboratory mouse. *Nat. Genet.* **43**, 648–655 (2011).
33. Bangham, J., Obbard, D. J., Kim, K.-W., Haddrill, P. R. & Jiggins, F. M. The age and evolution of an antiviral resistance mutation in *Drosophila melanogaster*. *Proc. R. Soc. B Biol. Sci.* **274**, 2027–2034 (2007).
34. Luijckx, P., Fienberg, H., Duneau, D. & Ebert, D. Resistance to a bacterial parasite in the crustacean *Daphnia magna* shows Mendelian segregation with dominance. *Heredity (Edinb)*. **108**, 547–551 (2012).
35. Thrall, P. H., Barrett, L. G., Dodds, P. N. & Burdon, J. J. Epidemiological and Evolutionary Outcomes in Gene-for-Gene and Matching Allele Models. *Front. Plant Sci.* **6**, 1–12 (2016).
36. Luijckx, P., Fienberg, H., Duneau, D. & Ebert, D. A matching-allele model explains host resistance to parasites. *Curr. Biol.* **23**, 1085–1088 (2013).
37. Fleckenstein, M. C. *et al.* A toxoplasma gondii pseudokinase inhibits host irg resistance proteins. *PLoS Biol.* **10**, 14 (2012).
38. Shwab, E. K. *et al.* The ROP18 and ROP5 gene allele types are highly predictive of virulence in mice across globally distributed strains of *Toxoplasma gondii*. *Int. J. Parasitol.* **46**, 141–146 (2016).
39. Xia, J. *et al.* Strain-specific disruption of interferon-stimulated N-myc and STAT interactor (NMI) function by *Toxoplasma gondii* type I ROP18 in human cells. *Parasitology* 1–32 (2020). doi:10.1017/S0031182020001249
40. Niedelman, W. *et al.* The Rhoptry Proteins ROP18 and ROP5 Mediate *Toxoplasma gondii* Evasion of the Murine , But Not the Human, Interferon-Gamma Response. *PLoS Pathog.* **8**, e1002784 (2012).
41. Lee, Y. *et al.* Initial phospholipid-dependent Irgb6 targeting to *Toxoplasma gondii* vacuoles mediates host defense. *Life Sci. Alliance* **3**, 1–16 (2019).
42. Boehm, U. *et al.* Two families of GTPases dominate the complex cellular response to IFN-gamma. *J. Immunol.* **161**, 6715–6723 (1998).
43. Gilly, M. & Wall, R. The IRG-47 gene is IFN-gamma induced in B cells and encodes a protein with GTP-binding motifs. *J. Immunol.* **148**, 3275–3281 (1992).
44. Lafuse, W. P., Brown, D., Castle, L. & Zwillling, B. S. Cloning and characterization of a novel cDNA that is IFN-γ-induced in mouse peritoneal macrophages and encodes a putative GTP-binding protein. *J. Leukoc. Biol.* **57**, 477–483 (1995).
45. Khaminets, A. *et al.* Coordinated loading of IRG resistance GTPases on to the *Toxoplasma gondii* parasitophorous vacuole. *Cell. Microbiol.* **12**, 939–961 (2010).
46. Hunn, J. P., Feng, C. G. & Howard, J. C. The immunity-related GTPases in

- mammals: a fast-evolving cell- autonomous resistance system against intracellular pathogens. *Mamm Genome* **22**, 43–54 (2011).
47. Howard, J. C., Hunn, J. P. & Steinfeldt, T. The IRG protein-based resistance mechanism in mice and its relation to virulence in *Toxoplasma gondii*. *Curr. Opin. Microbiol.* **14**, 414–421 (2011).
  48. Fentress, S. J. *et al.* Phosphorylation of immunity-related GTPases by a *Toxoplasma gondii* secreted kinase promotes macrophage survival and virulence. *Cell Host Microbe* **8**, 484–495 (2010).
  49. Steinfeldt, T. *et al.* Phosphorylation of mouse immunity-related gtpase (IRG) resistance proteins is an evasion strategy for virulent *Toxoplasma gondii*. *PLoS Biol.* **8**, (2010).

## 6. Key resources table

### Antibodies

Reagent or resource	Recognized antigen	Type	Dilution*			Source	Identifier
			IFI	WB	FC		
10E7 - Mouse Irga6	Recombinant Mouse Irga6	Mouse monoclonal	1:1000	1:1000		N/A	N/A
165 - Mouse Irga6	Recombinant Mouse Irga6	Rabbit serum		1:10000		N/A	N/A
2.1.2 - GRA7	<i>T. gondii</i> GRA7	Rat monoclonal	1:500			N/A	N/A
954 - Mouse tandem IRGs	Irgb1 C-terminal peptide CLSDLPEYW ETGMEL	Rabbit serum	1:2000	1:4000		N/A	N/A
Alexa Fluor® 488 Donkey Anti-Mouse	Mouse IgG	Polyclonal secondary	1:1000		1:300	Thermo Fisher	R37114; RRID:AB_2556542
Alexa Fluor® 555 Goat Anti-Rabbit	Rabbit IgG	Polyclonal secondary	1:1000			Thermo Fisher	A27039; RRID:AB_2536100
Alexa Fluor® 647 Goat Anti-Rat	Rat IgG	Polyclonal secondary	1:1000			Thermo Fisher	A-21247; RRID:AB_141778
Anti-Mouse IgG (Donkey) DyLight™ 800	Mouse IgG	Polyclonal secondary		1:10000		Rockland	610-745-002; RRID:AB_1660927
Anti-Rabbit IgG (Goat) DyLight™ 680	Rabbit IgG	Polyclonal secondary		1:10000		Rockland	611-144-002-0.5; RRID:AB_11182462
Anti-Tubulin	Recombinant full length protein corresponding to Human Tubulin aa 1-451	Mouse monoclonal		1:1000		Abcam	ab56676; RRID:AB_945996
TP3	<i>T. gondii</i> P30 protein/ SAG1	Mouse monoclonal			1:250	Santa Cruz	sc-52255; RRID:AB_630350

\*IFI: immunofluorescence; WB: western blot; FC: Flow Cytometry

\*\*Table modified from Alvarez et al., 2020

### Parasite strains

Reagent or resource	Genotype **	Origin	Year isolated	Geographic location	Reference	Note
AS28	Not specified	Mouse	1969	Brazil	Deane et al., 1971	
BV	Not specified	Goat	1975	Brazil	Chiari et al., 1985	
ME49	#1	Sheep	1965	USA	Lunde and Jacobs, 1983	
N	Not specified	Rabbit	1952	Brazil	Nóbrega et al., 1952	
Pru-tomato	#1	Human	N/A	N/A	Gregg et al., 2013	transgenic PRU strain expressing tdTomato
RH-YFP	#10	Human	N/A	N/A	Gubbels et al., 2003	transgenic RH strain expressing YFP

RHΔhxp <sup>prt</sup>	#10	Human	N/A	N/A	Roos et al., 1994	transgenic RH strain in which the hypoxanthine-xanthine-guanine phosphoribosyl transferase (HXGPRT) locus has been deleted
TgMmBr01	Not specified	Mouse	2014	Brazil	Unpublished data, Howard Lab	Isolated from a wild-caught mouse in Belem, Brazil.

\*Table modified from Alvarez et al., 2020; Müller, 2015; Ferreira et al., 2001

\*\*New nomenclature, see Shwab et al., 2013

## Chemicals, reagents, plasmids, kits and media

Reagent or resource	Source	Identifier
2-Mercaptoethanol	Gibco® by Thermo Fisher Scientific	21985023
4',6-Diamidino-2'-phenylindole (DAPI) - 1µg/ml	Life Technologies	D3571, RRID:AB_2307445
Accutase®	BioLegend	423201
BD Biosciences flow cytometry permeabilization buffer (BD Perm/Wash Buffer)	BD Biosciences	554723
Collagenase/Dispase, 100mg	Roche Diagnostics	50-100-3281
cOmplete Protease Inhibitor tablets	Roche	11697498001
Dimethyl sulfoxide	Sigma-Aldrich® (Merck)	41639
DNA-free™ Kit	Ambion®	AM1906
DNeasy Blood & Tissue Kit	Qiagen	50969504
DreamTaq™ Green DNA Polymerase	Fermentas GmbH	EP0712
Dulbecco's Phosphate Buffered Saline w/o Calcium w/o Magnesium 1X	Biowest	L0615-500
Dulbecco's modified Eagle's medium (DMEM), high glucose w/o L-glutamine w/o sodium pyruvate	Biowest	L0101-500
Enzyme BSED1 (BSA1)	Fermentas GmbH	10689330
Enzyme HinfI	Fermentas GmbH	10334580
FASTDIGEST ACII (SSII)	Fermentas GmbH	10819850
Fetal bovine serum, US origin , Virus and mycoplasma tested	PAN-Biotech	P30-1402
Fluorescein labeled Dolichos Biflorus Agglutinin (DBA)	Vector Technologies	FL-1031-2
Herculase II Fusion DNA Polymerase	Agilent Technologies	HPA600675
Iscove's Modified Dulbecco's Medium (IMDM) w/o L-Glutamine w/o Hepes	Biowest	L0192-500
L-glutamine 200mM	GRiSP Research Solutions	GTC03.0100
LIVE/DEAD® Fixable Violet Dead Cell Stain Kit, for 405 nm excitation	Molecular Probes® by Thermo Fisher Scientific	L34955
Midori Green Advance DNA Stain	Nippon Genetics	MG04
Mouse IFNγ (Stock solution: 200 µg/ml or 1000 U/µl)	Sigma-Aldrich® (Merck)	SRP3058
Nuclepore Membrane Circles, 25mm 3,0µm	Whatman	110612
Penicillin-Streptomycin Solution 100X	Biowest	L0022-100
Perm/Wash Buffer	BD Pharmingen™	554723
Plasmid DNA: pSV3-neo	ATTC	37150
ProLong® Gold Antifade Mountant	Life Technologies	P36934
RNeasy Mini Kit	Qiagen	74104
ScreenFect®A Transfection Reagent kit	InCella	S-3001
Sodium Heparin 5000 U.I./ml	B. Braun Medical Lda.	126/318463/0610

Sodium pyruvate 100mM	Sigma-Aldrich Química, S.A	S8636-100ML
Swin-Lok Holder 25mm	Whatman	420200
TaKaRa LA PCR Kit Ver. 2	Takara	RR013A
Toxocell Latex Kit	Biokit	T3000-4525
Trypsin-EDTA 10X	Biowest	X0930-100
α-Cellulose (CF-11)	Sigma-Aldrich®	C6429-500G

## Primers

Name	5' - 3' sequence	Description	Reference
101 F	GGTGAAACCCCGTCTCTACT	Alu element from primate 7SL RNA gene	Funakoshi et al., 2017
206 R	GGTTCAAGCGATTCTCCTGC	Alu element from primate 7SL RNA gene	Funakoshi et al., 2017
264740_AS_Fwd	GCTCCGATGTGTCTCGGATGC	TG_264740 (hypothetical protein) gene from <i>T. gondii</i>	This study
264740_AS_Rev	GACATCCTCCGCATGATATTCG	TG_264740 (hypothetical protein) gene from <i>T. gondii</i>	This study
293820_N_Rev	CCTTTCGCGACACGTACGCG	TG_293820 (calpain family cysteine protease domain-containing protein) gene from <i>T. gondii</i>	This study
293820_N_Fwd	GAGTACTGGGAGACGACCAGG	TG_293820 (calpain family cysteine protease domain-containing protein) gene from <i>T. gondii</i>	This study
294270_BV_Fwd	GTGCTGGGGAGGATGAGCATG	Histone arginine methyltransferase PRMT4/CARM1	This study
294270_BV_Rev	TTCCTCTGTTTTGCATTCTTC	TG_294270 (Histone arginine methyltransferase PRMT4/CARM1) gene from <i>T. gondii</i>	This study
321410_AS_N_Fwd	GGAAAAATAGAAAAACCAAATG	TG_293820 (hypothetical protein) gene from <i>T. gondii</i>	This study
321410_AS_N_Rev	AGCGCGGAAAAACATGCATCC	TG_293820 (hypothetical protein) gene from <i>T. gondii</i>	This study
alt. SAG2 multiplex F	GGAACGCGAACAATGAGTTT	alt. SAG2 gene from <i>T. gondii</i>	Su et al., 2006
alt. SAG2 multiplex R	GCACTGTTGTCCAGGGTTTT	alt. SAG2 gene from <i>T. gondii</i>	Su et al., 2006
alt. SAG2 nested F	ACCCATCTGCGAAGAAAACG	alt. SAG2 gene from <i>T. gondii</i>	Su et al., 2006
alt. SAG2 nested R	ATTTGACACGCGGGAGCAC	alt. SAG2 gene from <i>T. gondii</i>	Su et al., 2006
B1FCA	TGCTAGACAAATAATGAGTGTCTCC	Irgb1 gene coding region - Forward	This study
B1RCA	AGGTTAGTACCTCACAGTCCA	Irgb1 gene coding region - Reverse	This study
GRA6 nested F	TTTCCGAGCAGGTGACCT	GRA6 gene from <i>T. gondii</i>	Su et al., 2010
GRA6 nested R	TCGCCGAAGAGTTGACATAG	GRA6 gene from <i>T. gondii</i>	Su et al., 2010
GRA6 multiplex F	ATTTGTGTTTCCGAGCAGGT	GRA6 gene from <i>T. gondii</i>	Su et al., 2006
GRA6 multiplex R	GCACCTTCGCTTGTGGTT	GRA6 gene from <i>T. gondii</i>	Su et al., 2006
H16227R	GGAGTTGCAGTTGATGTGTGATAGT	Human mitochondrial DNA	Malmstrom et al., 2005
Irgb2_66F_b	CTGGAAACACTTTGCCACG	Irgb2 gene coding region - Reverse	Lilue et al., 2013
Irgb2_66F_f	CTGGACTCTGCGCTTTTATTGG	Irgb2 gene coding region - Forward	Lilue et al., 2013
Irgd_17B_b	AAGCACTGTTCTTGTTCAGG	Irgd gene coding region - Reverse	Lilue, 2012

Irgd_17B_f	ATGGATCAGTTCATCTCAGC	Irgd gene coding region - Forward	Lilue, 2012
Irgm1_655_b	CCTCTCAGAGAATCTAAAACCC	Irgm1 gene coding region - Reverse	Lilue, 2012
Irgm1_655_f	CTGCCGATTTCGATTCATAAAC	Irgm1 gene coding region - Forward	Lilue, 2012
L16124F	CTGCCAGCCACCATGAATATT	Human mitochondrial DNA	Malmstrom et al., 2005
L358 multiplex F	AGGAGGCGTAGCGCAAGT	L358 gene from <i>T. gondii</i>	Su et al., 2006
L358 multiplex R	GCAATTTCTCGAAGACAGG	L358 gene from <i>T. gondii</i>	Su et al., 2006
L358 nested F	TCTCTCGACTTCGCCTCTTC	L358 gene from <i>T. gondii</i>	Su et al., 2006
L358 nested R	CCCTCTGGCTGCAGTGCT	L358 gene from <i>T. gondii</i>	Su et al., 2006
M1F	CACCACCAGCACCCAAAGCT	Mouse mitochondrial D-loop	Goios et al., 2007
M2R	GGCCAGGACCAAACCTTGTG	Mouse mitochondrial D-loop	Goios et al., 2007
purple20b2b1F	ATGGGTCAGACTTCCTCTTCTAC	PWK Irgb2b1 tandem coding sequence - Forward	This study
purple20b2b1R	TCACAGCTCCATTCTGTTTCCC	PWK Irgb2b1 tandem coding sequence - Reverse	This study
SAG3 multiplex F	CAACTCTCACCATTCCACCC	SAG3 gene from <i>T. gondii</i>	Grigg et al., 2001
SAG3 multiplex R	GCGCGTTGTTAGACAAGACA	SAG3 gene from <i>T. gondii</i>	Grigg et al., 2001
SAG3 nested F	TCTTGTCGGGTGTTCACTCA	SAG3 gene from <i>T. gondii</i>	Su et al., 2010
SAG3 nested R	CACAAGGAGACCGAGAAGGA	SAG3 gene from <i>T. gondii</i>	Su et al., 2010
Tailed-PrimerF-Irgb2b1	TTTCTGTTGGTGCTGATATGCCTG GACTCTGCGCTTTTATTGG	Samples barcoding Nanopore sequencing - Irgb2b1PWK allele - Forward	This study
Tailed-PrimerR-Irgb2b1	ACTTGCTGTGCTCTATCTTCTCAC AGCTCCATTCCTGTTTCCC	Samples barcoding Nanopore sequencing - Irgb2b1PWK allele - Reverse	This study
Tox-5	CGCTGCAGACACAGTGCATCTGGAT	3' end of the 529 bp fragment from <i>T. gondii</i>	Homan et al., 2000
Tox-8	CCCAGCTGCGTCTGTCTGGGAT	5' end of the 529 bp fragment from <i>T. gondii</i>	Reischl et al., 2003

## Cell lines

Reagent or resource	Source	Identifier
Hs27 cell line (Human)	ATTC	CRL-1634, RRID:CVCL_0335
T17 (Irgb2-b1KO)	Kindly provided by Dr. Tobias Steinfeldt, University of Freiburg, Germany.	N/A
T17+Irgb2-b1C57BL/6	Kindly provided by Dr. Tobias Steinfeldt, University of Freiburg, Germany.	N/A
T17+Irgb2-b1CIM	Kindly provided by Dr. Tobias Steinfeldt, University of Freiburg, Germany.	N/A
T17+Irgb2-b1PWK	Kindly provided by Dr. Tobias Steinfeldt, University of Freiburg, Germany.	N/A

### Mice used for *in vivo* experiments

Name	Sample location	<i>M. musculus</i> subsp.	Source	Inbreeding*	Identifier
C57BL/6J	Laboratory strain	<i>M. m. domesticus</i>	Instituto Gulbenkian de Ciencia	Inbred	RRID:IMSR_JAX:000664
CIM	Masinagudi, India	<i>M. m. castaneus</i>	University of Cologne - F. Bonhomme, Institut de Science de l'Évolution, Montpellier, France	Inbred	-
PWD/PhJ	Prague, Czech Republic	<i>M. m. musculus</i>	University of Cologne - J. Forejt, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic	Inbred	RRID:IMSR_JAX:004660
PERA/EiJ	Rimac Valley, Peru	<i>M. m. domesticus</i>	The Jackson Laboratory, Bar Harbor, Maine, USA	Inbred	RRID:IMSR_JAX:000930
PERC/EiJ	Rimac Valley, Peru	<i>M. m. domesticus</i>	The Jackson Laboratory, Bar Harbor, Maine, USA	Inbred	RRID:IMSR_JAX:001307
MANA/Nach	Manaus, Brazil	<i>M. m. domesticus</i>	Dr. Michael Nachman, University of California Berkeley, California, USA	F13	-
MANB/Nach	Manaus, Brazil	<i>M. m. domesticus</i>	Dr. Michael Nachman, University of California Berkeley, California, USA	F14	-

\*inbred for more than 20 generations=inbred, inbred for less than 20 generations=generations specified by number