



HAL
open science

Developmental barriers in *T. gondii*: MORC at the onset of epigenetic rewiring of the parasite's cell fate

Dayana Farhat

► **To cite this version:**

Dayana Farhat. Developmental barriers in *T. gondii*: MORC at the onset of epigenetic rewiring of the parasite's cell fate. Cellular Biology. Université Grenoble Alpes [2020-..], 2020. English. NNT : 2020GRALV014 . tel-03148920

HAL Id: tel-03148920

<https://theses.hal.science/tel-03148920>

Submitted on 22 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : Biologie cellulaire

Arrêté ministériel : 25 mai 2016

Présentée par

Dayana C. FARHAT

Thèse dirigée par **Mohamed-Ali HAKIMI**

préparée au sein du **Laboratoire CRI IAB - Centre de Recherche Epigenetics, Chronic Diseases, Cancer - Institute for Advanced Biosciences**
dans l'École Doctorale Chimie et Sciences du Vivant

MORC, un régulateur épigénétique au carrefour des trajectoires développementales du parasite *T. gondii*

Developmental Barriers in *T. gondii* : MORC At the Onset of Epigenetic Rewiring of the Parasite's Cell Fate

Thèse soutenue publiquement le **22 octobre 2020**,
Devant le jury composé de :

Monsieur MOHAMED-ALI HAKIMI

DIRECTEUR DE RECHERCHE, INSERM DELEGATION ALPES,
Directeur de thèse

Monsieur ARTUR SCHERF

DIRECTEUR DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE
MEUDON, Rapporteur

Monsieur THIERRY LAGRANGE

DIRECTEUR DE RECHERCHE, CNRS DELEGATION OCCITANIE EST,
Rapporteur

Monsieur GUILLAUME MOISSIARD

CHARGE DE RECHERCHE, CNRS DELEGATION OCCITANIE EST,
Examineur

Madame Shelley L. Berger

PROFESSEUR, UNIVERSITE DE PENNSYLVANIE - ETAS-UNIS,
Examinatrice

Madame ISABELLE TARDIEUX

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES,
Présidente



Preface

I pressed the button of a blank page, it scared me at first, am I going to get the blank page syndrome? But I had lost enough time, time that was spent checking the news, checking how many people died today, what is the total in this and that country and where the next hit country will be? But as much as this period was painful and anxiogenic, as much as it made me and all the world realize once again how important what we are doing is, most importantly how critical it is to publish a type of science that is the closest to the truth as we can get. We work in flasks and we spend years of our PhDs staring at a gel and at a screen, but what we are accomplishing is real, and what is happening to the planet today illustrates this reality, along with a big demonstration for the need for scientists, ones that are at first ethical, honest, and not after ego and names. I started walking this path believing that the truth lies in science, and I will keep walking this path with this torch in hand. Today, I am proud to have accomplished an honest work, one where we didn't allow ourselves to claim anything without triple checking, without giving our best to put on the table a full story, as full as I was capable of reaching in 3 years. I publish this work, and I write my manuscript holding as a first criteria a great respect for my fellow scientist readers, for their critical minds, and for the fuller truth we are all seeking.

30.March.2020

Few words

Having finished writing my PhD manuscript, I would like to thank Ali for his help revising this work. His guidance, support and humanity are aspects that were present both during this writing period and the 4 years of my PHD, which I will always appreciate. I had a mentor but also a friend. I highly enjoyed our late evenings discussions, the scientific ones as much as the philosophical and social ones. I learned a lot, and enjoyed the process. You were demanding but also patient, extreme but also flexible, a supervisor but first a human! Thank you for granting me the opportunity to be under your guidance, and helping me set my foot into the scientific career.

Alexandre, your perfectionism was contagious, your criticism created the needed balance for the optimism of Ali, and your attention for details as well as your lab meeting comments added significantly to my project and to my scientific interpretations.

Laurence, Dominique and Charlotte, I appreciate your much needed friendly and peaceful presence in the lab. Thank you for the technical support, as well as the encouragements.

Isabelle and Marie-Pierre, two strong successful women that granted me moments of their time for scientific discussions that went beyond my field, and enriched my knowledge. Also, I highly appreciate your implicit ways of checking on my well-being during these years.

To the big family of the Hakimi and the Tardieux lab, I say thank you for making this journey greater than a mere scientific one, the days would have been duller if it wasn't for the laughs we shared.

To Sheena and Georgios, our parallel paths helped lighten up our difficulties. I gained two genuine friends who supported me both in the lab and out, I am glad we had the chance to develop such a caring friendship.

I want to also thank my parents for their unconditional love and support, despite them not having enough knowledge about the burden of the job, they were always there for me. Thank you for raising me a curious child allowed to ask 'why' all day. My sister, I'm sorry that my job being abroad costed you a daily companion, still thank you for being such a loving and supportive sister throughout this time. My brother, your humor made the load of this journey lighter, and your words reminded me how worth it was my goal.

My little sisters, I wish I was more present to witness your development, and I hope my journey teaches you that women can make it on their own, and that education is the only key.

Last but not least, my Lebanese friends, who always supported and pushed me further, namely Mira being the example of a trustworthy and loving friend. And to Amani, being on the receiving end of your love and care granted me enough power to get throughout this difficult period.

At the end, I would like to thank myself, for providing enough internal peace and well-being to be able to work as hard as it was needed to reach a well-deserved accomplishment.

Much Appreciation and Love,

Dayana

02.September.2020

Contents

Preface	3
Few words	5
Contents	9
Table of figures	13
Abstract	15
I. Introduction	17
1. Evolutionary History of Apicomplexa	19
2. Evolution towards parasitism	20
3. Host range and transmission modes	22
4. Coccidians and Sarcocystidians life cycle	23
5. Evolution of Sex, recombination and meiosis	26
6. The evolutionary advantage of haploidy	29
7. An unusual mitosis typifies apicomplexans	30
8. The genome-free organelles characteristics of apicomplexans	32
9. A peculiar nuclear compartment	33
10. The mitochondria genome	36
11. The Apicoplast genome	37
12. Apetala Transcription Factors	40
a. Their Origins	40
b. Expansions and Discovery in <i>Apicomplexa</i>	41
c. AP2 Transcription factors in Land Plants	43
d. Structural perspective	43
e. Outside the DBD	45
f. AP2-containing proteins in <i>T. gondii</i>	45
g. AP2-containing proteins in <i>Cryptosporidium sp.</i>	46
h. AP2-containing proteins in <i>Plasmodium sp.</i>	47
i. ApiAP2s involved virulence and clonally variant families	47
ii. ApiAP2 involved in Developmental regulation	48
iii. ApiAP2s involved in Sexual Development regulation	49
13. Epigenetic weight in <i>Plasmodium</i> species	50
a. On Clonally variant multigene families	50

b.	On Sexual development-----	52
14.	<i>Toxoplasma gondii</i> Development and gene expression -----	53
a.	The Merozoite Stage-----	54
b.	The Enteroepithelial stages and Gametogenesis -----	55
c.	The Oocyst and Sporozoites -----	55
d.	The Bradyzoite stage in intermediate hosts -----	56
15.	Signaling through Chromatin and Epigenetics in <i>T. gondii</i> -----	57
a.	Methyltransferases-----	59
b.	Bromodomains and Acetyltransferases (focus on GCN5)-----	60
c.	The evolutionary history of HDACs in the phylum-----	62
16.	<i>T. gondii</i> HDAC3 -----	63
a.	A nuclear-resident HDAC with sensitivity to cyclic tetrapeptides inhibitors -----	63
b.	<i>T. gondii</i> HDAC3 versus GCN5b, balancing stage conversions-----	66
c.	TgHDAC3 a class I HDAC co-repressor -----	68
17.	MORC -----	70
a.	Evolution and divergence-----	70
b.	MORC proteins functions-----	72
c.	The DNA methylation discrepancy (Plant focus)-----	73
d.	Mechanistic aspects-----	74
i.	Nuclear bodies and multimers formation-----	74
ii.	ATP-dependent dimerization of ATPase modules -----	76
iii.	The CW mediated target recognition and ATPase regulation -----	78
iv.	DNA binding and genome compaction -----	80
18.	Knowledge gap and preliminary questions -----	84
II.	Results-----	87
1.	The paper (Farhat et al., 2020) -----	89
2.	Extended Data -----	105
3.	Supplementary Figures -----	123
III.	Discussion -----	133
1.	Poised State and Chromatin Targeting -----	135
a.	Poised State-----	135
b.	CW-mediated chromatin targeting of MORC-----	137
c.	PHD-mediated chromatin targeting of MORC-containing complexes-----	139
2.	DNA-based targeting and Primary AP2s-----	140
3.	Secondary AP2s mediated regulation of targeting and action -----	141

4.	Alternative means for developmental regulation of expression-----	146
a.	Translational repression and the pocket concept -----	146
b.	mRNA stability and maturity-----	147
c.	RNA Polymerase II pausing-----	148
5.	HDAC3 involvement in post-transcriptional and post-translational regulation	148
6.	HDAC3 and MORC: a non-exclusive relationship-----	149
7.	MORC in Plasmodium falciparum -----	150
8.	MORC and HDAC3 at telomeric repeats? -----	151
9.	Host impact -----	154
a.	Host-parasite interactions-----	154
b.	Rhoptry- and Dense granules-resident Effectors -----	155
c.	Developmental influence -----	158
10.	Cell cycle impact-----	159
11.	Metabolism impact -----	161
12.	Contractility and KELCH-----	163
13.	Last words... -----	164
IV.	Bibliography-----	167

Table of figures

Figure 1. Evolutionary History of Apicomplexa.	20
Figure 2. The complex life cycle of <i>T.gondii</i> .	24
Figure 3. <i>T. gondii</i> cell illustration and cell cycle (Endodyogeny).	32
Figure 4. Plastid evolution and fate	39
Figure 5. A model of DNA-induced, domain swapped dimerization and DNA looping.	44
Figure 6. Sequence alignment of HDAC3 homologues in Apicomplexan parasites and other organisms.	65
Figure 7. The T99 mutation impacts greatly the activity of TgHDAC3.	66
Figure 8. The identification of TgMORC co-purified with TgHDAC3.	70
Figure 9. Phylogenetic classification of MORC genes in plant and animal lineages.	72
Figure 10. Domain organization of MORC family members from Homo sapiens (Hs).	76
Figure 11. Structure of MORC3 ATPase-CW cassette in complex with AMPPNP and H3K4me3 peptide.	77
Figure 12. Schematic of MORC architecture in two plants models.	78
Figure 13. The human zinc finger CW domain-containing proteins.	80
Figure 14. MORC-1 acts via a mechanism of DNA loop entrapment to compact chromatin.	81
Figure 15. DNA Loop extrusion vs. Loop entrapment.	82
Figure 16. MORC KD results in a heterogenous protein expression.	137
Figure 17. MORC guides developmental trajectories recruiting downstream regulating pathways.	142
Figure 18. A proposed model.	143
Figure 19. MORC and HDAC3 are enriched on telomeric regions.	152
Figure 20. Nucleosome organization in telomeric chromatin.	153

Abstract

T. gondii has a complex life cycle typified by an asexual development taking place in vertebrate, and a sexual reproduction which occurs exclusively in felids and thereby is less studied. The developmental transitions rely on changes in gene expression patterns, and recent studies have assigned roles for chromatin shapers, including histone modifications, in establishing specific epigenetic programs for each given stage. Here, we identified *T. gondii* microorchidia (MORC) protein as an upstream transcriptional repressor of sexual commitment. MORC, in partnership with Apetala (AP2) transcription factors, was shown to recruit the histone deacetylase HDAC3, thereby impeding the chromatin accessibility of the genes predestined to be exclusively expressed in sexual stages. We found that MORC-depleted cells underwent marked transcriptional changes, resulting in the expression of a specific repertoire of genes, and revealing a shift from asexual proliferation to sexual differentiation. MORC acts as a master regulator that directs the hierarchical expression of secondary AP2 factors, with these latter potentially contributing to the unidirectionality of the life cycle. Thus, MORC plays a cardinal role in the *T. gondii* life cycle, and its conditional depletion offers a way to study the parasite's sexual development in vitro, and is proposed as an alternative to the requirement of cat infections.

I. Introduction

1. Evolutionary History of Apicomplexa

The story begins with an ancestral photosynthetic cyanobacterium that got engulfed by a previously non-photosynthetic eukaryotic protist, this endosymbiotic event gave rise to the Archaeplastida. Glaucophyta, Rhodophyta and Chlorophyta all originated from this primary endosymbiosis (Archibald, 2009).

A red alga engulfed by a eukaryotic heterotroph protist would be at the origin of the appearance of the SAR supergroup which includes Stramenopiles, Alveolates and Rhizaria lineages (Cavalier-Smith, 2009) (Figure 1). To note that the origin of plastids of the photosynthetic lineages is also thought to originate from the red alga (Green, 2011), however the number of secondary endosymbiosis events as well as plastid losses and horizontal gene transfers added levels of complication regarding the exact chronological order of events and of the corresponding ancestors. The Stramenopiles are chromists which are reported to being behind algal diversity and aquatic habitats (Bhattacharya et al., 2004) , and together with the Alveolates, they make up the monophyletic Chromalveolata group.

Alveolates include apicomplexans, dinoflagellates and ciliates. The Apicomplexa phylum is one of few wholly parasitic lineages of eukaryotes and, consisting of more than 6000 species, most of its species are parasitic on insects and mollusks with few causing diseases in metazoans (Levin, 1989). From the SAR supergroup, the Apicomplexa which includes the malarial parasite *Plasmodium sp*, *Theileria sp*, *Toxoplasma gondii* and *Cryptosporidium sp*, along with the Kinetoplastida (including *Trypanosoma sp* and *Leishmania sp*) from the Excavata supergroup, together shape the major proportion of parasitic protozoa species.

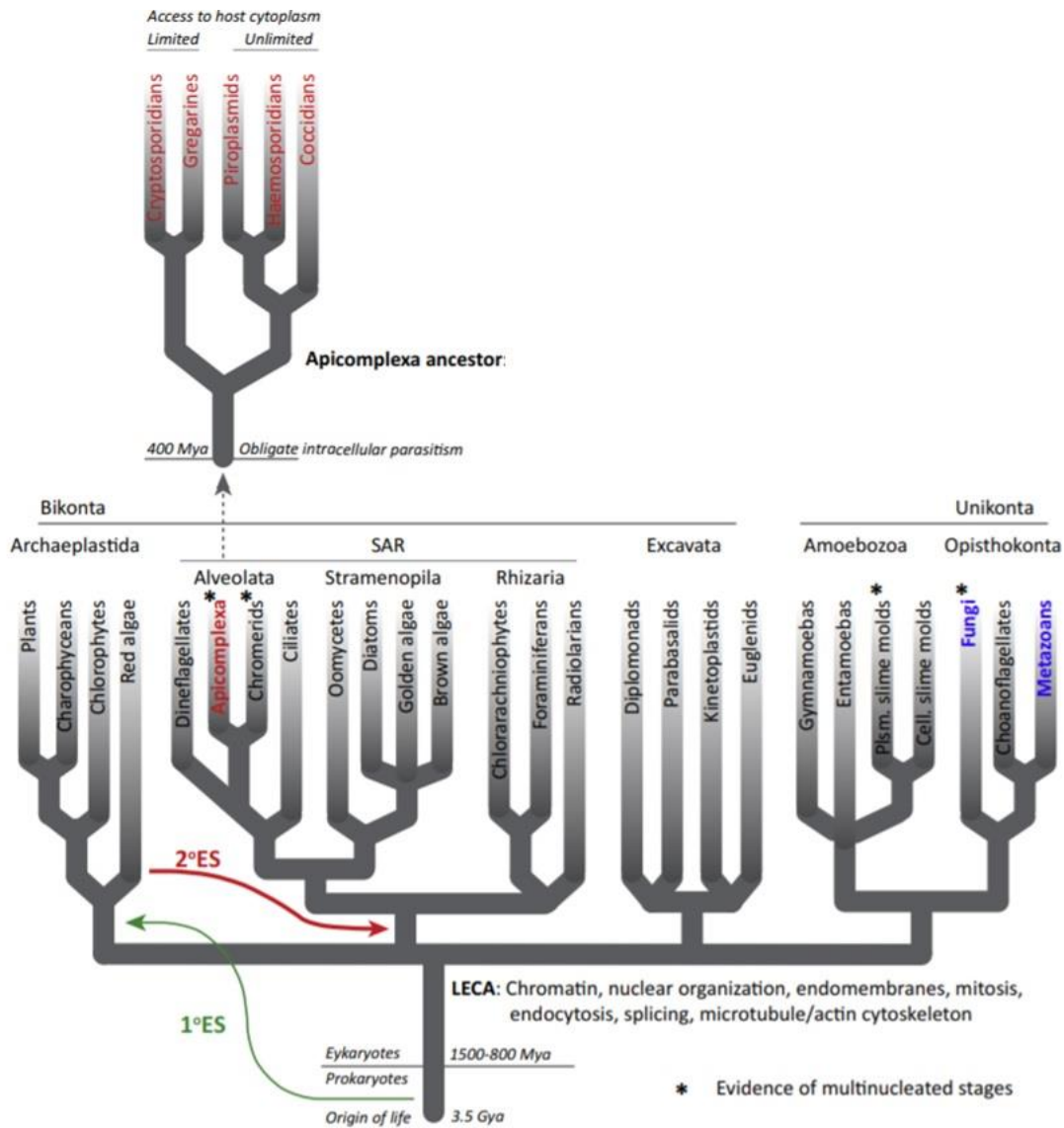


Figure 1. Evolutionary History of Apicomplexa.

Phylogenetic relationships of major eukaryotic groups are shown. Two bikont groups acquired plastids, Archaeplastida by primary endosymbiosis (1ES, green arrow) of ancestral cyanobacterium and the SAR group by secondary endosymbiosis (2ES, red arrow) of red algae. Adapted from (White and Suvorova, 2018).

2. Evolution towards parasitism

The emergence of parasitism in the phylum and thus the loss of the free-living style of their pre-parasitic ancestors, set a turning point in the evolution of these species. In the argument on the origin of parasitism, one of the developed ideas stated that this event was correlated more with protein loss than it was linked to the acquisition of novel structures. This theory was backed up by the estimation of more than 4000 genes in the pre-apicomplexan ancestor getting lost in the extant species (Woo et al., 2015).

Most of these genes are thought to have been needed for the free-living style, with some of them being involved in flagella motility, others in photosynthesis (Woo et al., 2015). The metabolic losses here would be of a significant weight in the transition to parasitism, especially as it is evident in the case of apicomplexans that a wider range of metabolic capabilities is exhibited by their closest free-living relatives, the Chrompodellids (Janouskovec and Keeling, 2016).

However, when comparing the genomes of Trypanosomatids and their free-living ancestor *Bodo saltans*, it was surprising to realize that the metabolic related genes reduction had preceded their transition to a parasitic life style (Janouskovec and Keeling, 2016). Such observations led the development of less simplified and clean-cut views for the progression towards parasitism. In fact, for a long time it was thought that the characteristics that are 'unique' to parasites would have been the drivers of the emergence of parasitism, however the relatively recent whole genome sequencing and advanced phylogenetic analysis made it possible to determine, on the genetical level whether such characteristics were adapted prior to the origin of parasitism or not. The genome of *Bodo saltans*, the free-living ancestor of Kinetoplastida and among the closest relatives of the trypanosomatid parasites, shares most of these characteristics once interpreted as adaptations to parasitism, showing that they might have evolved for different reasons.

Regarding Apicomplexans, their aforementioned closest cousins are heterotrophic, photosynthetic Chromerids (including *Chromera sp* and *Vitrella sp*) and predatory colpodellids (*Colpodellida*, *Alveolata*). Chrompodellids, a large sister group of apicomplexans, were seen to share a great number of characteristics, such as glideosome-associated proteins, oocyst wall proteins, microneme-like organelles which all seem to predate the emergence of parasitism (Janouškovec et al., 2015).

Structures involved in host invasion were also spotted in the free-living ancestors, such as pseudo-conoid and rhoptry-like organelles. However, it is believed that these organelles were needed for feeding the ancestors, thus they might have indirectly facilitated the emergence of parasitism as preconditions but not as previously thought parasite-specific adaptations.

Although it is common to talk about gene loss accompanying this important switch, one must not ignore the considerable expansion occurring in these same genomes, for example of their acquired diverse specific TFs. This compensation can also be seen in the fact that the overall regulatory input per protein-coding gene is comparable to other eukaryotes (Iyer et al., 2008a).

Also, once the parasitism is established, its persistence is likely linked to its adaptation to the host and its immunity. Having a host fighting back would increase the evolutionary rate of these species, with the greatest genome dynamism seen at the level of genes involved in the interaction with the host. Great amount of polymorphisms and lineage-specific innovations

would occur, and this is seen in the abundance of genes coding for effector proteins that are directed to either cell invasion or host immune modulation (Jackson, 2015). Therefore, the host responses would consist a great weight in driving the parasite genome evolution.

For instance, an arms race exists between the parasite *Toxoplasma gondii*, and one of its most significant hosts, the rodents. A great level of polymorphism was detected between the interacting surfaces of proteins from both the parasite and the host (Lilue et al., 2013), underlining a reciprocal selection pressures between these proteins and thus these species.

3. Host range and transmission modes

After acknowledging the weight of the host on the evolution of parasitism, mostly on its persistence and its subsequent adaptations to this new life style, it is important to go through one major limiting factor for the survival of the parasite, its transmission between hosts. Therefore, it is of interest to point out how the parasite would acquire its hosts and develop its life cycle.

Parasite can adopt a horizontal or a vertical mode of transmission, where the first is linked to a greater production of infectious particles and the second would be via the offspring of the parasite. However, a trade-off in transmission modes is needed in order to protect the host, first to prevent the loss in numbers of infectious particles, and second for the host to survive long enough until reproduction and an effective vertical transmission are possible (Alizon et al., 2009).

It is believed that the transmission modes of a parasite would evolve in a way to produce the greatest opportunities for its fitness at the least cost, with sexual transmission being perceived as the derived mode rather than ancestral one. In fact, one of the theories for understanding how the parasite acquired multiple hosts and complex life cycles, states that the original host would be the evolutionary more ancestral one, while another position sets the original host as the one where sexual reproduction would occur, which is the definitive host, and that the hosts where no sex can occur, would be acquired subsequently (Antonovics et al., 2017).

In reality, the incorporation of a new host might convert the definitive host into an intermediate one, with this event called upward incorporation, due to the parasites evolving to exploit the species predators. This mode seems to be selectively beneficial for the parasite evolution due to the decrease in its inbreeding, because of the higher potential for multiple infections to occur in a larger host, thus resulting in a greater genetic diversity (Rauch et al., 2005).

Another possibility would be that ecological and evolutionary perturbations would create a trophic vacuum, one which would constrain the parasite from directly infecting its definitive hosts, hence the need for adapting intermediate hosts to fill this gap and facilitate the transmission. In this case, a downward incorporation below the definitive host would occur,

involving the prey of this original host, mostly due to their geographical proximity and to their ingestion of parasitic transmission particles (Parker et al., 2015). That being said, the final complex cycle would be similar irrespective of the mode of the parasite acquiring its multiple hosts.

4. Coccidians and Sarcocystidians life cycle

This predator and prey relationship is utilized in the transmission between hosts of parasites displaying heteroxenous lifecycles, such as the family of Sarcocystidae, meaning flesh cyst forming parasites (Frenkel and Smith, 2003). This family includes the *Sarcocystis* and *Hammondia* species, along with *Neospora caninum* and *Toxoplasma gondii*, and they all form tissue cysts that confers them with a transmission mode via carnivorism.

Sarcocystidae are part of Coccidia which are classified within the phylum of Apicomplexa, along with Cryptosporidia and Haemosporidia that includes the *Plasmodium* species, the agents of malaria (Figure 1). The divergence of these coccidian species from their common apicomplexan ancestor is predicted to have occurred 400 Million years ago (Berney and Pawlowski, 2006).

Despite that the enteric non cyst-forming coccidian parasites do not display the two-host (heteroxenous) life cycle (Webster, 2010), yet they share the characteristic of possessing environmentally resilient oocyst which allow for a fecal-oral transmission mode (Frenkel and Smith, 2003).

However, it seems that between the Sarcocystidae, *T. gondii* stands out in its ability to bypass the sexual phase of the life cycle, and to be transmitted between intermediate hosts through oral ingestion of infected tissues (Su et al., 2003), with this feature probably being behind the broad host range this parasite acquired during its evolution.

In fact, feline hosts being the only definitive hosts for this parasite, were proven not be essential to maintain transmission of *T. gondii*, underlining the weight of other modes that this parasite adapted. The congenital mode was also determined to be sufficient to maintain a transmission of *T. gondii* in rats, when these latter were trapped in an environment devoid of oocysts (Webster, 1994).

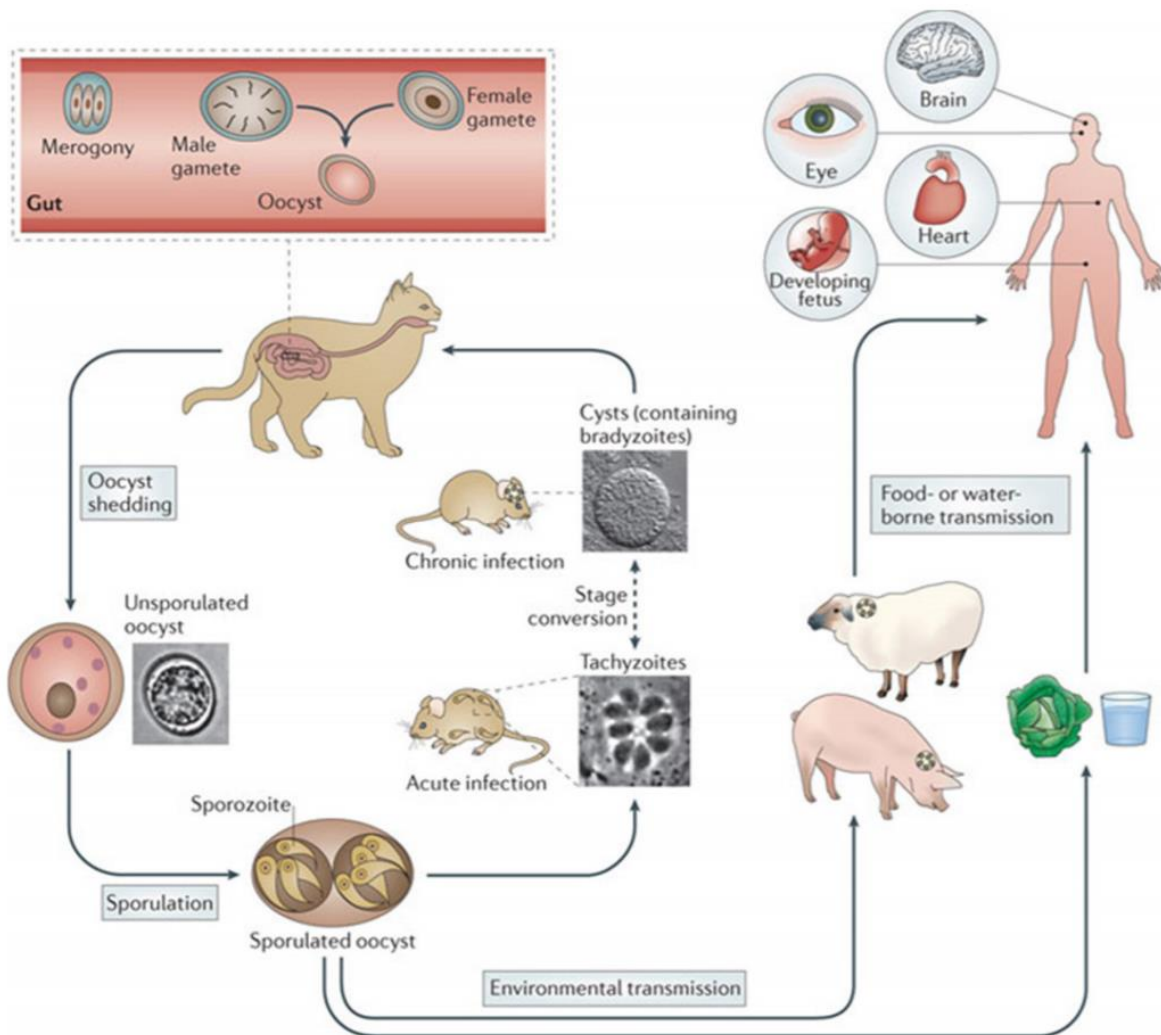


Figure 2. *The complex life cycle of T.gondii.*

Cats are the definitive host where sexual replication takes place. Following replication within enterocytes of the gut (a process known as merogony), male and female gametes are formed within the host cell, as described previously. Fusion of gametes leads to the formation of diploid oocysts that are shed in cat's feces and undergo meiosis in the environment to yield eight haploid progeny. Oocysts are capable of surviving in the environment for long periods of time and can contaminate food and water, providing a route of infection for intermediate hosts. In the intermediate host (shown here as rodents) asexual replication occurs. Acute infection is characterized by fast replicating tachyzoites that disseminate throughout the body. In response to the host immune response, slow-growing bradyzoites form cysts in deep tissues (e.g. brain) leading to long-term chronic infection. Ingestion of tissue cysts via omnivorous or carnivorous feeding can lead to transmission to other intermediate hosts or to cats, which re-initiates the sexual phase of the life cycle. *Adapted from (Hunter and Sibley, 2012)*

Despite evolving several ways for its persistence, yet the efficiency of its transmission within each different host changes depending on the route. For instance, it is best transmitted in the cat through carnivorism, namely ingesting tissue cysts, as it was reported that cats would shed millions of oocysts after ingesting even one bradyzoite from tissue cysts, noting that ingesting any of the other infectious stages would result in shedding oocysts, yet not in the same frequencies nor quantities as well as with a longer pre-patent period. Additionally,

intermediate hosts would get infected by ingesting even one oocyst, yet more than 100 oocysts would not infect cats (Dubey, 2005, 2009; Dubey and Frenkel, 1972).

These observations are linked to the fact that, not only this generalist pathogen had optimized its transmission modes based on its different hosts, but also each of its parasitic stages would specialize on the appropriate host.

First, let's go through the development of the *T. gondii* within its intermediate hosts. The list includes most warm-blooded organisms. The greatest evolutionary weight would fall on the intermediate hosts that hold great potential for the transmission of the parasite, hence, a potential of being a prey for Felidae. Therefore, despite the lack of intermediate host specificity that this parasite displays, yet unless they play regular parts in the felines food chain, they would be considered accidental intermediate hosts, as the case for humans.

In fact, the innate immune pathways that are employed during the control of the parasite infection, is quite distinct between the human and the mice, as these latter have been for a long time under selective pressure from the parasite, unlike the humans which became relevant hosts relatively recently, after the cat domestication.

Within the non-definitive hosts, *T. gondii*, like their fellow Sarcocystidae parasites, exists in both a highly invasive tissue-barriers crossing disseminating stage with the adopted form of tachyzoite and a residence stage relying on persistent tissue cysts containing bradyzoite forms.

The tachyzoites are the actors behind the acute phase of infection, due to their rapidly dividing nature, which assigns them with a role in the horizontal transmission, through a greater production of infectious particles. These organisms are highly adapted to successfully invade the host cells, scavenge most of their needs in nutrients, amino and fatty acids from that environment (Dou et al., 2014), and to modulate the immune system of the host into the optimal conditions for their persistence. (Hakimi et al., 2017)

The ability to form tissue cysts following the ingestion by the intermediate hosts, of oocysts shed by the definitive host, is as mentioned earlier, the characteristic that made Sarcocystidia a group on its own, allowing its parasitic species to be transmitted through carnivorousism.

The tissue-cyst forming stage is at the crossroad between the asexual and the sexual phases of the sarcocystidyan *T. gondii* parasite heteroxenous life cycle (Figure 2). The aforementioned trade-off existing between vertical and horizontal transmission modes is decided at this phase. Bradyzoites have the ability to persist, asymptotically, in the tissues of the intermediate host, with potentially limited access to the host materials for sustenance, and they adapt low energy requiring conditions aligned with a very slow rate of division.

To note, that in the absence of a potent immune response, these bradyzoites transition back to tachyzoites, which in most cases cause the destruction of the host cells (Weiss and Dubey, 2009) and to the loss of the balancing mechanisms that the parasite acquired during its

evolution in order to protect its host long enough until a chance for a transmission through sexual modes arises.

The parasite must persist within its modulated host, until the uptake of the bradyzoites forms by the definitive host. A highly virulent strain killing its mouse host would be counter-selected, yet the selective pressure these two parties imply on each other is behind the existence of highly resistant mice adapted for such strains and allowing for the encystment and the completion of their life cycle.

Felidae ingest tissue cysts which release their bradyzoites in the stomach of their definitive host, these latter convert into a form, called merozoite, able to penetrate the intestinal epithelial cells where they initiate the asexual development of five morphologically distinct types designated A to E (Dubey and Frenkel, 1972).

The sexual cycle starts a couple days after the tissue cysts ingestion by the cat, and the gamonts are found mostly in the ileum of its small intestine. The fertilization is concluded by the formation of the oocyst wall. The steps of both asexual and sexual development within the definitive host are shared between the subclasses of the coccidian family, concluding with the rupture of epithelial cells and the discharge of the persistent immature oocyst into the lumen and the environment (Dubey, 2006).

5. Evolution of Sex, recombination and meiosis

The persistent oocyst shed in the environment grants sarcocystidians with a flexibility regarding the duration it takes for these parasites to produce sexual forms, one that vector borne-dependent parasites such as *Plasmodium sp.* would not profit from, that is because these latter would need to produce transmissible stages over a longer time to increase their chances of encountering a vector (Weedall and Hall, 2015).

The differences in the routes of transmission adopted by apicomplexan parasites, them being through vectors or through a direct fecal-oral route, are linked to their lifestyles and stages, as well as to their adapted sexual development.

Protists were once considered to be primitive to the higher eukaryotes, and were thought to be reproducing only asexually. For instance, it was only until the 1970s that it was acknowledged that *T. gondii* displays a sexual life cycle within a definitive Felidae host (Frenkel, 1970). Detecting sex in parasites was not a straightforward mission, however the advances made in unraveling the eukaryotic evolution placed the evolution of sex to have occurred as early as the last common ancestor of all Eukaryotes (Cavalier-smith, 2002, 2010).

The emergence of conventional sex in eukaryotes consisting of cell fusion, nuclear fusion and meiosis, as well as its maintenance, have been both matters of debate. It is believed that sex offers advantageous for the individual to evolve and respond to rapid environmental shifts .

With one of the most widespread environmental changes being generated by the dynamic interactions between hosts and parasites, it is thought their co-evolution is a driver for the persistence of sex (Lively et al., 1990). This idea has been adopted and developed in the Red Queen hypothesis, which states that sexual hosts are continually running (adapting) to stay in one place (resist parasites), and that sex would allow these populations to keep pace with their parasites by granting the offspring with potential new recombination of parasitic resistance alleles (Watve, 1998) . This can be observed in the high rates of evolution of genes coding for immune system and defense proteins (Kuma et al., 1995).

In fact, in the search for evidence of sex in a species, one of the indirect methods would be to look for patterns of genetic variation and recombination indicating events of outcrossing.

I briefly mentioned in the former chapter how acquiring a host that is larger in size would be beneficial for the evolution of the parasite in order to increase its genetic diversity through multiple infections, yet another limiting factor exists residing in the transmission rates of the parasite.

For instance, *Plasmodium species* display high transmission rates due to their sexual development being closely linked to their transmission cycle. This feature together with its mosquitoes-borne transmission route, allow for the high occurrence of multiple infections in hosts, therefore for more possibilities of recombination of the parasites (Weedall and Hall, 2015).

However, *T. gondii* displaying a broad range of hosts and a highly clonal population structure with a low sexual transmission intensity, as mentioned earlier regarding its ability to bypass the sexual phase of its life cycle. This possibly results in the different parasites rarely meeting in the host and thus the occurrence of less out-crossing events.

In spite of this, genetic crosses of *T. gondii* in its definitive Felidae host produced offspring of differing virulence (Herrmann et al., 2012), and regarding their evolutionary weight, crosses have been documented in *T. gondii* suggesting that the current populations were derived from genetic recombination of four ancestral lineages (Khan et al., 2007).

Unlike the great genetic diversity observed in strains of wild animals, limited numbers of clonal lineages are found in the human populations, mostly because of the aforementioned arms race being dampened. The types I, II and III are most recurrently found in European countries. Genome-wide SNPs sequencing showed that crosses between the type II with the ancestral “ α ” and “ β ” strains gave rise to the current types I and III, respectively (Boyle et al., 2006). To note that more advanced high-resolution sequencing suggest an extensive mosaic genetic architecture across diverse haplogroups, originating from broad recombination events (Lorenzi et al., 2016).

Searching for traces of possible out-crossing events can be an evidence for sex occurrence in a species, but not exhaustively, that is because recombination can occur in bacteria without true sex, as the evolution of sex arose much later than recombination, presumably at the onset of the eukaryotic emergence (Cavalier-smith, 2002, 2010) .

Therefore, a more reliable method would involve the identification of genes in the species genome with functions related to meiosis. For instance, during the attempts to unravel the sexual cycle of the coccidian parasite *Cryptosporidium parvum* based on searches of homologues for recognized sex related proteins, a homologue of HAP2 was identified (Tandel et al., 2019), a protein required for gamete fusion in a range of organisms including other Apicomplexans such as *Plasmodium sp.* (Liu et al., 2008) and *T. gondii* (Ramakrishnan et al., 2019).

As mentioned in the former parts revolving around the coccidian parasites and their life cycles, namely that of *T. gondii*, we mentioned that the elements of this family present similar developmental phases occurring within the definitive host, in both their asexual and sexual parts.

The number of asexual cycles depend on the species, and it seems that in *T. gondii*, the end of these cycles and the conversion to the sexual development, present more flexibility than in the other Coccidia as *Eimeria sp.* where this number is fixed (Speer and Dubey, 2005).

However, the trigger responsible for this conversion is yet to be defined. Also, the process through which a merozoite go through its development into a micro- or a macro-gametocyte, is also unknown. The terms female and male gametes are recognized as accepted nomenclature for the macro- and micro-gametocyte, respectively (Josling and Llinás, 2015) .

The fertilization of the gametes is preceded by their maturation.

Despite the little number of descriptions of micro-gametogony, it is known that the microgametocyte of *T. gondii* produces 15 to 30 male gametes (Ferguson et al., 1974). This number is considered relatively low compared to the female counterparts, which is controversial if compared to the universal features observed in plants and animals. To note that other electron microscopy-based studies of micro-gametogenesis in Coccidia suggests that this number can reach few hundreds (Scholtyseck et al., 1972).

The maturation of these organisms provides them with flagella aiding most apicomplexan male gametes in their movement, and while *T. gondii* male gametes present two flagella (Ferguson et al., 1974), the ones of *Cryptosporidium sp.* seem to be able to reach the female gamete despite their lack of any flagella (Tandel et al., 2019).

A fertilization event with a true internalization of a male gamete in the female gamete has not been clearly observed, although the close attachment of the two gametes has been

documented, but as mentioned earlier, one can rely in their search for sex on the traces of cross fertilized parasites and on the existence of fertilization related homologues as HAP2.

Meiosis directly follows fertilization and its intermediates with one, two and four nuclei were observable, along with labelling of the meiosis related DNA-repair protein, RAD51, recalling the occurrence of recombination and outcrossing. It is suggested that the meiotic process partially overlaps with the deposition of the oocyst wall (Tandel et al., 2019).

The oocyst wall formation for *T. gondii* appears to be shared between coccidians namely *Eimeria sp.* and *Cryptosporidium sp.*, resulting in a three-layered resistant wall (Ferguson et al., 1975).

To note that this resistant feature seems to be acquired during the development of the female gamete. It was reported in *Cryptosporidium sp.* that proteins with roles in oocyst wall synthesis were upregulated in these gametes. Not only does it provide the competence for the thick wall formation, the female gamete seems also to carry the components for the ability of the oocyst to persist in the environment with limited resources, as the stock of enzymes that are required for the metabolism of amylopectin and trehalose are upregulated in the female gametes.

Unlike HAP2 that is expressed by the male gametes, it seems that most conserved eukaryotic factors of meiotic recombination, such as DMC1 and HOP2, were expressed in females (Tandel et al., 2019), underlining the weight of the female gamete in this process.

6. The evolutionary advantage of haploidy

The fertilization of the female gamete by the microgamete produces a diploid zygote. Amongst Coccidians, this stage represents the only diploid state of the parasite which undergoes the meiotic division (Ferguson et al., 1974).

In fact, in all other life stages, parasites display their chromosomes in single sets. This haploidy state is featured in many unicellular organisms other than *T. gondii* and its fellow apicomplexans (Otto and Gerstein, 2008). For instance, many Fungi species can propagate for many generations as haploid cells (Zeyl et al., 2003), it is the dominant state for many algae, and exceptionally in some higher Eukaryotes as male ants and females of one mite species, where haploidy is reported to be constant throughout their life cycles (Normark, 2003).

As prokaryotic organisms and viruses present genomes consisting of a single DNA or RNA molecules, the idea of haploidy being the ancestral status of evolution, prevails.

Adopting haploidy by many organisms and maintaining it over their course of evolution suggests that it grants them with selective advantages. As no second copy exists to attenuate

the weights of mutations, a deleterious mutation would be more likely to be eliminated in haploid populations, whereas a beneficial mutation would spread more easily than in diploids where a beneficial allele is less efficiently fixed and propagated.

The lower mutation load makes it seem more evolutionary advantageous for parasitic organisms to adopt this chromosomal state which seems also to aid them in their host escape by preventing the risk of expressing double antigen alleles and getting cleared (Otto and Gerstein, 2008).

7. An unusual mitosis typifies apicomplexans

In most bisexual animals, haploidy oscillates with diploidy where the former exists only in the post-meiotic germline and the latter ensures the mitotic divisions. However, when haploid organisms display the ability for continuous divisions, asexual reproduction could prevail as the case in yeast dividing by fission or budding, and in other species that are able to achieve complete life cycles by clonal reproduction (Hong, 2010), or at least to meet their biotic mass requirements as the case of *T. gondii*.

The asexual division would start and end at the same point with invasion competent “zoite”, only with increased numbers. The processes for copying and segregating chromosomes during cell divisions in apicomplexans share their basic mechanisms with eukaryotic ancestors and extant species (White and Suvorova, 2018). Additionally, the progress of their cell cycle is controlled by orthologs of orthodox checkpoints proteins as cyclin-CdKs that are present throughout the phylum (Vleugel et al., 2012).

However, Apicomplexa parasites seem to possess some cell cycle-related particularities. For instance, *T. gondii* possess a relatively short replicative cycle lasting not more than 6 hours, and many species of the phylum display a lack of a G2 period (Figure 3), which seems not to be so uncommon in unicellular eukaryotes (Radke et al., 2001).

Their mitosis also displays notable features namely the intranuclear spindle and the maintenance of the nuclear envelope throughout mitosis (Hager et al., 1999). A bipolar microtubule residing at the nuclear periphery persists throughout the cell cycle providing a constant association to the centromeres, noting that conventional kinetochores are a conserved feature in apicomplexans (Brooks et al., 2011).

This bipolar type of chromosome segregation tool seems to be a LECA (Last Eukaryotic Common Ancestor) innovation. Similarly, the red alga spindle pole initiates the formation of microtubules and kinetochores facing a polar protrusion (Dave and Godward, 1982). In fact, the similarity of the spindle poles of the red alga and the coccidian spindle poles, was observed

many years before this alga got recognized as an ancestor in the evolutionary path of Apicomplexa.

This similarity is also observed at the nuclear envelope level, where in *T. gondii* and other Apicomplexa, a semi-closed mitosis takes place during which there is no occurrence of a full membrane disintegration nor does it stay fully intact (White and Suvorova, 2018). This form of mitosis which was likely to be adopted by LECA (Makarova and Oliferenko, 2016), consists of temporary breaks at the spindle pole which would close again during the nuclear division.

T. gondii adapted different modes of divisions adapted to their developmental stages and to the changing host-cell environments they encounter. These variations of asexual divisions differ mainly in the location of daughter parasites formation.

The majority of Apicomplexa undergo schizogony during which an invading zoite can give rise to many daughter cells, through multiple cycles of DNA replication and nuclear division, a multinucleated stage that would be followed by daughter cell formation, and ending by their budding from the mother cell (Ferguson et al., 2008). This process is used by *Plasmodium sp.* resulting in a multinucleate schizont before it gives rise to multiple zoites at once (Bannister et al., 2000).

T. gondii adopts the conventional apicomplexan schizogony during its asexual development in the cat intestine, here budding takes place at the plasma membrane.

However, Coccidia parasites display additional processes consisting of daughter formation within the mother cell cytoplasm, called endodyogeny and endopolygeny. Endodyogeny is applied by *T. gondii* in its intermediate hosts and results in the budding of two daughters after each round of DNA replication (Sheffield and Melton, 1968) (Figure 3).

Endopolygeny division is carried out by the *T. gondii* in the cat gut, noting that a different form of endopolygeny is adapted by *Sarcocystis sp.* which following replication, bypasses karyokinesis resulting in a polyploid nucleus followed by the budding of multiple daughters (Ca and Jp, 1999), with this process being possible due to the adaptation of a pair of MTOC complexes allowing independent control for karyokinesis and cytokinesis, thus enabling a synchronous release of multiple daughter parasites from an infected host cell (White and Suvorova, 2018).

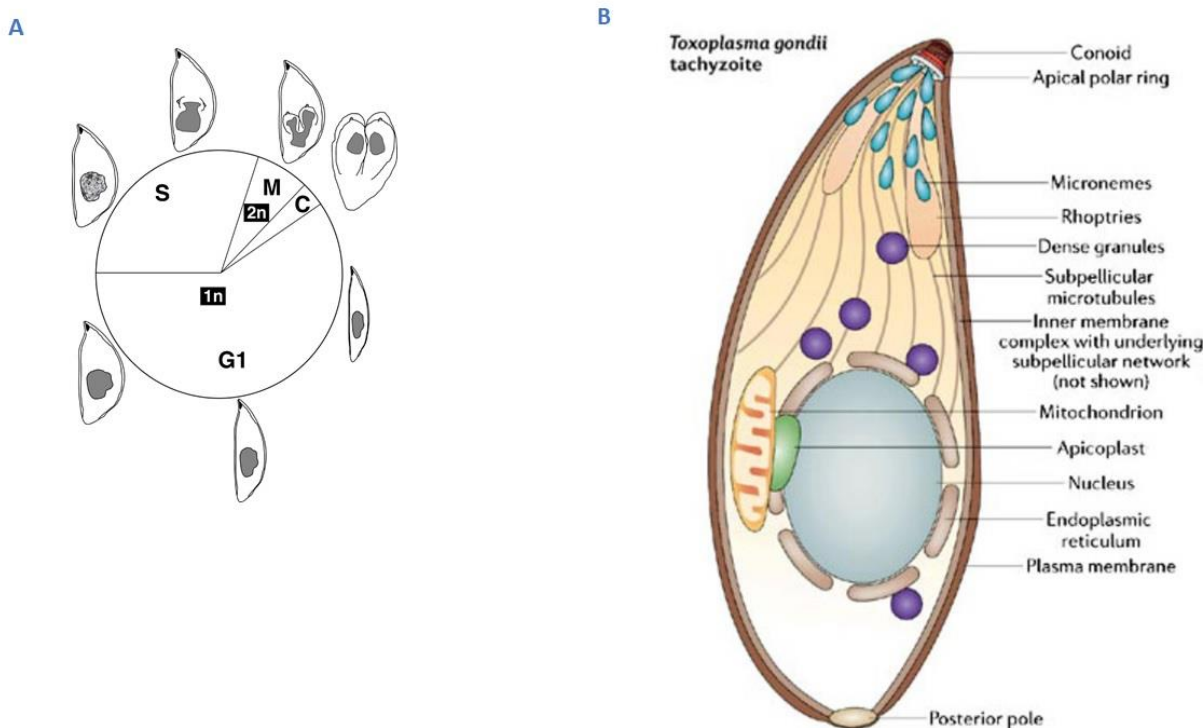


Figure 3. *T. gondii* cell illustration and cell cycle (Endodyogeny).

A) *T. gondii* cell cycle scheme of its endodyogeny mode of division. S, S phase; M, mitosis; C, cytokinesis. B) An illustration of the *T. gondii* tachyzoites showing the cells organelles and components.

8. The genome-free organelles characteristics of apicomplexans

The cell division ends by each daughter zoites emerging with a single nucleus and a full complement of organelles including an unique apicoplast, a single mitochondrion, a Golgi and multiple secretory organelles that seem to form *de novo* during the budding event (Figure 3).

Before going through the genome containing organelles of apicomplexans, namely of *Toxoplasma*, it is noteworthy to describe concisely the ones equipped for the purpose of invasion primarily, as well as of modulation of the host immune response. Bulbous rhoptries (ROP) and micronemes (MIC) organelles are packed at the anterior portion of the cell, whereas the dense granule (GRA) compartment is dispersed through the parasite cytosol (Tomavo et al., 2013).

These organelles comprise different families of secretory proteins destined to reach either the host cytosol or its nucleus, examples would include rhoptry-resident protein ROP16 that phosphorylates the host transcription factor STAT3 (Yamamoto et al., 2009), and dense granule proteins GRA16 and GRA24 that are able to modulate the host p53 and p38-alpha MAP kinase pathways, respectively (Bougdoor et al., 2013) (Braun et al., 2013) .

Before reaching the step of trafficking towards the host, the signal peptide loaded proteins would go firstly through the endoplasmic reticulum which would direct vesicles to the Golgi apparatus where the proteins would be sorted into their specific organelles prior to their discharge (Tomavo et al., 2013). A secondary host targeting motif would be required for the export across the parasite membrane into the host cell, a process well described in *T. gondii* (Hsiao et al., 2013) as well as in *Plasmodium sp.* (Spillman et al., 2015).

Some of these proteins undergo a process of cleavage by ASP5 prior to their export such as GRA15, GRA16 and TgIST (Hammoudi et al., 2015) (Gay et al., 2016) (Curt-Varesano et al., 2016). However, it seems that a protein complex residing on the parasitophorous membrane consisting of at least three proteins MYR1, 2 and 3 are the main actors in the export of dense granule proteins (Franco et al., 2016).

With most of these proteins displaying high abundance of intrinsically disordered regions, it seems that a positive evolutionary selection is weighing on *T. gondii* into adapting export machinery with preference for proteins with such regions, possibly as a mean to save energy as the secretion of IDR proteins does not require an active unfolding/refolding as is the case for chloroplast-destined proteins in Plants (Walker et al., 1996). The structural flexibility of these disordered proteins grants them with an additional advantage residing in their ability to acquire multiple protein partners (Hakimi et al., 2017), resulting in greater protein complexity that would serve the parasites in their arms race with their host, along with a probable easier immune evasion due to the higher rates of point mutations that these IDR secretor proteins display (Ma et al., 2014).

9. A peculiar nuclear compartment

Despite the different routes adapted by the various apicomplexans for the import of the secretory organelles proteins towards the host, these specialized structures and their respective effectors as well as transport mechanisms evolved in favor of the survival of these parasites within their hosts. Beyond harboring these highly efficient organelles, the majority of apicomplexans display another very special feature, consisting of having three genomes, which was for a long time thought to be a characteristic of plants and algae. In the following section I will elaborate on the organization and specificities of the three genomes-containing organelles namely the nucleus, mitochondria and the apicoplast.

The nucleus of Apicomplexa was not observed as one possessing any lamina structure (Koreny and Field, 2016), only an envelope with numerous nuclear pores and ribosomes bordering its external sides. The position of this basic eukaryotic structure seems to change in the case of *T. gondii* depending on the stage the parasite is in (Hager et al., 1999).

The nucleus of an organism harbors many of its secrets including the one telling its history and the extent of its cousin species divergence, as well as its own, from their common ancestor. The science behind tracing these events of divergence is based on determining the level of

synteny between the genomes of organisms belonging to one genus or in a larger extent to one phylum. A high synteny is defined as conserved content, order and chromosomal distributions of genomic *loci*.

Such levels of conservation were not found across the Apicomplexa phylum (DeBarry and Kissinger, 2011), however significant levels were detected between genomes of organisms in the same genus. Namely, for a long time this was believed to be the case for *T. gondii* and its fellow coccidian *Neospora caninum*, until a recent study challenged this paradigm and identified events of large chromosomal rearrangements between these two (Berna et al., 2020).

To note that such rearrangements would occur by chance, however their maintenance must be weighted by favorable selective pressures marked by the fact that rare synteny breaks could be observed within coding regions of these closely related species. Therefore, despite having half of their chromosomes structured similarly, they each display their particularities through the rest of their karyotype.

In fact, the karyotype of both these species had to be recently corrected. It was thought for a long time that *T. gondii* and *N. caninum* harbored 14 chromosomes, as is the case for *Plasmodium sp.*, until a 3D genome organization study reported an unusually high number of interactions occurring between chromosomes VIIb and VIII, with a great number of contacts between the right telomere of one and the left of the other (Bunnik et al., 2019).

This observation led to the suggestion that these two chromosomes could be physically linked, yet a recent study based on third generation sequencing technologies, i.e., *Nanopore*-based sequencing, settled the discrepancy and brought the confirmation of them being fused, thus reducing the karyotype of both *T. gondii* and *N. caninum* to 13 chromosomes instead of 14 (Berna et al., 2020).

Genome assembly artifacts are, in many cases, generated by the limitations of former sequencing technologies underlined by their inaccuracy of reading through repetitive regions. These latter are thought to account for more than 20% of the accurate genome of *T. gondii*. However, no transposable elements have been identified so far in apicomplexans genomes, unlike trypanosomatids that owe much of their genomic divergence to the presence of such elements in their genomes (DeBarry and Kissinger, 2011).

Intergenic regions do not stretch over long spans, with sometimes few hundreds of nucleotides separating two adjacent proteins encoding genes, creating a high density of transcriptional units (Kissinger and DeBarry, 2011).

Other non-coding regions consist of the telomeric and centromeric structures. The telomeric repetitive nucleotide sequences of many apicomplexan parasites seem to lack complete conservation with the one of vertebrate TTAGGG. The repeat sequence of *Plasmodium sp.* is GGGTTYA, with Y being either a T or a C (Scherf et al., 2001), which is different than the one of *T. gondii* TTTAGGG. The *T. gondii* telomeric sequence is identical to the one observed in

plants, thus adding a supplementary layer to the evolution story and divergence that occurred since the microalgae common ancestors of these two. Noting that *Cryptosporidium species* do not share this sequence (Liu et al., 1998), letting one suggest the possibility of the loss by this latter of the ancestral signature or that *T. gondii* acquired during the evolution the plant-like telomeric repetitions.

The centromeres of *T. gondii* that were mapped using CHIP of cenH3, did not show any nucleotides bias (Brooks et al., 2011), unlike *Plasmodium sp.* which display extreme A/T rich centromeres (Gardner et al., 2002). Yet both species structures were observed to be clustering within the nucleus, as *T. gondii* centromeres are confined to a single spindle pole body that is constitutively associated with the nuclear envelope throughout the cell cycle (Brooks et al., 2011).

Clustering of centromeres dominated the genome of *T. gondii*, yet not many other regions displayed such interactions. No Topologically Associated Domains (TADs) were described and no clustered organization of their families of species-specific multi-copy genes, nor their effectors was detected (Bunnik et al., 2019).

Such genome organization was however reported in *Plasmodium falciparum* which chromosomes bearing *var* genes seem to come together in 3D space and to form loops for sheltering the genes at a perinuclear position, with this organization possibly offering means for increased rates of recombination, thus generating greater diversity and coordination between the virulence genes (Scherf et al., 2008). Therefore, it is proposed that a higher level of genome organization flexibility can be observed in organisms that lack antigenic variation requirements, such as the case for *T. gondii*.

Such species-specific genes seem to represent the biggest share of protein-coding genes in the apicomplexan parasites. In fact, as mentioned earlier, major genomic rearrangements are observed through the phylum, thus as few as 1000 genes are conserved amongst them (DeBarry and Kissinger, 2011). Most of these shared genes have orthologs outside of the *Apicomplexa* phylum, suggesting their ancestral nature.

Apicomplexans present different numbers of protein-encoding genes ranging from around 3706 in *Babesia bovis* to 8322 in *T. gondii* (source VEuPathDB, release 48). These numbers are relatively reduced when compared to higher eukaryotes, however the same numbers are considered high when seen in the perspective of the relatively small genomes of the species of this phylum with the one of *T. gondii* accounting for merely 65-Mb increasing to 80-Mb after inclusion of the non-annotated repetitive elements, and those of the other phylum organisms ranging from 9- to 130-Mb (DeBarry and Kissinger, 2011). These aspects are all the more significant when compared to *Arabidopsis thaliana* carrying a genome of 135-Mb.

Noting that the nuclear genomes of the apicomplexans are heavily affected by the amount of acquired genes through their evolution via events of horizontal gene transfers as well as

intracellular transfers stemming from the mitochondrial genome and from the ancestral algal nuclear and plastid genomes.

10. The mitochondria genome

Apicomplexans harbor tubular organellar structures as their mitochondrions, usually a singular organelle with changes greatly at the level of its form and size based on the species and on the life-cycle stage (Keithly, 2008). Noting that unlike the mammalian structures which numbers vary independently of the cell cycle, the mitochondrion of apicomplexans and namely of *T. gondii* divides simultaneously with the cytokinesis (Nishi et al., 2008).

The mitochondrion carries its own genome, however most of the proteins destined to function at the organelle (e.g. its phage-like single-subunit RNA polymerase) are encoded in the cell nuclear genome and then imported to the former through translocation complexes that must have evolved in parallel to the intracellular transfer of mitochondrial genes (Mallo et al., 2018). This import machinery is also used to traffic tRNAs for synthesizing proteins within the organelle (Esseiva et al., 2004). Such gene transfers led to the reduction of the mitochondrial genome, which in Apicomplexa is as small as 7-kb which is recognized as the smallest mitochondrial genome of all organisms described so far (Gray et al., 2004).

It has been reported that many species of the phylum, including *Plasmodium sp*, *Theileria sp* and recently *T. gondii* display their mitochondrial DNA in linear fragments with no evidence of a detected circular molecule (Berna et al., 2020) (Preiser et al., 1996). As few as three protein-encoding genes have been retained in the mitochondrial genome, consisting of apocytochrome b, and 2 subunits of cytochrome c oxidase, with suggestions that their reservation in the organelle might be caused by their high level of hydrophobicity possibly defying any possible import from the nucleus (van Dooren et al., 2006).

Moreover, additional ORFs were detected within the genome, along with genes encoding proteins required for its maintenance and a part of the translation apparatus as the other part is imported as mentioned ahead (Esseiva et al., 2004). In fact, this powerhouse has been maintained in all Eukaryotes for its multi-functional nature. Fatty acids synthesis is essential for the fitness of any organism unless they are able to acquire their requirements from external sources as the case of *Theileria sp.* parasites which seem to have lost this metabolic pathway due to an abundance in nutrients caused by their direct access to the host cytoplasm (Seeber et al., 2008). For other parasites of the phylum, their mitochondria are at the base of the oxidative decarboxylation of pyruvate to acetyl-CoA known as the building blocks for FA.

Other functional aspects would consist of the biosynthesis of iron-sulfur clusters, a pathway that also seem to be affected by the different host environments, and additionally in *T. gondii* and *Plasmodium sp.* this organelle fulfills the requirement for the crucial pyrimidine biosynthesis (Seeber et al., 2008).

Despite sharing the same set of retained genes, however it seems that the gene order, distribution and order of the genes within the species of apicomplexan species are highly divergent, to the extent that the synteny is lost even between two strains of the same species, showcasing a great level of mitochondrial DNA sequence heteroplasmy between closely related species (Berna et al., 2020). This level of reshuffling wasn't observed in the apicoplast, as this other endosymbiotic-originating organelle display genomes that seem virtually identical between the species of the phylum (Berna et al., 2020).

11. The Apicoplast genome

Other than both originating from endosymbiotic events, the apicoplast and mitochondria seem to also form a physical association prior to their segregation into daughter cells (van Dooren et al., 2005). In fact, for a long time the apicoplast organelle was mistaken to be the mitochondria of apicomplexans, this dating back to the 1970s (Kilejian, 1975), until this dogma started getting solved by accepting that the linear genome of 7-kb in *Plasmodium sp.* belonged to mitochondria (Vaidya et al., 1989). The sequencing of the 35-kB molecule resulted in the identification of inverted repeats with duplicated sets of ribosomal RNAs, which at the time was a feature recognized as characteristic of plastid genomes (Gardner et al., 1991).

Moreover, electron micrographs on *T. gondii* helped to identify that the organelle was surrounded by what they thought at the time was two membranes, like the case observed in plants and microalgae (McFadden et al., 1996) (Köhler et al., 1997). This was the first time that it was suggested that apicomplexan parasites might have evolved from algal ancestors.

The apicoplast is a relic non-photosynthetic plastid-like organelle that seems to change in shape and location in the cell throughout the life cycle (Striepen et al., 2000). The plastid genomes are a small fraction of the size of the genomes of *Cyanobacteria* which is at the origin of the emergence of this organelle as will be discussed ahead. Still, the apicoplast genome represents one of the most reduced plastid genomes identified so far (Smith and Keeling, 2015).

Reports about the apicoplast being circular have recently been challenged claiming not finding any compelling evidence arguing in this favor, instead they suggested this genome exists as a linear molecule accounting for 60-kb instead of the presumed 35-kb (Berna et al., 2020). This feature was observed also in *Chromerida sp.* which seems to have lost the circular aspect of its plastid (Janouškovec et al., 2010).

Around 60 open reading frames stretch over this genome with most of them serving for its maintenance, corresponding to rRNAs and tRNAs, and with the exception of three identified protein encoding genes, namely *SufB*, *clpC* and a yet to be characterized gene (Berna et al., 2020). The aforementioned high level of conservation of the gene content of the apicoplast between species, seems to be common to most plastids as little genetic diversity was observed

between plastid-bearing species, despite the multiple events of endosymbiotic gene transfers that led to the movement of many genes to the nucleus independently in many lineages (Gould et al., 2008).

This pattern of genetic integration between the endosymbiont and the host, in parallel to adapting mechanisms for the protein import, make up the two criteria for an endosymbiont to be considered an organelle within its host. Noting that the trafficking of the nuclear-encoded proteins to plastids originating from one endosymbiotic event, as the case of the three microalgae, would require less complex machinery than the one targeting secondary plastids, as the case for plants and Apicomplexa (Patron and Waller, 2007).

The establishment of such import machinery is suggested to have occurred as the cyanobacterium got genetically integrated within its primitive eukaryotic host (Figure 4), which allowed for further transfer to occur, ending with the progressive degeneration of the cyanobacterium (Douglas and Raven, 2003).

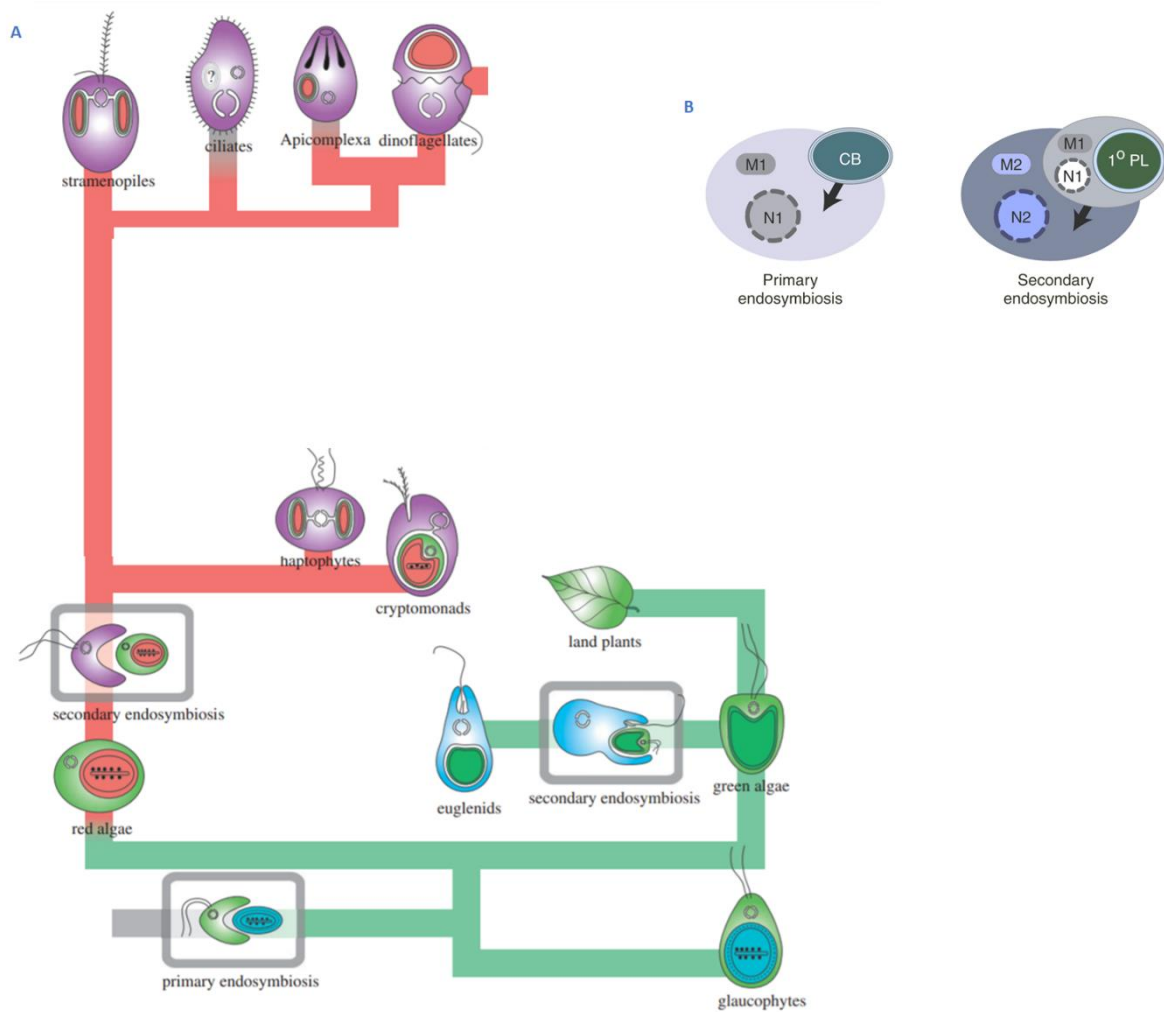


Figure 4. **Plastid evolution and fate**

A) Endosymbiosis events are boxed, and the lines are colored to distinguish lineages with plastids from the green algal lineage (green) or the red algal lineage (red). At the bottom is the single primary endosymbiosis leading to three lineages (glaucophytes, red algae and green algae). On the lower right, a discrete secondary endosymbiotic event within the euglenids led to their plastid. On the lower left, a red alga was taken up in the ancestor of chromalveolates. *Adapted from (Keeling, 2010)*

B) A cartoon depicting the primary endosymbiotic origin of plastids through the uptake of a double-membrane-bound cyanobacterium by a non-photosynthetic host eukaryote. Secondary endosymbiosis involves the engulfment of a primary-plastid-containing eukaryote by a second, non-photosynthetic eukaryote. Abbreviations: CB, cyanobacterium; M, mitochondrion; N, host nucleus; PL, plastid. *Adapted from (Archibald, 2009)*

A secondary degeneration also occurred when the endosymbiont algae transferred all of its genetic material to the secondary host nucleus (Figure 4), including the plastids proteins it once encoded, and all that remained from the alga was the plastid, which would be surrounded by four membranes corresponding to the host endomembrane, the engulfed alga plasma membrane and the primary plastid double membrane (Gould et al., 2008).

Despite the apicoplast organelle having most of its genes being encoded in the nucleus and then imported, yet the organelle is retained in most apicomplexans, except *Cryptosporidium* sp. (Abrahamsen et al., 2004), letting one ponder on the limiting factors behind the retention of the organelle with its remaining reduced genome.

The functional characterization of this organelle is resolved through the identification of the nuclear components destined for the metabolic pathways that are retained in the apicoplast.

Heme synthesis takes place in the organelle, yet cannot account for its global retention pattern in apicomplexans, as this pathway seems to be essential only in the liver and Anopheles developmental stages of *Plasmodium berghei*, whereas the parasite requirements in the blood stages can be met through their uptake from the host (Nagaraj et al., 2013).

The retention of the plastid was argued to be linked to its involvement to the biosynthesis of fatty acids such as the isoprenoids which are important lipid compounds, and their building blocks are IPP, products of the apicoplast (Botté and Yamaro-Botté, 2018). It is thought that *Cryptosporidium* sp. have lost their plastid because of the ease with which the parasites are able to scavenge its IPP requirements from the host (Gisselberg et al., 2013).

Cryptosporidium sp. stand out between Eukaryotes as one rare example for proving an event of plastid loss, where it offers enough genomic and structural evidence allow to say that no plastid is present, while offering enough evolutionary evidence that its ancestors contained a plastid through the Chromalveolata hypothesis (Keeling, 2008).

This hypothesis states that a single common endosymbiosis is behind the plastids present in the species of this reign (Cavalier-Smith, 1999). Yet the discrepancy about the origin of these plastids being from a red or green alga, was settled thanks to the sequencing of the plastid of *Chromera velia*, the photosynthetic sister to Apicomplexa, which enabled the demonstration that the plastid origin is of a red alga (Moore et al., 2008) (Keeling, 2010). The discovery of this species made it possible to assert that plastids of dinoflagellates and Apicomplexa share a common origin, despite the differences between them namely on the level of their plastid membranes number (Janouškovec et al., 2010).

The red alga origin makes some observations in *T. gondii* puzzling, as many features were witnessed to be similar to plants, although these latter arose from the green alga (Andrabi et al., 2018).

12. Apetala Transcription Factors

a. Their Origins

Plants and Alveolates were described to share many transcription factors (TFs), yet less than what plants and Stramenopiles share, which is noteworthy seeing that they both arose from

a single red alga endosymbiotic event in the ancestor of the Chromalveolata lineage, suggesting the possibility of a greater loss of plant-like genes occurring within this branch species as the apicomplexans and the ciliates (Armbrust et al., 2004) (Bhattacharya et al., 2004).

In fact, it was suggested that TFs that were once thought to be specific to plants have their homologs found in cyanobacterium, namely the ones harboring an Apetala (AP2) DNA binding domain (DBD) (Magnani et al., 2004). Similar proteins, recognized as ApiAP2, with one or more copies of this AP2 DBD were identified in apicomplexan species, even in the ones that lost their apicoplast as *Cryptosporidium sp.* (Balaji et al., 2005).

The gap in determining the evolutionary origin of these AP2s domains arose from their apparent absence in the red algae (Magnani et al., 2004). It was proposed that this was either due to a subsequent loss of the domain in this organism or that plants have acquired this domain via horizontal transfers from bacterial or viral infections, an idea that is opposed by the existence of this domain in the green alga suggesting an early evolution of the AP2 domain in plants.

The fact that homologues of the AP2 domain were found in bacteriophages and bacteria, led to the suggestion initiated ahead about it originating from cyanobacteria. An observation arguing in this favor is the lack of introns in the majority of plant AP2 TFs further supporting their prokaryotic nature (Magnani et al., 2004).

As mentioned earlier, the cyanobacterium endosymbiont got progressively degenerated and transferred its genetic content to the algae nucleus, herein these TFs can still be traced in the nucleus of the green algae as well as in the nucleus of species belonging to haptophytes and Stramenopiles, both included in the Chromalveolata reign (Thiriet-Rupert et al., 2016).

The AP2 domain of cyanobacterium was part of an HNH endonuclease which seem to have diverged towards losing the HNH domain and retaining the AP2 domain (Magnani et al., 2004). This pattern seems to be conserved in the evolution of many TFs which would emerge from a recycling event of inactive mobile elements that have lost their catalytic activity of the transposase domain yet retained the DBD (Babu et al., 2006).

Newly acquired TFs seem to spread in their new genome via transposition and homing processes (Magnani et al., 2004), and unlike proteins that would have settled into highly conserved roles during their eukaryotic residency, these former TFs would be rapidly diverging and adapting in a lineage-specific manner. Lineage specific architectures were observed of this AP2 domain in the *Phytophthora* parasite where it was found in a combination with an HDAC domain and a PHD finger domain (Iyer et al., 2008a).

b. Expansions and Discovery in *Apicomplexa*

In *Apicomplexa*, the AP2 domain went through great expansions, unlike in Ciliates where very low numbers are detected (Balaji et al., 2005). Such genes expansions would have to be

accompanied by functional diversification in order for the duplicated genes to be maintained, considering that redundant functions being lost would not cause phenotypic changes and thus no real selective pressure.

This holds true for TFs harboring a conserved DNA binding domain (DBD), which despite them sharing secondary and tertiary structures required for binding DNA, they display diversity in their recognition sequences and thus in their functions as TFs.

Such diversity can be achieved through mutations occurring on the residues required for base recognition, but which wouldn't influence the protein folding, unlike the ones that are relevant for the DNA binding mode which undergo more positive pressure (Yamasaki et al., 2013). Also, acquiring additional domains through recombination events can help achieve such functional expansion. The combination of structural domains could be achieved not only on the same protein, but also through interactions with other proteins, them being activators or repressors, which would grant different functional fates for TFs that harbor the same DBDs.

I noted earlier that the AP2 domain seem to have originated from a cyanobacterium homing endonuclease, and their expansion must have occurred through transposition and homing processes as the case for many other TFs families. However, it must be noted that not much evidence exist supporting the apicomplexan extant species having retro transposable elements, other than in some early branching species as it seems that these elements were lost from the apicomplexan genomes (Roy and Penny, 2007). Nevertheless, versions of the AP2 domain that are independent of the integrases or transposases were recently identified in bacteria and it is suggested that they might represent the progenitors of the AP2 TFs of apicomplexans (Iyer and Aravind, 2012).

AP2 harboring proteins were described in Apicomplexa in 2005, in alignment with the acknowledgment of the scarcity of families of specific TFs in this phylum where very few examples of conserved DNA binding domains of specific TFs were identified, such as Myb, C2H2 zinc fingers, High motility group b, PREBP, and E2F that is lost in all but *Cryptosporidium* species (Aravind et al., 2003) (Komaki-Yasuda et al., 2013). The deficiency both in numbers and variation of specific TFs would seem incompatible with the elevated gene regulation requirements that are emanating from the high number of protein-encoding genes, as well as from the various environmental pressures that a parasitic life style bestow on these species.

It is suggested that some AP2 domains have evolved in a species-specific manner, in parallel to their regulatory targets involved in parasitic processes such as invasion or immune evasion, and that other ancestral domains that span the phylum might be responsible for regulating basal processes such as replication and translation (Behnke et al., 2010). In fact, it is reported that the species of *Plasmodium* and of *T. gondii* witnessed the greatest events of lineage-specific amplifications with these accounting for more than 20 out of the 50 domains in the case of *Plasmodium spp*, and for around 60 out of the 90 domains identified in the case of coccidian parasites namely of *T. gondii* (Oberstaller et al., 2014).

c. AP2 Transcription factors in Land Plants

The common Chromalveolata ancestor is reported to having about 9 to 18 AP2 domain containing TFs, that are originating from the cyanobacterium via the multiple endosymbiotic events, yet this assumption is relatively recent, as for a long time it was thought that the rhodophyte progenitor is at the origin of this transfer, as discussed in the earlier section.

Also, this family of proteins was thought to be exclusive to the plant kingdom until homologs of the APETALA2/ethylene-responsive factor (AP2/ERF) were identified in various organism, including protists, slime molds, bacteria, viruses and even animals with 12 homologs identified in a bat species (UPI000944DD98).

However, it holds true that the AP2/ERF is one of the most represented families in plants with the AP2 domains being the second most common DBD after the Myb domain in *Arabidopsis thaliana* (Riechmann and Meyerowitz, 1998). The AP2 subfamily is one constituent of the large family, along with ERF, DREB (dehydration-responsive element binding) and RAV (related to ABI3/VP1) subfamilies, with their various proteins being involved in responses to abiotic stresses and development of flowers and seeds, as well as in ethylene response (Mizoi et al., 2012) (Aida et al., 2004). Ethylene is a plant hormone that impacts plant development, and genes that are induced by this hormone are recognized through their Ethylene responsive elements (or GCC box) by the ERF proteins (Broglie et al., 1989) (Fujimoto et al., 2000). Proteins of this subfamily contain one AP2 domain, unlike the AP2 subfamily which possess two copies of the 70 amino acid domain.

It's been reported that the AP2 domain would occur on the same protein with other DBD such as B3, and that the two domains would bind their respective motifs with great affinity and specificity despite the various spacings occurring between the motifs, and it has been suggested that this might be reinforced by the highly flexible structure that connects the two domains (Kagaya et al., 1999).

An early study had already reported the double AP2 domain binding-mediated DNA sequence recognition, where neither AP2 repeat of the ANT protein was able of binding the longer than usual identified specific sequence (Nole-Wilson and Krizek, 2000).

d. Structural perspective

The DNA binding mode of this domain has been identified through structural determination of AtERF1 (pdb: 3GCC) revealing three β strands contacting the DNA backbone and specific bases within the major groove with the strands being stabilized by a C-terminal α -helix that does not contact DNA (Allen et al., 1998).

This structure placed the AP2/ERF domain as a part of a structural superfamily which groups the DBD of AP2 with the ones of Tn916 integrase (Wojciak et al., 1999), the I-integrase and

the methyl-CpG-binding domain (Ohki et al., 1999). To note that no apparent sequence similarity is found between these domains (Magnani et al., 2004).

The resemblance in the three-dimensional structures of the AP2 domain and of proteins of bacterial origins recalls the evolution traces of this domain which were further elaborated in the former chapter. Unlike plants, the apicomplexan domains seem to share little sequence identity with each other. This diversity is reflected by the fact the ApiAP2 TFs have also expanded their range of recognition motifs (Linas et al., 2008) as they present greater variability of DNA sequences than the plants AP2/ERF TFs most of which bind the canonical GCC box (Sakuma et al., 2002).

A crystallographic analysis of the AP2 domain of PF14_0633 from *P. falciparum* in complex with DNA (pdb: 3IGM) reported a similar yet not identical structural motif as that of AtERF1 and those of bacterial integrases. The secondary structure of the apicomplexan domain seems to differ from the plant domain, in that it was found in a homodimer where the α -helix of one monomer aligns with the β -sheet of the other monomer (Lindner et al., 2010).

As described in plants regarding the co-operative binding of two AP2 domains within the same protein, it was revealed in PF14_0633 that a similar binding mode exists, but unlike plants, it seems that in this case the two domains would dimerize and this dimerization would be triggered by their binding to DNA. This last statement was underlined by size exclusion chromatography of the AP2 domain in the absence or presence of DNA. The eluted fraction in the absence of DNA corresponded to a monomer (Lindner et al., 2010).

This domain-swapping mechanism is responsible for the aforementioned structure consisting of the α -helix from one APiAP2 aligning with the β -strands of the other APiAP2 (Figure 5). The DNA-binding dimerization trigger could induce conformational rearrangements of the remainder of the protein or its interaction partners (Lindner et al., 2010). To note that the dimerization ability of these domains permits a higher level of functional diversification and thus a greater expansion of this family.

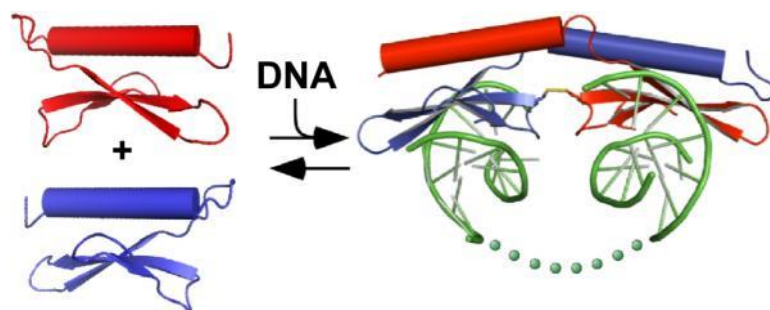


Figure 5. A model of DNA-induced, domain swapped dimerization and DNA looping.

DNA binding stimulates formation and/or stabilization of the domain-swapped dimer, which in turn loops out intervening DNA between the two binding sites (illustrated as dots). Dimerization may allow portions of the

proteins to undergo functional rearrangement of protein interaction/transcriptional regulatory domains. Adapted from (Lindner et al., 2010)

e. Outside the DBD

It is suggested that the structural rotations of the C-terminal α -helix would expose any additional domains, outside the DBD, to facilitate protein-protein interactions (Lindner et al., 2010). No noteworthy sequence homology was detected in the rest of the ApiAP2 between proteins (Jeninga et al., 2019). Few additional domains have been identified as a zinc-finger domain, and Acyl-CoA-N-acetyltransferase, a pentapeptide-repeat-like domain and an AT-hook (Jeninga et al., 2019) that might contribute to DNA binding but not in a sequence-specific way (Aravind and Landsman, 1998).

Many ApiAP2 TFs in apicomplexan parasites also contain an ACDC domain, the function of which is unknown, and which is exclusively found in ApiAP2s, at a rate of one copy per protein unlike the AP2 domain, and it is mostly found on the C-terminus of the protein, which explains its acronym, **AP2-Coincident Domain** mainly at the **C-terminus** (Oehring et al., 2012).

The function of the ApiAP2 protein as a repressor or activator is suggested to be weighed by the auxiliary domains it presents or on the interactions it is able to establish.

f. AP2-containing proteins in *T. gondii*

Functional characterization has been assigned to relatively few ApiAP2 in the coccidian *T. gondii*, which harbors 68 AP2 domain-containing proteins, out of them 24 have been reported to be cell cycle-regulated (Behnke et al., 2010).

Out of the 24, the TgAP2X-5 (5th AP2 protein on Chr. X) was reported to indirectly regulate the expression of virulence factors that are expressed in S and M phases in the cell cycle. Its action is accomplished through its regulation of the targeting of another ApiAP2 TgAP2XI-5. This latter with its additional trans-activator domain, is the alleged activator of these factors (Walker et al., 2013a), yet its binding to its targets promoters require the cooperative contribution of both interacting proteins (Lesage et al., 2018).

The intermediate part of the life cycle of *T. gondii* gained the most interest in studying the AP2 proteins of this parasite, mostly due to the ease with which the invasive non-cell discriminating tachyzoite and tissue-cyst bradyzoite stages can be manipulated for lab assays, and secondly because the interconversion between these two stages is responsible for the acute and chronic stages of the toxoplasmosis disease, respectively, and thus is behind the opportunistic nature of this parasite and the host life-threatening outcomes of the disease.

Still, no AP2 “master regulator” was identified to be responsible for the parasite differentiation into bradyzoites, but some AP2s were characterized as secondary contributors

to this interconversion which seems to involve an intermediate stage, where switching parasites didn't yet reach their mature bradyzoites fate (Hong et al., 2017).

One of the presumed actors of this intermediate stage is AP2IX-9, the deletion of which led to an increase in the cyst formation, assigning it with a bradyzoite repressor role, one acting at the transcriptional level as it was found to be associated with promoters of bradyzoite-specific genes (Radke et al., 2013). The promoters of the same set of genes were observed to be bound by another factor AP2IV-3, the deletion of which reduced the tissue cyst formation; thus it was suggested that the two antagonistic actors might be either competing or operating sequentially in order to regulate the bradyzoite gene expression (Hong et al., 2017).

This regulation seems to involve additional actors namely AP2XI-4 as an activator (Walker et al., 2013b) and AP2IX-4 as a repressor (Huang et al., 2017). AP2IX-4 was recently found to be associated with AP2XII-2, with these two sharing identical cell cycle expression patterns beginning at the S phase, peaking at mitosis and dropping after the nuclear division (Srivastava et al., 2020). Their impact on the bradyzoite differentiation is linked to their weight for a proper cell cycle progression, as it was previously shown that the initiation of this differentiation path occurs during the S/M phase with a slowing of growth and a delay in S-phase (Radke et al., 2003).

Despite the lack of extensive research on the AP2 repertoire of *T. gondii* and their functions, it holds true that this family of protein got more expansions in the coccidian parasites than in *Plasmodium* species for example, which can be related to the wider intermediate host ranges of coccidian parasites, as well as to the smaller genome size of *Plasmodium* species.

In the realm of the search for functional homology of ApiAP2s between species of the phylum, it was observed that many DNA-binding motifs are conserved, however, with very few exceptions (Pieszko et al., 2015), no exhaustive proof was presented in favor of a parallel conservation of the biological function or of similar sets of genes being targeted.

g. AP2-containing proteins in *Cryptosporidium* sp.

Before delineating the functional characterization of some of ApiAP2s in *Plasmodium* sp., it is worthy to outline some of the specificity of the other coccidian *Cryptosporidium*. The AP2s of this parasite display a lack of diversity regarding their binding sequences, compared to other parasites of the phylum. In fact, a high level of redundancy is suggested to be present between these proteins where different motifs could be recognized by the same protein.

It seems that this parasite is less reliant on the AP2 family of specific TFs and has put up more functional weight on other specific TFs family, namely the ancient eukaryotic conserved E2F/DP1 transcription factors which are absent, or more plausibly, lost in all other apicomplexans noting that the binding motifs of these proteins are amongst the most

overrepresented sequences upstream of the *Cryptosporidium parvum* genes (Oberstaller et al., 2014).

This expanded family of proteins granted *Cryptosporidium* sp. with a ratio of 1 TF to 340 target genes, which is much higher than the one corresponding to *Plasmodium falciparum* consisting of 1 TF to 800 genes (Templeton et al., 2004).

h. AP2-containing proteins in *Plasmodium* sp.

i. ApiAP2s involved virulence and clonally variant families

The regulation of genes in *Plasmodium* species gained great interest especially after realizing its weight in aiding this parasite in its strategic host immune system evasion, which is based in part on the families of proteins with features of antigenic variation, such as the *var* genes bearing mutually exclusive expression of a critical virulence factor for malaria, the Erythrocyte Membrane Protein1 (PfEMP1) (Scherf et al., 1998) (Deitsch et al., 2001). The low rate of switching of the PfEMP1 variants is a factor in pathogenesis as well as in the parasite persistence.

The multicopy *var*, *rifin*, *stevor* and *surfin* gene families all show clonally variant gene expression and they provide *Plasmodium* species with greater adaptive phenotypic plasticity and better responses to their hosts environments (Scherf et al., 2008).

An ApiAP2 protein has recently been described as a driver of virulence in the rodent malaria parasite, *P. berghei*. Proof was brought on the impact of a single SNP occurring within the DBD of ApiAP2, on the regulation role of this latter. The SNP results in the switch of a hydrophobic aromatic aa, phenylalanine, into a hydrophilic serine aa (Otto et al., 2014), which was sufficient to alter the binding of the DBD and resulted in a differential expression of many genes, mostly involved in invasion, antigenic variation and immune evasion (Akkaya et al., 2020).

The single polymorphism granted one strain (S aa) of *P. berghei* over the other (F aa), with the ability of evading triggering an early IFN- γ parasite-specific immune response in the host, rendering this protein one of the first ApiAP2 to be reported to fulfill a role in the host-pathogen interaction, noting that the ortholog in *P. falciparum* seems to share the same target DNA motif, and is thought to have conserved this beneficial immune evading SNP (Campbell et al., 2010).

Var gene silencing regulation was reported to be involving an ApiAP2 in *P. falciparum*, named PfSIP2. This protein binds exclusively SP2 arrays upstream of subtelomeric *var* genes, and is reported to harbor two AP2 domains, both required for a proper binding of the bipartite SP2 motifs, recalling the AP2 binding mechanism described in the double AP2 DBD containing ANT protein in plants. Additionally, this ApiAP2 is shown to require a specific proteolytic event in

order to activate the DNA-binding activity. PfSIP2 is thought to play a role in chromosome end biology and heterochromatin formation (Flueck et al., 2010).

Also, in *P. falciparum*, PfAP2-Tel co-localizes with telomeric clusters at the nuclear periphery, as mentioned ahead regarding the genome organization of *Plasmodium sp.* (Sierra-Miranda et al., 2017). This telomere-specific protein harbors an AP2 domain with an unorthodox structure as it lacks one of the three β -sheet recognized to be essential for DNA binding. The ortholog of this protein in *P. berghei* is AP2-Sp3 (Modrzynska et al., 2017), but it has not yet been assigned with telomer-associated functions.

In an assay to predict binding motifs using recombinant AP2 domains, the promoters of *var* genes were predicted to be bound by the DBD of AP2-exp protein (Campbell et al., 2010), yet this observation didn't get backed up by the transcriptomic effects of this protein, the mutation of which yielded alterations in the expression of members of clonally variant gene families but not of the *var* gene family (Martins et al., 2017), recalling similar events reported in *A. thaliana* where the binding of an AP2 on a gene promoter and the transcriptional outcome didn't form a linear link (Yant et al., 2010).

Instead, AP2-exp was shown to regulate the expression of RIFINs and STEVORs, with this protein being the first described to fulfill a specific regulatory function that is not dependent on its AP2 DNA binding domain, but on its C-terminal (Martins et al., 2017), yet it is unknown how this specificity is accomplished.

It seems that divergent C-terminus parts could provide distinct functions for proteins harboring similar DNA binding domains, as is thought to be the case for the ortholog of AP2-exp in *P. berghei* AP2-Sp, which together share DBD homology of 92%, and yet this latter is reported to rather be implicated in the transcription of developmentally regulated genes (Yuda et al., 2010).

ii. [ApiAP2 involved in Developmental regulation](#)

The majority of *Plasmodium sp.* genes are known to be induced in narrow windows, as gene expression is tightly linked to the development of the parasite and its transition between stages and hosts (Otto et al., 2010). *Plasmodium* shares many features of the complex life cycles of apicomplexan that we initiated in earlier parts with intermediate and definitive hosts and oocysts formation, yet unlike coccidian the cycle doesn't include any tissue-cyst forming stage.

The parasite alternates between a vertebrate host, which can be humans or rodent depending on the parasite species, and an anopheline mosquito vector, where fertilization and meiosis occur exclusively. In all other stages, the parasite divides asexually, with the multiplications inside red blood cells being responsible for the disease symptoms. In this blood stage, very

few parasites would initiate their development into gametocytes, awaiting their uptake by a female mosquito, where fertilization would take place generating diploid zygotes.

20 to 30 hours are enough for the zygotes to transform into motile ookinetes which subsequently form oocysts. Many rounds of asexual divisions would produce a mass of sporozoites which migrate to the salivary glands of the mosquito in order for them to be injected and transmitted into the next mammalian host.

Like the aforementioned Ap2-Sp, many ApiAP2s in *Plasmodium sp.* were characterized to being specifically expressed in a given stage, as well as being drivers of developmental pathways. AP2-Sp is exclusively expressed in the sporozoite stage and was reported to regulate transcription in this stage, its disruption resulted in the loss of sporozoite formation. But unlike AP2-Exp, this protein relies greatly on its DBD in fulfilling its regulatory role, as it was shown that replacing the AP2 DBD in a protein belonging to another stage, by the DBD of AP2-Sp led to the induction of the targets of this latter (Yuda et al., 2010).

AP2-Sp3 is another AP2 involved in the proper development of sporozoites, yet it acts on the level of their maturation, as its loss led to a motility defects and failure of the sporozoites to reach the salivary glands (Bushell et al., 2017).

Other examples involved in developmental control include AP2-O involved in ookinete development (Yuda et al., 2009), and AP2-L for liver stage maturation (Iwanaga et al., 2012), yet the mechanisms regulating the seldom events of sexual development of this parasite greatly occupy the community.

iii. [ApiAP2s involved in Sexual Development regulation](#)

As detailed in the earlier sections regarding the modes of transmission adopted by a parasite, a trade-off between the horizontal and the vertical modes is needed, so that the greatest chances for the fitness of the parasite are met at the least cost. The *Plasmodium* parasite is no exception to this evolutionary path, herein investing most of its energy into mass production of infectious particles, and keeping limited the transmission via the offspring, as the conversion of the parasites into the non-invasive gametocytes doesn't build up to more than 5% of the population.

The whole genome sequencing of *P. berghei* lines that lost their ability to produce gametocytes, revealed mutating SNPs occurring within a gene harboring a single AP2 domain (Sinha et al., 2014). The discovery of the PbAP2-g was accompanied in the same year by the identification of its counterpart in *P. falciparum* (Kafsack et al., 2014). The transcriptional alterations following the ablation of *PfAP2-G* were focused on early gametocytes genes, the upstream regions of which were found to be enriched by the PfAP2-G cognate DNA motif. The same study reported the link between the expression levels of PfAP2-G and the levels of

gametocyte formation, stating the heterochromatic state in which this locus is found in the majority of the parasites, with transcription getting permitted in minor cases.

Other than this protein directly auto-regulating its own expression, many observations rose describing the mode of action of this protein as well as its upstream regulatory events. The switch triggering the expression of AP2-G was thought to be adjusted by the action of AP2-G3 as the loss of the latter resulted in the decrease of the former expression (Zhang et al., 2017). This factor is thought to act upstream of AP2-G as a sensor for the environment signals in view of its double localization in the nucleus and the cytoplasm, as well as its expression in asexual parasites as well as in gametocytes.

Downstream of AP2-G, there seem to be an additional factor ensuring the maturation of the gametocytes. Disrupting AP2-g2 would nearly abolish any sign of gametocytes being transmitted to the mosquitoes (Modrzynska et al., 2017). It is thought that this protein would take charge of repressing the asexual gene expression during the differentiation process. Moreover, AP2-g2 is predicted as versatile repressor in different stages including sporozoite genes, and its repression is alleviated by the activation counter effect of AP2-G during the sexual differentiation, yet in other stages it remains unknown which antagonistic actor would be involved.

The regulation of such crucial protein function and thus of the pathway in which its involved, would generally involve multiple means and routes to achieve non stochastic outcomes and strict transcriptional signature transitions required for proper development of the parasite.

In the case of AP2-G, the gametogenesis seems to involve not only its fellows G2 and G3, but also a partner ApiAP2, namely AP2-I, which itself is regulated through post-translational modifications on residues within one of its AP2 DBD (Josling et al., 2020), as was reported for many ApiAP2 that can have their DNA binding be altered through protein acetylation (Cobbold et al., 2016). The regulation of gametogenesis exceeds the transcriptional level, as female gametocytes were shown to adopt translational repression of stored mRNAs, with no evidence for proteins being found for more than 500 transcripts (Lindner et al., 2019). This allows the parasite to guarantee a proactive production of mRNA and a just-in-time production of the corresponding proteins.

13. Epigenetic weight in *Plasmodium* species

a. On Clonally variant multigene families

The proportion of TFs to the total number of genes is suggested to be linked to the level of complexity in an organism, therefore the number of specific TFs in *Plasmodium species* would seem extremely low in comparison to other eukaryotes, especially with the requirements for a tightly regulated gene expression that are imposed by their multi-hosts parasitic life style.

60 potential TFs among 5800 genes is a small proportion to be enough for such complex expression programs (Coulson et al., 2004), leaving space for additional mechanisms based on the epigenetics means. In *Plasmodium sp.*, epigenetic regulation of gene expression holds its biggest weight in heterochromatin-based gene silencing, as in antigenic variation genes and the control of gametocyte conversion rates.

The maintenance of a singular *var* gene choice was reported to be dependent not only on ApiAP2 transcription factors but also on chromatin actors, underlining an adapted epigenetic signature which is built on one hand, on the heterochromatic inheritance relying on the deacetylating action of the histone deacetylase 2 PfHda2 on the H3K9 residue (Coleman et al., 2014) allowing for its methylation by PfSETvs/2 (Ukaegbu et al., 2014), and the concomitant recruitment of the proteins heterochromatin protein 1 (PfHP1) for gene silencing (Flueck et al., 2009).

On the other hand, the one active *var* gene promoter would be bound by the histone acetyltransferase PfMYST (Miao et al., 2010), the action of which would aid in recruiting the RNA Polymerase II to this site, and the expression of this gene in particular seems to be boosted by an intron-derived long non-coding RNA (lncRNA) (Amit-Avraham et al., 2015). Two lncRNAs have been reported to be derived from a bidirectional promoter within the intron in between two exons of each *var* genes, with the sense lncRNAs, when expressed, acting as a silencer, and the antisense as an enhancer for the activation of the *var* gene in question (Epp et al., 2009).

The transcription of this gene would be poised for reactivation in the following cycle of replication, and this poised chromatin state is in part fulfilled by the H3K4 methylation action of PfSET10 (Volz et al., 2012).

Similarly to *T. gondii*, which holds remarkably few constitutive heterochromatin sites throughout its genome, a great part of the *P. falciparum* chromatin, and particularly the chromatin embedding clonally variant members of multigene families, was found to be in a poised state (Karmodiya et al., 2015).

In fact, the overall profiles of histone activation and repression marks remaining similar on active and inactive genes, as in the case of two activation marks H3K9ac and H3K4me3 being enriched within intergenic areas of both expressed euchromatic and silenced heterochromatic genes, but with their levels varying significantly (Karmodiya et al., 2015), rendering the fate of genes a subject of the winning mark of the binary ratio which is weighed by repressive marks such as H3K36me3 and H4K20me3 (Duraishing and Skillman, 2018). To note that the co-occupancy of H3K9ac and H3K4me3 was not arising from cellular heterogeneity, as sequential ChIP-assays proved (Karmodiya et al., 2015).

Moreover, it is proposed that active marks are widespread throughout the genome and that repressive marks appear to be specific for the virulence genes. For instance, H3K9me3 is devoted to marking variant antigen encoding genes, abandoning its recognized role as a

general repressive mark on heterochromatic regions, such as at centromeres where its function seems to be fulfilled instead by the enrichment of H4K31me1 (Sindikubwabo et al., 2017).

H3K9me3 plays a role in tethering the *var loci* within the nuclear periphery. The activation of a particular gene requires the acetylation of the same lysine in order for it to move out of the perinuclear centers (Lopez-Rubio et al., 2009). Yet as this mark is enriched at all families of variant genes and them being independently regulated, it was thought to be insufficient to explain the mechanisms behind their regulation, ones that, nowadays we can affirm, rely greatly on the action of the aforementioned ApiAP2s factors.

Considering the few parts of the genome being set into a constitutive heterochromatic state with most of the genes being either active or poised, and the level of chromatin compaction that is relatively looser than in other organisms with a large proportion of its genome being constitutively acetylated (Ay et al., 2015), along with this parasite lacking an orthodox H1 linker histone (Sullivan et al., 2006), one would reflect on the tightness of the epigenetic checkpoints that are providing this parasite with its great success in persisting in the bloodstream and proliferating towards accomplishing its life cycle development.

b. On Sexual development

These aforementioned chromatic checkpoints involving balanced ratios of histone PTMs are also strongly linked to developmentally controlled stage specific genes transcription. *Plasmodium* species evolved great plasticity regarding their cell fate decisions and their differentiation into the distinct life cycle stages (Waters, 2016).

It has been described that only few parasites are primed for undergoing sexual differentiation (roughly 5% of the population). It appears that features of epigenetic inheritance have been supporting the switch to the gametocyte formation with the commitment to the sexual stage occurring one replicative cycle prior to the gametocyte formation and this decision seemingly being transmitted to the next cycle.

Many chromatic actors were reported to be involved in this process such as the histone deacetylase PfHda2 (Coleman et al., 2014) shown to be controlling indirectly the frequency of switching from asexual to sexual development through orchestrating the repression of the expression of the acknowledged activator of gametogenesis PfAP2-g. PfAP2-g is also reported to be regulated by another epigenetic regulator, PfHP1, of which the knockdown resulted in the increase of gametocyte induction (Brancucci et al., 2014a), also resulted in the simultaneous activation of all *var* genes, here underlining the common mechanisms regulating the rate of sexual commitment and the virulence gene expression.

The PfHP1-bound heterochromatin was also reported, as initiated ahead, to be mobilized under the action of PfHda2 for nucleating heterochromatin formation, with the binding of PfHda2 to histone tails plausibly occurring upstream of that of PfHP1 for their deacetylation to take place prior to their methylation.

The maintenance of the heterochromatic state which is in part established by the binding of PfHP1 to H3K9me3 at PfAP2-g, gets disrupted by the action of the gametocyte development protein 1 (GDV1) in evicting PfHP1 from the methylated lysine (Filarsky et al., 2018), rendering this protein an upstream activator of sexual differentiation. GDV1 offers another example for a gene being regulated through a noncoding RNA derived from its own locus.

14. *Toxoplasma gondii* Development and gene expression

The 'just in time' concept described in the gene regulatory patterns of *Plasmodium falciparum*, is one shared with *T. gondii*. About 3000 out of the 8322 protein-encoding genes display cyclical profiles with two major transcriptional waves consisting of a subset of genes with a maximum of expression at G1 and the function of which is metabolic-related, and genes that would peak at S/M stages being involved in daughter budding and egress (Behnke et al., 2010).

However, the regulation of genes is broader than its dependency on the cell cycle phasing, with a great fraction of genes being stage-developmentally regulated albeit a scarcity of knowledge of the molecular events governing their expression dynamics.

The ahead delineated ApiAP2 proteins hold a layer of control of such events, yet many evidences point out to additional actors based on the epigenetics means. As described in *Plasmodium sp.*, the weight of such mechanisms is underlined mostly by the paucity both in numbers and variety of specific TFs, relatively to the high number of protein-encoding genes.

Moreover, a high density of genes is observed in *T. gondii* with many of them displaying overlaps in the 5'UTR of one gene and the 3'UTR of another, letting one ponder about the actors involved in the accessibility to the genes in question and their tight regulation, especially when taking into account that physical proximity doesn't appear to withhold a big weight in the matter with some genes having less than 1500 bp apart, displaying distinct differential expression patterns (Kissinger and DeBarry, 2011; Lyons et al., 2002).

Also, a scattered distribution between chromosomes is observed for not only development-specific genes, but also genes encoding proteins involved in similar biochemical pathways, such as the families of secreted proteins (Bunnik et al., 2019), as previously described regarding the genome of *T. gondii*.

The complexity of this parasite's gene regulation programs is greatly dependent on its previously described multi-host life style, and is regarded as being highly established, at least in part, and is the one we're going to focus on for now, at the transcriptional level.

Prior to the development of the genome wide RNA-sequencing, a limited set of developmentally-regulated genes had been identified, utilizing techniques based on small microarrays and EST analysis (Manger et al., 1998), with the gene sets mostly belonging to the

tachyzoites and bradyzoites stages and very few from oocysts, mainly due to limitations in accessing the cat enteric sexual and asexual developmental stages.

a. The Merozoite Stage

The *in vivo* access difficulty was behind the merozoite being the least studied developmental stage. This rapidly dividing asexual form is at the onset of sexual development and of gametes generation. The high expansion rate is shared between this form and the tachyzoite, however the former remains confined to the cat enterocyte monolayer whereas tachyzoite retains its exclusive ability for invading a broad range of cell types in a variety of hosts. Distinct expression profiles have been identified between these two forms, underpinning their unique adaptation to the tissues they develop within, as well as to their subsequent developmental fates (Hehl et al., 2015).

A merozoite transcriptional profile was generated using multiple cat infected-enterocytes. Applying an 8 and above fold difference a threshold of, this high-throughput RNA-seq based study identified merozoite and tachyzoite stage exclusivity of 312 and 453 genes, respectively.

Except for some fine tunings and few adaptations, most probably caused by the enterocyte's low oxygen tension, there doesn't seem to be great differences in expression of genes related to metabolism, noting that the position of the parasite close to the stream of nutrient of the host offers an advantage in scavenging nutrients at this stage.

A large set of surface antigen (SRS, SAG1-Related Sequence) proteins displayed a remarkable stage-specific expression, with 52 members out of 111 being present in merozoite, versus 14 being exclusive to tachyzoite (Hehl et al., 2015). These SRS proteins are reported to be involved in host-parasite interactions, namely in the attachment process and the promotion of host immune responses (Jung et al., 2004). It is also suggested that the merozoite-specific set of SRS proteins could be involved in the gamete development process.

The majority of dense granule- and rhoptry-resident proteins, recognized for their remodeling of the host immune responses, as well as their MYR1 transporter (Franco et al., 2016), are not expressed in merozoite (Hehl et al., 2015). The absence of ROP16 and GRA15 proteins suggests that pro-inflammatory signaling pathways such as NF- κ B are not targeted at this stage, yet it can't be excluded that paralogs specific to merozoite might be performing GRA and ROP related functions. Also, the parasite requires some level of migration through the intestinal contents following their egress within the gut, hence there should be proteins fulfilling the motility role of the seemingly absent conventional microneme proteins. It must be noted that the residence of the parasite in its merozoite form in the enterocytes is relatively short, hence a lower need for a drastic modulation of the host cell.

The small intestine would also serve as a setting for the merozoite to differentiate into gametes. In fact, the transcriptome of the aforementioned merozoites fraction might itself include minor inputs from sexual stages (Hehl et al., 2015).

b. The Enteroepithelial stages and Gametogenesis

However, another survey focused on uncovering the genes that are exclusive to micro and macrogametes, with samples taken from the small intestine epithelium of infected cats (Ramakrishnan et al., 2019). Their comparative RNA-seq approach pinpointed genes encoding for proteins with probable functions in sexual reproduction and fertilization, such as HAPLESS 2 (HAP2) (Liu et al., 2008). Five EnteroEpithelial Stages (EES) were set to group genes belonging to different points of developmental progression going from EES1 representing very early, to EES5 representing very late (Ramakrishnan et al., 2019).

The technical variabilities caused by the asynchronous nature of coccidian development set the last stage as the most resourceful for an accurate identification of the molecules most involved in gametogenesis development. This last stage is also at the onset of oocyst development initiation, and it is noteworthy that proteins involved in oocyst wall formation seem to be expressed and stocked already in macrogametocytes (Ramakrishnan et al., 2019) (Walker et al., 2015), recalling events of translational developmental control seen in *Plasmodium species* (Lindner et al., 2019).

c. The Oocyst and Sporozoites

Consistent with such observations, are overlaps detected in an Affymetrix microarray-based analysis of oocyst stage-specific genes (Fritz et al., 2012). They report the presence of RNA of genes involved in the prior feline enterocyte developmental stages such as meiotic recombination genes, thus it was suggested that meiosis might not be fully accomplished at the time when oocysts are shed, yet this also reflects the technical limitations of such assays measuring the RNA abundance and not nascent RNA creating this gap between the active transcription taking place and the residual transcripts detection.

However, the transcriptome becomes more robust as the oocyst proceed in their development (from day 0 to day 10). The initially noninfectious oocysts, when exposed to warmth and oxygen, would form two sets of four mature haploid and infectious sporozoites. The transcriptomic analysis provided data that were relevant to the requirements for the maturation and robustness of oocysts and to the infectivity of sporozoites.

In fact, it was reported that formation of sporozoites, was in alignment with the increase expression of secreted and surface-antigen proteins, such as SRS28 or SporoSAG (Radke et al., 2004), with rhoptries and micronemes being absent from immature oocysts (Fritz et al., 2012), consistent with them being needed for the moment sporozoites would encounter their host cell that is represented solely by the intestinal epithelial cells to be invaded.

As SRS are recognized for their immune provoking abilities, it was suggested that they might also play a role in the shedding and dispersal of oocysts through the stimulation of intestinal inflammation and diarrhea. Oocysts can persist for months, if not years, in the environment before being ingested, thus their need for an energetic supply that seems to be fulfilled by amylopectin granules (Guérardel et al., 2005), which were also reported to ensure the viability of sporozoite in the fellow coccidian *Cryptosporidium* (Jenkins et al., 2003). This relative dormancy in the environment must be followed by a rapid availability of proteins needed for restarting a competent infection in the new host, which was reflected by the up-regulation of many ribosomal protein encoding genes in the late stages of oocysts development (Fritz et al., 2012).

d. The Bradyzoite stage in intermediate hosts

As described when delineating the heteroxenous life of *T. gondii*, this parasite has the ability to divide and survive by means of horizontal transfer. Yet, for sex and thus genetic diversity to occur, the life cycle must be completed in both its intermediate and definitive hosts. Thus, the development of mature oocysts containing infectious sporozoites invade their next intermediate host, which would trigger immune responses inducing the encystment of the parasite in the brain and muscle cells, and their lifelong persistence. This immune control is at the onset of the duration of infection of the intermediate host, with this persistence being crucial for the subsequent ingestion by a feline and the occurrence of sexual development discussed ahead.

It must be noted here that a considerable weight would lay on the aforementioned intermediate host, considering that a host that could not be ingested by a feline would not grant the parasite much developmental purpose (Gazzinelli et al., 2014).

Therefore, most studies delineating the development aspects of the cyst stage would rely on mice studies, with these latter not merely serving as an *in vivo* model, but as the most evolutionarily significant host for *T. gondii* that must balance its virulence in order to allow for a proper immune response permitting its encystment (Lilue et al., 2013).

Hence the choice of the parasite strain in such studies is critical, as highly virulent strains would kill laboratory mice which could not resist the acute infection phase. This rose questions about the evolution of such strains and whether they lost their ability for differentiation within mice, until it became clear that wild mice species exist that have evolved their immune systems in an adapted way to respond to such highly virulent strains (Lilue et al., 2013).

Therefore, in most *in vivo* studies of bradyzoites development, the low virulent type II strain would be used and not the type I that present highly virulence for lab mice and doesn't provide great representation of this stage.

A recent *in vivo* transcriptomic-analysis-based study used RNA collected from forebrains of mice infected with type II strains of *T. gondii* and presented an innovation in not purifying

parasites from the brain tissue (Pittman et al., 2014). The transcriptional progression from acute (10 dpi) to chronic (28 dpi) stages of infection was followed. An interesting observation was the parasite downregulating tachyzoite specific sets of SRS markers and upregulating one set specific to the bradyzoite. Microneme proteins were abundant enough to suggest events of active invasion occurring during this stage, once thought to be fully dormant. Evidence of the relative dynamism of these cyst forms has been detected, on their replicative abilities as well as on their interaction with their hosts and the active gene transcription in correspondence with the required responses to the host environment.

For instance, many proteins have been previously detected in this stage with roles in the metabolism of oxygen radicals and DNA repair, answering to the long-term exposure to the host reactive metabolites (Manger et al., 1998). Nevertheless, the parasite still rely on a storage of amylopectin granules in its tissue persistence endeavor, with further changes in its carbohydrate metabolism being evidenced in its expression of bradyzoite-specific isoforms of lactate dehydrogenase (LDH2), and enolase-1 (ENO1) (Denton et al., 1996), which have been detected in early studies based on cDNA libraries from brain cysts (Yang and Parmley, 1997) (Yahiaoui et al., 1999).

Such markers, along with cyst wall markers, as the DBA (Dolichos biflorus agglutinin) that presents high affinity for the acetyl galactosamine within the wall structure (Zhang et al., 2001), served as detection tools at the onset of the studies of cysts development in tissue cultures.

15. Signaling through Chromatin and Epigenetics in *T. gondii*

Taken together, these studies underline the statement initiated ahead, regarding the weight of gene transcriptional regulation on the complexity of the developmental program, with differentiation being, at least in part, regulated at this level. I briefly mentioned how such tight control of stage specific transcripts is thought to be established via epigenetic means, hereafter the humble progress made into understanding such mechanisms must be elucidated, with evidence pointing toward a combinatorial action gathering many elements including mainly ApiAP2s, histone modifications, and chromatin actors.

Apicomplexans share a common histone structure with other eukaryotes, with a conserved globular domain (Nardelli et al., 2013), and it is not surprising to witness them displaying common chromatin architectures as the packaging of eukaryotic genetic information into chromatin was an innovation already led by LECA, allowing its nascent eukaryote genome to expand and to gain new genes and paralogs (Omodeo, 2010).

The packaging of chromatin in *T. gondii* holds its peculiarity, in part because of the suggested lack of a conventional H1 linker protein (Sullivan et al., 2006), and the absence of the common

detected DNA methylation (Gissot et al., 2008). The 3D structure of the chromatin shows very few clustered regions mostly at centromeres as previously mentioned, and no apparent topological domains were detected, except for relatively close contacts within chromosomes (Bunnik et al., 2019).

In fact, little knowledge is achieved regarding the establishment of chromatin territories, namely of the few heterochromatin regions. The conservation of the CenH3 variant in this parasite allowed to visualize its centromeres and to assess the putative enrichment of other epigenetic marks, hence H4K20me3 was detected at these areas (Sautel et al., 2007). H3K9me3 was enriched at centromeres as well as other constitutive heterochromatin regions (Brooks et al., 2011), unlike *Plasmodium falciparum* which as mentioned ahead, has privileged this mark to only the subtelomeric regions with embedded clonally variant genes (Lopez-Rubio et al., 2009).

The observed abundant lysine acetylation points out to a dominant transcriptionally active state of the chromatin. The weight of the histone acetylation state on the normal development of parasites preceded the studies linking epigenetics to transcriptional control. In fact, the apicidin cidal effect on apicomplexan parasites as *T. gondii* and *P. falciparum* was demonstrated back in the nineties, before its characterization as a histone deacetylase inhibitor (Darkin-Rattray et al., 1996).

It is increasingly accepted that the chromatin of *T. gondii*, like *Plasmodium sp.*, that was once thought to be generally active, is in many cases in a poised state, hence the preponderance of acetylation has its promoting effect on transcription getting fine-tuned by the co-enrichment of repressive marks. This bivalent chromatin sets straight the controversies of the early reports of acetylation and active marks on genes with no detectable expression.

In *T. gondii*, the co-enrichment was reported between H3K14ac and H3K9me3 (Sindikubwabo et al., 2017). This last mark was not associated in other organisms with genes in poised states, but with constitutive heterochromatin (Nakayama et al., 2001), yet most of the genes embedded in such chromatin state in the parasite were found to be developmentally regulated (Sindikubwabo et al., 2017), in alignment with similar bimodal patterns being enriched at inactive promoters in mouse embryonic stem cells, pointing up that in stem cells the reported combinatory consist of H3K14ac/H3K27me3 (Karmodiya et al., 2012), or of H3K4me3/H3K27me3 (Matsumura et al., 2015).

To note that the PRC2 catalyzed H3K27me3 modification, that is recognized for its role in the cell type-specific repression of genes (Lee et al., 2006b), is an absent mark from *T. gondii* and *P. falciparum*, as well the case for EZH2, in accordance with an apparent lack of enhancer elements in these organisms (Toenhake et al., 2018).

This poised state of chromatin grants great plasticity for an organism as *T. gondii* which must answer rapidly to the changing environments in which it finds itself, by achieving the distinct protein requirements for each developmental stage.

One way through which this state can achieve its poised fate is through the interference with the RNA machinery elongation, as the case for H4K31me1 enriched *loci* (Sindikubwabo et al., 2017), or with the promoter proximal pausing of RNA Polymerase II, wherein developmentally-regulated promoters could be loaded with Pol-II and awaiting their activation (Williams et al., 2015). Another way would be for this state to interfere with the recruitment of chromatin writers and readers to the genes in question. The crosstalk between acetyl- and methyltransferase complexes could produce allosteric effects for each other. For instance, the acetylation of lysine 18 on Histone H3 was often detected along with a methylation of the adjacent arginine 17 (Saksouk et al., 2005).

a. Methyltransferases

Before delving into the acetyltransferase responsible for the first mark, it must be noted that *T. gondii* displays an expanded repertoire of methyltransferases, especially when compared to yeast or to its fellow apicomplexan *P. falciparum*. *T. gondii* carries 22 SET-containing KMTs, 2 Disruptor of telomeric silencing (Dot)-related enzymes all targeting lysines, and 5 PRMT-like enzymes targeting arginines, with both families reported to withhold functions in gene expression in this parasite.

TgCARM1 is the one behind the methylation of H3R17 in *T. gondii*, with the substrate specificity being shared with its human counterpart, and the enrichment being detected at promoters of active genes (Saksouk et al., 2005).

Also, amongst the five, is PRMT1, which catalyzes the methylation of H4R3 and was thought to be involved in gene regulation through this histone modification (Saksouk et al., 2005). Yet this protein was reported to also act at the level of the parasite cell division and centrosome biology (El Bissati et al., 2016). It seems rational to ponder on the expanded roles of methyltransferases in this parasite, in accordance with their distinct duplication, also notwithstanding the correlated expansion of demethylases in this parasite. In fact, the orthologs of PRMT1 in yeast and humans are nuclear proteins, whereas it is detected primarily in the cytoplasm in *T. gondii*, which is consistent with that most of its substrates are non-histone proteins.

As mentioned ahead, *T. gondii* carries within its genome 22 duplicates of the Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) domain, which is shared between lysine methyltransferases, with many enzymes presenting structural homology to the Set8, that was once thought to be exclusive to metazoans (Sautel et al., 2007). It is believed that the ancestor of *Apicomplexa* has acquired their SET subfamily from their vertebrate hosts through lateral

transfer, however some evidence points out the unlikelihood of such event based on the time of appearance of mammals, and argues in favor of the apicomplexan parasitizing nematodes and acquiring this protein (Kishore et al., 2013).

Regardless of its origin, TgSet8 displays a unique ability to catalyze not only the monomethylation of H4K20, as the case for metazoans, but also its di and tri-methylation (Sautel et al., 2007). This broader substrate activity seems to be derived from a single amino acid gain-of-function mutation within the substrate specific domain. This methylation mark, was as mentioned ahead found enriched within centromeres, but also was suggested to play an indirect role in the development towards the bradyzoite stage through its role in maintaining a proper cell cycle progression (Sautel et al., 2007).

Methylation and acetylation are independent yet complementary in their action. The first statement can be observed at the level of proteins in eukaryotes which tend to not display both a methylase and an acetylase domain, nor a methylase and a demethylase and same for acetylation, suggesting a relative modification stability.

However, despite their preference of acetylases co-occurring with bromodomains, and methylases with PHD fingers, eukaryote proteins still display fusion events between a methylation mark writer and an acetylation mark reader and the other way around (Iyer et al., 2008a).

Such crosstalk underlines their complementary process. Examples were identified in apicomplexan, such as the SET1 protein in *T. gondii*, and its homologue in *P. falciparum*, which both seem to gather their lysine methyltransferase activity, with a number of PHD finger domains, as well as an N-terminal bromodomain (Jeffers et al., 2017). Nevertheless, it is suggested that such protein fusions, even the simple writer/reader combinations, evolved earlier for acetylation processes than for methylation (Iyer et al., 2008a).

b. Bromodomains and Acetyltransferases (focus on GCN5)

The critical weight of acetylation on gene regulation has been tackled in apicomplexans, as well as in kinetoplastids which seem to be relying on histone acetylation in regulating their polycistronic transcripts (Thomas et al., 2009).

Non-histone proteins were reported to be subject to acetylation in both *T. gondii* and *P. falciparum* in different subcellular compartments, yet the best characterized acetylation-mediated protein-protein interaction would consist of the recognition of histone tail lysine acetylation by corresponding readers as the bromodomains that hold a layer in the gene expression regulation process.

The Bromodomain was first described in *Drosophila* within a SWI-SNF ATPase homologue with gene expression activation role (Tamkun et al., 1992). Seven proteins in *P. falciparum* were found to hold such domains, against twelve in *T. gondii* with the first identified bromodomains being held in the acetyltransferase GCN5 protein family. Members of the GCN5

acetyltransferases superfamily can be traced back to LECA (Iyer et al., 2008a), and homologues have been identified in a wide range of Eukaryotes. Apicomplexans possess homologues of the main HAT classes, including GCN5 and MYST, but no presence of the mammals PCAF family, was detected. Considering that invertebrates and other apicomplexans harbor one copy of GCN5, it is safe to say that *T. gondii* display a special feature by carrying a double isoform of the protein (Bhatti et al., 2006). To note that two GCN5 acetyltransferases are also detected in its close relative *Neospora caninum*.

TgGCN5a and TgGCN5b don't show signs of complete redundancy, with only few histone acetylation specificities being shared between the two. Functional studies also suggested their involvement in different processes with GCN5a thought to be required for appropriate responses to alkaline stresses (Naguleswaran et al., 2010). Their functional distinction is in correlation with different substrate specificities, as TgGCN5a was reported to hold high selectivity for H3K18, a mark mediated by a PCAF protein in humans (Daujat et al., 2002), while GCN5b presents a more prototypical repertoire of lysine targets, with its ability to catalyze the acetylation on H3K9, K14 as well as K18 (Bhatti et al., 2006; Saksouk et al., 2005), noting that it was assigned as the HAT for a core histone acetylation of H4K31, with this latter enriched at active genes promoters (Sindikubwabo et al., 2017).

Both GCN5s in *T. gondii* possess one C-terminal bromodomain and one KAT domain, similarly to the one GCN5 present in *P. falciparum* (Fan et al., 2004). It is noteworthy that the bromodomain of this latter share a higher level of similarity with that of TgGCN5b than that of the similarity shared between the two isoforms (Jeffers et al., 2017).

The *P. falciparum* GCN5 was first studied in view of the yeast homologue, where they brought proof of homology of their HAT domains, underlined by successful complementation experiments of the yeast mutated HAT by the plasmodium counterpart, marked by a conservation of substrates specificity at lysines 8 and 14 of H3 (Fan et al., 2004).

However, the divergence between the two proteins was witnessed at the level of their interaction with the corresponding ADA2 coactivator proteins where no complementation proved functional. In *T. gondii*, GCN5b carries an ability to interact with either one of the two identified ADA2 homologs, while GCN5a associates with only one (Bhatti et al., 2006).

The specificity and distinction in the additional interactors' repertoire of the apicomplexan GCN5 proteins is suggested to occur at the level of their N-terminal sequences, which display high levels of divergence. This protein-protein interaction mediating N-terminal extension is one that is absent in most early eukaryotes, and present in mammals but with lack of the variation aforementioned in apicomplexans (Dixon et al., 2011).

Functional characterization of GCN5b suggests its involvement in differential gene expression control, with focus held onto the range of interactor proteins and multi-proteins complexes, as the one gathering this protein with the aforementioned arginine methyltransferase CARM1 (Saksouk et al., 2005). The fact that many of the substrates of GCN5 are non-histone proteins (Jeffers and Sullivan, 2012), increases the amount of proteins detected in biochemical assays and thought as solid interactors. Nevertheless, GCN5b is reported to interact with different ApiAP2s (Wang et al., 2014), plausibly to fulfill a certain level of gene targets specificity as GCN5 KATs lack DNA binding domains, as well as to attain a fine-tuned regulatory activity.

Such adjustments are also established through the acetylation balance at same sites, generated from the antagonistic activities of KATs and KDACs.

c. The evolutionary history of HDACs in the phylum

It is suggested that at least one HDAC related protein was present in LECA (Iyer et al., 2008a), with phylogenetic evidence pointing towards the HDAC/RPD3, along with the Sir2 superfamily, being derived from prokaryotes proteins mainly involved in metabolic processes (Leipe and Landsman, 1997).

Copies of the three classical HDACs are found in apicomplexans, with many of their features being shared with fellow Eukaryotes and others that seem to be non-canonical and unique to the phylum and to their close ancestors.

The Silent information regulator (Sir2) superfamily is represented in most apicomplexans by two Sirtuins A and B. These proteins are recognized to have expanded their functions broader than their regular activities of histone deacetylase and ADP-ribosyl transferase, to be involved in fatty acid metabolic processes, as seen in *Plasmodium falciparum* (Zhu et al., 2012), where they also evolved redundant functions related to the regulation of antigenic variation, a process delineated earlier.

Some remarkable fusion events can be witnessed through these proteins, as a patchwork can be seen, for instance on PfSir2A, between its first HDAC domain being analogous to bacteria, and the second sharing features with plants HDACs. A similar pattern can also be seen on the *P. falciparum* gametogenesis regulator PfHDA2 (Kanyal et al., 2018).

The roots of PfHDA2 can be traced earlier than the apicomplexans emergence, to reach extents as early as the divergence of alveolates. Along with its *T. gondii* homologue HDAC1, these proteins display an alveolate-phylum-specific adaptation consisting of a unique combination of the HDAC domain with an inositol polyphosphate multikinase (IPMK) domain at their C-termini.

To note, that the protein displaying the highest deacetylation domain homology to TgHDAC1, from the fellow alveolate and photosynthetic species of Chromerids, wasn't detected as presenting such co-occurrence, which could have been lost later in the evolution of these species.

Another unusual family of class II histone deacetylase has been detected in apicomplexans, the proteins of which display a distinctive version of HDACs containing N-terminal ankyrin repeats. *T. gondii* possesses one protein presenting such aspects, TgHDAC5, which seems to share these properties with HDAC5 of *P. falciparum* as well as with *C. parvum* CpHDAC3. One copy can also be detected in *Chromera*, yet it seems that this co-occurrence can be traced to primitive chlorophyte algae, as well as haptophytes (Iyer et al., 2008a; Rider and Zhu, 2009). However, no orthologous genes were apparent in rhodophyte algae and higher plants, adding an additional compelling layer to the attempts of uncovering the evolutionary traces of apicomplexans.

These repeats widely known to be involved in mediating protein-protein interactions, were characterized in *C. parvum* as being essential for the deacetylating activity of CpHDAC3 (Rider and Zhu, 2009).

An apicomplexan-phylum-unique feature is pretty conserved amongst its species, pointing at its involvement in specific essential pathways, and it consists of a family of HDAC proteins displaying an Alanine-Threonine insertion at the catalytic pocket within the second catalytic pocket loop (Bougdoor et al., 2009). To note that *C. parvum* presents a homolog of this HDAC with a TT insertion instead of an AT (Figure 6). Tracing this insertion to *Chromera sp.*, one would detect three enzymes with homologous HDAC domains, and with AT, QT, GT insertions. Hence, it seems that the T residue holds the most critical weight in this evolutionary adaptation, as will be delineated ahead regarding the *T. gondii* homologue HDAC3.

16. *T. gondii* HDAC3

a. A nuclear-resident HDAC with sensitivity to cyclic tetrapeptides inhibitors

T. gondii maintained 5 HDACs, most of them marking their early emergence and conservation, namely the aforementioned TgHDAC1, TgHDAC5 and TgHDAC3.

At the time these enzymes were identified, the realm of studying acetylation impact on gene regulation in *T. gondii* was focused on acetylases as GCN5, probably because of the technical difficulties that can occur due to the dynamic nature of the deacetylation process, as in most cases, cross-talks occur between deacetylating a site and its subsequent methylation. However, as mentioned earlier, the weight of acetylation on the fitness of the parasite was established early on with the identification of the cidal apicidin and HC-toxin as two HDAC inhibitors (Darkin-Rattray et al., 1996).

The differentially acetylated state of genes expressed in a stage-specific manner, was observed back in 2005, when the Hakimi lab optimized the ChIP assays to adapt the technique to the use in *T. gondii* parasites (Saksouk et al., 2005). At the time, only a limited set of bradyzoite-specific genes had been identified, such as a few immuno-fluorescence available cyst wall markers, and some were identified based on Expressed Sequence Tags (EST) analysis

and *in vitro* alkaline induction methods. Moreover, most developmental stages were technically inaccessible.

They managed to map the acetylation state of H3 and H4 at putative transcription initiation sites of a set of genes that were recognized as either house-keeping, or specific to either tachyzoite or bradyzoite, and they brought primary proof that the putative cis-regulatory elements of a set of bradyzoite-specific genes, including a tandemly arrayed cluster of SRS genes, were in a hypoacetylated state prior to the alkaline-stimulated differentiation process, which became acetylated afterwards. The opposite was observed for the set of tachyzoite-specific genes, with no alterations observed at the level of the house-keeping genes (Saksouk et al., 2005).

They reasoned that this hypoacetylation would be mediated by an HDAC enzyme, obviously based in the nucleus. Their attempts to identify the localization of the 5 identified HDACs were based on generating parasites expressing ectopically recombinant tagged copies of the homologues. This allowed them to claim TgHDAC3 as the only exclusively nuclear-based enzyme. To note that such approaches can yield observations that don't mimic the endogenous reality, due to altered expression conditions, as well as possible design and fusion caused folding defects that could alter the protein targeting and its observed localization.

Nevertheless, they brought ChIP-based proof that TgHDAC3 was recruited upstream of the putative promoters of the set of bradyzoite-specific genes, consistent with their H3 and H4 hypoacetylated state. At the opposite, TgGCN5b was shown to bind acetylated tachyzoite-specific genes.

Therefore, HDAC3 seems to be the only nuclear enzyme *T. gondii* possesses, while mammals present 8 different nuclear-targeted counterparts. The sites of histones that can be modified are greatly conserved between these species, hence there seems to be a bigger weight on TgHDAC3 to fulfill diverse deacetylating requirements, and to expand its catalyzing abilities to a wider set of targets. This could explain the highly essential aforementioned Alanine-Threonine insertion that this enzyme displays at its catalytic pocket, of which the increase in capacity would allow TgHDAC3 to bind its targets with greater flexibility regarding their size and their surrounding peptidic contexts.

The wider apicomplexan-specific catalytic pocket would be behind the inefficiency of the relatively small hydroxamate-based HDACi TSA to reside in such pockets, hence its potency tests on *T. gondii* being low (Strobl et al., 2007), unless it was used at really high concentrations, which can cause metabolic instability due to its non-selective binding of other zinc-containing proteins (Kozikowski et al., 2007).

The most effective tested HDACi in *T. gondii* proved to be the cyclic tetrapeptides as FR235222, as well as HC-toxin and Apicidin (Bougdour et al., 2009), especially when compared

to short-chain fatty acids and thiol based HDACi which both displayed poor activity in inhibiting the proliferation of the parasite (Jones-Brando et al., 2003) (Hailu et al., 2017).

The FR235222 compound which has been isolated from fermentation broth of *Acremonium* species (Mori et al., 2003), brought up abnormal parasitic cell cycles of many apicomplexan species, as *T. gondii*, *P. falciparum* and *P. berghei*, as well as *N. caninum*, with this block in growth being more pronounced than when using the clinically available pyrimethamine drug (Bougdour et al., 2009).

The deacetylase inhibition activity of FR235222, was proven by the observed H4 hyperacetylation, yet the targeted enzyme was not yet established, however the nuclear setting of TgHDAC3 had placed it as the suspected target.

Forward genetic approaches were conducted to unravel the mode of action of FR235222, based on EMS-based chemical mutagenesis which generated FR235222-resistant parasites that all had distinct point mutations in HDAC3 occurring at their T99 residue (Bougdour et al., 2009). Two lineages were isolated that were resistant to the cell-cycle blocking effect of the drug, yet still presenting growth defects, one displaying a mutation of the Threonine by an Alanine, and the other by an Isoleucine, hence the resistance conferring mutations were ones diminishing the polarity of the T residue (Figure 6).

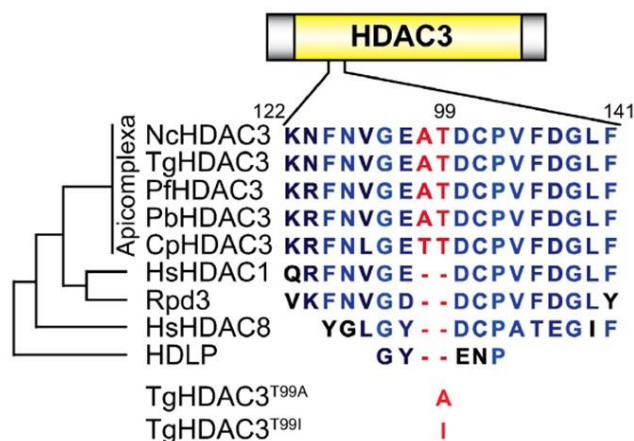


Figure 6. Sequence alignment of HDAC3 homologues in Apicomplexan parasites and other organisms.

The region (from amino acids 122–141 of TgHDAC3) surrounding the point mutation identified in TgHDAC3-resistant mutants is shown. Point mutations identified in the *T. gondii* FR235222-resistant mutants are shown at the bottom. Rpd3 and HDLP are the HDAC homologues in *Saccharomyces cerevisiae* and the hyperthermophilic bacterium *Aquifex aeolicus*, respectively. Abbreviations: Cp, *Cryptosporidium parvum*; Hs, *Homo sapiens*; Nc, *N. caninum*; Pb, *P. berghei*; Pf, *P. falciparum*; Tg, *T. gondii*. Adapted from (Bougdour et al., 2009)

It is noteworthy that the loss of T for an I residue was enough to bring about hyperacetylation at the promoter region of a bradyzoites reporter gene, based on scanning ChIP assays (Figure 7). The FR235222 treatment of the parasite induced the expression of a bradyzoite-specific SRS, as seen by immunofluorescence. This protein induction was also seen when staining parasites displaying the T99 mutation (Bougdour et al., 2009) (Figure 7).

Therefore, the T residue seems to carry the catalytic activity of HDAC3, while the A residue of the AT insertion, could play a structural spacing role for the setting of the essential T residue, which explains why *Cryptosporidium* as well as *Chromera species*, displayed insertions with differential primary residues, but with the conserved T, which seem to be selectively pressured considering, for instance, the potential involvement of TgHDAC3 in the development of the parasite and its life cycle accomplishment.

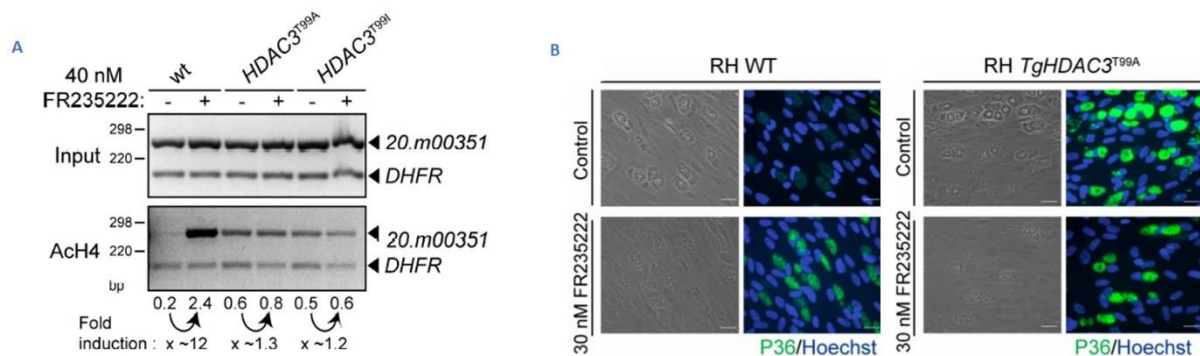


Figure 7. The T99 mutation impacts greatly the activity of TgHDAC3.

A) Scanning ChIP experiments showing the effects of FR235222 on Ach4 levels in the presence of the TgHDAC3 WT, TgHDAC3T99A, and TgHDAC3T99I alleles at the promoter region of the bradyzoite-specific gene 20.m00351. B) The T99A mutation induces the expression of the p36 bradyzoite-specific gene to the same extent that FR235222 treatment would. Adapted from (Bougdour et al., 2009).

b. *T. gondii* HDAC3 versus GCN5b, balancing stage conversions

ChIP coupled to microarray hybridization-based assays identified a set of genes regulated by the action of TgHDAC3. Almost a third of the genes displaying H4 (K5,K8,K12,K16) hyperacetylation following the FR235222-mediated inhibition of the enzyme were associated to stage-specific ESTs, with reporter genes pointing at the transcriptional regulatory nature of the inhibitory drug. Also, the inhibition of HDAC3 induced the expression of bradyzoite-specific SRS to levels comparable with the aforementioned alkaline-based bradyzoite-inducing conditions (Bougdour et al., 2009).

However, many aspects of this HDAC3-based regulation were unraveled, like the discrepancies of the gene expression observed between the mRNA and protein levels. Also, many of the bradyzoite-specific markers available at the time, were not detected. It must be highlighted that the ESTs-based libraries poorly represent *T. gondii* mRNA diversity and a great number (around 50%) of the genes predicted in this essay to be regulated by HDAC3 were not found

to be correlated with any EST, further underlining the importance of the development of the genome wide RNA based transcriptomics, delineated earlier.

The gene expression activation effect of the TgHDAC3 inhibitor was considered to be mainly through disrupting the H4 acetylation balance, yet this disruption was suggested to be limited to facultative heterochromatin regions (Bougdour et al., 2009), and that the methylation of H4K20 at constitutively silent heterochromatin (Sautel et al., 2007) might play a part in the prevention of the acetylation spread following the inactivation of HDAC3.

The parasite relying on acetylation dynamics for its developmental transcriptional transitions, grants it with an advantage on the level of its flexibility, as such processes are relatively fast, compared to methylation-based processes. The expanded poised state of the *T. gondii* chromatin has been initiated ahead as a plasticity asset for the parasite in its requirement for fast adaptations to the changing cells and hosts environments.

As mentioned earlier, this bivalent chromatin seems to be established, in part, by the co-enrichment of H3K14ac and H3K9me3. It has not been established whether HDAC3 is acting on the deacetylation of either of these sites, as most studies focused on its impact on their global levels and not on their site-specific alterations. Yet, both of these residues have PTMs catalyzed by TgGCN5b which is considered to play an antagonistic role against HDAC3 at sites like H4K31, a residue embedded within the globular domain of histone H4 at which the balance between acetylation and methylation is linked to poised state dependent gene activation (Sindikubwabo et al., 2017).

For the parasite to respond rapidly, it must employ actors that are able to interpret external signals, many of which can be of a metabolic nature. The link between HDAC enzymes and the metabolic state can be addressed from many angles.

Sirtuins and HDACs require NAD⁺ and Zn²⁺ ions for their activities, respectively, therefore they can play parts as sensors of the intracellular metabolic states. Also, Acetyl-CoA metabolism has been reported to be linearly linked to developmental processes in *T. gondii*, by not only impacting histone accessibility, but also by weighting on the activity and interactions of proteins as AP2 transcription factors and histone-modifying enzymes as GCN5b (Kloehn et al., 2020). The non-histone proteins targeting by acetylases and deacetylases in *T. gondii* has been reported by mapping their effects on proteome-wide acetylome (Hakimi *et al.*, personal communication), which extends further the complexity of the cellular regulations implicating these enzymes.

Also, in view of the evolutionary history of these proteins being derived from bacterial acetoin-hydrolyzing enzymes, their chromatin-focused regulatory roles would be considered a eukaryotic innovation. However, histones should not be placed as entities separate from metabolic processes.

In fact, only recently that the histone H3-H4 tetramer has been studied in the realm of its structural conservation in Archaea that lacks requirements for histone-based epigenetic regulation or for chromatin compaction. It is impressive to realize that the tetramer carries a primitive oxidoreductase function (Attar et al., 2020), which might have permitted a DNA-protecting, non-toxic copper utilization and transition in the oxidizing conditions which accumulated at times close to the emergence of eukaryotes (Anbar, 2008).

c. TgHDAC3 a class I HDAC co-repressor

Functional characterization of HDAC proteins have mainly been tackled in the sphere of multi-protein complexes. The involvement of this family of enzyme on many processes has been witnessed in mammals, including cell cycle progression, and cell differentiation with HDACi being placed as anti-cancer epi-drugs (Bolden et al., 2006; Haumaitre et al., 2009). The action of HDACs on non-histone targets can be exemplified by its deacetylation of the p53 family member p63 at least in the context of epidermal development, as well as of p53 itself, with HDAC1/2 acknowledged to act in some contexts through both alteration of protein expression as well as of protein activity (LeBoeuf et al., 2010).

There seem to be a high level of functional divergence between the individual HDAC of different organisms (Haberland et al., 2009), as well as little substrate specificity between individual enzymes (Riester et al., 2007), hence pointing at the weight of the co-factors in the matter, as it is suggested that the specific corepressor complexes they belong to, would control the substrate selection.

Class I HDACs are recognized to be recruited to their corresponding repression complexes through the mediation of proteins displaying ELM2 and SANT domains (Ding et al., 2003) (Lee et al., 2006a), as their lack for conventional NLS is acknowledged, thus their need for interacting with additional actors.

This class presents the most documented roles in regulation of gene expression, namely HDAC1 and 2 which are found within different multi-protein complexes such as NuRD (Nucleosome Remodeling and Deacetylation) which combines the deacetylating activity with the one of helicases (Xue et al., 1998), and as the CoREST (Co-Repressor for Element-1-Silencing Transcription factor) which combines it with histone demethylation through LSD1/BHC110 / KDM1A (Hakimi et al., 2002) (Lee et al., 2005).

The human HDAC3 was identified as a constitutive core complex constituent of the SMRT (Silencing Mediator of Retinoid and Thyroid receptors) complex. TBL1 was also confirmed to co-elute within the same high molecular weight complex, yet HDAC3 didn't seem to interact directly with TBL1 but through the mediation of SMRT (Guenther et al., 2000).

In harmony with these protein-interaction-based functional characterizations, the role of TgHDAC3 was studied also at the biochemical level. Early size exclusion chromatography-

based pull-down assays established that TgHDAC3 co-purifies with a set of proteins, together forming a high molecular-mass complex of about 1 million Daltons, namely the *T. gondii* CoRepressor Complex (TgCRC) (Saksouk et al., 2005). The robustness of the complex was underlined by the stringent purification conditions (salinity up to 500 mM KCl). To note that the Flag-tagged HDAC3 protein was one that is ectopically expressed, thus the interacting proteins would be divided between this protein and the endogenous counterpart, thus limiting the indisputable nature of the detected protein repertoire.

The complex contained homologues of proteins that have been already reported to interact with HDACs. For instance, the NLS-deficient TgHDAC3 co-purified with a protein containing a SMRT-like domain which is suggested to act on the activity or on the nuclear targeting of the enzyme. A TBL1 homologue was also identified, with which the interaction was behind naming the newly characterized enzyme at the time TgHDAC3, as it was believed that it is a type I HDAC that belongs to only one complex, unlike the human HDAC1 and 2 which are able to associate with different complexes. Yet, the 64% identity of TgHDAC3 to hHDAC1 argues against this title, as well as the fact that the pull down of TBL1 didn't include some of the polypeptides identified when purifying HDAC3, hinting to the involvement of this latter in other TBL1 unrelated complexes (Saksouk et al., 2005).

Some of these polypeptides co-purifying with HDAC3 were novel uncharacterized components and were suggested to account for the co-activation of the enzyme or the co-establishment of a transcriptionally repressive chromatin state, including ones that were named at the time TgCRC350 and TgCRC230, based on their molecular weight.

TgCRC350 was predicted to carry an NLS, yet at the time AP2 proteins weren't yet identified in apicomplexans (Balaji et al., 2005), so they couldn't identify its DNA-binding domain as the one of AP2s. It cannot be excluded that other AP2 proteins interact with TgHDAC3 in the real context, but as mentioned earlier, they were most probably taken up by the endogenous copy.

TgCRC230 was identified through SMART analysis as a protein carrying a bipartite NLS, along with six KELCH repeats and a HATPase domain, which they didn't know at the time belonged to the family of MORC proteins that were characterized few years later in Plants (Kang et al., 2008).

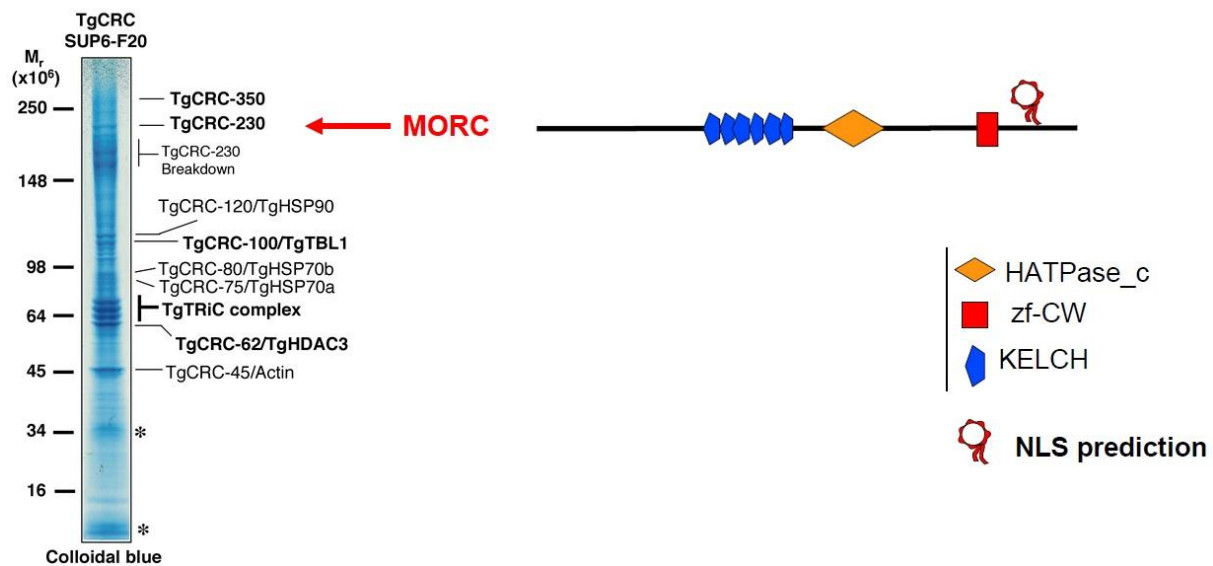


Figure 8. The identification of TgMORC co-purified with TgHDAC3.

Colloidal blue staining of a Trichloroacetic acid (TCA) precipitated Sizing fraction of TgHDAC3 chromatography purification. Representative domain architectures of *T. gondii* MORC domains. Adapted from (Saksouk et al., 2005) and (Farhat et al., 2020).

17. MORC

a. Evolution and divergence

Contextualizing MORC proteins would require first delineating their evolutionary traces, as well as their functional characterization, and in the following section to detail what is until the moment known about their mechanism of action. The attempts for finding these proteins and their homologs are based on aligning their N-terminal ATPase module. This module consists of a combination of a GHKL domain and an S5 domain.

The GHKL domain owes its name to its occurrence on different proteins, namely the prokaryotic topoisomerase related Gyrase, HSP90, histidine Kinase, and a DNA repair protein of the MutL family (Dutta and Inouye, 2000) (Kang et al., 2008). The ribosomal protein S5 domain 2-like, is usually found along with the GHKL domain, as is the case for the latter enzymes, and it seems to provide a conserved basic residue that can functionally replace the arginine fingers of phosphohydrolase reaction (Burroughs et al., 2006). Some prokaryotic homologs lack the S5 fusion but seem to possess a domain that shares a similar secondary structure as the S5 (Iyer et al., 2008b).

However, both eukaryotic and prokaryotic versions of the GHKL family contained conserved motifs required for adenosine and phosphate binding, placing the proteins harboring such structures as putative active enzymes capable of ATP hydrolysis, which is likely as the common

ancestor of these proteins appears to have been involved in the catalysis of DNA structural organization (Iyer et al., 2008b). The association between the GHKL+S5 modules and nucleases/helicases appear to have arose early in the bacterial evolution.

Proteins possessing such fusions in prokaryotes are found in bacteria that are distantly related, as cyanobacteria, proteobacteria, and archaebacteria. It was seen that the MORC homologs within the same lineage of proteobacteria do not cluster together, suggestive of lateral transfer of these genes between prokaryotes. Also, the eukaryotic proteins are thought to have been acquired through a single lateral transfer event from a bacterial source, one that belongs to eubacteria and not to archaea (Iyer et al., 2008b). One would ponder on whether the primary endosymbiosis of a cyanobacteria by the primitive eukaryote, as delineated in the first chapters, would be behind the wide distribution of the MORC proteins in most eukaryotes except fungi, in apicomplexans, as well as in plants.

The number of MORC genes differs among the lineages of plants, with the green unicellular algae possessing only one copy, while at least five members are detected in most of the monocots and three to four in the basal angiosperms. It is suggested that these proteins were subjected to distinct duplication events during the evolution of land plants (Dong et al., 2018) (Figure 9).

For the duplicate genes to be retained in the genome they usually exhibit different functional fates. For instance, paralogous gene pairs in plants species have been detected to experience transcriptional subfunctionalization, as in they would have expression patterns that are divergent and tissue-specific. Such pairs were seen in *Arabidopsis thaliana*, soybean and rice and are thought to have been subjected to different selection pressures after their duplication.

Most MORC proteins in plants were seen to exhibit higher levels of expression within reproductive tissues, as in flowers and seeds, than in other organs, suggesting their possible involvement in the development of plant reproductive tissues (Dong et al., 2018). Functional characterization in plants will be delineated ahead.

Not as many divergence events seem to have occurred in animal MORC compared to the plants counterparts, with one duplication event seeming to have occurred during the evolution of vertebrates producing multiple gene copies, as most species contained four MORC members, except for rats and mice possessing five counterparts (Figure 9).

In animals, MORC1 and 2 can be evolutionary grouped together as they differ from MORC3 and 4 by having a coiled coil domain separating their GHKL and S5 domains. This discrepancy will be further delineated along the chapter in the realm of its mechanistic impact.

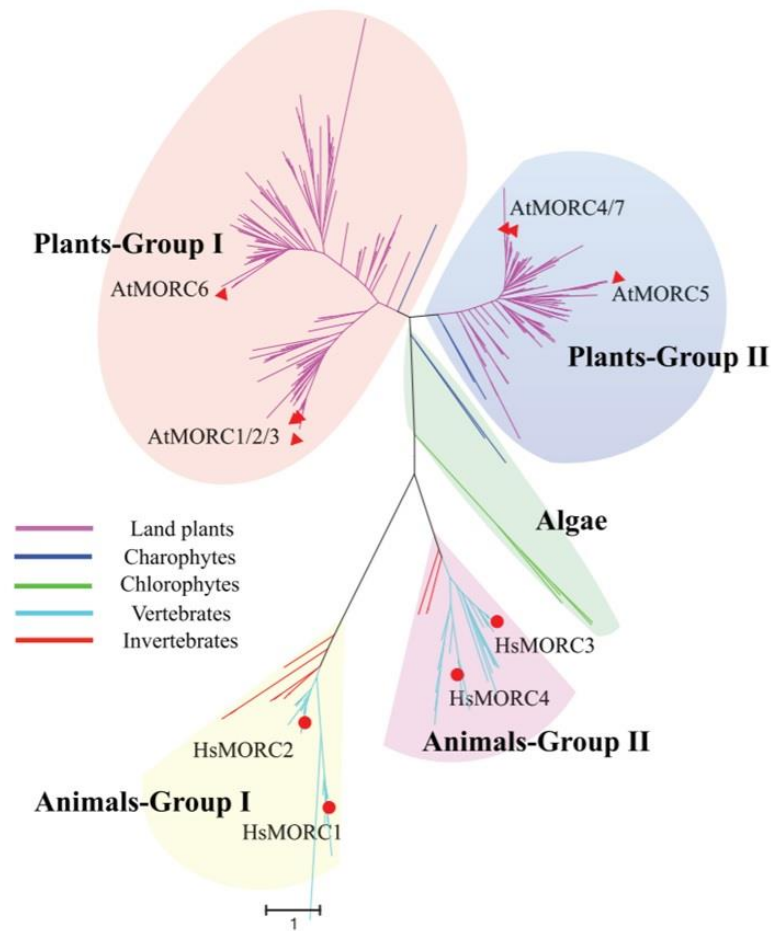


Figure 9. *Phylogenetic classification of MORC genes in plant and animal lineages.*

Plant MORCs encountered several rounds of gene duplication events in the long-term evolutionary courses, eventually generating large-scale expansion in most species. MORC genes in plants and animals can be topologically classified into two major groups, respectively, which is designed as Plants-Group I, Plants-Group II and Animals-Group I, Animals-Group II. The green algae were located in the base of these two plant groups. Seven and four canonical MORC genes in *Arabidopsis thaliana* and human were labeled with red triangle and circle, respectively. Adapted from (Dong et al., 2018)

b. MORC proteins functions

In parallel with its expression across tissues, the function of the human MORC2 has been widely characterized. It has been reported to act along HUSH to bind L1 retrotransposons, mostly the ones located within transcriptionally permissive euchromatic environment. MORC2 would act on the chromatin compaction at these *loci*, concomitant with a role in promoting the maintenance of high levels of H3K9me3 (Liu et al., 2018).

This protein was reported to act as well on the cytosolic level, as an interaction was detected between MORC2 and ATP-citrate lyase (ACLY), a lipogenesis related enzyme that catalyzes the formation of acetyl-CoA. It was shown that MORC2 is in fact regulating the activity of ACLY, placing MORC as an actor in the adipocyte differentiation (Sánchez-Solana et al., 2014).

While MORC4 is suggested as a potential biomarker for B-cell lymphomas (Liggins et al., 2007), MORC3, also known as NXP2, is involved in the induction of cellular senescence. It is reported as an activator of p53, which seems to implicate a localization-based regulation through the recruitment of this latter to promyelocytic leukemia protein (PML) nuclear bodies (Takahashi et al., 2007) (Mimura et al., 2010).

The mammalian *MORC1* holds the status of the first identified *MORC* gene. In fact, the family of *Micrororchidia* proteins owes its name to the mice deficient for *MORC1* which led to aberrant and small testicles. An insertional mutation of the gene encoding this protein which is highly expressed in male germline, was reported to cause arrest in the meiosis process, and mice holding homozygous mutations were infertile, hence assigning MORC1 with a role in germ cells development regulation (Inoue et al., 1999; Watson et al., 1998).

The *Caenorhabditis elegans* single copy of MORC was also detected as an effector in transgenerational fertility through its role in the maintenance of small interfering RNA (siRNA)-dependent chromatin organization. MORC-1 of *C. elegans* was suggested to act as a protector against the spread of euchromatic state, thus maintaining the H3K9me3-dependent repressive state of a subset of siRNA targets, and preventing disorganization and genome decondensation of germline chromatin (Weiser et al., 2017).

Similarly to its detected role in silencing repetitive transgene in *C. elegans* (Moissiard et al., 2012), the mice MORC1 was also reported to act at a similar level, regulating the repression of transposons in the germline, thus preventing their expression at such developmental stages when repression is widely relieved. The regulation of such elements plays a role in the meiotic alterations observed in infertile mice (Pastor et al., 2014).

The impact of MORC1 on transposon repression in the male germline was suggested to display similar patterns to the ones observed following the disruption of DNA methyltransferases (Aravin et al., 2008), placing it as an epigenetic regulator, the effect of which extends for longer than its own downregulation.

Despite the phenotypic similarities observed in mice deficient in *MORC1* or in *DNMTL3*, yet whole genome bisulfite sequencing brought proof of MORC1 not acting directly on the level of neither de novo nor maintenance of DNA methylation (Pastor et al., 2014). It was proposed to promote the DNA methylation of a variety of transposons in the germline, but the mechanism behind this effect was not revealed.

c. The DNA methylation discrepancy (Plant focus)

Such discrepancies of whether the mechanism of action of MORC proteins was linked or not to DNA methylation, have been argued and documented in Plants.

In 2012, Moissiard *et al.*, stated that their Bisulfite sequencing (BS-seq) detected little evidence for changes in methylation occurring following the alterations of *A. thaliana* *AtMORC1* and *AtMORC6*, at neither the global level, nor at the derepressed transposable

elements. An absence of repressive histone marks at their reporter genes was also noted (Moissiard et al., 2012).

Considering that transposons and repeats are in part regulated through RNA-directed DNA methylation (RdDM), attempts were carried to unravel whether MORC is acting on such level. This aforementioned study proposed that MORC proteins might be acting downstream of the RdDM pathway (Moissiard et al., 2012). On the other hand, a study from the same year, detected reduction in DNA methylation at their reporter *loci* following the mutation of *AtMORC6* (Lorković et al., 2012), reinforced by a following study also stating their observations of DNA methylations alterations at RdDM target sites, with mutants defective in the RdDM pathway being found to be also defective for *AtMORC6* (Brabbs et al., 2013).

It was proposed that the discrepancy might be due to locus specific effects and the choice of reporter genes, suggesting that *AtMORC6* might be influencing DNA methylation at a distinct subset of *loci*.

It was until 2016 that the Pr. Jacobsen SE's lab performed a hexuple mutant of *Arabidopsis thaliana* MORC proteins and eliminated the disparity, thus stating that the repression exerted by MORC proteins was uncoupled from DNA methylation, and that they didn't act downstream of the RdDM pathway (Harris et al., 2016). However, they recognized that the silencing of a small subset of RdDM target *loci* (about 5%) thought to be poised for reactivation, required the action of MORC for their proper DNA methylation.

The DNA-methylation uncoupled mechanism of action of MORC proteins was strengthened by the repressing activity of MORC1 from *C. elegans* which is devoid of DNA methylation (Simpson et al., 1986), supporting that the MORC proteins might be exerting their ancient silencing roles through different mechanisms, which will be detailed at once.

d. Mechanistic aspects

i. Nuclear bodies and multimers formation

AtMORC6 was suggested to repress its targets consisting of transposons, through modulating higher order compaction of pericentromeric heterochromatin, as it was observed, along with *AtMORC1*, to form nuclear bodies adjacent to chromocenters, and their alterations led to the de-condensation of pericentromeric regions (Moissiard et al., 2012).

Chromocenters in *A. thaliana* constitute chromosomal territories dense in transposons-rich pericentromeric heterochromatin (Fransz et al., 2002). Such compaction makes it that pericentromeric regions don't display many interactions with the rest of the genome, yet High-throughput Chromosome Conformation Capture (Hi-C) analyses also showed that mutating *AtMORC6* increased such aberrant connections with euchromatic arms (Feng et al., 2014; Moissiard et al., 2012).

The physical interaction between *AtMORC6* and *AtMORC1* was detected by immunoprecipitation coupled to mass spectrometry (MS) analyses, and it was suggested that

MORC proteins exist primarily *in vivo* as dimers. The heterodimerization of the two homologs was recognized, yet it was later seen that AtMORC6 also interacts with AtMORC2, the closest homolog to AtMORC1 (80% aa identity), yet in a mutually exclusive manner, which is linked to these two proteins acting redundantly at a set of Transposable Elements similar to that of AtMORC6, placing this one as the epistatic factor in these functional heterodimers (Moissiard et al., 2014).

Homodimerization was also reported in *A. thaliana*, with AtMORC4 and AtMORC7 both forming homomeric complexes *in vivo*, displaying concentrated nuclear bodies that are adjacent to and not within the chromocenters, as described earlier. These proteins were seen to be acting in a partially redundant manner, yet with AtMORC7 exerting a stronger effect, at repressing a large set of protein-encoding genes involved in pathogen response. Consistent with a gene regulatory function, and unlike the first pair described, AtMORC4 and 7 were also uniformly distributed throughout the nucleoplasm (Harris et al., 2016).

MORCs are able to form nuclear bodies, both in plants and in animals. For instance, MORC-1 of *C. elegans* was observed to localize into such heterochromatic structures (Weiser et al., 2017). Another example is the aforementioned human MORC3 recruiting p53 to PML nuclear bodies. It was suggested that MORC3 homodimerizes and that alterations in any of the protein domains would impact this process (Mimura et al., 2010).

I briefly stated that animal MORC1 and 2 display in addition to their C-terminal coiled coil domain, one that is triple stranded separating their GHKL and S5 domains (Koch et al., 2017). MORC3 and 4, as well as all other known MORC proteins display only one CC region at their C-termini (Figure 10).

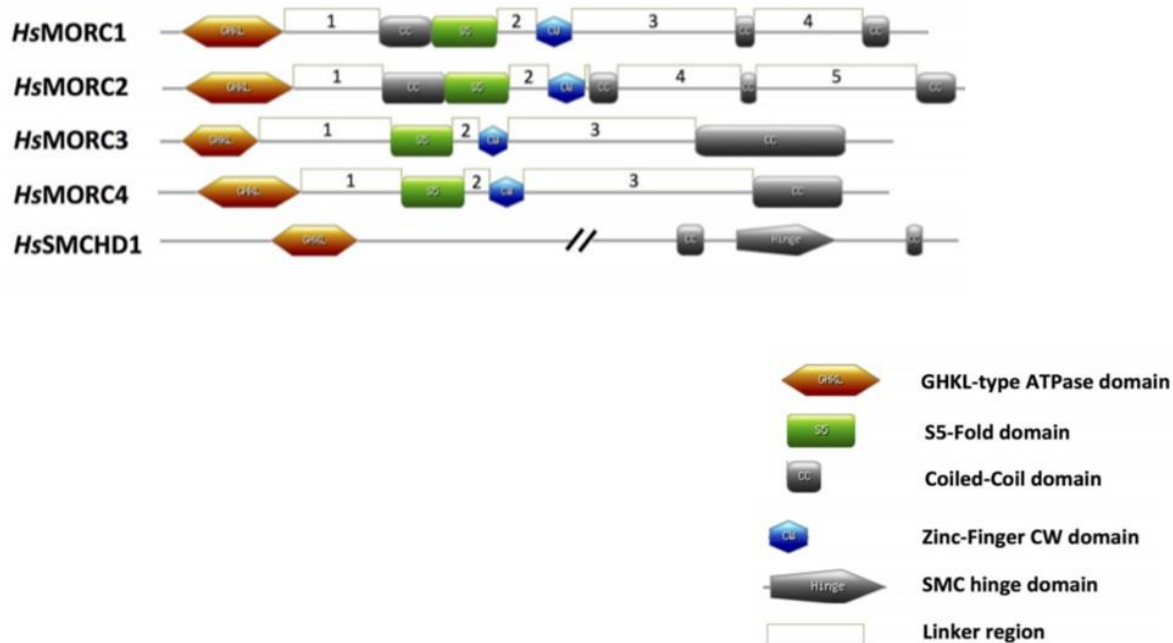


Figure 10. Domain organization of MORC family members from *Homo sapiens* (Hs).

Human MORC1 and MORC2 are different from their counterparts by their additional Coiled coil domain inserted within the ATPase module. Adapted from (Koch et al., 2017)

The function of this region was characterized in MORC3 suggesting its requirement for a proper homodimerization of the protein and thus its localization to the PML regions (Mimura et al., 2010). The disruption of such structure has been documented also in plants where its deletion suppressed the *in vitro* homodimerization of the tomato and potato MORC1 proteins (Manosalva et al., 2015).

The C-terminal coiled-coil domain is suspected to mediate constitutive homodimerization in GHKL-ATPase family members (Dutta and Inouye, 2000). For instance, ATP-free MORC3 homodimerizes through this domain, thus forming its open state structure. However, the closed state of the dimer would require ATP binding for the N-terminal ATPase domains to come together. Mutations impairing the ATP binding were seen to disrupt the nuclear body formation of human MORC3 (Mimura et al., 2010).

The CC-mediated multimerization is behind the detected modest events of MORC proteins in their dimerized state, even in the absence of ATP.

ii. ATP-dependent dimerization of ATPase modules

The team of Pr. Jacobsen SE attempted to understand the molecular behavior of MORC proteins, and to confirm their dependence on ATP for stable dimerization, in a manner resembling that of GHKL proteins (Li et al., 2016).

Their crystallography-based structure of MORC monomers in the presence of a non-hydrolysable ATP analogue showed a symmetrical dimer between the ATPase domains. Each

monomer would bind one molecule of phosphoaminophosphonic acid-adenylate ester (AMPPNP)-Mg²⁺ within their active site, and extensive interactions would form between the two, with aromatic side chain residues from one monomer making hydrophobic contact with the ones from the other (Figure 11). These residues are conserved and necessary for an active ATPase enzyme adopting a GHKL fold (Corbett and Berger, 2003).

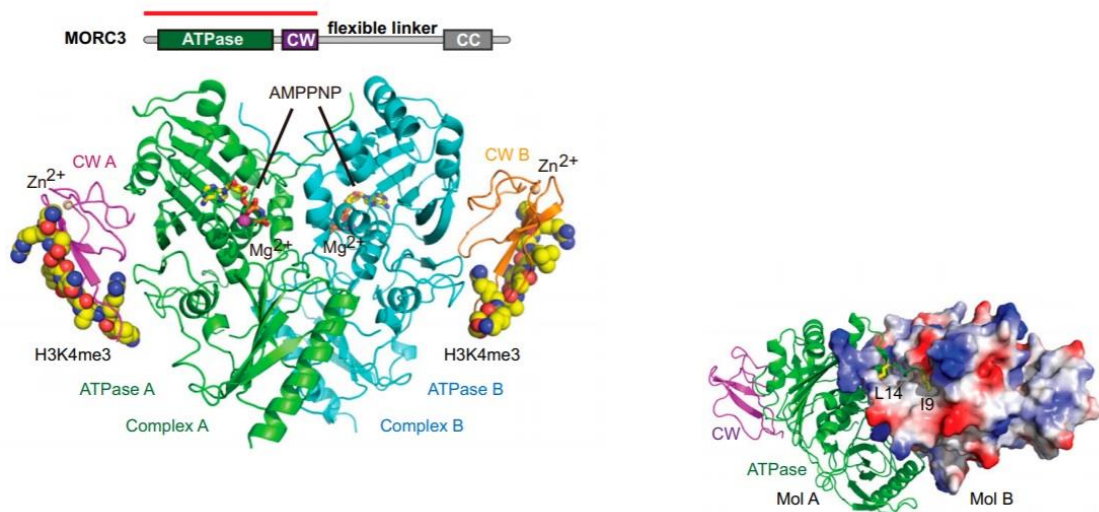


Figure 11. Structure of MORC3 ATPase-CW cassette in complex with AMPPNP and H3K4me3 peptide.

On the left: Overall crystal structure of MORC3–AMPPNP–H3K4me3 complex. The crystallized region is delineated by the red line. The ATPase and CW domains are shown in ribbon representation and colored green and magenta in monomer Mol A and cyan and orange in monomer Mol B, respectively. The bound peptides are shown in space filling representation, whereas the AMPPNP molecules are in stick representation. On the right : The MORC3 dimer demonstrating the extensive interactions between the two monomers, with Ile9 and Leu14 residues positioned within two small hydrophobic pockets. *Adapted from (Li et al., 2016)*

Hence, they confirmed that the ATPase domain of MORC3 is in a dimer state when in a complex with a non-hydrolysable ATP analog. Native mass spectrometry was used to assess the extent of dimer formation in the presence of such nucleotides.

The incubation of MORC3 with no ligand or with ATP were similar with the ratio tilting towards a higher monomeric state, yet with ATP slightly more dimers were detected. This ratio shifted strongly towards a 90% dimer formation, when the protein was incubated with non-hydrolysable ATP analogs, underlining that MORC3, and possibly other MORC homologs, use ATP for inducing their dimer formation and its hydrolysis for dimer dissolution (Li et al., 2016).

it is noteworthy to mention that isoleucine residue 9 was established in this assay, as being essential for mediating monomer-monomer contacts and stabilizing the dimer (Figure 11), as its mutation shifted the ratio in favor of a monomeric state, regardless of the nucleotide in incubation. Thus, the mechanism of dimerization of MORC resembles greatly that of other

GHKLs, as this residue seem to be functionally conserved across phyla (Corbett and Berger, 2003).

iii. The CW mediated target recognition and ATPase regulation

The canonical MORC proteins of human all possess a CW-type Zinc Finger (ZF), downstream of the ATPase module. Only a small subset of plants MORC were detected to display a similar domain (Figure 12). The crystal structure of MORC3 ATPase-CW bound to AMPPNP and in complex with a short peptide surrounding the K4 residue of histone H3, provided insights on the binding and targeting of MORC proteins (Li et al., 2016).

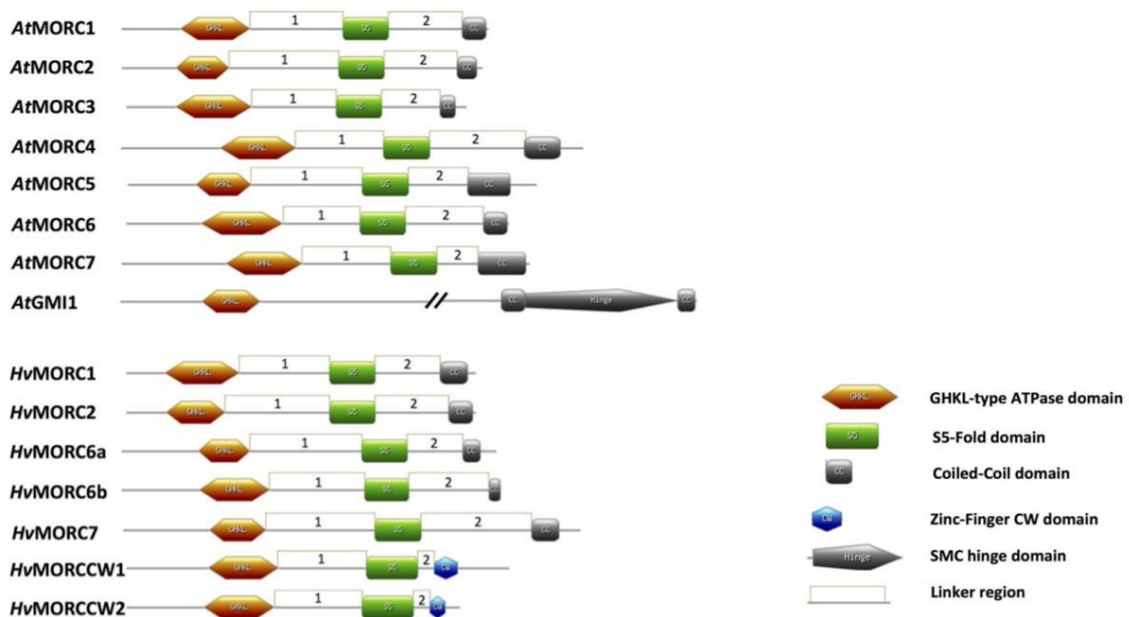


Figure 12. Schematic of MORC architecture in two plants models.

Domain organization of MORC family members from Arabidopsis thaliana (At), Hordeum vulgare (Hv) are shown. All canonical animal MORC contain a CW domain as seen in figure 10, yet only few plants display this domain. No plant MORC carry the additional CC domain within their ATPase module as the case of the human MORC1 and 2. Adapted from (Koch et al., 2017)

In fact, a study showed that MORC proteins don't seem to hold DNA sequence preferences, as it was able to bind *in vitro* to different DNA templates (Kim et al., 2019), which is consistent with plants MORC acting at transposons that are composed of a diverse array of sequences, thus its targets selection was thought to rely on chromatin.

The CW-ZF was named after its conserved cysteine and tryptophan residues (Berg et al., 2003).

Two conserved tryptophan residues make up the building blocks of the aromatic surface pocket of the CW domain, serving as the methyl lysine recognition cage (Hoppmann et al.,

2011). The right wall tryptophan is conserved in all human CW proteins including the four MORC proteins as well as ZCWPW1, ZCWPW2 and LSD2 (Figure 13), and is considered to be essential for the binding of the methylated lysine, while the floor one which is absent in MORC1 and 2, is thought to enhance the binding affinity (Liu et al., 2016).

The human MORC3 was identified as an H3K4me3 reader and was detected *in vivo* to be enriched at promoters marked by H3K4me3. The mutation of the wall tryptophan of CW was enough to disrupt the targeting of MORC3 to the chromatin (Li et al., 2016).

This domain has been documented to carry the ability to bind methyl deposited on H3K4 irrespective of the methylation state, and it was suggested that this is due to the MORC3 CW displaying a negatively charged end-wall residue, which is not conserved in other CW domains. Such residues would mediate hydrogen bonds with H3K4 when it is a low methylation state thus increasing the electrostatic attraction forces (Li et al., 2016) (Andrews et al., 2016). It is believed that the chromatin-binding mode of MORC4 would resemble that of MORC3 as they share high sequence similarity.

On the other hand, MORC1 and 2 form a subfamily of their own, one that as mentioned earlier displays a coiled coil domain within their ATPase module, in addition to an aberrant CW domain that is unable to bind H3K4 in any of its methylation states (Liu et al., 2016), suggesting that their chromatin engagement relies on elements distinct from the ones of MORC3. There must be other mechanisms that ensure the targeting of the MORC proteins that lack either a histone-binding CW, or the domain all together, as is the case for most plants MORC (Koch et al., 2017).

It must be noted that the CW domain could be forming assemblies with other homologous sub-structures as PHD and GATA-like zinc fingers (Argentaro et al., 2007). Thus, the aberrant CW of MORC2 might be able to function within such assemblies that allows for greater affinity than the one detected in isolation. Also, proteins possessing H3K4me-reading CW domains were identified in *A. thaliana* (Hoppmann et al., 2011), hinting to possible interactions or merely recruiting events occurring between the CW-lacking plants MORCs and such proteins.

Nevertheless, despite its inability to mediate recognition of H3K4me3, The CW domain of MORC2 was reported to be required for its aforementioned function at the level of HUSH regulation. The CW domain of MORC proteins seems to be establishing critical interactions with the ATPase module, as weakening these interactions altered its fine-tuned hydrolysis rates and abolished the function of MORC2 (Douse et al., 2018).

Moreover, such interactions were documented in MORC3, with CW reported to regulate the binding of the ATPase module to DNA. In fact, the ATPase activity of MORC3 was suggested to be enhanced by DNA, and the MORC3 CW was seemingly able to establish two mutually exclusive interactions, one with H3 and the other with the ATPase, where binding to the

methyated H3 would free the ATPase module and the inhibition of its catalytic activity would be alleviated (Andrews et al., 2016) (Figure 13).

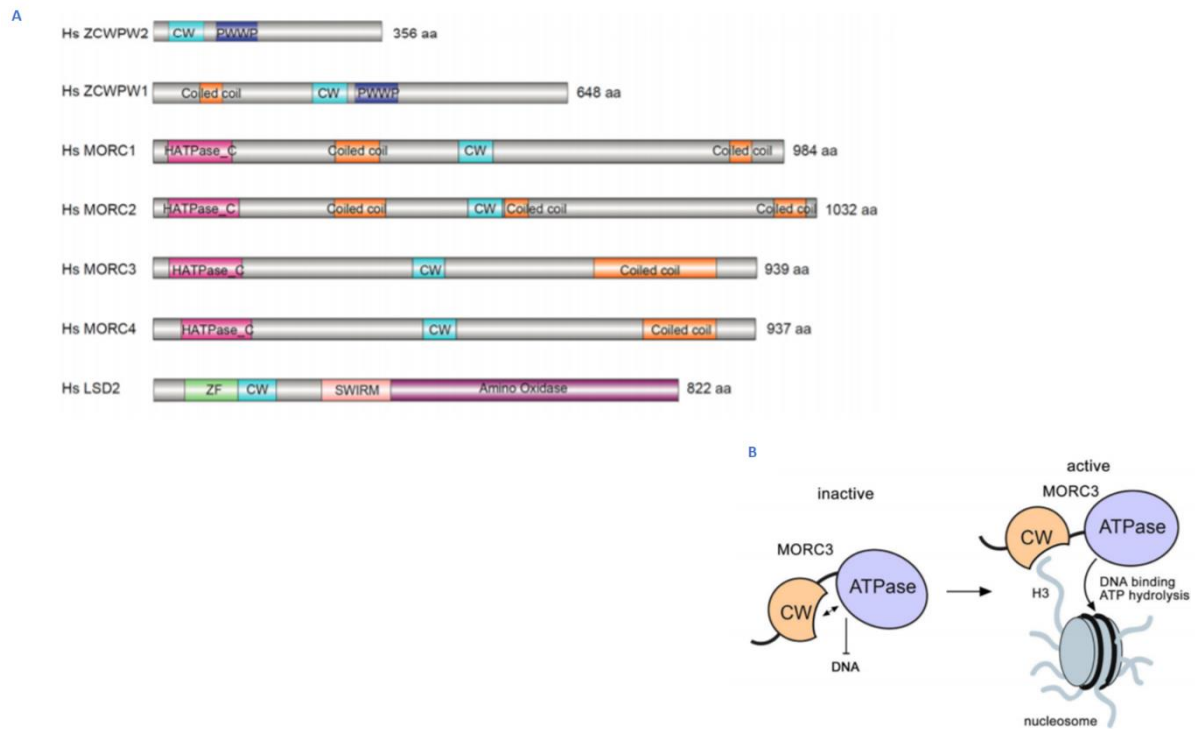


Figure 13. The human zinc finger CW domain-containing proteins.

A) Schematic representation of the domain structure of these proteins. B) CW negatively regulates the DNA-dependent MORC3 ATPase activity likely through hindering DNA binding of the ATPase domain. Adapted from (Liu et al., 2016) and (Andrews et al., 2016).

iv. DNA binding and genome compaction

After having described the aspects of MORC proteins at the level of their nuclear bodies, their ATP-dependent dimerization abilities, and their targeting to the chromatin, I will proceed into delineating what was reported regarding the ability of these proteins to act on the compaction of DNA which is in most cases behind their earlier stated functions.

The *C. elegans* MORC-1 was used as a model of a mechanistic *in vitro* study of the MORC means for compacting DNA. This protein shares a similar structure with the human MORC3, with one coiled coil domain in their C-terminus, and a CW domain downstream their ATPase module. As mentioned earlier, the disruption of this protein led to visible genome decondensation in the germline, in a DNA methylation-independent manner, as the species lack such modifications.

The proposed model of the study, in a nutshell, states that MORC-1 bind their targets via DNA entrapment, then undergo multimerization upon the binding to form intracellular condensates and topologically entrap *foci* for establishing a higher compaction (Kim et al., 2019). (Figure 14)

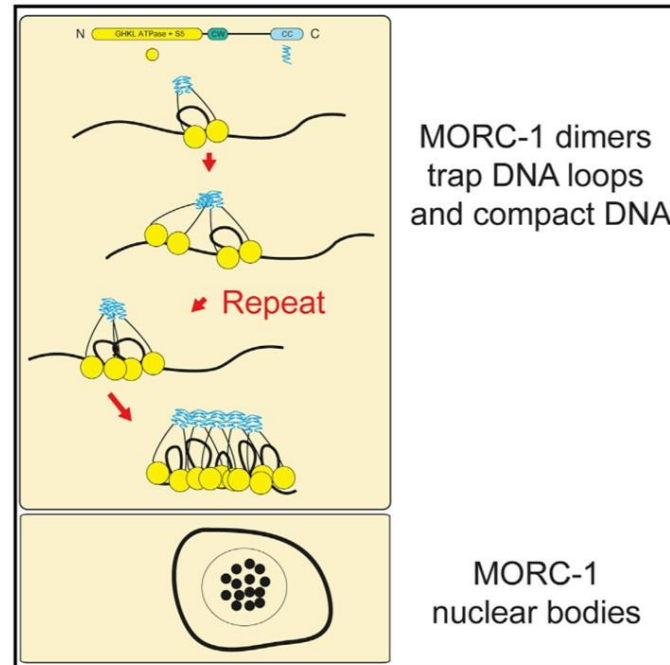


Figure 14. MORC-1 acts via a mechanism of DNA loop entrapment to compact chromatin.

Caenorhabditis elegans MORC-1 traps DNA loops. Recruitment of additional MORC-1s causes further loop trapping and DNA compaction. MORC-1 assemblages become topologically entrapped on DNA. MORC-1 forms discrete foci *in vivo*. Adapted from (Kim et al., 2019).

This study re-confirmed initial findings, regarding the slight ability of MORC proteins to dimerize even in the absence of ATP, and their exhibition of their highest dimerization rates when incubated with non-hydrolysable ATP-analogs (Li et al., 2016). However, in this assay, this was interpreted with a focus on the impact of such processes on the compaction of DNA.

Hence, it was stated that the inability to hydrolyze ATP, and therefore its constant binding, yielded the highest DNA compaction rates, as it promoted the most stability of the protein on the DNA, acting against the opening of the ATPase dimers, thus preventing the dissociation of DNA from MORC1 (Kim et al., 2019). The fact that the protein was able to compact DNA without the addition of ATP suggests that MORC-1 binds DNA through a loop-trapping mechanism and not a loop extrusion one as this latter involved directional motor activity that requires active ATP processing, as seen for yeast condensin (Ganji et al., 2018) (Figure 15).

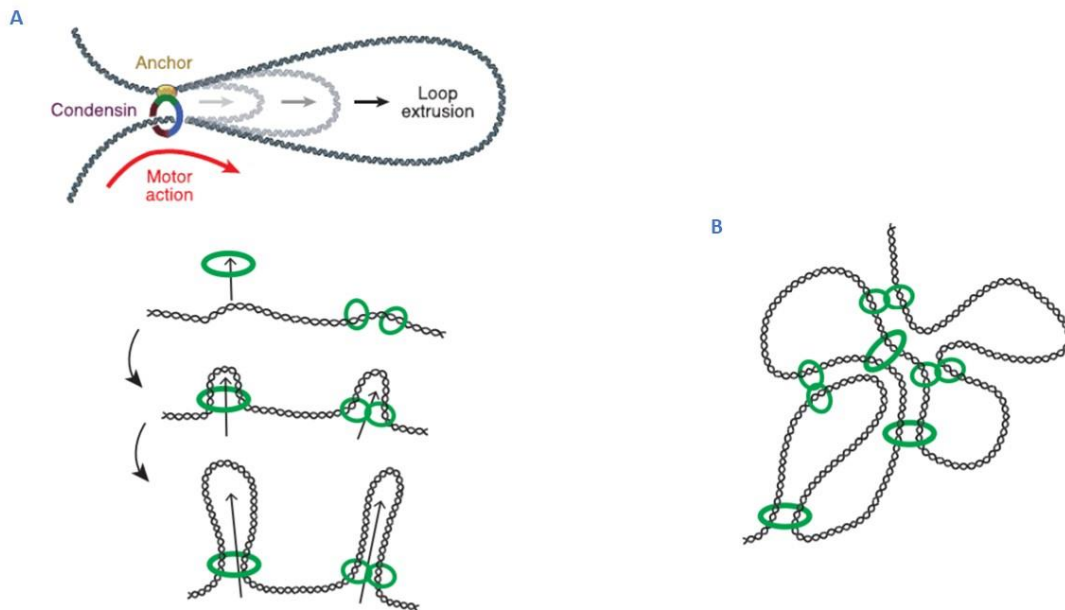


Figure 15. DNA Loop extrusion vs. Loop entrapment.

A) Loop extrusion : DNA is trapped within one or two SMC complexes and the DNA further extrudes mediated by an ATP-dependent motor activity. B) Loop entrapment: Random crosslinking of DNA together by trapping two strands inside the SMC complex rings whether it's one or two rings from two interacting complexes. Adapted from (Ganji et al., 2018) and (Eeftens and Dekker, 2017)

The idea that MORC ATPases would encircle their substrates was defended using single molecule experiments, where a buffer flow was applied. It was observed that the protein would move with the flow, and slide off the free end of DNA to be released. This observation was further challenged with the attachment of quantum dots at the free ends, which consistently showed the retention and accumulation of the protein at these free ends (Kim et al., 2019).

Furthermore, MORC-1 incubated with plasmid DNAs of various topologies followed by the application of high-salt washes, showed that the protein retained open circular and supercoiled forms yet not linear DNA. This is consistent with the observations of Smc5/6 interacting with circular and not linear DNA at high-salt concentrations, underlining their DNA-binding through topological entrapment (Kanno et al., 2015).

However, MORC-1 proteins themselves were observed to act similarly to quantum dots at retaining their own release, when used at higher concentrations, thus static DNA entrapping foci would serve as roadblocks for mobile MORC-1 foci. The static foci of labelled proteins would grow brighter, as initial protein binding events yielding additional interactions of MORC proteins (Kim et al., 2019).

This is consistent with the MORC proteins forming nuclear *puncta* in cells, and it also explains the apparent preference of these proteins for longer DNAs over shorter ones, as the longer substrates would be able to load more proteins and thus increase their concentration enough for them to assemble and prevent their own dissociation from the DNA molecules.

Therefore, the stable chromatin compaction takes place, as initiated, through the initial encircling and entrapment of DNA, the subsequent multimerization of the proteins and the formation of compacted entrapped DNA *foci* ([Figure 14](#)).

18. Knowledge gap and preliminary questions

If I were to describe in clear sentences the knowledge gap in understanding the development of *T. gondii*, I would say that the community is facing a great lack of information regarding the molecular mechanisms underlying its sexual development, as the parasite in these stages is far less accessible than it is in its asexual intermediate host cycle.

One other gap would revolve around the weight of the epigenetic mechanisms in this parasite, which seem to have extended and diversified its repertoire of PTM's writers and readers, in a way that is evolutionary sophisticated enough to allow us to assume or at least imagine that chromatin structure and its remodeling must be carrying great responsibility in ensuring the success of this species.

Also, the interplay between such processes and the transcription machinery has not been fully elucidated, nor the weight of this latter on the developmental commitment. Also, not clear are the mechanisms behind the translation of the external signals that the parasite receives from its host and environment throughout its life, and their weight in shaping the responses of the parasite and its developmental switches.

No master regulator has yet been identified for neither developmental pathways, other than a few AP2 TFs that were characterized, yet none displayed a clean-cut phenotype on the transcription nor on the development of the parasite, hinting on the complexity of the molecular reprogramming of such commitments.

The genome wide analysis techniques and new genetic tools becoming widely available permitted the expansion of our ability to tackle questions that were once too difficult to pursue.

I will outline in a synthetic way the preliminary questions that we aimed to solve in the beginning of this project, which were later broaden depending on the subsequent findings.

These would include the functional characterization of the unique MORC protein found in *T. gondii*. A diverse set of functions were attributed to this family of proteins, as detailed earlier, yet it is obvious that they were studied in a limited set of organisms, namely in *Arabidopsis thaliana*, in mammals, and in *C. elegans*. Our contribution would thus bring additional evolutionary and functional knowledge to the MORC community by providing insights on this new player.

Also, as the protein was identified as a partner of TgHDAC3, an enzyme which weight on the transcriptional regulation of the parasite was already witnessed, we sought to understand the secrets behind this interaction and whether these two proteins were equally contributing to the fate of the parasite, with hopes of having identified a novel HDAC co-repressor complex in *Apicomplexa*.

We wanted to tackle the mechanistic aspects behind the putative function of MORC (unknown at the time), and whether a crosstalk exist between this protein and any of the 62 AP2 DBD-containing proteins found in *T. gondii*, especially considering that one of these factors had been co-purified initially with TgHDAC3.

Also, the domain architecture of this protein would let one wonder about its involvement in any epigenetic pathways. Aside from its interaction with an HDAC enzyme, the TgMORC displayed a structure that is similar to the one of *C. elegans* and HsMORC3 and 4, all of which sharing a CW-ZF domain with PTM reading abilities, thus pushing us to pursue the potential chromatin related function of our protein.

II. Results

1. The paper (Farhat et al., 2020)



A MORC-driven transcriptional switch controls *Toxoplasma* developmental trajectories and sexual commitment

Dayana C. Farhat¹, Christopher Swale¹, Céline Dard¹, Dominique Cannella¹, Philippe Ortet^{1,2}, Mohamed Barakat², Fabien Sindikubwabo¹, Lucid Belmudes³, Pieter-Jan De Bock³, Yohann Couté³, Alexandre Bougdour¹ and Mohamed-Ali Hakimi¹✉

***Toxoplasma gondii* has a complex life cycle that is typified by asexual development that takes place in vertebrates, and sexual reproduction, which occurs exclusively in felids and is therefore less studied. The developmental transitions rely on changes in the patterns of gene expression, and recent studies have assigned roles for chromatin shapers, including histone modifications, in establishing specific epigenetic programs for each given stage. Here, we identified the *T. gondii* microorchidia (MORC) protein as an upstream transcriptional repressor of sexual commitment. MORC, in a complex with Apetala 2 (AP2) transcription factors, was shown to recruit the histone deacetylase HDAC3, thereby impeding the accessibility of chromatin at the genes that are exclusively expressed during sexual stages. We found that MORC-depleted cells underwent marked transcriptional changes, resulting in the expression of a specific repertoire of genes, and revealing a shift from asexual proliferation to sexual differentiation. MORC acts as a master regulator that directs the hierarchical expression of secondary AP2 transcription factors, and these transcription factors potentially contribute to the unidirectionality of the life cycle. Thus, MORC plays a cardinal role in the *T. gondii* life cycle, and its conditional depletion offers a method to study the sexual development of the parasite in vitro, and is proposed as an alternative to the requirement of *T. gondii* infections in cats.**

The parasite *Toxoplasma gondii* is a prevalent single-celled eukaryote that causes toxoplasmosis, a food-borne zoonotic disease that has life-threatening potential in cases of immune-suppressed individuals and of unborn fetuses. *T. gondii* has a dynamic life cycle that alternates between multiple host species. The asexual development takes place in a wide range of mammals and birds, which are identified as the intermediate hosts, whereas the sexual reproduction occurs exclusively in felids, which are identified as the definitive hosts¹.

Detailed transcriptome and proteome studies conducted throughout the life cycle of *T. gondii* have uncovered that large numbers of mRNAs and proteins are exclusively expressed in either one of the developmental stages. *Toxoplasma* has evolved an extremely dynamic as well as precise genetic reprogramming system, yet the mechanisms that link gene expression and stage transitions remain poorly understood, mainly because of the limited in vivo access to the sexual stages and also because of current methodological limitations in maintaining these stages in cell culture.

A complex gathering several AP2 transcription factors (TFs), which emerged in the Apicomplexa phylum as key regulators of life cycle transitions^{2,3} along with the histone deacetylase 3 (HDAC3), was suspected to operate as a chromatin-engaged molecular motor that drives stage-specific gene expression in *T. gondii*^{4,5}. Interestingly, HDAC3 co-purified with a microorchidia (MORC) homologue in *T. gondii*⁴ that, to date, has not been further studied. MORC proteins in plants and animals act on chromatin compaction^{6–12}, and in this study we establish that *T. gondii* MORC is a master regulator of the developmental trajectories throughout the

parasites life cycle, including changes that occur throughout the course of sexual commitment.

Results

MORC, a highly conserved ATPase protein, forms complexes that assemble HDAC3 and distinct AP2-related TFs. The MORC family is typified by the aggregation of a gyrase, Hsp90, histidine kinase, and MutL (GHKL) and S5 domains, which together form a catalytically active ATPase module^{13,14} (Fig. 1a). MORC is highly conserved across eukaryotes, with the exception of fungi, yet is present in apicomplexan parasites, suggesting that it has a relatively early origin in nucleated cells¹⁵ (Fig. 1b, Supplementary Fig. 1). From an evolutionary point of view, the apicomplexan version of the GHKL ATPase domain is actually closer to that of animals than of plants (Fig. 1b). Apicomplexans have a unique version of the MORC ATPase¹⁵ that is rimmed by Kelch-repeat β -propellers and a CW-type zinc-finger domain; the CW-type zinc-finger domain has been reported to be a histone PTM reader¹⁶ that is present in the metazoan MORC proteins as well (Fig. 1a,c).

MORC was originally shown to copurify with HDAC3 in tachyzoites⁴. Immunofluorescence analysis of intracellular parasites revealed that HDAC3 and MORC diffusely spread throughout the nucleus in a punctate pattern and they frequently overlapped (Fig. 1d). The extremely strong and specific association between MORC and HDAC3 was further confirmed by reverse immunoprecipitation and mass-spectrometry(MS)-based proteomic analysis, using knock-in parasite lines expressing a tagged version of MORC (Fig. 1e–h, Supplementary Table 1). Together, these data indicated

¹Institute for Advanced Biosciences (IAB), Team Host-Pathogen Interactions and Immunity to Infection, INSERM U1209, CNRS UMR5309, University Grenoble Alpes, Grenoble, France. ²BIAM-LEMIRE, UMR 7265, CEA, CNRS, University Aix-Marseille, St-Paul-Lez-Durance, France. ³University Grenoble Alpes, CEA, INSERM, Grenoble, France. ✉e-mail: mohamed-ali.hakimi@inserm.fr

that there is cooperative action between multiple TFs (co-elution of a set of 10 AP2-related TFs; Extended Data Fig. 1a) embedded within a high-molecular-mass chromatin complex (~500–800 kDa; Fig. 1e) that couples lysine deacetylation with ATPase-mediated chromatin-remodelling machinery^{6,7,12}.

MORC and HDAC3 colocalize extensively in the genome. To further explore how MORC and HDAC3 partner in vivo, we examined their genome-wide distributions using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq; GSE136060). Here and previously¹⁷, in parallel, we generated high-resolution profiles of six permissive and repressive histone post-translational modification (PTM) marks in the cystogenic type II (Pru) strain. Overall, the inspection of these individual ChIP-seq tracks revealed a strong overlap between the binding sites of the HDAC3 and MORC cis-tromes (Fig. 2a, Supplementary Fig. 2a,b), and we found that about 90% of the peaks were located at intergenic regions (Fig. 2b). A closer look at the gene level revealed that MORC and HDAC3 consistently overlapped with hypoacetylated chromatin, probably resulting from the deacetylation activity of HDAC3 (Supplementary Fig. 2c,d). At transcribed genes, MORC and HDAC3 seem to act as barriers to demarcate active chromatin regions typified by distinct acetylation levels, or to polarize promoter regions, thereby probably forming blueprints for the acquisition and maintenance of pre-determined gene expression profiles throughout the tachyzoite cell cycle (Supplementary Fig. 2c,d). At many other regions, MORC and HDAC3 were coenriched near to silenced genes that were embedded in chromatin that was marked by both H3K9me3 (a repressive mark) and H3K14ac (an activation mark), a dual PTM signature that was recently reported as a poised chromatin state characteristic of developmentally regulated genes in *T. gondii*¹⁷ (Supplementary Fig. 2e). In this respect, MORC acts as a chromatin repressor that insulates neighbouring genes by potentially blocking the influences of transcriptional activators.

HDAC3 binds to chromatin in a MORC-dependent manner. We next examined whether MORC is required for the recruitment of HDAC3 to the chromatin by investigating the genome-wide occupancy of the deacetylase in the absence of MORC. Owing to the essential requirement of MORC in *T. gondii* (fitness score = -4.69)¹⁸, we used the auxin-inducible degron (AID) system¹⁹, which we improved (Extended Data Fig. 2a, Supplementary Table 2), to acutely and reversibly deplete the MORC protein below detectable levels. Treatment of the edited parasites with indole-3-acetic acid (IAA) triggered a highly specific and near-complete clearance of the pool of MORC-mAID-HA protein after 2 h and 16 h in the RH and Pru strains, respectively (Fig. 2c,d, Extended Data Fig. 2b,c).

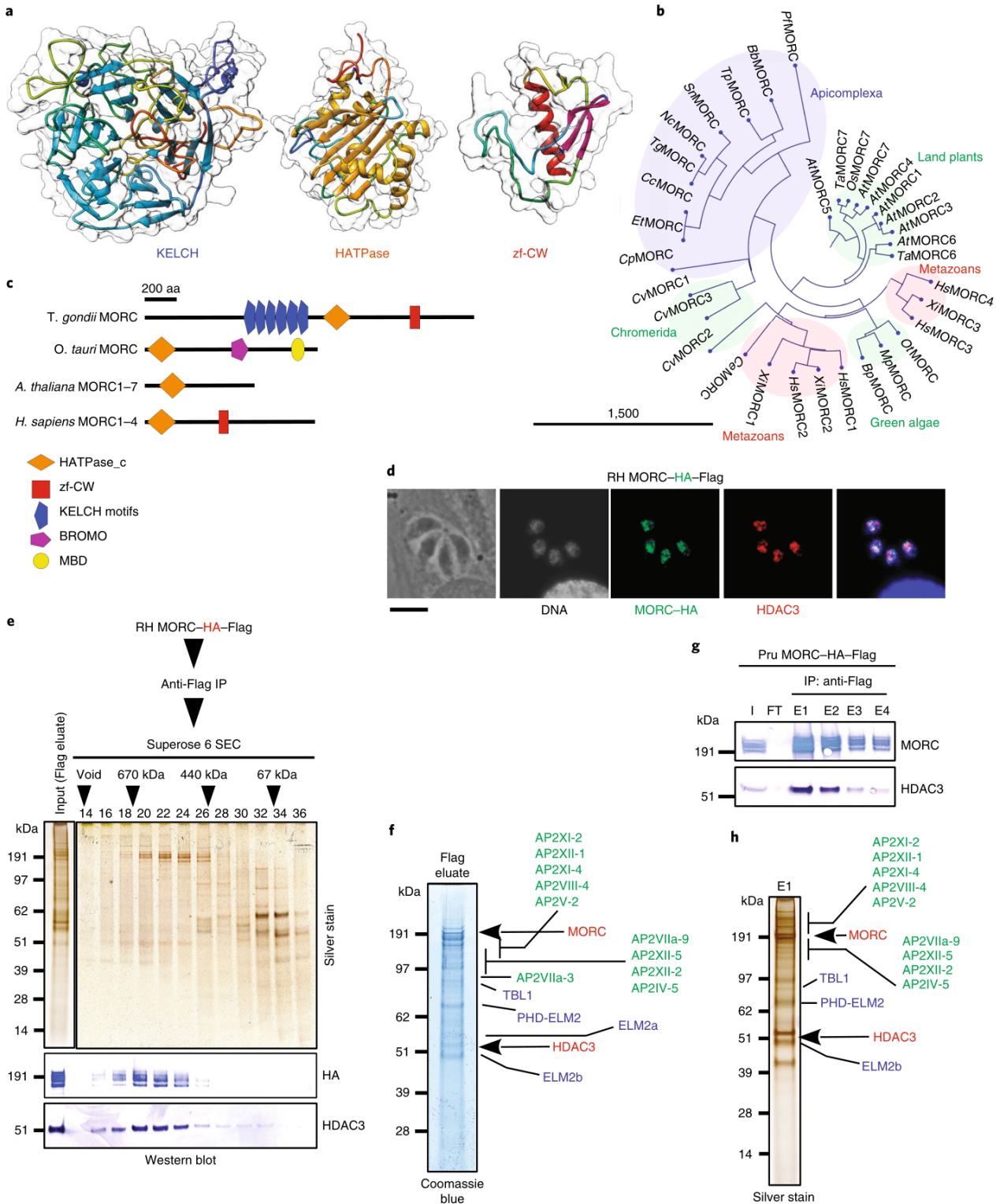
After degradation of MORC, HDAC3 protein levels remained relatively stable, at least at early times following the depletion (Fig. 2d); no apparent change in the nuclear location of HDAC3 was detected (Fig. 2e). To further investigate the functional interplay between MORC and HDAC3, we then performed ChIP-seq in the context of MORC conditional knock-down (KD). As expected, addition of IAA was accompanied by a near-complete loss of MORC chromatin enrichment compared with untreated cells (Fig. 2f,g). Concomitant with the loss of MORC, we detected a severe reduction in the binding of HDAC3 to chromatin, with a minimal retention of HDAC3 at few loci (Fig. 2f,g). All of these loci were initially enriched for MORC; this observation could therefore be attributed to the distinct immunoprecipitation abilities of the different antibodies, thereby stressing the dependency of HDAC targeting on its interaction with MORC. Notably, the loss of HDAC3 on chromatin produced a substantial and selective increase in histone acetylation nearby, underlining the deacetylation function of the enzyme (Extended Data Fig. 2d).

Depletion of MORC phenocopies the inhibition of HDAC3 in inducing the expression of targeted genes. To define the effects of MORC depletion on transcription, we conducted detailed time-course measurements of mRNA levels using RNA-sequencing (RNA-seq), after the depletion of MORC in both RH and Pru strains of *T. gondii* (Supplementary Table 3; GEO Serie, GSE136123). In parallel, we investigated the transcriptional alterations caused by the exposure of the parasites to a sublethal concentration of FR235222, which is a specific inhibitor of HDAC3 activity⁵ (Fig. 3a, Extended Data Fig. 3a). The most evident effect of either depletion of MORC or inhibition of HDAC3 was that a sizeable fraction of mRNAs (corresponding to ~1,181 genes; Fig. 3b) accumulated gradually overtime, regardless of the strain type. We found that 1,647 genes, hereafter referred to as cluster 1, were specifically silenced by MORC (Fig. 3b, Supplementary Table 4). We then investigated whether the robust expression of *T. gondii* genes after the depletion of MORC was also reflected by the abundance of corresponding proteins by monitoring proteome changes after addition of either IAA or FR235222 (Fig. 3c, Extended Data Fig. 3b). For several genes, the changes in mRNA levels were strongly correlated with changes in protein levels (Fig. 3d, Supplementary Table 5). We corroborated these data by generating several knock-in parasite lines to show that MORC depletion or HDAC3 inhibition was able to elicit the expression of epitope-tagged target proteins (Extended Data Fig. 3c). Importantly, the combined ChIP-seq and RNA-seq analysis revealed, at the genome-wide level, that the majority of genes in cluster 1 were direct targets of MORC and its partner HDAC3 (Fig. 3e). Effectively, the addition

Fig. 1 | MORC, a highly conserved ATPase protein, interacts with HDAC3 in *T. gondii*. **a**, MORC has three predicted structural domains: KELCH motifs, a conserved HATPase catalytic domain and a C-terminal CW-type zinc-finger domain (zf-CW); secondary structures are shown as cartoons in blue, red and orange. Structural predictions were modelled using the Phyre2 server and are represented by an enclosed surface cartoon diagram using UCSF Chimera. **b**, The relationships between MORC proteins (Supplementary Fig. 1). The MORC domains have arisen early in eukaryotic evolution before the divergence of plants and animals. In this context, the apicomplexan MORC family shares its speciation event with metazoans but not with land plants. Apicomplexan parasites have only one gene copy that encodes MORC, whereas other species have an expanded family. *At*, *Arabidopsis thaliana*; *Ta*, *Triticum aestivum*; *Os*, *Oryza sativa*; *Hs*, *Homo sapiens*; *Xl*, *Xenopus laevis*; *Ce*, *Caenorhabditis elegans*; *Cv*, *Chromera velia*; *Tg*, *T. gondii*; *Nc*, *Neospora caninum*; *Pf*, *Plasmodium falciparum*; *Cp*, *Cryptosporidium parvum*; *Sn*, *Sarcocystis neurona*; *Et*, *Eimeria tenella*; *Cc*, *Cyclospora cayetanensis*; *Bb*, *Babesia bovis*; *Tp*, *Theileria parva*; *Ot*, *Ostreococcus tauri*; *Mp*, *Micromonas pusilla*; *Bp*, *Bathycoccus prasinos*. **c**, Representative domain architectures of *T. gondii* MORC domains and of related proteins. MORC is encoded by TGME49_305340. **d**, The nuclear location of MORC (green) and HDAC3 (red) in human primary fibroblasts (HFFs) infected with parasites expressing an HA-Flag-tagged copy of MORC. Cells were co-stained with Hoechst DNA-specific dye. Scale bar, 5 μm. **e**, MORC forms a partnership with HDAC3. Size-exclusion chromatography of MORC-containing complexes after Flag-affinity selection. The fractions were analysed on silver-stained SDS-PAGE gels (top) and analysed using western blots to detect MORC-HA-Flag (anti-HA antibodies) and HDAC3 (bottom). SEC, size-exclusion chromatography. **f**, MS-based proteomic analysis of Flag elution identified MORC and its partners. The identities of the proteins are indicated on the right. **g,h**, The purification was repeated using HFFs infected with a type II (Pru) strain expressing an HA-Flag-tagged copy of MORC. Flag-affinity eluates were analysed using western blot (**g**) to detect MORC-HA-Flag and HDAC3, and using MS-based proteomic analysis (**h**) to identify shared complexes between the type I (RH) and type II (Pru) strains. For **d-h**, all data are representative of three independent biological experiments.

of IAA was followed by a near-complete loss of MORC enrichment and a concomitant reduction of HDAC3 occupancy in the vicinity of the 5' ends of the cluster-1 genes (Fig. 3e); nevertheless, it is worth mentioning that we found that a significant fraction of the genes of cluster 1 ($n = 498$, cluster 1b) display a lack of nearby

occupancy of MORC as well as of HDAC3 (Supplementary Fig. 3, Supplementary Discussion). Taken together, our results provide strong evidence that MORC, in partnership with HDAC3, has a crucial and broad role in silencing the transcription of a precise set of genes in tachyzoites.



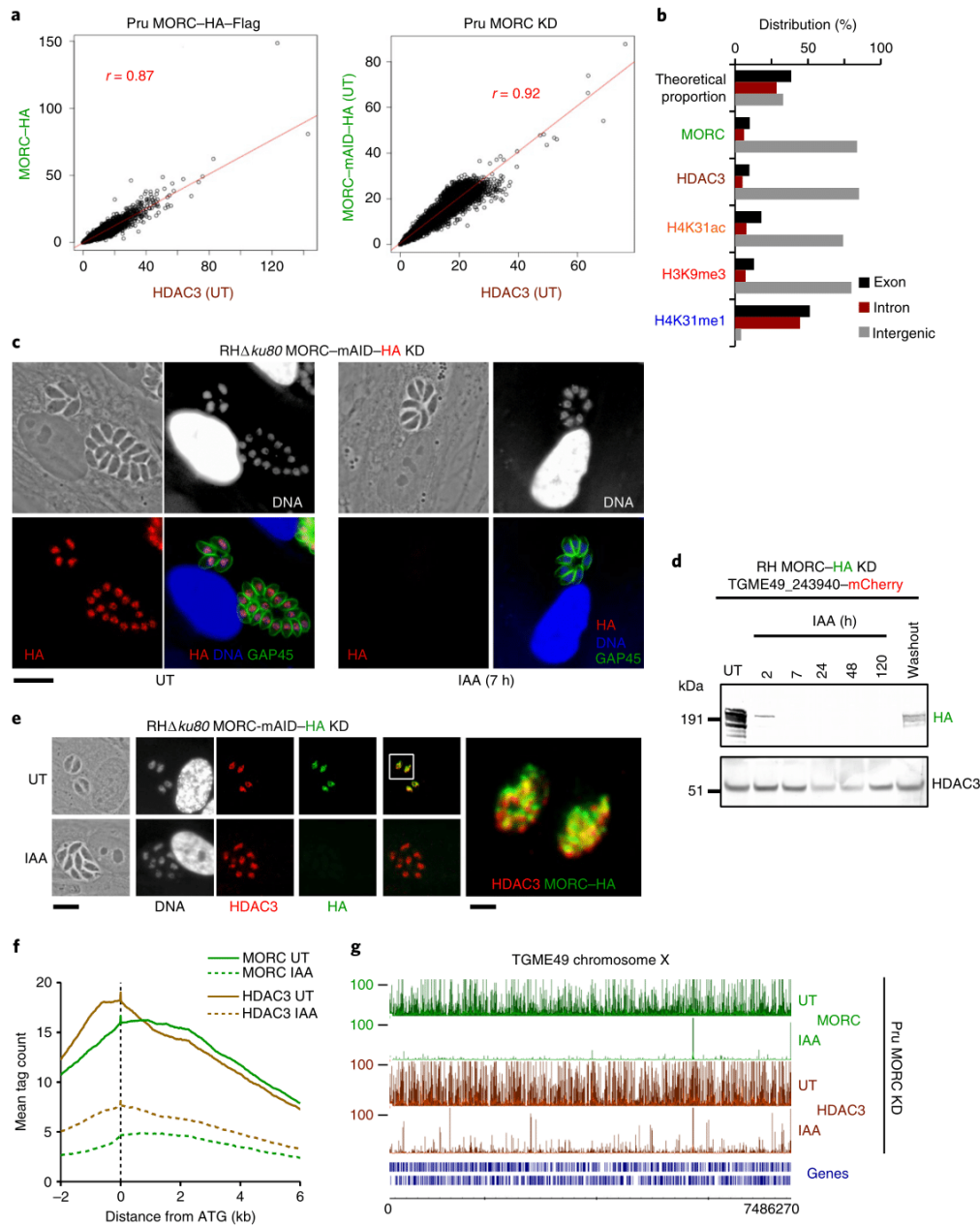


Fig. 2 | Recruitment of HDAC3 to chromatin is mediated by MORC. **a**, Comparison of the enrichment levels between MORC and HDAC3. ChIP-seq data from HFFs infected with Pru MORC-HA-Flag (left) and Pru MORC-mAID-HA KD (right) parasites that were untreated (UT) were analysed. The average tag count of the enrichment is shown on the x and y axes ($n = 8,322$ genes). Pairwise correlations were calculated using the Pearson correlation coefficient (r). **b**, Distribution of MORC-, HDAC3- and PTM-occupied regions relative to the *T. gondii* reference genome annotation. **c**, MORC-mAID-HA protein expression levels after 7 h of adding IAA, displayed by IFA on cells infected with RH MORC-mAID-HA. Fixed and permeabilized parasites were probed with antibodies against MORC-mAID-HA (red) and GAP45 (green). Scale bars, 10 μ m. **d**, Time-course analysis of the expression levels of MORC-mAID-HA. The samples were taken at the indicated time periods after addition of IAA and were probed with antibodies against HA and HDAC3. IAA-treated (3 d) parasites were also washed, incubated with fresh media in the absence of IAA (2 d) and analysed using western blot. The same experiment was repeated three times and a representative blot is shown. **e**, Coexpression of MORC-mAID-HA (green) and HDAC3 (red) in RH MORC KD left untreated or treated with IAA for 16 h. Right, magnification of nuclear areas marked by HDAC3 and MORC. Scale bars, 2 μ m (right) and 10 μ m (left). **f**, Genome-wide MORC and HDAC3 occupancy profiles in Pru MORC KD parasites that were either untreated or treated with IAA for 30 h. The average-signal profiles of each protein were plotted across a -2 kb to $+6$ kb region with respect to the start codon of the nearest *T. gondii* gene. The average tag count of the enrichment is shown on the y axis. **g**, Integrated genome browser view of MORC and HDAC3 enrichment across *T. gondii* X chromosome before and after MORC degradation, assessing the extent of the effect of MORC on the HDAC3 chromatin occupancy. Read density is shown on the y axis. The experiments in **c** and **e** were conducted three times and representative images are displayed.

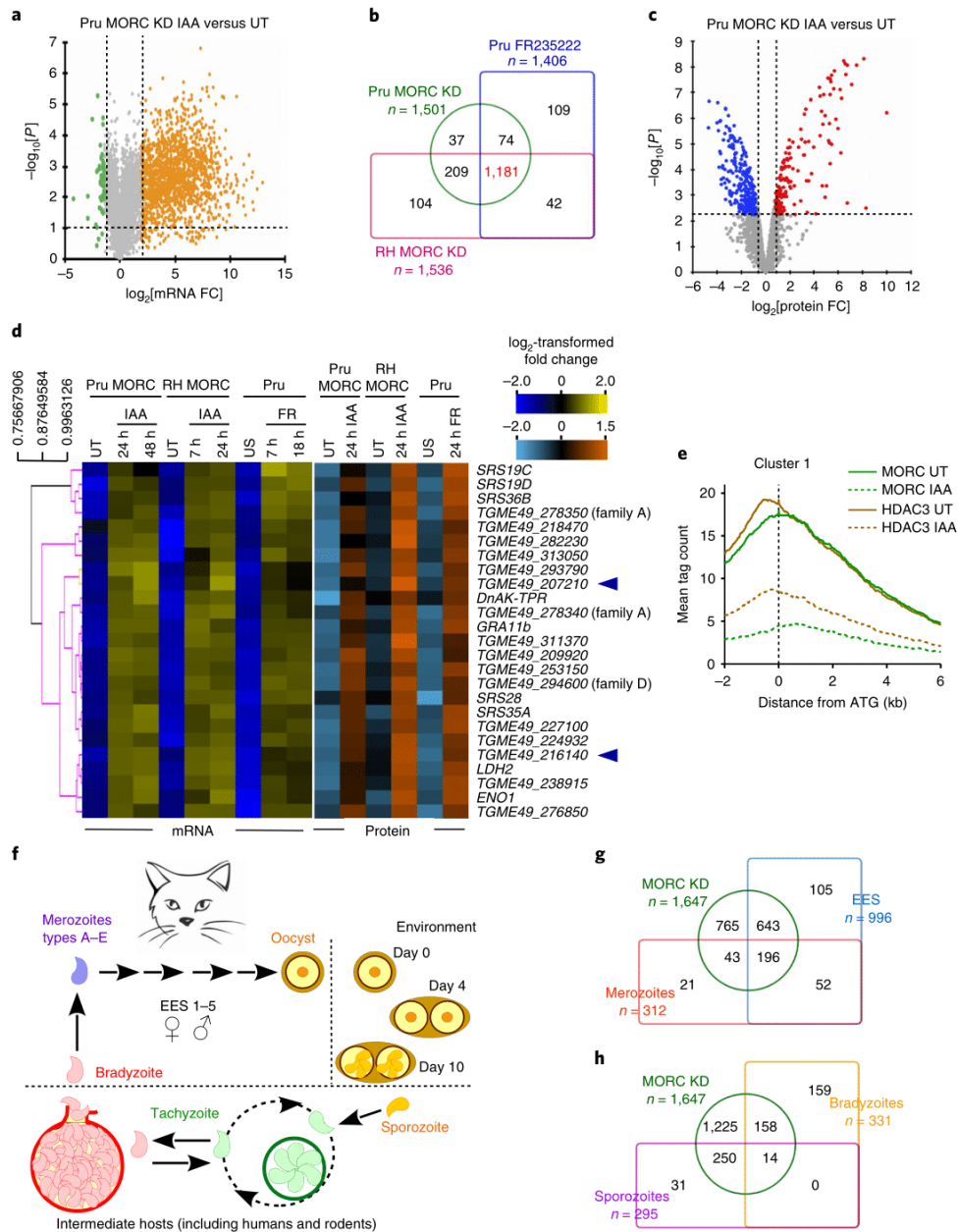


Fig. 3 | Depletion of MORC phenocopies the inhibition of HDAC3 by inducing the expression of sexual-stage-specific genes. **a**, Gene expression variation between untreated versus IAA-treated (24 h) Pru MORC KD parasites ($n = 8,914$; Supplementary Table 3). The orange and green dots indicate the number of significantly upregulated and downregulated genes, respectively, using an Bonferroni-adjusted $P < 0.01$ and a fold change (FC) of ± 2 as the cut-off threshold corresponding to each comparison. Vertical dashed lines indicate two-fold up- and downregulation. Horizontal dashed line indicates the cut-off value. **b**, Comparison of upregulated mRNAs after depletion of MORC in RH and Pru strains as well as in the context of HDAC3 inhibition by FR235222. **c**, Changes in protein expression between Pru MORC KD parasites that were either untreated or treated with IAA for 24 h ($n = 1,673$ proteins; Supplementary Table 5). Overexpressed proteins after depletion of MORC are indicated in red and underexpressed proteins are shown in blue. Benjamini-Hochberg FDR = 1.02%; $P \leq 0.00631$. Vertical dashed lines indicate $\log_2[\text{FC}] \geq 0.8$ or ≤ -0.8 . Horizontal dashed line depicts the cut-off value. **d**, Heat map showing hierarchical clustering analysis (Pearson correlation) of the selected cluster-1 genes and their corresponding proteins that are differentially regulated after the depletion of MORC for different combinations of strain and induction. Parasites were left unstimulated (US) or treated with FR235222 (FR) to inhibit HDAC3. The colour scale indicates \log_2 -transformed fold changes. Arrowheads highlight genes of interest. **e**, Genome-wide MORC and HDAC3 occupancy profiles at cluster-1 genes in Pru MORC KD parasites that were either untreated or treated with IAA for 30 h. The average-signal profile of each protein was plotted over a -2 kb to $+6$ kb region with respect to the start codon of the nearest *T. gondii* gene. The average tag count of the enrichment is shown on the y axis. **f**, Scheme of *T. gondii* life cycle, showing the different stages of the lytic asexual cycle (bottom) and of the sexual stages (top left), with the differentiation events that occur in the corresponding hosts and environments. **g, h**, Comparison of upregulated mRNAs after depletion of MORC (cluster 1, $n = 1,647$ genes) and the RNAs detected in the merozoite and EES (**g**) and sporozoite and bradyzoite (**h**) life cycle stages (data source, ToxoDB).

MORC and HDAC3 silence merozoite-primed sexual commitment. We have provided strong evidence that cluster-1 genes are not initially expressed in tachyzoites, revealing that MORC has a repressive role during this stage and enabling us to infer that the cluster-1 genes might be expressed at other times during the life cycle (Fig. 3f). Interestingly, comparative RNA-seq analysis using existing datasets generated from early and mature bradyzoites²⁰, merozoites²¹, longitudinal studies on enteroepithelial stages (EESs)²² or three developmental stages of oocysts²³ revealed that the depletion of MORC predominantly induced the expression of sexual-stage-specific genes (~55% of cluster 1, including merozoite genes) and, to a lesser extent, sporozoite- (15%) and bradyzoite-specific (~10%; Fig. 3g,h, Supplementary Table 4) genes. The enteroepithelial cycle starts 2 d after the ingestion of tissue cysts by the cat, during which the merozoites initiate gamete formation¹ (Fig. 3f). Although tachyzoites and merozoites are both fast-replicating and metabolically similar, they can be readily distinguished by their peculiar transcriptional signatures; these signatures are typified by 312 transcripts that are exclusively expressed in merozoites, whereas 453 genes are strictly expressed during the tachyzoite stage^{21,24}. Notably, ~85% of the merozoite-specific transcripts ($n=268$ out of 312) were induced after the depletion of MORC (Fig. 4a,b, Supplementary Table 4). This correlation was also observed in the abundance profiles of proteins in the same cluster, including genuine merozoite markers such as GRA11a and GRA11b²⁵ and a large set of merozoite-restricted surface proteins (SRS), which were predicted to promote gamete development and fertilization²¹ (Fig. 4b). To better document this phenotype, we performed a targeted knock-in within the MORC-mAID-HA strain of the reporter gene *TGME49_243940* for which expression was predefined as being restricted to the merozoite stage (Fig. 4b). The IAA treatment of these edited parasites led first to the near-complete clearance of MORC as expected and second to the concomitant accumulation of *TGME49_243940*-mCherry (Fig. 4c). Similarly, the systematic CRISPR-mediated gene disruption of the partners of MORC resulted in a high level of *TGME49_243940*-HA induction (Extended Data Fig. 4a), which was also triggered by chemical inhibition of HDAC3 (Extended Data Fig. 4b). As predicted, in the context of MORC-depleted cells, a substantial reduction in the chromatin occupancies of both HDAC3 and MORC was observed near to the 268 predefined merozoite genes (Fig. 4d,e). The loss of HDAC3 binding often resulted in increased levels of histone acetylation at the promoter regions, thereby inducing the transcriptional activation of merozoite-specific genes, including *TGME49_243940* (Fig. 4e). This acetylation pattern was mostly observed at the genes that are directly regulated by MORC. Together, these data position MORC as a central and early checkpoint of sexual commitment.

MORC and HDAC3 maintain a great number of sexual-stage genes in a sustained repressive chromatin state. After their release from the schizonts, merozoites initiate gamete formation (Fig. 3f). A recent comparative transcriptomic study of sampled EESs collected from kittens defined five stages with distinct transcriptional profiles (EES1–5) and ranging from the early merogony to the late sexual stages, which include gametes and oocysts²². Notably, 51% of the genes that were found to be significantly upregulated in the MORC-depleted cells are defined as being expressed strictly during the sexual cycle—well above the EES-specific proportion of the entire genome (estimated at ~12%)²² (Fig. 3g, Supplementary Table 4). Cluster 1 is greatly enriched for a subset of genes that encode micro- and macrogamete-specific proteins holding putative important roles in fertilization and oocyst wall formation (Extended Data Figs. 4c,d and 5a,b).

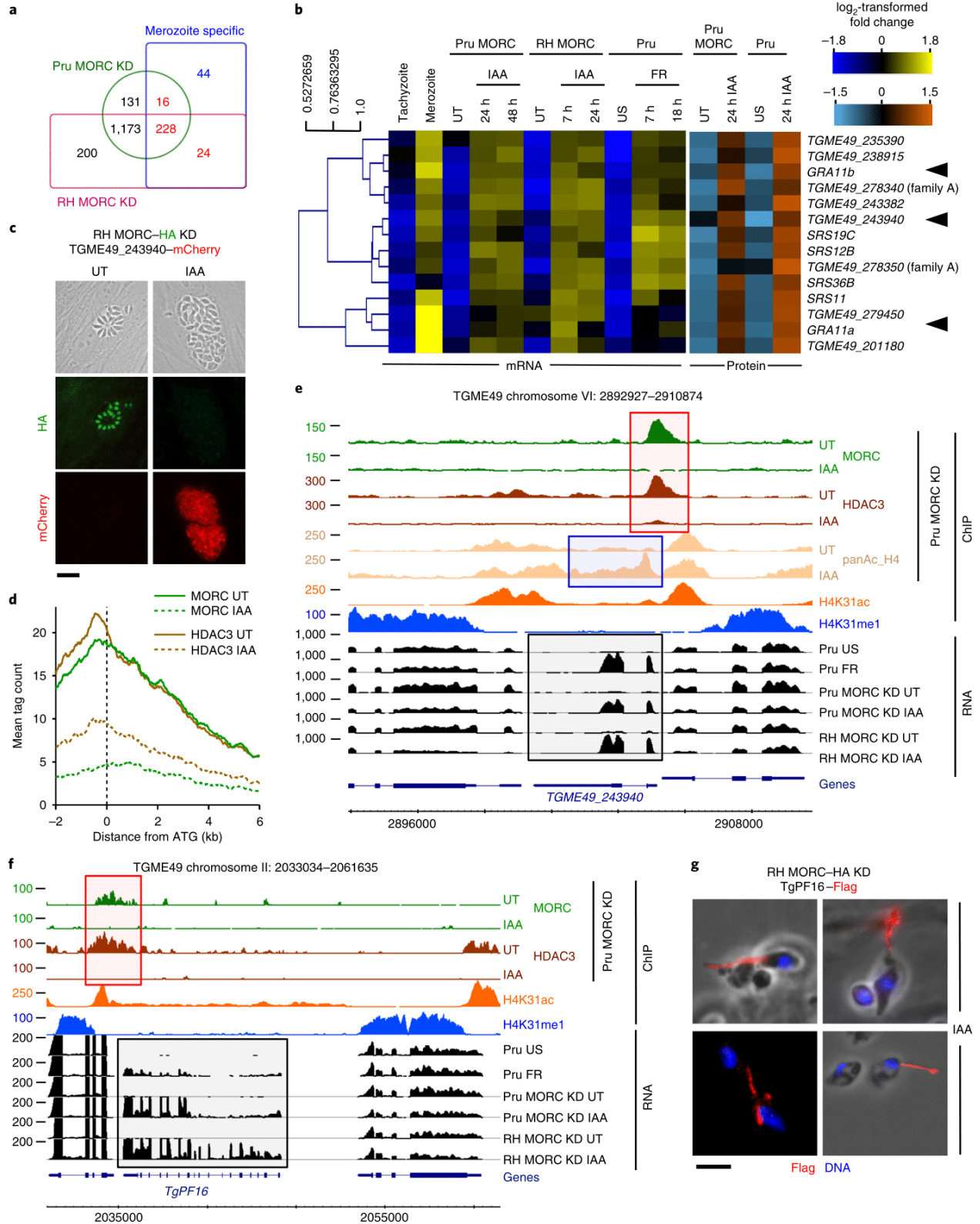
The male gamete of *T. gondii* is characterized by the presence of two flagella that resemble, in many aspects, the flagellum of eukaryotic sperm²². The majority, if not all, of the proteins behind the main features of microgametes, such as those involved in the axonemal cytoskeleton (one basal body protein, 3 radial spoke proteins and 11 dyneins) and in the intraflagellar transport (6 IFT proteins and 9 kinesins), were labelled as being regulated targets of MORC (Extended Data Fig. 6). *TgPF16*, a sperm-associated antigen^{26–28}, displayed direct MORC–HDAC3-mediated gene silencing (Fig. 4f). To examine the localization of *TgPF16*, we transiently expressed *TgPF16* fused to a C-terminal Flag tag within the MORC-mAID-HA parasite line. After addition of IAA, *TgPF16* was induced and found to be associated with growing cytoplasmic axonemes and the protein was occasionally detected along the emerging flagella of a presumably male gamete with a patchy distribution resembling the previously described PbPF16 labelling in *Plasmodium*²⁸ (Fig. 4g). MORC also regulates *Hapless 2* (*HAP2*), the expression of which was shown to be restricted to male gametes in different organisms, ranging from protists to flowering plants and insects^{29,30} (Extended Data Fig. 5c). *HAP2* is required in both *Plasmodium* and *Toxoplasma* for the fusion of gametes and the subsequent fertilization^{22,31,32}. Although the sexual development occurs in several variations, it seems that the alternation between the haploid and diploid states, the progression from one differentiated stage to the other and the maintenance of the epigenetic signature of a given stage are strictly controlled by MORC and HDAC3.

MORC and HDAC3 silence oocyst-specific gene expression. We next explored whether depletion of MORC stimulated—with a magnitude as strong as the aforementioned observations—the expression of oocyst-specific genes. We compared our data with previously reported transcriptomes stemming from three time

Fig. 4 | Degradation of MORC, similar to the inhibition of HDAC3, induces the expression of merozoite-restricted transcripts. **a**, The overlap between the MORC-regulated genes in the RH and Pru strains, and the 312 genes of which expression was reported to be exclusive to merozoite. **b**, Heat map showing mRNA hierarchical clustering analysis (Pearson correlation) of selected merozoite-specific genes and their corresponding proteins that were differentially regulated after depletion of MORC for different combinations of strain and induction. The colour scale indicates \log_2 -transformed fold changes. Arrowheads highlight genes of interest. **c**, IFA of HFFs that were infected with parasites harbouring a merozoite-specific reporter gene (*TGME49_243940*) endogenously tagged with mCherry within the RH MORC-mAID-HA lineage. Fixed and permeabilized parasites were probed with antibodies against HA (green) and mCherry (red). Scale bar, 10 μm . **d**, Genome-wide occupancy profiles of MORC and HDAC3 at merozoite-specific genes ($n=268$) in Pru MORC KD parasites that were either untreated or treated with IAA for 30 h. **e,f**, Density profiles for representative regions of chromosomes VI (**e**) and II (**f**). Chromosomal positions are indicated on the x axis. The ChIP-seq profiles were obtained using antibodies directed against various histone marks, HDAC3 and HA (MORC detection), and chromatin sampled from a Pru MORC KD strain that was either untreated or treated with IAA for 24 h. RNA-seq data for different combinations of strain and induction are shown in black. Read density from ChIP-seq data and RPKM values for RNA-seq data are shown on the y axis. The merozoite- and axonemal-specific genes *TGME49_243940* (**e**) and *TgPF16* (*TGME49_297820*; **f**) are shown in dark blue. The red box indicates the MORC and HDAC3 enrichment peak, the blue box indicates the acetylation rate of histone H4 and the black box indicates mRNA levels of the gene of interest. The experiment was repeated independently twice with similar results. **g**, IFA of HFFs that were infected with parasites harbouring an endogenously Flag-tagged axonemal-specific reporter gene (*TgPF16*) within the RH MORC-mAID-HA lineage. Fixed and permeabilized parasites were co-stained with Flag (red) and Hoechst DNA-specific dye (blue). Scale bar, 5 μm . The experiments in **c** and **g** were conducted more than three times and representative images are displayed.

points, representing immature (day 0), maturing (day 4) and mature (day 10) stages of oocyst development²³ (Fig. 3f). Notably, ~89% ($n=264$ out of 295) of the genes defined as being strictly expressed

in the oocysts were found to be induced in the MORC-depleted cells, well above the oocyst-specific proportion of the entire genome (corresponding to roughly 4%)²³ (Fig. 3h, Supplementary Fig. 4,



Supplementary Table 4). In this respect, the acute degradation of MORC was sufficient to induce almost all of the proteins that are usually present in fully sporulated oocysts (Fig. 5a), a result that we substantiated by detecting, in the MORC-depleted cells, a G1/19 monoclonal-antibody-reactive sporozoite-specific protein³³ (Fig. 5b). This was also corroborated by our proteomic analysis, which identified, with high confidence, SporoSAG³⁴—the hallmark surface antigen in sporozoites (Fig. 5a, Supplementary Table 5). The SporoSAG locus displays the canonical bivalent chromatin state ‘poised’ for transcription, along with a local co-occupancy of MORC and HDAC3 (Extended Data Fig. 7a). At this locus and at all of the loci that share the specificity for this same stage, we observed a concomitant loss of MORC and of its partner HDAC3 from the chromatin after the depletion of MORC (Fig. 5c, Extended Data Fig. 7). Notably, the induction of SporoSAG expression reached an unmatched level (mRNA levels > 8,000 reads per kb of transcript per million mapped reads (RPKM)) following either the sustained degradation of MORC or the inhibition of HDAC3 (Extended Data Fig. 7a). A larger group of proteins that showed similar expression patterns were uncovered (details are provided in the Supplementary Discussion, Supplementary Fig. 4 and Extended Data Fig. 7), emphasizing a critical role of MORC and HDAC3 in the development of oocysts, a key stage for the spread of *T. gondii*.

MORC depletion triggers a tachyzoite-to-bradyzoite switch, a mandatory step towards sexual development. Once in an enterocyte of a cat, the bradyzoite converts into a merozoite, distinctly priming the initiation of the in vivo sexual cycle (Fig. 3f). Although we show that MORC acts as a repressor at strategic points along the life-cycle continuum, it still remained to be determined whether the commitment into merozoite, which is triggered in vitro by depletion of MORC, also requires the compulsory passage through the bradyzoite stage¹. In this respect, we identified a substantial, albeit limited, number of bradyzoite-specific genes ($n=172$; Fig. 3h) of which the expression was induced in the MORC-depleted cells (Extended Data Fig. 8a). These genes encode surface antigens (BAG1, BSR4/p36 and BRP1), metabolic enzymes (LDH2 and ENO1) and cyst-wall-associated proteins (CST1, CC2 and BCLA; Fig. 5d,e, Extended Data Fig. 8 and Supplementary Fig. 5a–c) and they seem to be regulated by MORC in a similar manner to the genes described above. In the vicinity of their promoters, these genes display co-binding of MORC and HDAC3, which disappeared with the loss of MORC, and resulted in the transcriptional induction of these genes (Extended Data Fig. 8b, Supplementary Fig. 5a–c). Furthermore, we created transgenic parasite strains that express tagged versions of the bradyzoite-specific enzymes ENO1 and LDH2, and showed that their expression was silenced in an HDAC3-dependent manner (Supplementary Fig. 5a,b). The cyst wall is highly glycosylated

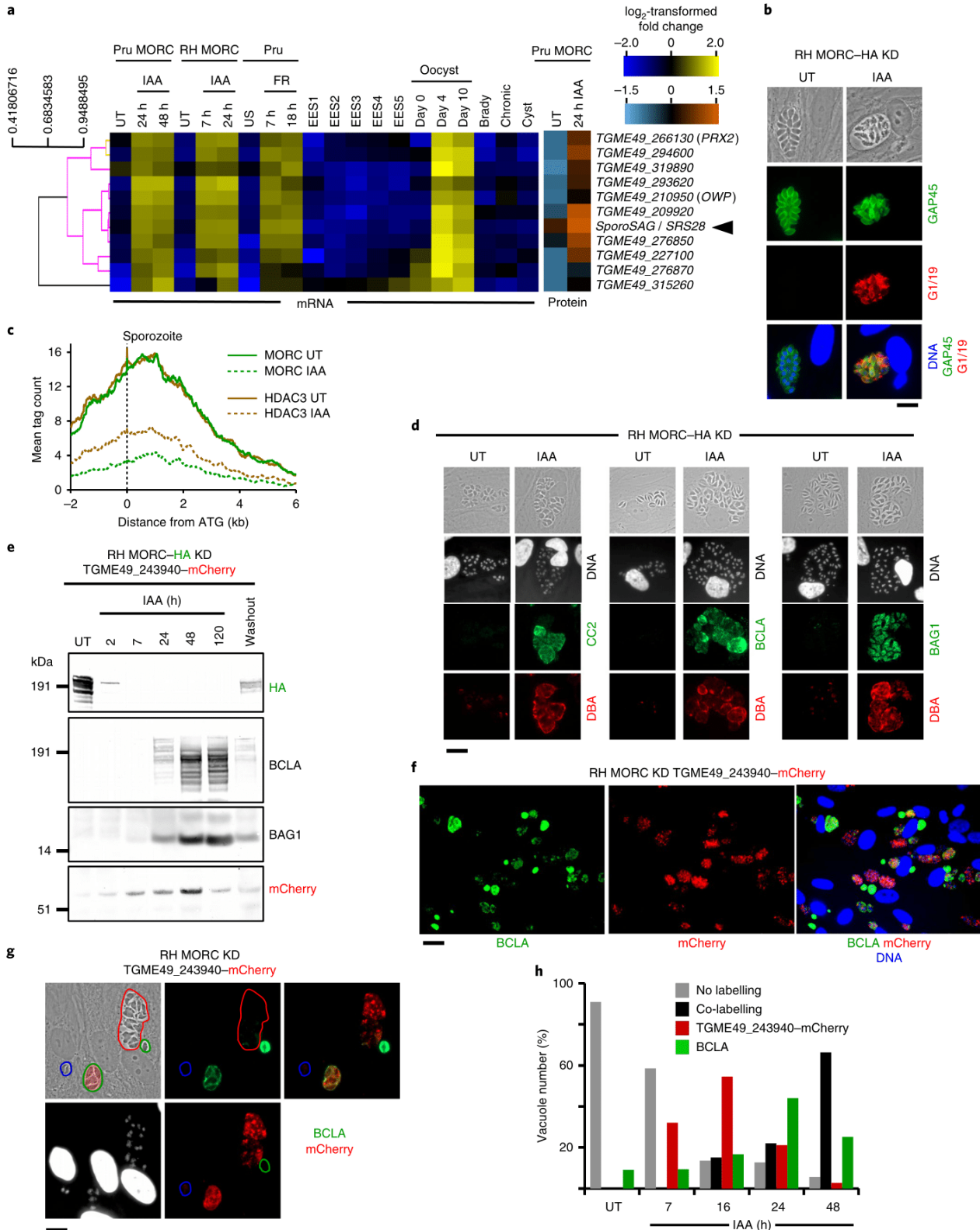
and stains easily with the *Dolichos biflorus* agglutinin (DBA) lectin. We observed a clear accumulation of DBA at the MORC-depleted vacuoles, probably as a consequence of the induction of *CST1-SRS44*—a gene known to encode a cyst wall protein that is responsible for the reactivity of DBA³⁵ (Fig. 5d). Furthermore, the induction of BCLA—a hallmark of latent tissue cysts (C. Dard et al., unpublished observations)—was compared with the induction of the merozoite marker TGME49_243940. Interestingly, the level of TGME49_243940 increased at the very beginning of MORC degradation, peaked at 48 h and declined sharply afterwards, whereas BCLA was not detectable until 24 h after the complete clearance of MORC, and its expression level was sustained overtime (Fig. 5e). Immunofluorescence analysis of individual vacuoles revealed non-homogeneous labelling following the degradation of MORC; we observed a great number of vacuoles expressing either BCLA or TGME49_243940–mCherry in an almost mutually exclusive manner (Fig. 5f–h). This observation is consistent with the initial findings that roughly half (159 out of 331; Fig. 3h) of bona fide bradyzoite genes were still repressed in the MORC-depleted cells. We speculate that this partial induction of bradyzoite-specific genes could be due to a repressor acting downstream of the MORC pathway. In this regard, MORC-depleted cells expressed AP2IX-9, a transcriptional repressor that was initially suspected to restrict the commitment of *Toxoplasma* to developing mature bradyzoite tissue cysts³⁶ (Fig. 6a). These data provide a solid baseline for a model in which the degradation of MORC induces AP2IX-9 to prevent the merozoite from converting back into a bradyzoite, therefore elucidating the molecular mechanisms behind the maintenance of the unidirectionality of the sexual stage development (Fig. 3f).

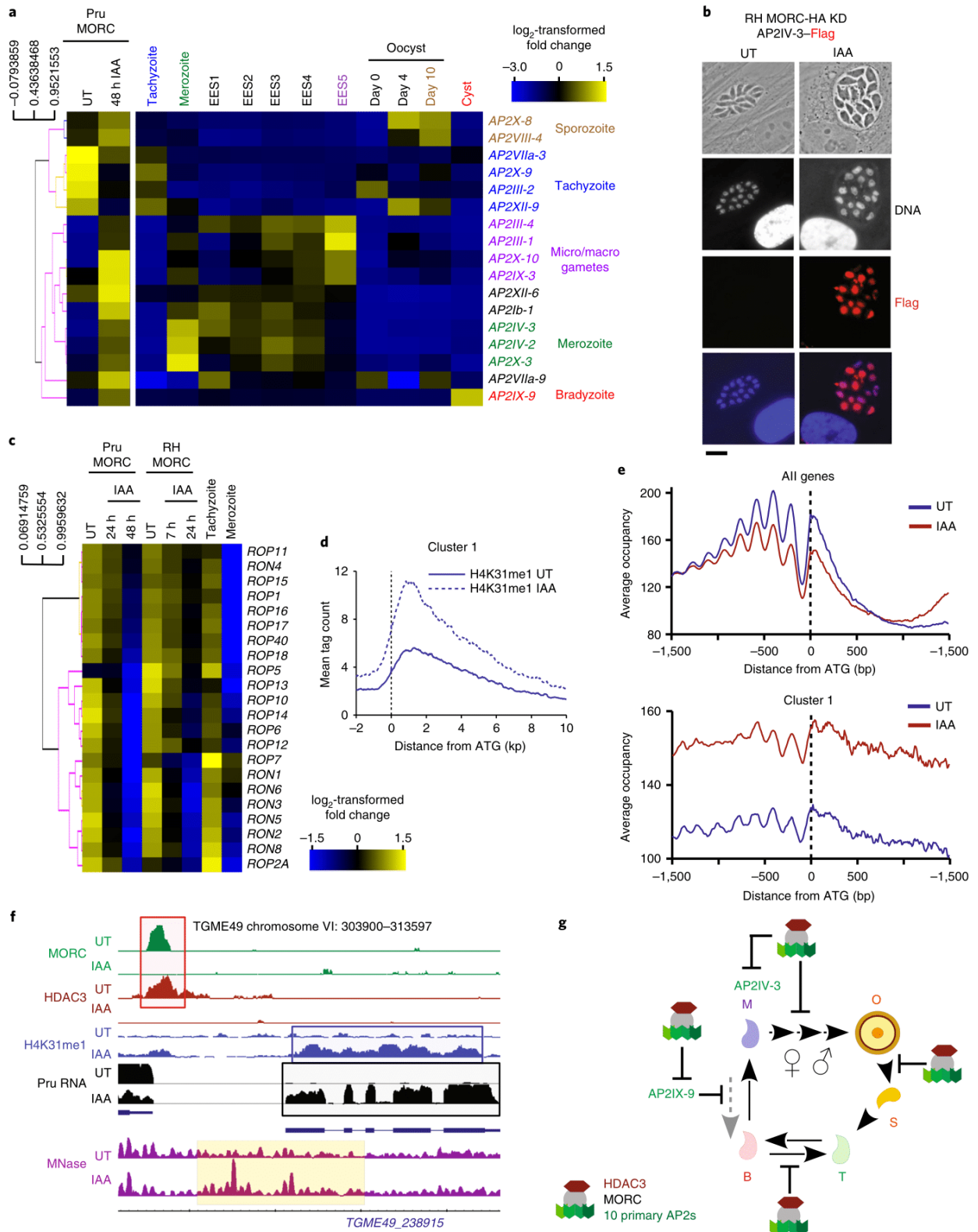
MORC depletion induces a downstream network of secondary transcriptional regulators to guide developmental trajectories. In addition to balancing bradyzoite gene expression, the depletion of MORC was shown to concomitantly restrain sets of genes that were defined as being expressed exclusively in tachyzoites and not in merozoites²¹ (Extended Data Fig. 9, Fig. 6c). In view of the small number of genes found to be repressed in the absence of MORC ($n=126$ genes; Fig. 3a, Supplementary Table 3), this transcriptional output could result from secondary TFs. In fact, four AP2 factors (Fig. 6a, blue) displayed the same restricted mRNA patterns, both in the context of MORC depletion and in the merozoite stage, and they might therefore be acting as transcriptional activators in tachyzoites. In our current model, the MORC-controlled developmental trajectories are probably directed downstream by the hierarchical expression of other regulators that, in turn, dictate the setting of a distinct predetermined transcriptional program for a given stage (Fig. 6g). For example, three MORC-regulated AP2 factors were strictly expressed in merozoites (Fig. 6a, green). A representative

Fig. 5 | MORC regulates developmental transitions at multiple checkpoints. **a**, Heat map showing mRNA hierarchical clustering analysis (Pearson correlation) of selected oocyst/sporozoite-specific genes and their corresponding proteins that were differentially regulated after the depletion of MORC for different combinations of strain and induction. The abundance of their transcripts in the various stages of development, namely tachyzoite, bradyzoite/cyst, merozoite, EES and oocyst is shown. The colour scale indicates log₂-transformed fold changes. Arrowheads highlight genes of interest. **b**, Expression of a sporozoite-specific protein after depletion of MORC in the RH strain was measured by IFA. Parasites were probed with antibodies against G1/19 (monoclonal-antibody-reactive sporozoite-specific protein; red) and GAP45 (green). **c**, Genome-wide MORC and HDAC3 occupancy profiles at sporozoite-specific genes ($n=264$) in Pru MORC KD parasites that were either untreated or treated with IAA for 30 h. **d**, Expression levels of bradyzoite- and cyst-specific markers after depletion of MORC in the RH strain were measured by IFA. Fixed and permeabilized parasites were probed with antibodies against CC2, BCLA or BAG1 (green) along with DBA lectin (red). **e**, Time-course analysis of the expression levels of bradyzoite-specific (BCLA and BAG1) and merozoite-specific (TGME49_243940) markers after depletion of MORC–mAID–HA. Samples were taken at the indicated time periods after addition of IAA and were probed with antibodies against HA, BCLA, BAG1 and mCherry. The experiment was repeated three times and a representative blot is shown. **f,g**, IFA of HFFs that were infected with MORC–HA–mAID RH parasites expressing a targeted knock-in merozoite-specific gene (TGME49_243940). Immunofluorescence analysis of individual vacuoles after depletion of MORC, with a co-staining using a bradyzoite marker (BCLA, green) or a merozoite marker (red, TGME49_243940–mCherry). **h**, The differential expression of the two markers was assessed by counting the number of vacuoles ($n=100$) containing TGME49_243940–mCherry, BCLA or both markers over a time course. The experiments in **b**, **d**, **f** and **g** were conducted more than three times and representative images are displayed. Scale bars, 10 μ m (**b** and **g**); and 5 μ m (**d** and **f**).

of these potential regulators, AP2IV-3—which is clearly induced in the MORC-depleted cell (Fig. 6b, Extended Data Fig. 10a)—shares a homologous DNA-binding domain with AP2-G (Extended Data

Fig. 10b,c), the master transcriptional regulator of gametocytogenesis in *P. falciparum*³⁷, therefore underlining a possible convergence of function within the phylum. After depletion of MORC, the





serial expression of as many as 17 AP2 TFs, including TFs of which expression is restricted to micro- and macrogametes or sporozoites (Fig. 6a), suggests a sequential transcriptional modulation that is

mediated by AP2 TFs throughout the life cycle (Fig. 6g). Although the functions of AP2 TFs as activators or repressors remain to be determined, they still represent tenable candidates for

Fig. 6 | MORC guides developmental trajectories recruiting downstream regulating pathways. **a**, Heat map showing mRNA hierarchical clustering analysis (Pearson correlation) of AP2 TFs shown to be regulated by MORC depletion. The abundance of their transcripts in the various stages of development, namely tachyzoite, bradyzoite/cyst, merozoite, EES and oocyst stages, is displayed. The colour scale indicates \log_2 -transformed fold changes. The AP2 factors were colour-coded according to the colour of the life stage during which they are most expressed. **b**, IFA of HFFs that were infected with parasites harbouring an endogenously Flag-tagged merozoite-specific AP2IV-3 within the RH MORC-mAID-HA lineage. Fixed and permeabilized parasites were probed with antibodies against Flag (red) and Hoechst DNA-specific dye (blue). Scale bar, 10 μ m. The experiment was conducted more than three times and representative images are displayed. **c**, Heat map showing mRNA hierarchical clustering analysis (Pearson correlation) of selected MORC-regulated genes for different combinations of strain and induction. The presented genes belong to the set that was predefined as exclusively expressed in tachyzoites and repressed in merozoites. The colour scale indicates \log_2 -transformed fold changes. **d**, Genome-wide H4K31me1 enrichment profiles at cluster-1 genes in Pru MORC KD parasites that were either untreated or treated with IAA for 30 h. **e**, The average nucleosome occupancy plotted on the basis of the relative distance to the nearest annotated ATG, using Danpos2 (ref. 49) to define the most likely position of the nucleosomes. Nucleosomes occupancy at all *T. gondii* genes (top) or on the MORC-regulated cluster-1 genes (bottom) in the context of MORC depletion are shown. **f**, Density profiles are displayed at the merozoite-specific *TGME49_238915* gene. The occupancies of MORC and HDAC3, the H4K31 methylation enrichment, and the corresponding transcript level (black) are all displayed before and after the depletion of MORC. Similarly, the surrounding nucleosome occupancy and distribution (purple) are displayed on the basis of the MNase-seq data that were conducted before and after MORC degradation. Chromosomal positions are indicated on the x axis. The experiment was repeated independently twice with similar results. **g**, A proposed model describing the mechanism that underlies the MORC-complex-mediated modulations of the *T. gondii* life cycle. B, bradyzoite; M, merozoite; O, oocyst; S, sporozoite; T, tachyzoite.

understanding the priming for a permissive or for a transcriptionally silent state of chromatin, respectively. In this regard, chromatin opening was synergistically enhanced by the deposition of H4K31me1 on gene bodies after the depletion of MORC (Fig. 6d,f, Supplementary Fig. 6a–d), a mark that is associated with the progression of the elongating RNA polymerase II in *T. gondii*¹⁷. Nucleosome positioning also plays a critical role in the regulation of gene transcriptional³⁸. In fact, genome-wide nucleosome mapping enabled us to explore the genomic properties of the chromatin after degradation of MORC. Overall, nucleosome distribution was not affected by the depletion of MORC; however, the occupancy of nucleosomes—especially in 5' upstream gene regions (Supplementary Fig. 7a–f)—increased considerably within the cluster-1 gene group (Fig. 6e), and this correlated positively with an increase in the mRNA levels of the downstream genes (Fig. 6f). Importantly, the micrococcal nuclease (MNase)-protected footprints gave rise to a similar increase in patterns regarding the pan-acetylated histone H4 ChIP-control (Supplementary Fig. 7g). This is counterintuitive, as TFs would usually bind to nucleosome-depleted regions, and the genes with higher expression would tend to have greater degrees of nucleosomes depletion at their promoter regions. However, some exceptions have been reported, such as the pioneering factors that bind preferentially to regions of high nucleosome occupancy^{38–40}. This apparent increase of occupancy is further deliberated in the Supplementary Discussion. These results considered, we provide evidence that MORC is a critical determinant of nucleosome positioning; however, whether this effect stems from its intrinsic ATP-dependent remodelling activity, as reported in human and plant cells^{6,10,41}, remains to be determined.

Discussion

Here we provide compelling evidence that MORC is a crucial regulator on multiple levels throughout the life cycle of *T. gondii*. The acute degradation of MORC enabled us to map, in great detail, the gene expression changes that occur during the sexual development of *T. gondii*, revealing a role for MORC and for HDAC3 in gene silencing. We generated a model in which MORC establishes multiple partnerships with HDAC3, along with primary AP2 TFs (Extended Data Fig. 1a), to control genetic regulatory circuits; these genetic regulatory circuits consist of clusters of secondary AP2 TFs (Fig. 6a) that, in turn, influence the gene expression of the particular stage that they belong to (Fig. 6g). The diversity of the MORC–AP2 TF complexes might therefore hold a layer of regulation of the differential cell fate switches. How MORC is implicated in the cell fate decision remains enigmatic. This could be adjusted by the external

cues from the different host cell environments that the parasite faces throughout its life cycle, potentially leading to better adapted genetic reprogramming of the different stage-specific signatures (see Supplementary Discussion).

The study of the sexual stages of *T. gondii* is associated with technical difficulties, and the demand for alternatives to cat models is increasing owing to the rising ethical challenges⁴². Within the same vein of stem cell research, this study paves the way for in vitro epigenetic reprogramming of the *T. gondii* sexual development, using MORC as a cornerstone.

Methods

Parasites and human cell culture. HFFs (ATCC CCL-171) were cultured in Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 10% heat inactivated fetal bovine serum (Invitrogen), 10 mM HEPES buffer pH 7.2, 2 mM L-glutamine and 50 μ g ml⁻¹ of penicillin and streptomycin (Invitrogen). Cells were incubated at 37 °C in 5% CO₂. The *Toxoplasma* strains that were used in this study (listed in Supplementary Table 2) were maintained in vitro by serial passage on monolayers of HFFs. The cultures were free of mycoplasma, as determined by qualitative PCR.

Reagents. The following primary antibodies were used in the immunofluorescence, immunoblotting and/or ChIP assays: rabbit home-made anti-BCLA, rabbit anti-TgHDAC3 (RRID, AB_2713903), rabbit anti-H4K31ac (RRID, AB_2811024), rabbit anti-H4K31me1 (RRID, AB_2811025), rabbit anti-TgGAP45 (gift from D. Soldati, University of Geneva), rat anti-CC2 (gift from W. Bohne and U. Gross), mouse anti-TgBAG1, mouse anti-G1/19 (gift from G. Schares, Friedrich-Loeffler-Institut), mouse anti-HA tag (Roche; RRID, AB_2314622), rabbit anti-HA Tag (Cell Signaling Technology; RRID, AB_1549585), rabbit anti-mCherry (Cell Signaling Technology; RRID, AB_2799246), rabbit anti-Flag (Cell Signaling Technology; RRID, AB_2798687), rabbit anti-H3K14ac (Diagenode; RRID, AB_2713906), H3K9me3 (Millipore; RRID, AB_916348), H3K4me3 (Diagenode; RRID, AB_2616052), rabbit anti-acetyl-histone H4, pan (Lys 5, 8 and 12; Millipore; RRID, AB_310270). Immunofluorescence secondary antibodies were coupled with Alexa Fluor 488 or Alexa Fluor 594 (Thermo Fisher Scientific). Secondary antibodies used in western blotting were conjugated to alkaline phosphatase (Promega) or horseradish peroxidase.

Auxin-induced degradation. Depletion of MORC–AID–HA was achieved using IAA (Sigma-Aldrich, 45533) as described previously¹⁹. A stock of 500 mM IAA dissolved in 100% ethanol at 1:1,000 was used to deplete mAID-tagged proteins at a final concentration of 500 μ M. Mock treatment consisted of an equivalent volume of 100% ethanol at a final concentration of 0.0789%, w/v. To monitor the degradation of AID-tagged proteins, parasites grown in HFF monolayers were treated with auxin or ethanol alone for different time intervals at 37 °C. After treatment, parasites were collected and analysed using immunofluorescence or western blotting. Complete clearance of MORC was successful after 2 h and 16 h of IAA treatment in the type I (RH) and type II (Pru), respectively. For the RNA-seq analysis, IAA induction in type I took place for both 7 h and 24 h before the RNA extraction, and in type II the timings were fixed to 24 h and 48 h after treatment. For the proteome-wide analysis, the RH strain was treated for 24 h and the Pru

strain was treated for 36 h. For the ChIP-seq analysis, the RH strain was treated for 24 h in RH strain and the Pru strain was treated for 32 h.

HDAC3 inhibition using FR235222. For the RNA-seq assays, FR235222 was used at a final concentration of 25 nM for 7 h and 18 h; for the immunofluorescence assay of reporter genes, FR235222 was used at 100 nM for 18–24 h; and, for the proteome, the concentration was fixed at 25 nM and proteins were extracted 24 h after treatment.

Immunofluorescence microscopy. *T. gondii* infecting HFF cells grown on coverslips were fixed in 3% formaldehyde for 20 min at room temperature, permeabilized with 0.1% (v/v) Triton X-100 for 15 min and blocked in phosphate-buffered saline (PBS) containing 3% (w/v) BSA. The cells were then incubated for 1 h with primary antibodies followed by the addition of secondary antibodies that were conjugated to Alexa Fluor 488 or 594 (Molecular Probes). Nuclei were stained for 10 min at room temperature with Hoechst 33258 (2 µg ml⁻¹) in PBS. After four washes in PBS, the coverslips were mounted onto a glass slide with Mowiol mounting medium, and the images were acquired using a fluorescence ZEISS Aptomte.2 microscope and processed using the ZEN software (Carl Zeiss).

Western blot analysis. Immunoblot analysis of protein was performed as described previously⁴⁵. In brief, ~10⁷ cells were lysed in 50 µl lysis buffer (10 mM Tris-HCl, pH 6.8, 0.5% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (v/v), 10% glycerol (v/v), 1 mM EDTA and protease inhibitors cocktail) and sonicated. Proteins were separated using SDS-PAGE, transferred to a polyvinylidene fluoride membrane (Immobilon-P; EMD Millipore) by liquid transfer and western blots were probed using the appropriate primary antibodies followed by alkaline phosphatase or horseradish peroxidase-conjugated goat secondary antibodies. Signals were detected using NBT-BCIP (Amresco) or an enhanced chemiluminescence system (Thermo Fisher Scientific).

Plasmid construction. A list of the plasmids and primers for genes of interest (GOIs) that were used in this study is provided in Supplementary Table 2. To construct the vector pLIC-GOI-HAFlag, the coding sequence of GOI was amplified with the primers LIC-GOI-Fwd and LIC-GOI-Rev, using *T. gondii* genomic DNA as a template. The resulting PCR product was cloned into the pLIC-HA-Flag-dhfr or pLIC-mCherry-dhfr vectors using the ligation independent cloning method⁴³. The plasmid pTOXO_Cas9-CRISPR was described previously⁴⁴. We cloned 20-mer oligonucleotides corresponding to specific GOIs using the Golden Gate strategy. In brief, primers GOI-gRNA-Fwd and GOI-gRNA-Rev containing the sgRNA targeting GOI genomic sequence were phosphorylated, annealed and ligated into the pTOXO_Cas9-CRISPR plasmid linearized with BsaI, leading to pTOXO_Cas9-CRISPR:sgGOI. Only two transgenic components are needed to implement the auxin-inducible decon (AID) system—a plant auxin receptor called transport inhibitor response 1 (TIR1) and a POI tagged with an AID. We engineered a type I RHΔ*ku80* and a type II line of *T. gondii* to stably express *Tir1* from *O. sativa* tagged with Ty and controlled by a promoter, which we selected for a moderate expression of the chimeric protein. The plasmid pModProm1-Tir1-TY1-3DHFR (the DNA sequence of which is provided in Supplementary Table 2) was DNA-synthesized and then cloned into pUC57-Simple using Genscript. The chimeric construct was inserted within the UPRT locus. We also created a pLIC vector containing a DNA block that was codon-optimized for *T. gondii* expression with the mAID from *A. thaliana* auxin-responsive protein IAA17^{566–5133}, as defined previously¹⁹, in frame with an HA tag (Supplementary Table 2).

Transfection of *T. gondii*. *T. gondii* strains were electroporated with vectors in cytomix buffer (120 mM KCl, 0.15 mM CaCl₂, 10 mM K₂HPO₄, 10 mM KH₂PO₄, pH 7.6, 25 mM HEPES pH 7.6, 2 mM EGTA and 5 mM MgCl₂) using a BTX ECM 630 machine (Harvard Apparatus). Electroporation was performed in a 2 mm cuvette at 1.100 V, 25 Ω and 25 µF. When needed, the following antibiotics were used for drug selection: chloramphenicol (20 µM), mycophenolic acid (25 µg ml⁻¹) with xanthine (50 µg ml⁻¹), pyrimethamine (3 µM) or 5-fluorodeoxyuracil (10 µM). Stable transgenic parasites were selected with the appropriate antibiotic, single-cloned in 96-well plates by limiting dilution, and verified by immunofluorescence assay or genomic analysis.

Chromatographic purification of MORC-containing complex. *T. gondii* extracts from RHΔ*ku80* or PruΔ*ku80* cells stably expressing HA-Flag-tagged MORC, were incubated with anti-Flag M2 affinity gel (Sigma-Aldrich) for 1 h at 4 °C. Beads were washed with 10-column volumes of BC500 buffer (20 mM Tris-HCl pH 8.0, 500 mM KCl, 20% glycerol, 1 mM EDTA, 1 mM dithiothreitol, 0.5% NP-40 and protease inhibitors). Bound polypeptides were eluted stepwise with 250 µg ml⁻¹ Flag peptide (Sigma-Aldrich) diluted in BC100 buffer. For size-exclusion chromatography, protein eluates were loaded onto a Superose 6 HR 10/30 column equilibrated with BC500. Flow rate was fixed at 0.35 ml min⁻¹, and fractions of 0.5 ml were collected.

MS-based interactome analyses. Protein bands were excised from gels stained with colloidal blue (Thermo Fisher Scientific) before in-gel digestion using

modified trypsin (Promega, sequencing grade) as previously described⁴⁵. The resulting peptides were analysed by online nanoliquid chromatography (UltiMate 3000 and UltiMate 3000 RSLCnano (Thermo Fisher Scientific) for the RH and Pru strains, respectively) coupled to tandem MS (MS/MS; LTQ-Orbitrap Velos Pro and Q-Exactive Plus (Thermo Fisher Scientific) for the RH and Pru strains, respectively). The peptides were sampled on a 300 µm × 5 mm PepMap C18 precolumn and separated on a 75 µm × 250 mm C18 PepMap column (Thermo Fisher Scientific) using gradients of 25 min. MS and MS/MS data were acquired using Xcalibur (Thermo Fisher Scientific). Peptides and proteins were identified using Mascot (v.2.6) through concomitant searches against the *Toxoplasma gondii* database (ME49 taxonomy, v.30, downloaded from ToxoDB⁴⁶), the UniProt database (*H. sapiens* taxonomy, March 2019 version), the classical contaminant database and the corresponding reversed databases. Trypsin was chosen as the enzyme and two missed cleavages were allowed. Precursor and fragment mass error tolerances were set at 10 ppm and 0.6 Da, respectively. Peptide modifications allowed during the search were: carbamidomethyl (C, fixed), acetyl (protein N-term, variable) and oxidation (M, variable). The Proline software (<http://proline.profiptomics.fr>) was used to filter the results: conservation of rank 1 peptide spectrum matches (PSMs), PSM homology threshold ≤ 0.01, PSM score ≥ 25 and a minimum of 1 specific peptide per identified protein group. Proline was then used to perform a compilation and grouping of the protein groups identified in the different samples. Proteins from the contaminant database were discarded from the final list of identified proteins. MS1-based label-free quantification of the protein groups was performed using Proline to infer intensity-based absolute quantification values that were used to rank identified *Toxoplasma* proteins in the interactomes. The MS proteomics data have been deposited to the ProteomeXchange Consortium through the PRIDE partner repository with the dataset identifier PXD016846.

MS-based proteome-wide analyses. HFF cells were grown to confluence and infected with type I (RHΔ*ku80* or RHΔ*ku80* MORC KD) or type II (PruΔ*ku80* or PruΔ*ku80* MORC KD) strains before lysis in 8 M urea and 50 mM HEPES. Extracted proteins were reduced using 20 mM of dithiothreitol for 1 h at 37 °C before alkylation with 55 mM of iodoacetamide for 45 min at room temperature in the dark. The samples were then diluted using ammonium bicarbonate to obtain a urea concentration of 4 M. Proteins were digested with LysC (Promega) at a ratio of 1:200 for 4 h at 37 °C. The samples were diluted again using ammonium bicarbonate to obtain a urea concentration of 1 M. Proteins were then digested with trypsin (Promega) at a ratio of 1:200 overnight at 37 °C. The resulting peptides were purified using C18 reverse phase chromatography (Sep-Pak C18, Waters) and dried before fractionation by tip-based strong cation exchange (3M Empore). For this, peptides were dissolved in 5% acetonitrile, 1% trifluoroacetic acid (TFA) and eluted in 4 fractions (F1: 100 mM ammonium acetate, 20% acetonitrile (ACN) and 0.5% formic acid; F2: 175 mM ammonium acetate, 20% ACN and 0.5% formic acid; F3: 375 mM ammonium acetate, 20% ACN and 0.5% formic acid; F4: 80% acetonitrile and 5% ammonium hydroxide) before desalting using C18 reverse-phase chromatography (Ultra-Micro SpinColumns, Harvard Apparatus). Technical triplicates were performed. NanoLC-MS/MS analyses were performed using an Ultimate 3000 RSLCnano coupled to a Q-Exactive Plus (Thermo Fisher Scientific). Peptides were sampled on a 300 µm × 5 mm PepMap C18 precolumn and separated on a 75 µm × 250 mm PepMap C18 column (Thermo Fisher Scientific) using a gradient of 120 min. MS and MS/MS data were acquired using the Xcalibur software (Thermo Fisher Scientific). RAW files were processed using MaxQuant⁴⁸ v.1.6.2.10. Spectra were searched against the *Toxoplasma gondii* database (ME49 taxonomy, v.30 downloaded from ToxoDB⁴⁶), the UniProt database (*H. sapiens* taxonomy, April 2019 version), the frequently observed contaminants database embedded in MaxQuant⁴⁷ and the corresponding reverse databases. Trypsin was chosen as the enzyme and two missed cleavages were allowed. Precursor and fragment mass error tolerances were set at their default values. Peptide modifications allowed during the search were: carbamidomethyl (C, fixed), acetyl (Protein N-term, variable) and oxidation (M, variable). The minimum number of peptides and razor + unique peptides was set to 1. Maximum false-discovery rates (FDRs) were set to 0.01 at the PSM and protein levels. The match between runs option was activated. Statistical analyses were performed using ProStaR⁴⁹. Peptides and proteins that were identified in the reverse and contaminant databases or matched to human sequences were discarded. Only proteins quantified in at least three replicates of one condition were conserved. After log₂ transformation, protein intensities were normalized using the summed intensities of a pool of proteins for which transcripts were found to be invariable in RNA-seq dataset (0.91 < fold change < 1.1). Missing values were then imputed (slsa method for partially observed values and DetQuantile with quantile set to 2.5 and factor set to 1 for missing entirely in the condition). Statistical testing was conducted using limma. Differentially abundant proteins were sorted using the following cut-offs: log₂-transformed fold change of ≥ 0.8 or ≤ -0.8 and *P* values were allowed to reach an FDR of ~1% according to the Benjamini-Hochberg estimator. The MS proteomics data have been deposited to the ProteomeXchange Consortium through the PRIDE partner repository with the dataset identifier PXD016845.

ChIP and next-generation sequencing in *T. gondii*. HFF cells were grown to confluence and infected with type II (Pru Δ Ku80 or Pru Δ Ku80 MORC KD) strains. The collected intracellular parasites were crosslinked with formaldehyde (final concentration, 1%) for 8 min at room temperature and the crosslinking was stopped by addition of glycine (final concentration, 0.125 M) for 5 min at room temperature. Cross-linked chromatin was lysed in ice-cold lysis buffer (50 mM HEPES KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.125% Triton X-100 and protease inhibitor cocktail) and sheared in shearing buffer (1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0 and protease inhibitor cocktail) by sonication using a Diagenode Biorupter. The samples were sonicated for 16 cycles (30 s on and 30 s off) to an average size of 200–500 bp. Immunoprecipitation was performed using sheared chromatin, 5% BSA, protease inhibitor cocktail, 10% Triton X-100, 10% deoxycholate, DiaMag Protein A-coated magnetic beads (Diagenode) and antibodies against histone PTM or protein of interest. A rabbit IgG antiserum was used as a control mock. After overnight incubation at 4 °C on a rotating wheel, chromatin–antibody complexes were washed and eluted from beads by using the iDeal ChIP-seq kit for Histones (Diagenode) according to the manufacturer's protocol. Samples were decrosslinked by heating for 4 h at 65 °C. DNA was purified by using the IPure kit (Diagenode) and quantified using Qubit Assays (Thermo Fisher Scientific) according to the manufacturer's protocol. For ChIP-seq, purified DNA was used to prepare libraries and then sequenced by Arraystar.

Library preparation, sequencing and data analysis by Arraystar. ChIP-seq library preparation was performed according to Illumina's protocol for preparing samples for ChIP sequencing of DNA.

Library preparation. DNA (10 ng) of each sample was converted to phosphorylated blunt-ended using T4 DNA polymerase, Klenow polymerase and T4 polymerase (NEB). An A base was added to the 3' end of the blunt phosphorylated DNA fragments using the polymerase activity of Klenow (exo minus) polymerase (NEB). Illumina's genomic adapters were ligated to the A-tailed DNA fragments. PCR amplification was performed to enrich ligated fragments using Phusion High Fidelity PCR Master Mix with HF Buffer (Finnzymes Oy). The enriched product of ~200–700 bp was cut out from the gel and purified.

Sequencing. The library was denatured using 0.1 M NaOH to generate single-stranded DNA molecules, and loaded onto channels of the flow cell at a concentration of 8 pM and amplified in situ using TruSeq Rapid SR cluster kit (GD-402–4001, Illumina). Sequencing was performed by running 100 cycles on an Illumina HiSeq 4000 system according to the manufacturer's instructions.

Data analysis. After the sequencing platform generated the sequencing images, the stages of image analysis and base calling were performed using the Off-Line Basecaller software (OLB v.1.8). After passing Solexa CHASTITY quality filter, the clean reads were aligned to the *T. gondii* reference genome (Tgo) using BOWTIE (v.2.1.0). Aligned reads were used for peak calling of the ChIP regions using MACS v.1.4.0. Statistically significant ChIP-enriched regions (peaks) were identified by comparison of two samples, using a *P*-value threshold of 10^{-5} . The peaks in each sample were then annotated by the overlapped gene using the newest *T. gondii* database. The EXCEL/BED format file containing the ChIP-enriched regions was generated for each sample.

Data visualization. The mapped 100-bp reads represent DNA fragments enriched by the ChIP experiment. Any region of interest in the raw ChIP-seq signal profile can be directly visualized in a genome browser. We use resolution intervals of 10 bp (10-bp bins) to partition the stacked reads region, and count the number of reads in each bin. All of the 10-bp resolution ChIP-seq profiles of each sample are saved as UCSC WIG format files, which can be visualized in the *T. gondii* Genome Browser (http://protists.ensembl.org/Toxoplasma_gondii/Info/Index). All of these raw and processed files can be found at Gene Expression Omnibus (GEO) Series GSE136060.

RNA-seq and sequence alignment. Total RNAs were extracted and purified using TRIzol (Invitrogen) and the RNeasy Plus Mini Kit (Qiagen). RNA quantity and quality were measured using a NanoDrop 2000 (Thermo Fisher Scientific). RNA integrity was assessed using standard non-denaturing 1.2% Tris–borate–EDTA (TBE) agarose gel electrophoresis. The ribosomal large subunits from *Toxoplasma* and the host cells was used to verify that the ratio between *Toxoplasma* RNA versus host RNA was equivalent between the different biological samples, therefore indicating that the samples had equivalent infection rates. For each condition, total RNAs from two independent biological replicates were pooled to prepare cDNA libraries, which were then sequenced using Illumina technology in a single replicate dataset. RNA-seq was performed by GENEWIZ. In brief, the RNA quality was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies) and Illumina TruSeq RNA library preparation and sequencing reagents were used according to the manufacturer's recommendations (Illumina). The samples were paired-end multiplex sequenced (2 × 125 bp) using an Illumina HiSeq 2500 platform and generated at least 70 million reads for each sample. The RNA-seq reads (FASTQ) were processed and analysed using the Lasergene Genomics Suite

v.14 (DNASTAR) with the default parameters. The paired-end reads were uploaded onto the SeqMan NGen (v.14, DNASTAR) platform for reference-based assembly using either the *Mus musculus* genome package (GRCm38.p3) or the *Toxoplasma* Type II ME49 strain (ToxoDB-24, ME49 genome) as the reference template. The ArrayStar module (v.14, DNASTAR) was used for normalization, differential gene expression and statistical analysis of uniquely mapped paired-end reads using the default parameters. The quantification and normalization of expression data were calculated using the RPKM normalization method. All of these raw and processed files are provided at GEO Series GSE136123.

MNase digestion. Formaldehyde crosslinking was performed for 15 min and the optimal MNase digestion time was determined. Nuclei were mildly sonicated to free the crosslinked nucleosomes from the nuclear membranes. Soluble chromatin was collected and decrosslinked directly (MNase-seq) or used for ChIP. Chromatin was then resuspended in 1 ml of prewarmed buffer M containing 320 mM of sucrose and 1 mM CaCl₂, after which 1 U MNase (Sigma-Aldrich) was added. The MNase incubation time was set to 15 min at 37 °C and the reaction was stopped by adding EDTA (10 μ l of 0.5 M per 1 ml of extract). DNA was extracted from nucleosomes by eliminating using proteinase K for 1 h at 37 °C, followed by a standard phenol–chloroform–isoamyl alcohol DNA extraction and the DNA was then NaCl–ethanol precipitated with the addition of a final glycogen precipitation step. DNA was then purified using the IPure kit (Diagenode) and quantified using Qubit Assays (Thermo Fisher Scientific) according to the manufacturer's protocol. Purified DNA was used to prepare libraries and was then sequenced by Arraystar. Positioned nucleosomes were called using the DANPOS (v.2.2.1) as described previously¹⁹.

Phylogenetic analysis of HATPase-domain-containing MORC proteins. The HATPase sequences were identified by a blastP analysis using the Swiss-Prot database. The sampled HATPase domains were aligned using ClustalW. The neighbour-joining tree was computed using ClustalX.

Software and statistical analyses. Volcano plots, scatter plots and histograms were generated using Prism 7. Sample sizes were chosen according to previous literature. All of the experiments were performed in biological replicates to enable statistical analyses. No method of randomization was used. All of the experiments were performed in independent biological replicates as stated for each experiment in the manuscript. All corresponding treatments and control samples from ChIP-seq and RNA-seq were processed at the same time to minimize technical variation. The investigators were not blinded during the experiments. The experiments were performed in biological replicates and provided consistent statistically relevant results.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data supporting the findings of this study are available from the corresponding author on reasonable request. The RNA-seq demultiplexed FASTQ files and gene-wise quantifications have been deposited in the NCBI Gene Expression Omnibus (GEO) and are accessible under accession number GSE136123. The ChIP-seq data have been deposited in the GEO under accession number GSE136060. The MS proteomics data have been deposited in the ProteomeXchange Consortium through the PRIDE partner repository with the dataset identifiers PXD016846 (MORC interactome) and PXD016845 (proteome-wide analyses). Source data for Figs. 1, 2 and 5 are presented with this paper.

Received: 22 August 2019; Accepted: 23 January 2020;

Published online: 24 February 2020

References

- Dubey, J. P., Lindsay, D. S. & Speer, C. A. Structures of *Toxoplasma gondii* tachyzoites, bradyzoites, and sporozoites and biology and development of tissue cysts. *Clin. Microbiol. Rev.* **11**, 267–299 (1998).
- Balaji, S. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.* **33**, 3994–4006 (2005).
- Kim, K. The epigenome, cell cycle, and development in *Toxoplasma*. *Annu. Rev. Microbiol.* **72**, 479–499 (2018).
- Saksout, N. et al. Histone-modifying complexes regulate gene expression pertinent to the differentiation of the protozoan parasite *Toxoplasma gondii*. *Mol. Cell. Biol.* **25**, 10301–10314 (2005).
- Bougdour, A. et al. Drug inhibition of HDAC3 and epigenetic control of differentiation in Apicomplexa parasites. *J. Exp. Med.* **206**, 953–966 (2009).
- Moissiard, G. et al. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **336**, 1448–1451 (2012).
- Lorković, Z. J. MORC proteins and epigenetic regulation. *Plant Signal. Behav.* **7**, 1561–1565 (2012).

8. Harris, C. J. et al. *Arabidopsis* AtMORC4 and AtMORC7 form nuclear bodies and repress a large number of protein-coding genes. *PLoS Genet.* **12**, e1005998 (2016).
9. Weiser, N. E. et al. MORC-1 integrates nuclear RNAi and transgenerational chromatin architecture to promote germline immortality. *Dev. Cell* **41**, 408–423 (2017).
10. Pastor, W. A. et al. MORC1 represses transposable elements in the mouse male germline. *Nat. Commun.* **5**, 5795 (2014).
11. Tchasovnikarova, I. A. et al. Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in *MORC2*. *Nat. Genet.* **49**, 1035–1044 (2017).
12. Kim, H. et al. The gene-silencing protein MORC-1 topologically entraps DNA and forms multimeric assemblies to cause DNA compaction. *Mol. Cell* **75**, 700–710 (2019).
13. Inoue, N. New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum. Mol. Genet.* **8**, 1201–1207 (1999).
14. Iyer, L. M., Abhiman, S. & Aravind, L. MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol. Direct* **3**, 8 (2008).
15. Iyer, L. M., Anantharaman, V., Wolf, M. Y. & Aravind, L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* **38**, 1–31 (2008).
16. Andrews, F. H. et al. Multivalent chromatin engagement and inter-domain crosstalk regulate MORC3 ATPase. *Cell Rep.* **16**, 3195–3207 (2016).
17. Sindikubwabo, F. et al. Modifications at K31 on the lateral surface of histone H4 contribute to genome structure and expression in apicomplexan parasites. *eLife* **6**, e29391 (2017).
18. Sidik, S. M. et al. A genome-wide CRISPR screen in *Toxoplasma* identifies essential apicomplexan genes. *Cell* **166**, 1423–1435 (2016).
19. Brown, K. M., Long, S. & Sibley, L. D. Plasma membrane association by N-acylation governs PKG function in *Toxoplasma gondii*. *mBio* **8**, e00375-17 (2017).
20. Pittman, K. J., Aliota, M. T. & Knoll, L. J. Dual transcriptional profiling of mice and *Toxoplasma gondii* during acute and chronic infection. *BMC Genom.* **15**, 806 (2014).
21. Hehl, A. B. et al. Asexual expansion of *Toxoplasma gondii* merozoites is distinct from tachyzoites and entails expression of non-overlapping gene families to attach, invade, and replicate within feline enterocytes. *BMC Genom.* **16**, 66 (2015).
22. Ramakrishnan, C. et al. An experimental genetically attenuated live vaccine to prevent transmission of *Toxoplasma gondii* by cats. *Sci. Rep.* **9**, 1474 (2019).
23. Fritz, H. M. et al. Transcriptomic analysis of *Toxoplasma* development reveals many novel functions and structures specific to sporozoites and oocysts. *PLoS ONE* **7**, e29998 (2012).
24. Behnke, M. S., Zhang, T. P., Dubey, J. P. & Sibley, L. D. *Toxoplasma gondii* merozoite gene expression analysis with comparison to the life cycle discloses a unique expression state during enteric development. *BMC Genom.* **15**, 350 (2014).
25. Ramakrishnan, C., Walker, R. A., Eichenberger, R. M., Hehl, A. B. & Smith, N. C. The merozoite-specific protein, TgGRA11B, identified as a component of the *Toxoplasma gondii* parasitophorous vacuole in a tachyzoite expression model. *Int. J. Parasitol.* **47**, 597–600 (2017).
26. Smith, E. F. PF16 encodes a protein with armadillo repeats and localizes to a single microtubule of the central apparatus in *Chlamydomonas* flagella. *J. Cell Biol.* **132**, 359–370 (1996).
27. Sapiro, R. et al. Male infertility, impaired sperm motility, and hydrocephalus in mice deficient in sperm-associated antigen 6. *Mol. Cell. Biol.* **22**, 6298–6305 (2002).
28. Straschil, U. et al. The armadillo repeat protein PF16 is essential for flagellar structure and function in plasmodium male gametes. *PLoS ONE* **5**, e12901 (2010).
29. Clark, T. HAP2/GCS1: mounting evidence of our true biological EVE? *PLoS Biol.* **16**, e3000007 (2018).
30. Mori, T., Hirai, M., Kuroiwa, T. & Miyagishima, S. The functional domain of GCS1-based gamete fusion resides in the amino terminus in plant and parasite species. *PLoS ONE* **5**, e15957 (2010).
31. Liu, Y. et al. The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev.* **22**, 1051–1068 (2008).
32. Angrisano, F. et al. Targeting the conserved fusion loop of HAP2 inhibits the transmission of *Plasmodium berghei* and *falciparum*. *Cell Rep.* **21**, 2868–2878 (2017).
33. Gondim, L. F. P. et al. Characterization of an IgG monoclonal antibody targeted to both tissue cyst and sporocyst walls of *Toxoplasma gondii*. *Exp. Parasitol.* **163**, 46–56 (2016).
34. Radke, J. R. et al. Identification of a sporozoite-specific member of the *Toxoplasma* SAG superfamily via genetic complementation: *T. gondii* sporozoite developmental antigens. *Mol. Microbiol.* **52**, 93–105 (2004).
35. Tomita, T. et al. The *Toxoplasma gondii* cyst wall protein CST1 is critical for cyst wall integrity and promotes bradyzoite persistence. *PLoS Pathog.* **9**, e1003823 (2013).
36. Radke, J. B. et al. ApiAP2 transcription factor restricts development of the *Toxoplasma* tissue cyst. *Proc. Natl Acad. Sci. USA* **110**, 6871–6876 (2013).
37. Poran, A. et al. Single-cell RNA sequencing reveals a signature of sexual commitment in malaria parasites. *Nature* **551**, 95–99 (2017).
38. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
39. Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.* **339**, 258–266 (2010).
40. Teif, V. B. et al. Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.* **19**, 1185–1192 (2012).
41. Li, D.-Q. et al. MORC2 signaling integrates phosphorylation-dependent, ATPase-coupled chromatin remodeling during the DNA damage response. *Cell Rep.* **2**, 1657–1669 (2012).
42. Wadman, M. Closure of U.S. *Toxoplasma* lab draws ire. *Science* **364**, 109 (2019).
43. Bougdour, A. et al. 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).
44. Sangaré, L. O. et al. Unconventional endosome-like compartment and retromer complex in *Toxoplasma gondii* govern parasite integrity and host infection. *Nat. Commun.* **7**, 11191 (2016).
45. Salvetti, A. et al. Nuclear functions of nucleolin through global proteomics and interactomic approaches. *J. Proteome Res.* **15**, 1659–1669 (2016).
46. Gajria, B. et al. ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Res.* **36**, D553–D556 (2007).
47. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
48. Wiczorek, S. et al. DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **33**, 135–136 (2017).
49. Chen, K. et al. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).

Acknowledgements

We thank the developers of the ToxoDB.org Genome Resource, ToxoDB and EuPathDB as part of the National Institutes of Health/National Institutes of Allergy and Infectious Diseases (NIH/NIAID)-funded Bioinformatics Resource Center. We also thank G. Communie for his help computing Danpos software. This work was supported by the Laboratoire d'Excellence (LabEx) ParaFrap (grant no. ANR-11-LABX-0024), the Agence Nationale pour la Recherche (Project HostQuest, grant no. ANR-18-CE15-0023 and ProFl grant no. ANR-10-INSB-08-01), the European Research Council (ERC Consolidator grant no. 614880 Hosting TOXO to M.-A.H.) and Fondation pour la Recherche Médicale (FRM grant no. FDT201904008364 to D.C.F.).

Author contributions

M.-A.H. supervised the research. M.-A.H. and D.C.F. generated the genetic tools, performed and analysed all of the genome-wide studies (RNA-seq, ChIP-seq and MNase assay), conducted most of biochemical experiments and the immunofluorescence assays. C.D., D.C. and F.S. generated biochemical reagents. L.B. and P.-J.D.B. performed MS analysis. Y.C. guided the MS experiments. C.S. performed structural modelling and analysed the MNase assay. A.B. computed and analysed the RNA-seq data. P.O. and M.B. analysed the ChIP-seq data. M.-A.H. and D.C.F. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41564-020-0674-4>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41564-020-0674-4>.

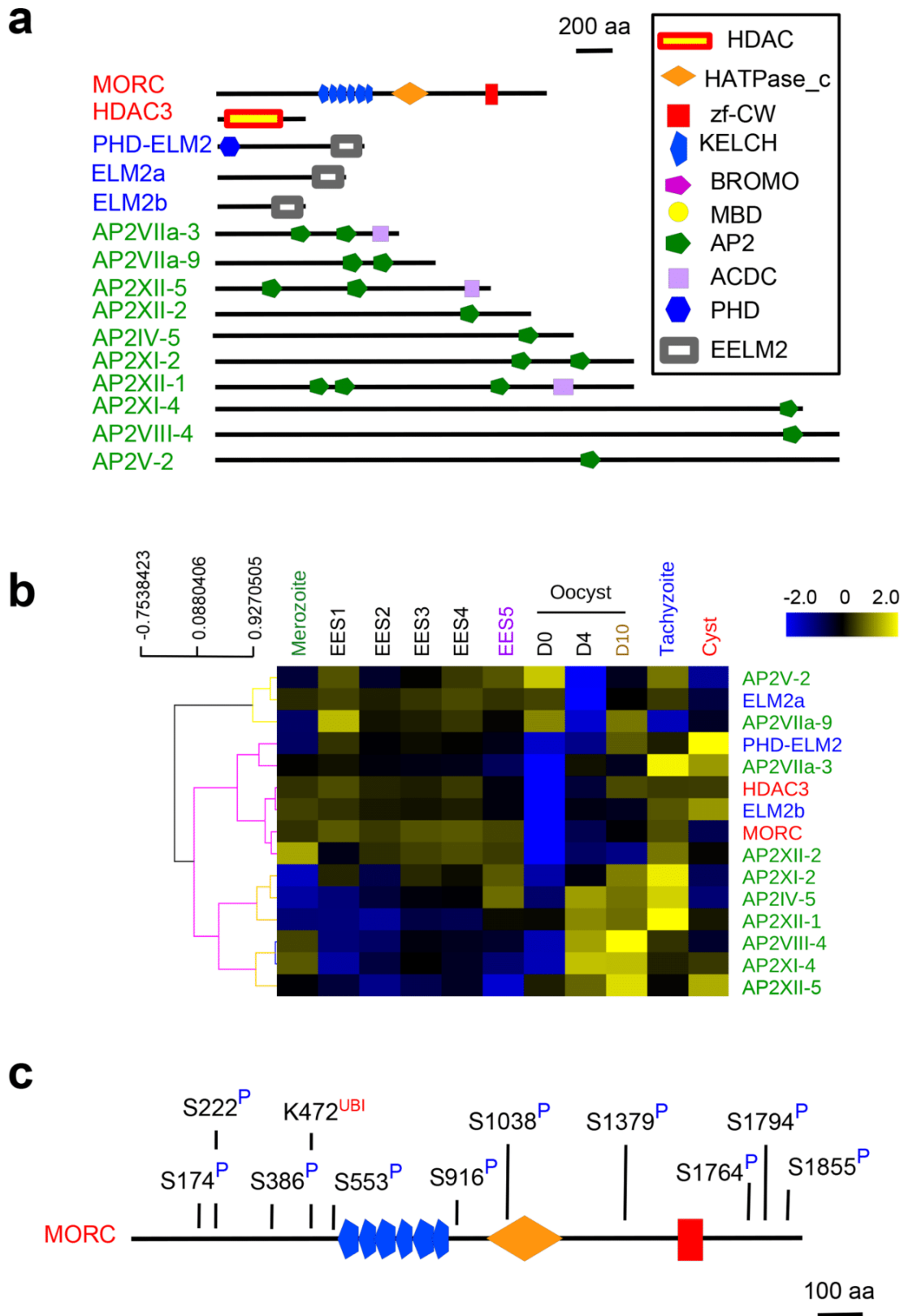
Correspondence and requests for materials should be addressed to M.-A.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

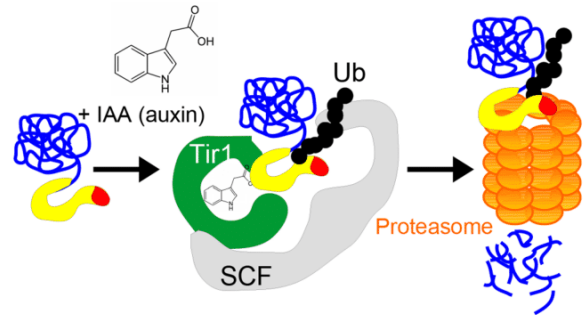
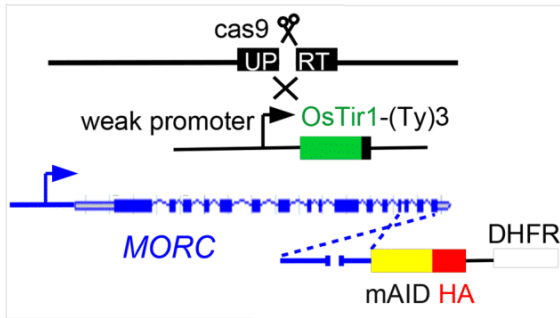
2. Extended Data



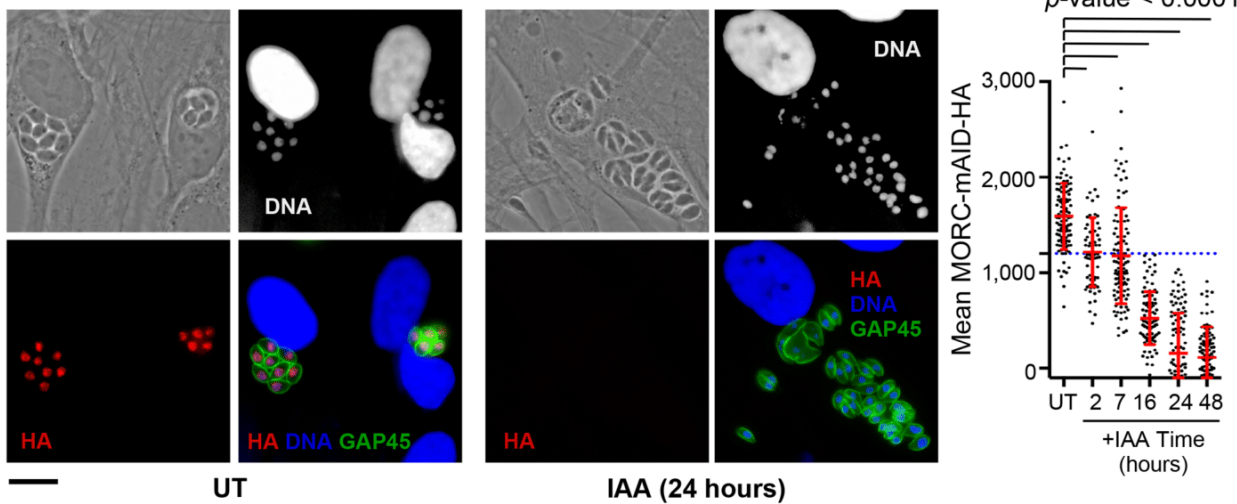
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Domain architectures and expression levels during the life cycle of MORC and partners. **a**, Representative domain architectures of *T. gondii* MORC and partners identified by mass spectrometry-based proteomic are shown. Domains were predicted by SMART and PFAM: ELM2 (Egl-27 and MTA1 homology 2), PHD (plant homeodomain), and AP2 (APETALA2) **b**, Heatmap representation of MORC and partners gene expression in different life cycle stages (source: ToxoDB). Gene expression values were mean log2 transformed and median centered for clustering. Transcriptomic data from tachyzoite, merozoite, longitudinal studies on enteroepithelial stages (EES1 to EES5), immature (day 0), maturing (day 4) and mature (day 10) stages of oocyst development and cysts from chronically infected mice were used. **c**, Mapping of domains and identified phosphorylation and ubiquitination sites detected by mass-spectrometry (source: ToxoDB).

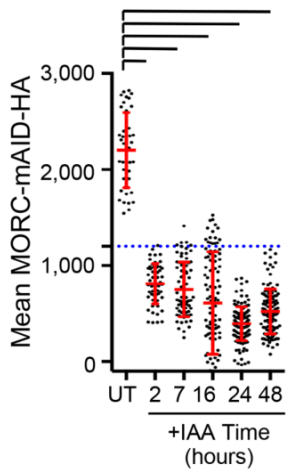
a RH $\Delta ku80$ or Pru $\Delta ku80$ Tir1-Ty MORC-mAID-HA KD



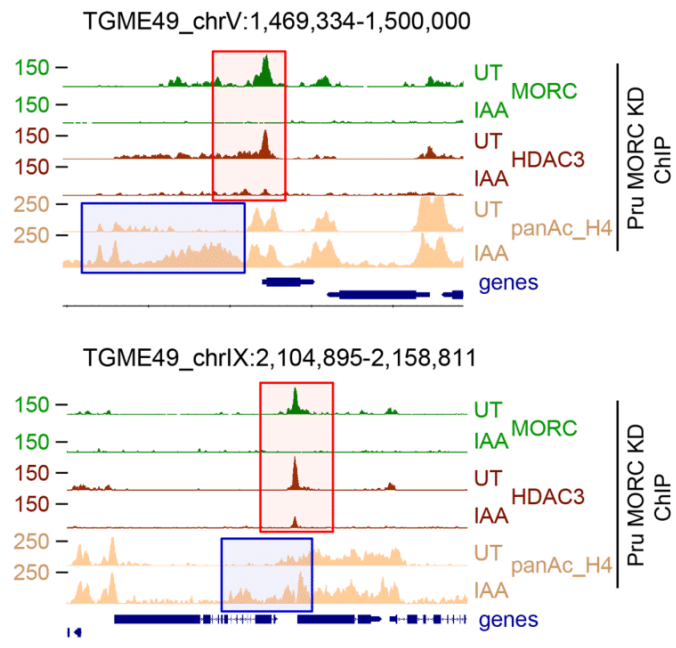
b Pru $\Delta ku80$ MORC-mAID-HA KD



c $p < 0.0001$

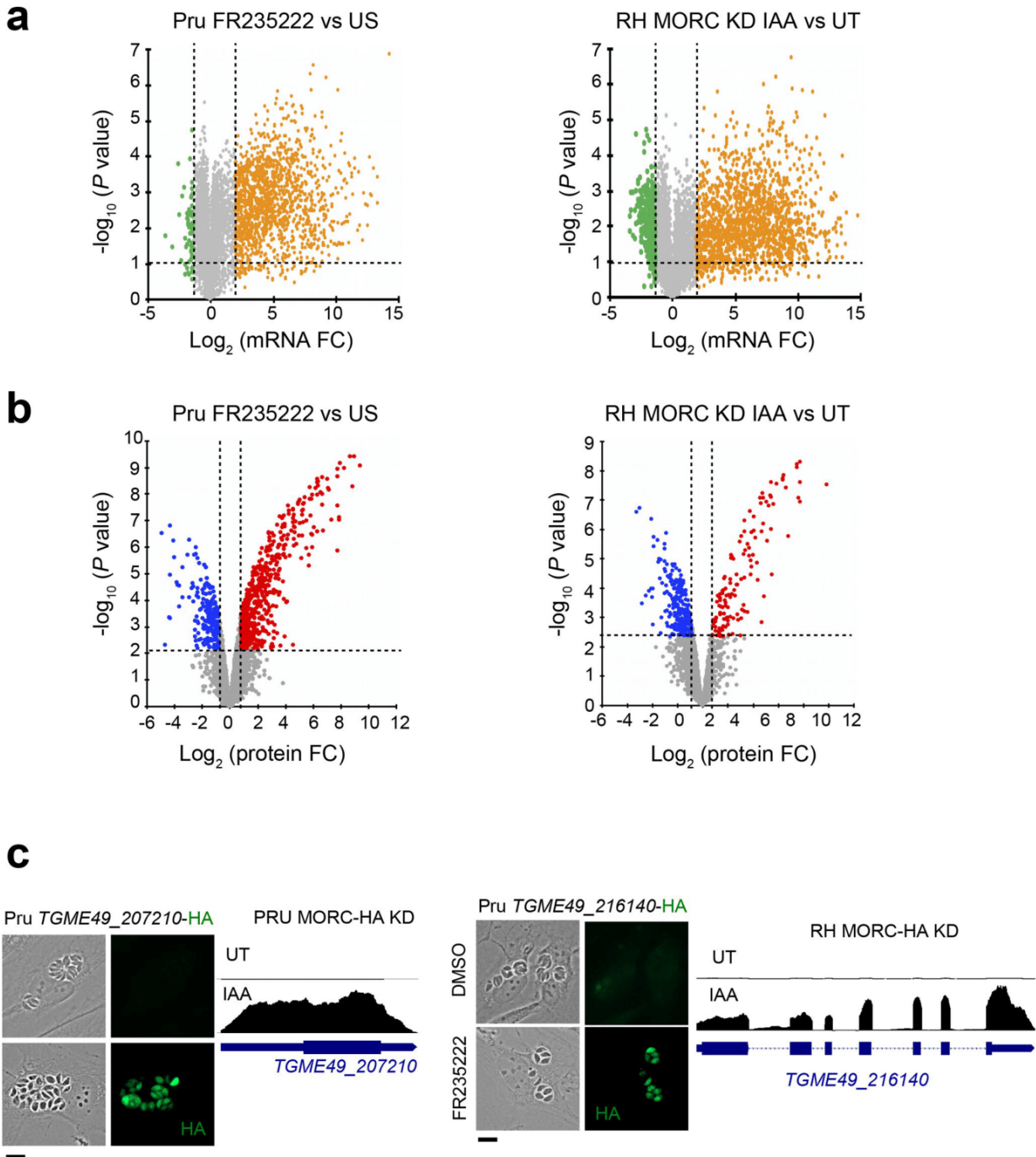


d

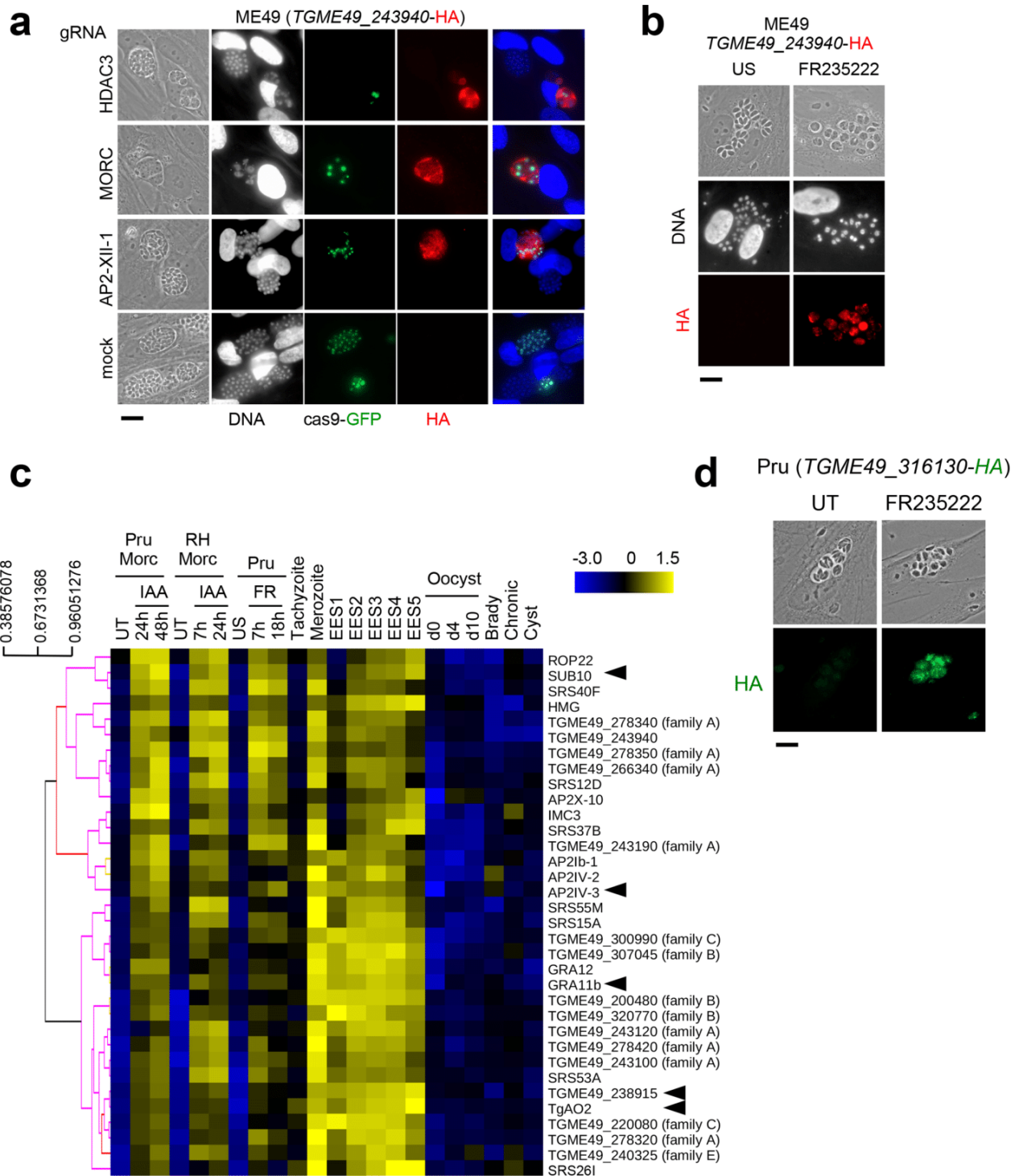


Extended Data Fig. 2 | See next page for caption.

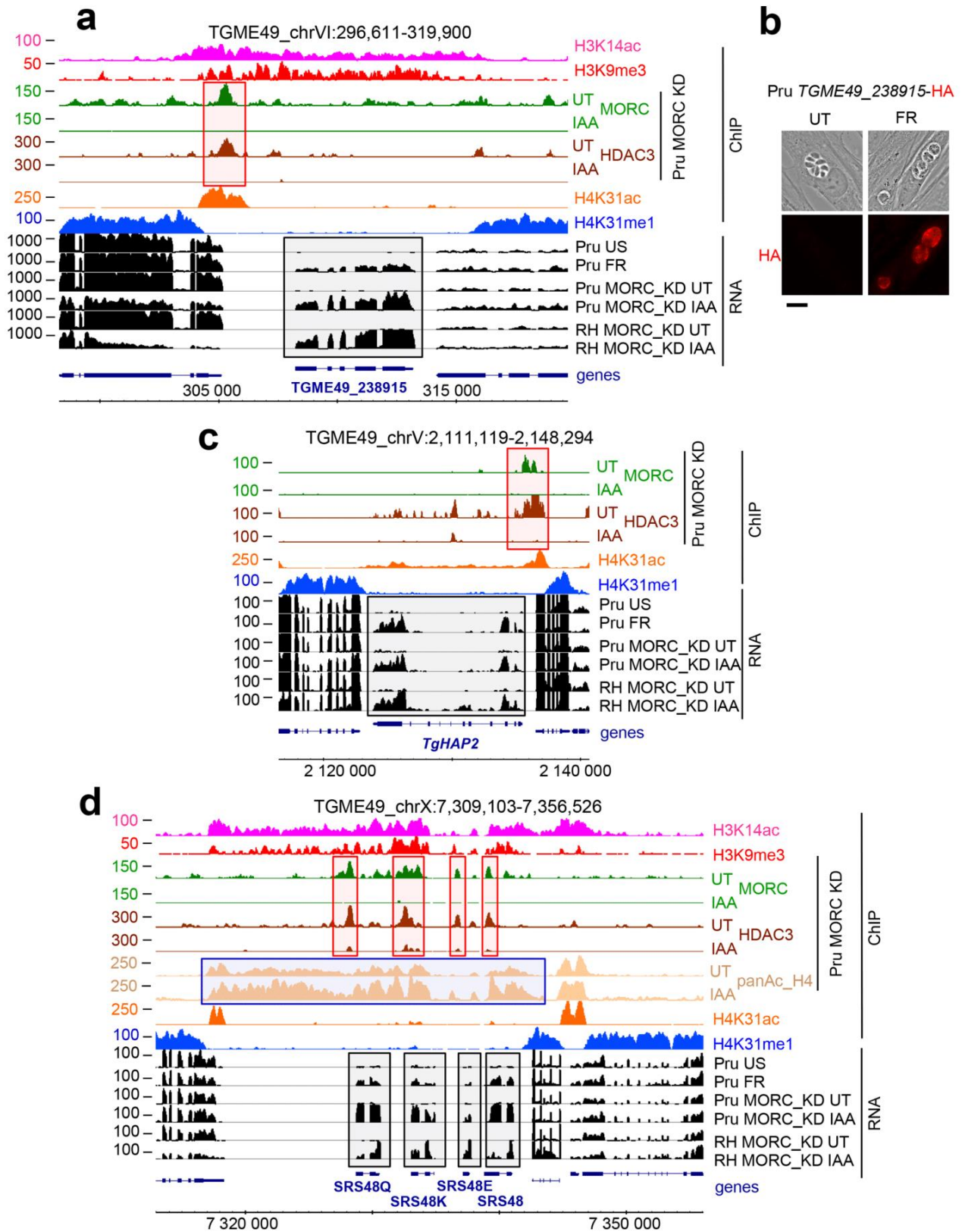
Extended Data Fig. 2 | mAID-based MORC inducible KD system successfully ablated its expression on the protein post-translational level in both type I (RH) and type II (Pru) strains. a, Auxin-inducible degradation system for controlling protein stability in *T. gondii*. We first engineered RH and Pru strains ectopically expressing the plant auxin receptor called transport inhibitor response 1 (TIR1). We chose the UPRT locus to integrate TIR1 under the control of a promoter allowing a mild expression of the protein tagged to Ty. We utilized a mini-AID (mAID) tagging LIC system for conditional MORC depletion. The resulting cell lines are referred to as MORC-KD hereafter. Conditional depletion of MORC-mAID-HA is reliant on auxin (IAA), TIR1, and the proteasome. **b,** Depletion of MORC-mAID-HA, upon addition of IAA for 16 hours, was measured by IFA in cells infected with Pru MORC KD. Fixed and permeabilized parasites were probed with MORC-mAID-HA (red) and GAP45 (green). MORC-depleted cells displayed strong inhibition of proliferation when compared to untreated cells. Yet, this growth defect phenotype was completely reversed upon IAA washout, indicating that MORC depletion while impeding the cell cycle progression does not kill the parasites. Graph on the right, *in situ* quantification of nuclear MORC-mAID-HA using IFA. The horizontal bars represent the mean \pm s.d. of the nuclear MORC intensity from three independent experiments ($n = 50$ nuclei per dot). The p -values were calculated using one-way ANOVA. Scale bar, 10 μ m. **c,** *in situ* quantification (related to Fig. 2c) of nuclear MORC-mAID-HA using IFA in RH MORC KD as described above. **d,** Smoothed and background-subtracted tag density profiles are displayed over representative regions of Chr. V (top) and X (bottom). The ChIP-seq profiles were obtained with antibodies directed against pan-acetylated histone H4, HDAC3 and HA (MORC detection) from chromatin sampled from a Pru MORC KD strain left untreated (UT) and or treated with IAA for 24 hours. The experiment was repeated independently twice with similar results.



Extended Data Fig. 3 | MORC protein depletion, and HDAC3 inhibition induces gene expression. **a**, Volcano plots showing gene expression differences identified from comparison of UT versus FR235222 in Pru MORC KD (left graph) and of UT versus IAA (24 hours) in RH MORC KD (right graph) (n=8914 genes, Supplementary Table 3). The orange and green dots indicate the number of significantly up- and down-regulated genes, respectively, using adjusted $p < 0.01$ (Bonferroni-corrected) and ± 2 -fold change as the cut-off threshold corresponding to each comparison. X-axis showing \log_2 fold change, Y-axis showing $-\log_{10}(p \text{ value})$. **b**, Volcano plots illustrating changes in protein expression between UT versus FR235222 in Pru MORC KD (left graph, n=2566 proteins, Benjamini-Hochberg FDR = 1%, $p\text{-value} \leq 0.00794$) and of UT versus IAA (24 hours) in RH MORC KD (right graph, n=2139 proteins, Benjamini-Hochberg FDR = 1.01%, $p\text{-value} \leq 0.00501$). Overexpressed proteins upon MORC depletion are indicated in red and under-expressed ones in blue. **c**, Pru strains were engineered to endogenously epitope tag with HA two MORC-regulated genes, *TGME49_207210* and *TGME49_216140*. Expression was monitored following HDAC3 chemical inactivation by FR235222 and HA staining. Scale bars, 10 μm . Experiments were conducted more than three times and representative images are displayed.

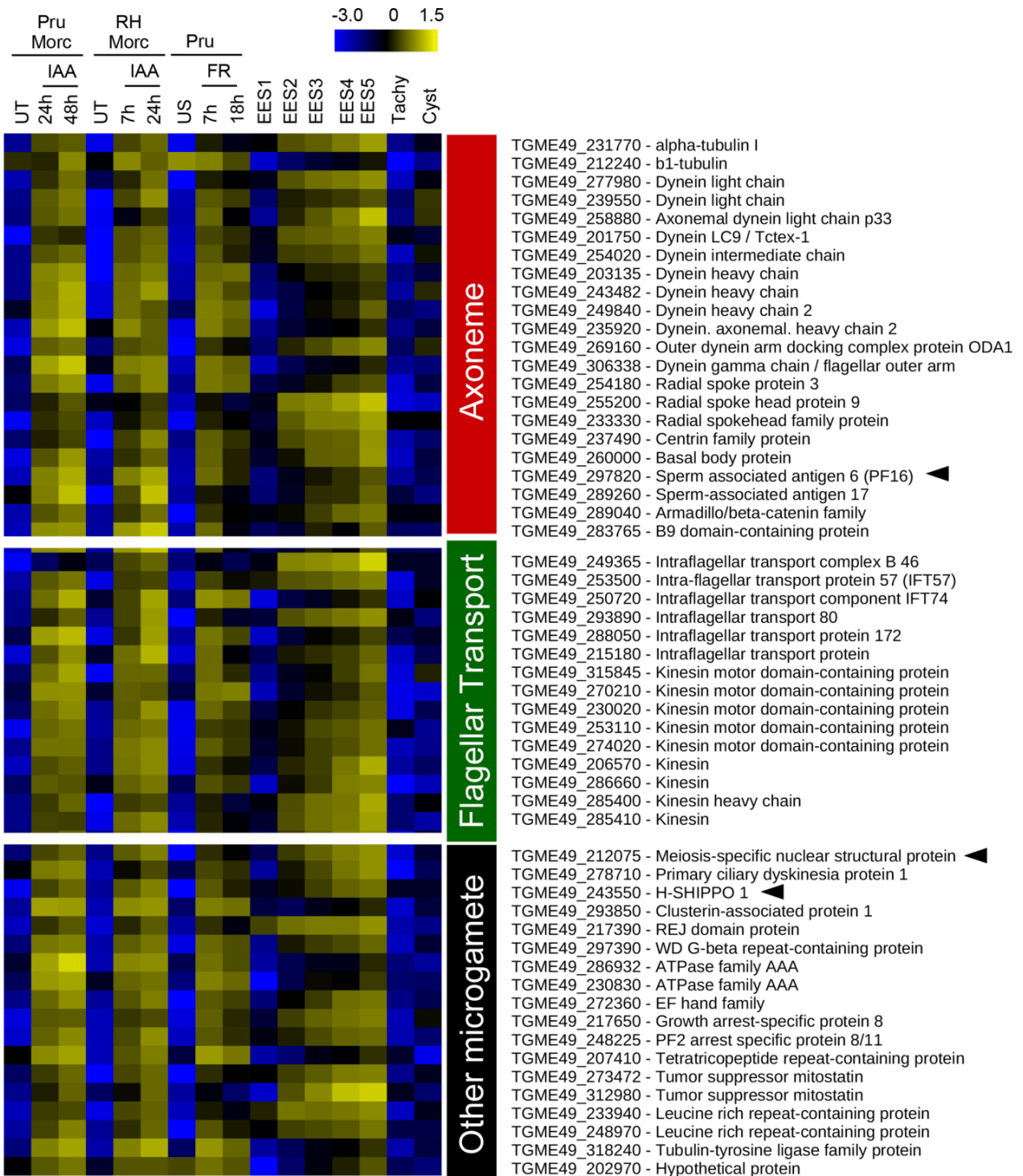


Extended Data Fig. 4 | MORC alongside HDAC3 represses the expression of sexual stages-specific genes. **a**, A type II (ME49) strain was engineered to endogenously express a HA tag version of the merozoite gene *TGME49_238915*. Genetic inactivation of HDAC3, MORC and AP2-XII-1 promote *TGME49_243940* protein expression (in red). The efficiency of genetic disruption in Cas9-expressing parasites was monitored by cas9-GFP expression (in green). Scale bar, 5 μ m. **b**, Expression of *TGME49_243940* was monitored following HDAC3 chemical inactivation by FR235222 and HA staining (in red). Scale bar, 10 μ m. **c**, Heatmap showing hierarchical clustering analysis of selected MORC-regulated genes from cluster 1 through different strain/induction combinations, including the abundance of their transcripts in the various stages of development, namely tachyzoite, bradyzoite/cyst, merozoite, enteroepithelial stages (EES) and oocyst stages. The color scale bar indicates log₂ fold changes. **d**, Expression of *TGME49_316130* was monitored following HDAC3 chemical inactivation by FR235222 and HA staining (in red). Scale bar, 10 μ m. Experiments in **a**, **b** and **d** were conducted more than three times and representative images are displayed.

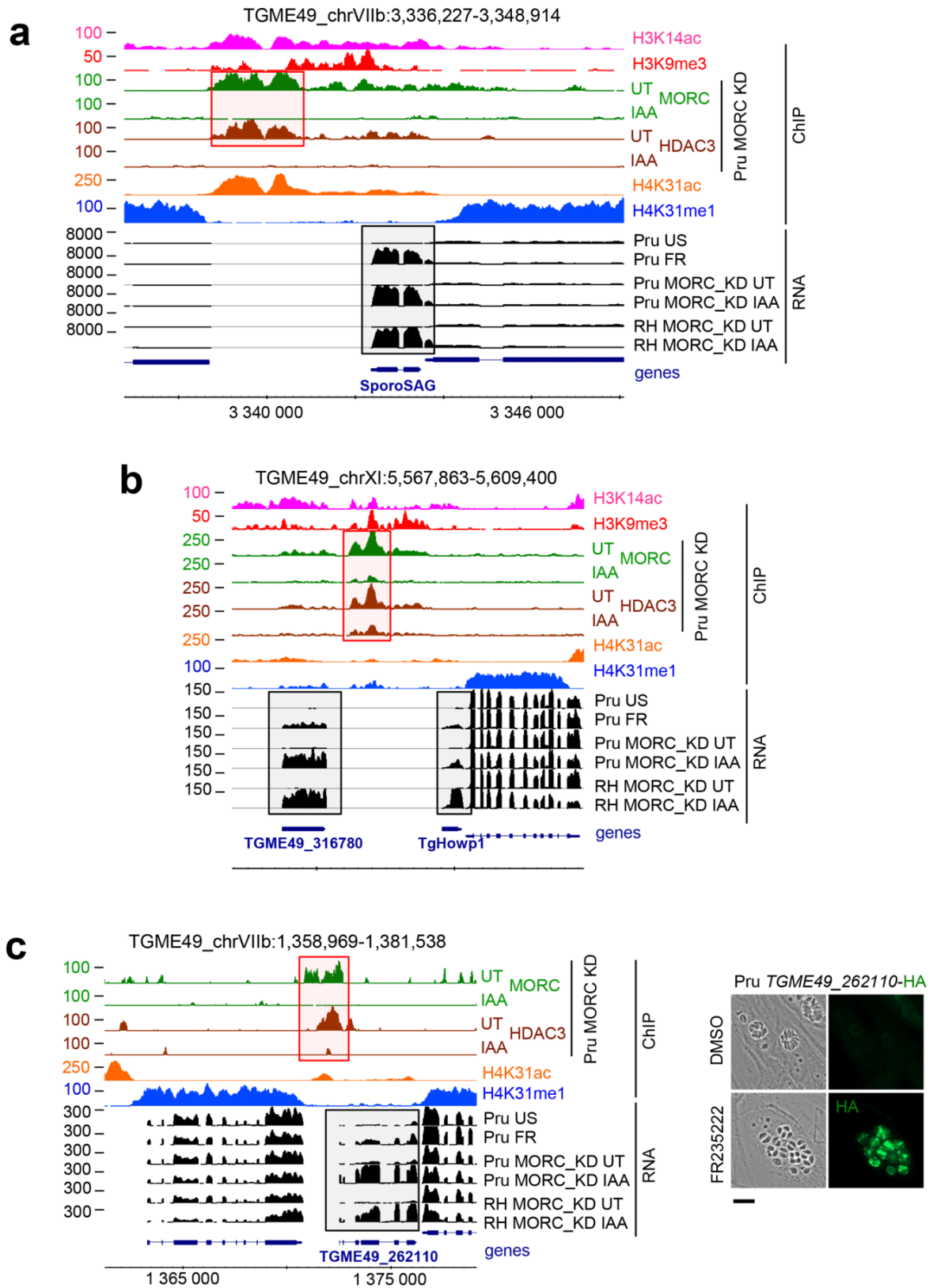


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | MORC KD derepresses proteins involved in merozoite and fertilization. **a**, Smoothed and background-subtracted tag density profiles are displayed over Chr. VI. The ChIP-seq profiles were obtained with antibodies directed against various histone marks, HDAC3 and HA (MORC detection) from chromatin sampled from a Pru MORC KD strain left untreated (UT) and or treated with IAA for 24 hours. RNA-seq data through different strain/induction combinations are shown in black. The y-axis depicts read density from ChIP-seq and RPKM values for RNA-seq data. The merozoite-specific gene *TGME49_238915* is shown in dark blue. **b**, Pru strain was engineered to endogenously expressed HA epitope-tagged *TGME49_238915*. Expression was monitored following HDAC3 chemical inactivation by FR235222 and HA staining. Scale bar, 10 μm . Experiment was conducted more than three times and representative images are displayed. **c**, Density profiles are displayed on Chromosome V at the *TgHAP2* (*TGME49_285940*) locus. **d**, Density profiles are displayed on Chromosome X around a merozoite-specific cluster of tandemly repeated genes (*SRS48* family). (**a**, **c-d**) The experiment was repeated independently twice with similar results. Chromosomal positions are indicated on x-axis.



Extended Data Fig. 6 | MORC regulates microgamete related genes including those coding for flagella components. Heatmap representation of RNA-seq data portraying the number of genes involved in microgamete biology, alongside their levels of expression upon MORC depletion/HDAC3 inhibition through different strain/induction combinations. The abundance of their transcripts in the various stages of development, namely tachyzoite, bradyzoite/cyst, merozoite, enteroepithelial stages (EES) and oocyst stages was displayed. The color scale bar indicates log2 fold changes. The genes were divided into sets grouping together genes involved in axonemal cytoskeleton, and genes that harbor potential domains involved in intraflagellar transportation.



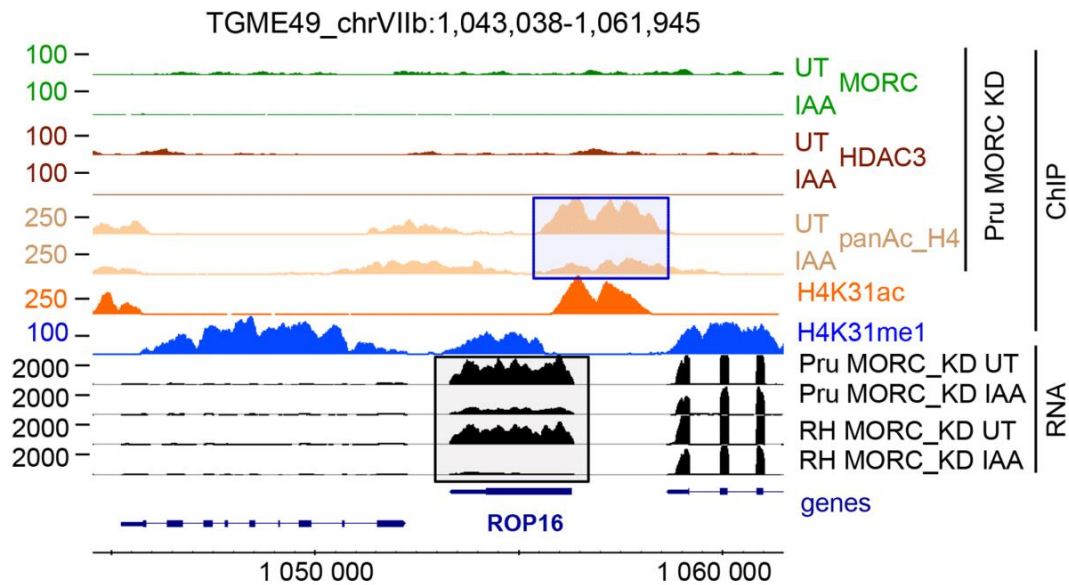
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | MORC depletion induces the expression of genes involved in oocyst wall formation and coding for sporozoite-specific

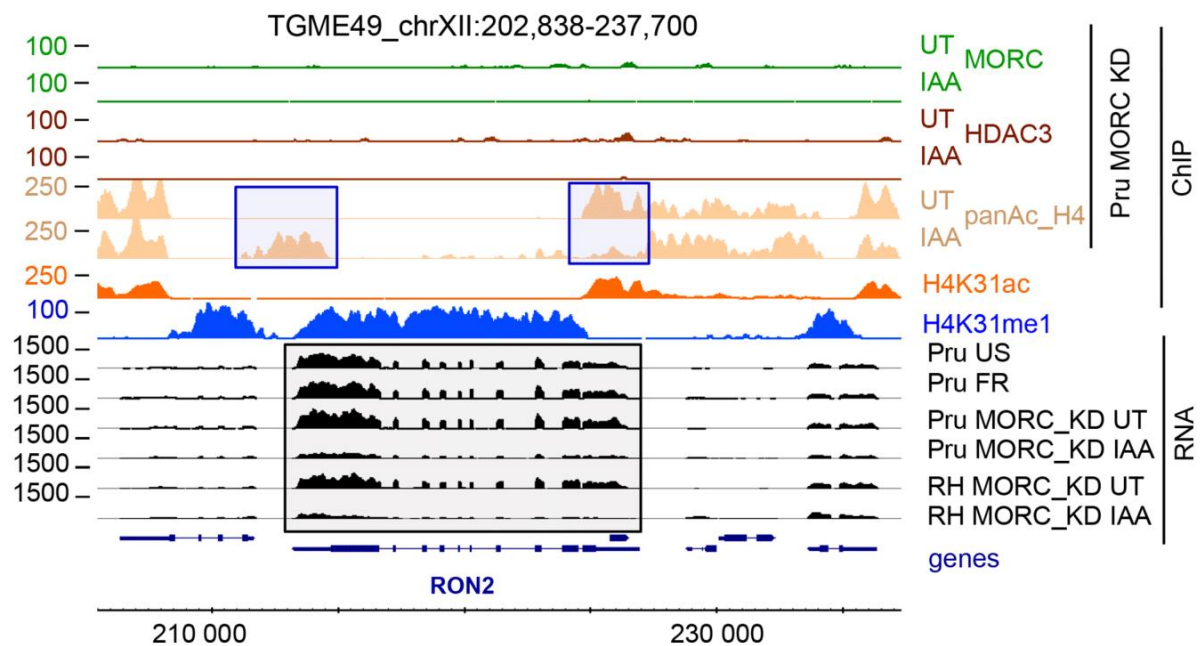
markers. a, Density profiles are displayed on Chromosome VIIb at the *SporoSAG/SRS28* (*TGME49_258550*) locus, encoding for the hallmark surface antigen in sporozoite. The ChIP-seq profiles were obtained with antibodies directed against various histone PTMs, HDAC3 and HA (MORC detection) from chromatin sampled from a Pru MORC KD strain left UT and or treated with IAA for 24 hours. RNA-seq data through different strain/induction combinations are shown in black. The y-axis depicts read density from ChIP-seq and RPKM values for RNA-seq data. **b,** Density profiles are displayed on Chromosome XI at the *TgHowp1* (*TGME49_316890*) locus. TgHOWP1 is a representative example of proteins expressed in the early stages of wall formation as revealed by the accumulation of its transcripts in both late developmental EES5 and in immature (D0) oocyst (ToxoDB data). **c,** Density profiles are displayed on Chromosome VIIb at the *TGME49_262110* locus. Expression of *TGME49_262110* was monitored following HDAC3 chemical inactivation by FR235222 and HA staining (in green). Experiments was conducted more than three times and representative images are displayed. Scale bar, 5 μm . (**a-c**) The experiment was repeated independently twice with similar results. Chromosomal positions are indicated on x-axis.

Extended Data Fig. 8 | MORC depletion derepresses partly bradyzoite-specific genes. **a**, Heatmap showing hierarchical clustering analysis of selected bradyzoite-specific and MORC-regulated genes from cluster 1 through different strain/induction combinations. The abundance of their transcripts in the various stages of development, namely tachyzoite, bradyzoite/cyst, merozoite, enteroepithelial stages (EES) and oocyst stages was displayed. The color scale bar indicates log₂ fold changes. **b**, Density profiles are displayed on Chromosome VIIb at the *Bag1* (*TGME49_259020*) locus, encoding for the hallmark surface antigen in bradyzoite. The ChIP-seq profiles were obtained with antibodies directed against various histone PTMs, HDAC3 and HA (MORC detection) from chromatin sampled from a Pru MORC KD strain left UT and or treated with IAA for 24 hours. RNA-seq data through different strain/induction combinations are shown in black. The y-axis depicts read density from ChIP-seq and RPKM values for RNA-seq data. The experiment was repeated independently twice with similar results. Chromosomal positions are indicated on x-axis.

a

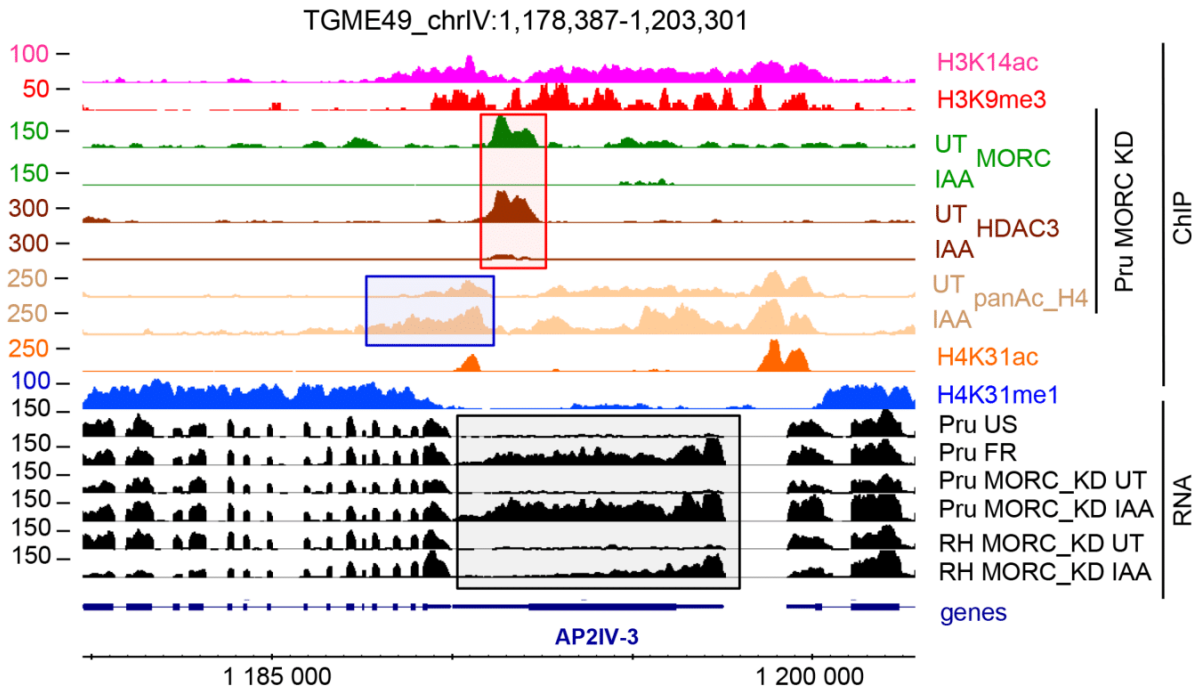


b

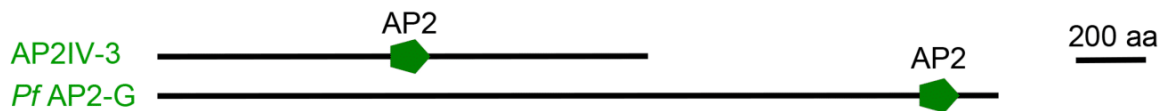


Extended Data Fig. 9 | MORC depletion causes indirect repression of tachyzoites specific genes. a-b, Density profiles are displayed at the tachyzoite-specific *ROP16* (a) and *RON2* (b) genes. The ChIP-seq profiles were obtained with antibodies directed against various histone marks, HDAC3 and HA (MORC detection) from chromatin sampled from a Pru MORC KD strain left UT and or treated with IAA for 24 hours. RNA-seq data through different strain/induction combinations are shown in black. The y-axis depicts read density from ChIP-seq and RPKM values for RNA-seq data. The experiment was repeated independently twice with similar results. Chromosomal positions are indicated on x-axis.

a



b



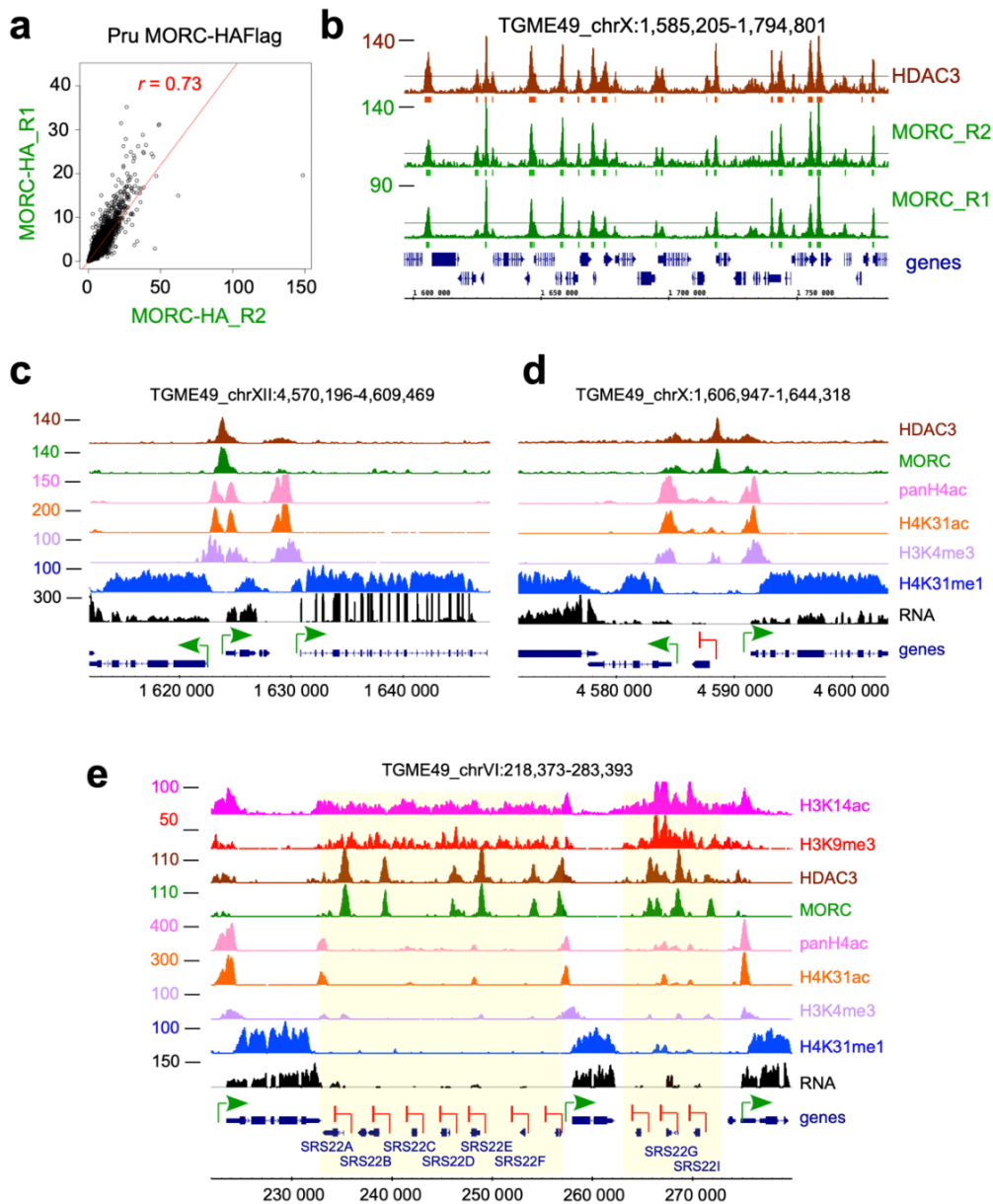
c

AP2IV-3 594 PFRGVYFDTRRGNLSWRC SWK VHGKKRSRSWSVMKFGMERARAKAIAERMK 644
 P V+ DT RG+ SWRC W +G++ S++++V +FG + A AI ++K
 Pf AP2-G 2159 PIH SVWKDTTRGHCSWRCRW WENGRRLSKNFNVKRF GNDGALRMAITMKLK 2209

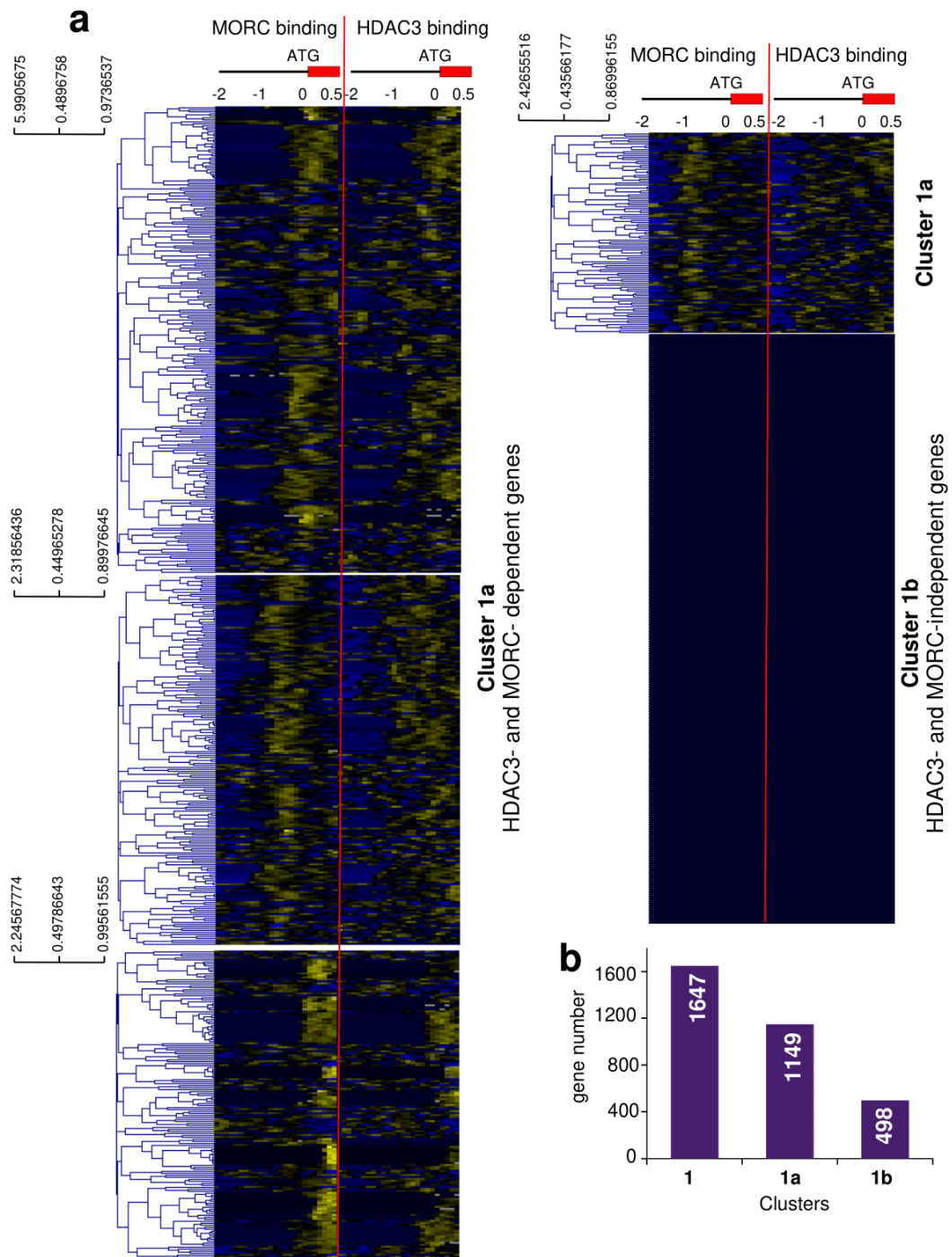
Extended Data Fig. 10 | MORC regulates the expression of AP2IV-3, a *Plasmodium falciparum* AP2-G homologous protein. **a**, Density profiles are displayed at the merozoite-specific AP2IV-3 gene. The ChIP-seq profiles were obtained with antibodies directed against various histone marks, HDAC3 and HA (MORC detection) from chromatin sampled from a Pru MORC KD strain left UT and or treated with IAA for 24 hours. RNA-seq data through different strain/induction combinations are shown in black. The y-axis depicts read density from ChIP-seq and RPKM values for RNA-seq data. The experiment was repeated independently twice with similar results. Chromosomal positions are indicated on x-axis. **b**, Representative domain architectures of *T. gondii* AP2IV-3 and *P. falciparum* AP2-G, displaying the approximate AP2 domain position within the protein sequence. **c**, Protein alignment of the AP2IV-3 and AP2-G respective AP2 domains, with the amino acid homology shown in red.

3. Supplementary Figures

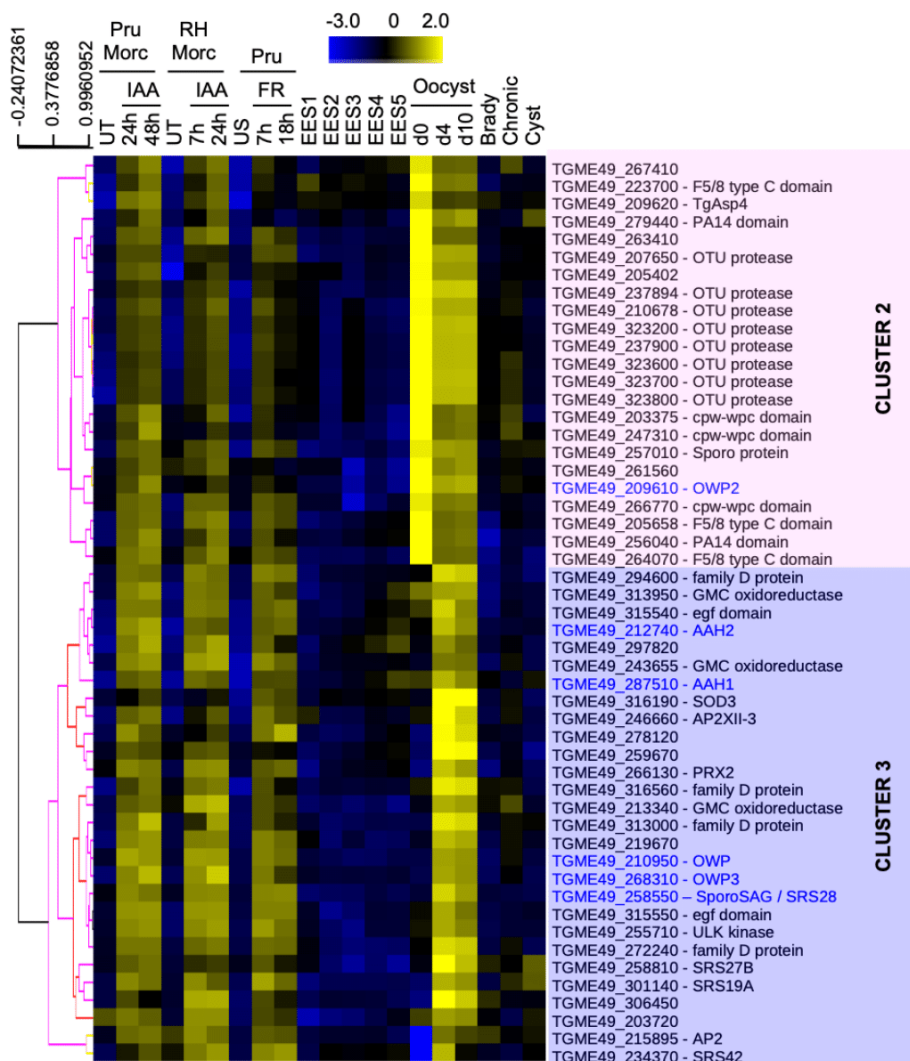
accession number: A0A3B6SGZ3; sequence aligned: 218-330). *Oryza sativa* OsMORC (UniProtKB accession number: B9EZX5; sequence aligned: 142-263). *Homo sapiens*: HsMORC1 (UniProtKB accession number: Q86VD1; sequence aligned: 23-123); HsMORC2 (UniProtKB accession number: Q9Y6X9; sequence aligned: 24-134); HsMORC3 (UniProtKB accession number: Q14149; sequence aligned: 26-165); HsMORC4 (UniProtKB accession number: Q8TE76; sequence aligned: 45-154). *Xenopus laevis*: XIMORC1 (UniProtKB accession number: A0A1L8H790; sequence aligned: 23-120); XIMORC2 (UniProtKB accession number: A0A1L8HPW9; sequence aligned: 24-124); XIMORC3 (UniProtKB accession number: Q6NRH8; sequence aligned: 26-123). *Caenorhabditis elegans* CeMORC (UniProtKB accession number: Q23243; sequence aligned: 394-485). *Chromera velia*: CvMORC1 (EuPathDB accession number: Cvel_117.t1-p1; sequence aligned: 907-1016); CvMORC2 (EuPathDB accession number: Cvel_4364.t1-p1; sequence aligned: 160-277); CvMORC3 (EuPathDB accession number: Cvel_4837.t1-p1; sequence aligned: 1535-1653). *Toxoplasma gondii* TgMORC (UniProtKB accession number: S8EVD3; TGME49_305340; sequence aligned: 1088-1204). *Neospora caninum* NcMORC (UniProtKB accession number: A0A0F7V117; sequence aligned: 1088-1201). *Plasmodium falciparum* PfMORC (UniProtKB accession number: Q8IKF6; PF3D7_1468100; sequence aligned: 1283-1408). *Cryptosporidium parvum* CpMORC (UniProtKB accession number: Q5CPM2; sequence aligned: 767-888). *Sarcocystis neurona* SnMORC (EuPathDB accession number: SN3_02700325; sequence aligned: 1627-1742). *Eimeria tenella* EtMORC (EuPathDB accession number: ETH_00019430-t26_1-p1; sequence aligned: 845-960). *Cyclospora cayetanensis* CcMORC (EuPathDB accession number: LOC34617800-t45_1-p1; sequence aligned: 872-987). *Babesia bovis* BbMORC (EuPathDB accession number: BBOV_III010280-t26_1-p1; sequence aligned: 615-729). *Theileria parva* TpMORC (EuPathDB accession number: TP02_0119-t26_1-p1; sequence aligned: 678-800). *Ostreococcus tauri* OtMORC (UniProtKB accession number: A0A1Y5IAL6; sequence aligned: 35-141). *Micromonas pusilla* MpMORC (UniProtKB accession number: C1MH37; sequence aligned: 9-98). *Bathycoccus prasinos* BpMORC (UniProtKB accession number: K8F270; sequence aligned: 102-190).



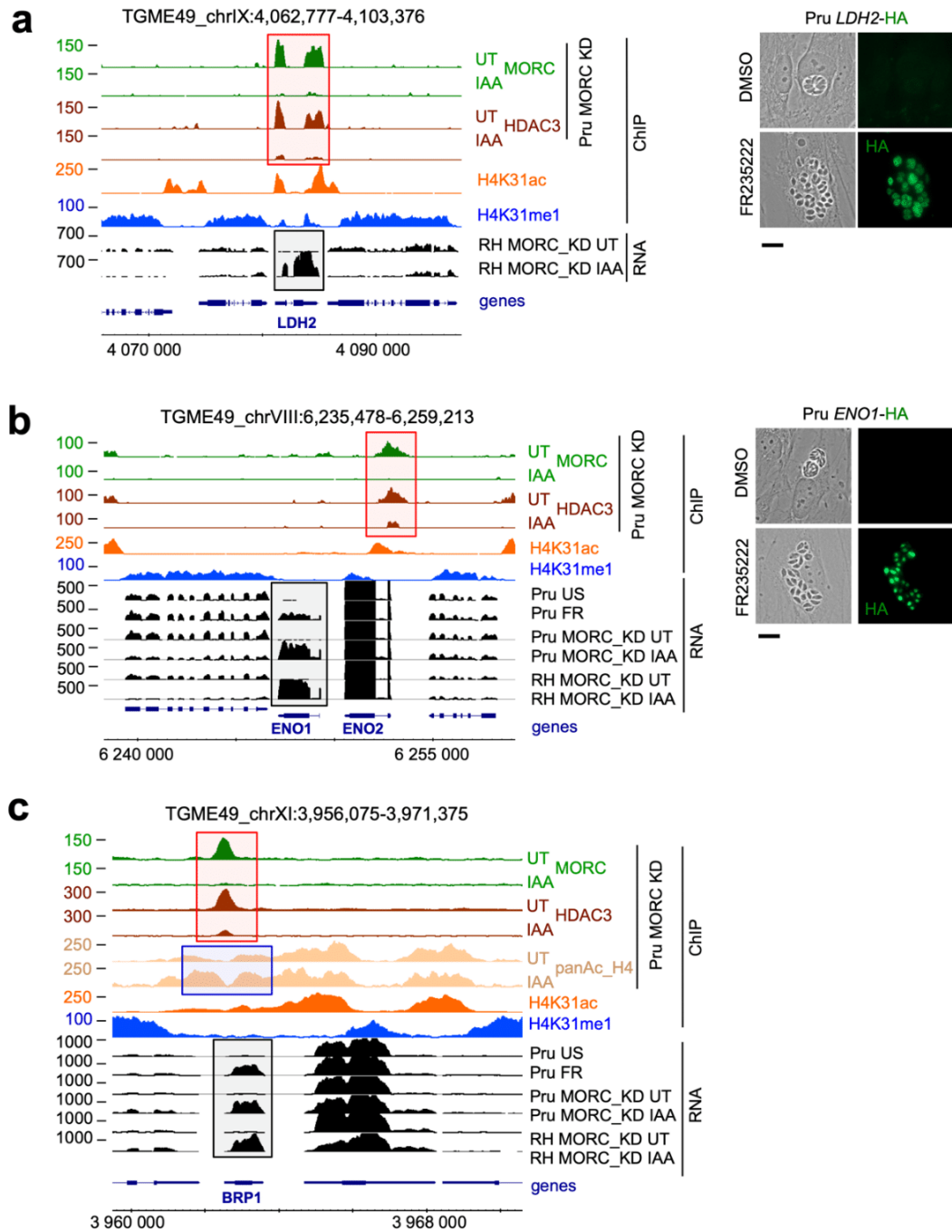
Supplementary Fig. 2 | MORC and HDAC3 co-localize at hypoacetylated regions. We noticed a low variability and a high degree of similarity in the read coverage between the technical replicates, regardless of the antibodies used for immunoprecipitation. **a**, Scatterplot comparing the enrichment difference of MORC ChIPseq replicates. The x- and y-axis show the average tag count of the enrichment ($n=8322$ genes). Pairwise correlations were calculated using Pearson correlation coefficient (r). **b**, A zoomed-in view of Chr. X region showing the distribution of MORC ChIP-seq replicates (R1 and R2) and HDAC3. Overall, we observed a high degree of similarity in the read coverage between MORC and HDAC3. **c-e**, Close-up view of Chr. XII (**c**), X (**d**) and VI (**e**) showing ChIP-seq reads for MORC, HDAC3 and various histone marks as well as RNA-seq data (in black) and predicted genes (in dark blue). The y-axis depicts read density from ChIP-seq and RPKM (Reads Per Kilobase of transcript per Million mapped reads) values for RNA-seq data. **e**, MORC and HDAC3 bind near silenced genes (red stop sign) embedded in a poised state chromatin, marked dually by H3K14ac and H3K9me3. Green arrows indicate the expressed gene. (**b-e**) The experiment was repeated independently twice with similar results. Chromosomal positions are indicated on x-axis.



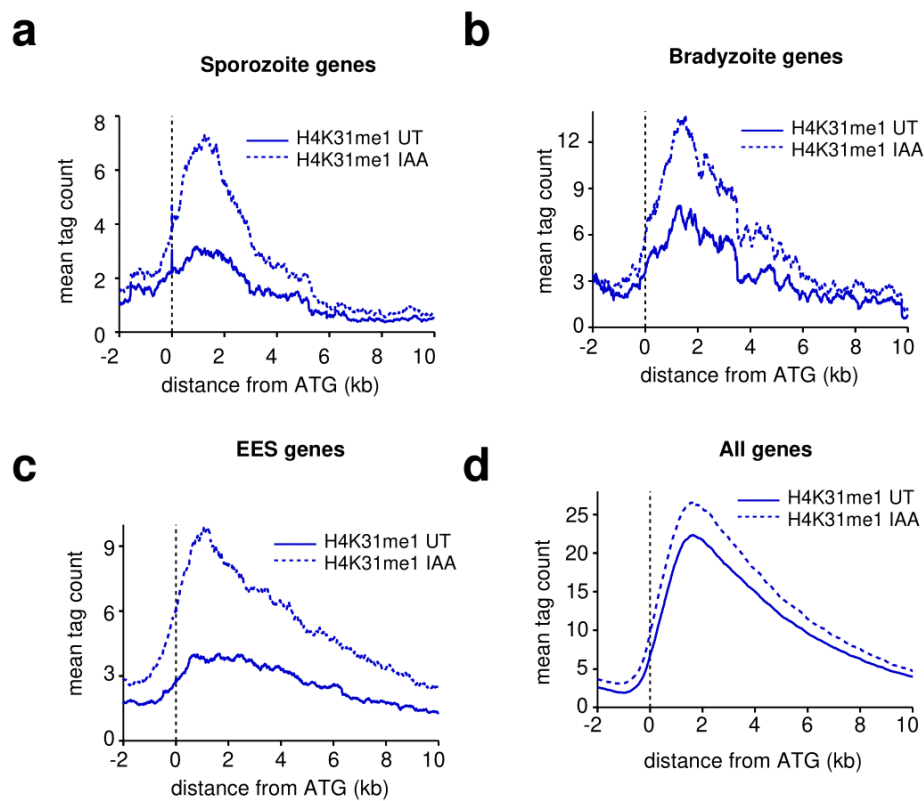
Supplementary Fig. 3 | MORC and HDAC3 bind, in a cis-regulating manner, to more than a thousand genes. a, Heatmap showing hierarchical clustering analysis of all the genes from cluster 1 based on the detected binding events of MORC (on left) and of HDAC3 (on right). The enrichment mapping of these 2 proteins spans regions of approximately 2-kb upstream and 500-bp downstream from the genes ATG sites. The mean log₂ gene expression values were median centered for the 1647 genes (more than fourfold; $p < 0.05$, unpaired t-test) that were transcriptionally up-regulated upon MORC depletion (cluster 1); the genes were then clustered by hierarchical clustering based on a Pearson's correlation. Cluster 1a was shown to group genes with an enrichment of MORC and HDAC3 nearby its regulating elements thus defining these latter as directly regulated genes, while Cluster 1b gathers the genes that lack any nearby enrichment of these 2 proteins. The complete set of genes is listed in [SI Table 4. b,](#) Graph displaying the number of directly vs indirectly MORC regulated genes from the total of the Cluster 1 genes.



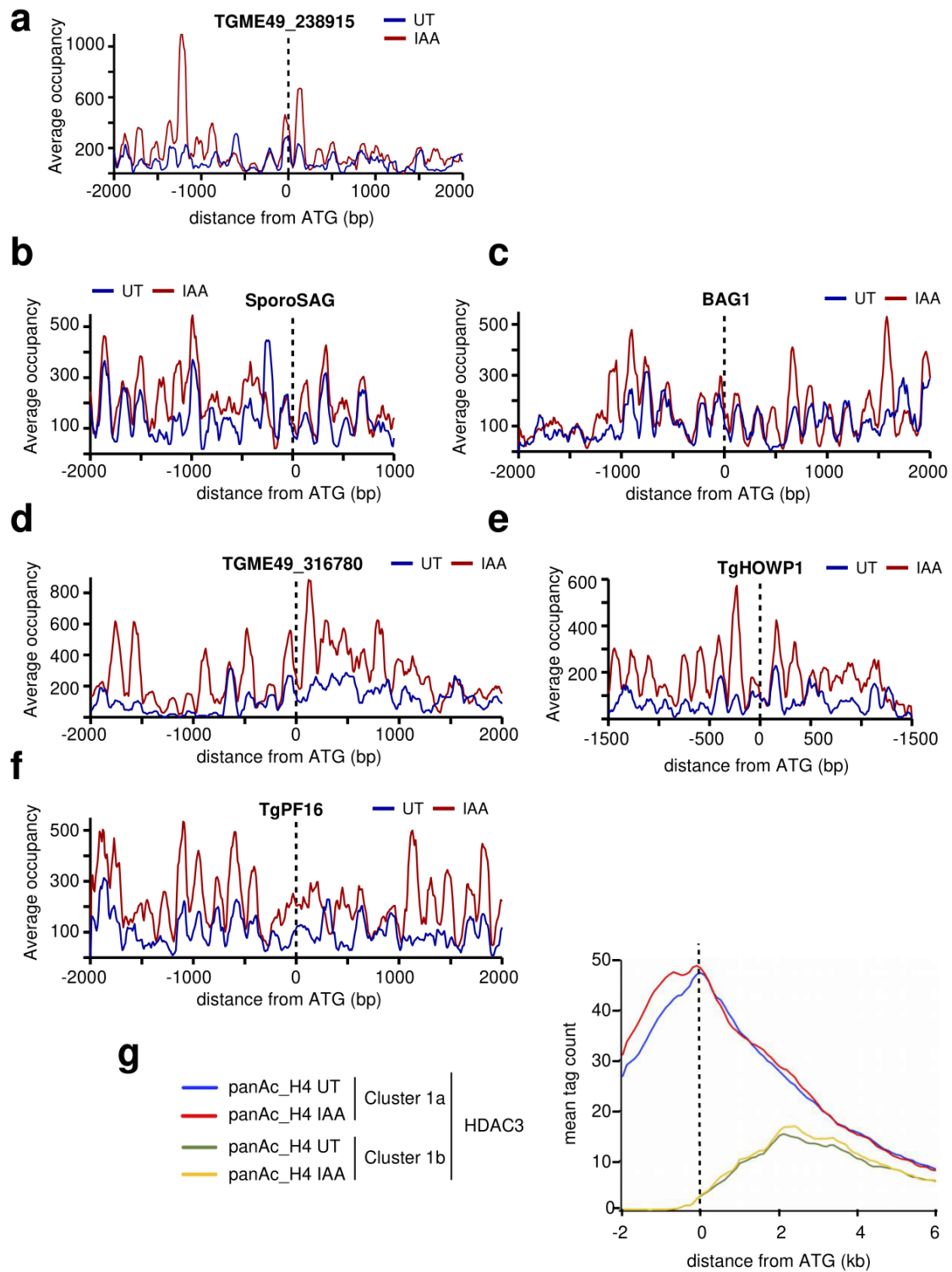
Supplementary Fig. 4 | MORC depletion derepresses genes belonging to different stages of oocysts maturation. Heatmap showing hierarchical clustering analysis (Pearson's correlation) of selected oocyst-specific and MORC-regulated genes from cluster 1 (n=1647) through different strain/induction combinations. The abundance of their transcripts in the various stages of development, namely tachyzoite, bradyzoite/cyst, merozoite, enteroepithelial stages (EES) and oocyst stages were displayed. The color scale bar indicates log₂ fold changes. Two distinct clusters were defined: cluster 2 displays genes with peaked expression at D0 followed by a slight decrease over time, while cluster 3 aggregates the genes whose expression was low at D0 then accumulated gradually to peak at D10.



Supplementary Fig. 5 | MORC depletion and HDAC3 inhibition induce bradyzoite-specific gene expression. **a**, Density profiles are displayed on Chromosome IX at the bradyzoite-specific enzyme LDH2 (*TGME49_291040*) locus. The ChIP-seq profiles were obtained with antibodies directed against various histone PTMs, HDAC3 and HA (MORC detection) from chromatin sampled from a Pru MORC KD strain left UT and or treated with IAA for 24 hours. RNA-seq data through different strain/induction combinations are shown in black. The y-axis depicts read density from ChIP-seq and RPKM values for RNA-seq data. On the right, immunofluorescence images showing the expression of LDH2 following HDAC3 chemical inactivation by FR235222 and HA staining (in green). Scale bar, 5 μ m. **b**, Density profiles are displayed on Chromosome VIII at the bradyzoite-specific enzyme ENO1 (*TGME49_268860*) locus. On the right, immunofluorescence images showing the expression of ENO1 following HDAC3 chemical inactivation by FR235222 and HA staining (in green). Scale bar, 5 μ m. **c**, Density profiles are displayed on Chromosome XI at the bradyzoite- and merozoite-specific BRP1 (*TGME49_314250*) locus. Immunofluorescence experiments in **a** and **b** were conducted more than three times and representative images are displayed. (**a-c**) The experiment was repeated independently twice with similar results. Chromosomal positions are indicated on x-axis.



Supplementary Fig. 6 | MORC depletion induces chromatin opening through deposition of H4K31me1, a histone mark associated with RNA polymerase II progression. a-d, Genome-wide H4K31me1 enrichment profiles at selected sets of genes in Pru MORC KD left UT and treated with IAA for 30 hours. The figure is displaying the set of genes established as sporozoites specifically expressed (**a**), the bradyzoites set (**b**), the genes that were grouped as sexual stages specifically expressed (**c**), and finally, on all the annotated *T. gondii* genes (**d**). The average signal profiles of each protein were plotted over a -2 kb to +10 kb region with respect to *T. gondii* genes ATG. The y-axis shows the average tag count of the enrichment.



Supplementary Fig. 7 | MORC depletion is altering the nucleosome occupancy of MORC's target genes. a-f, Average nucleosome occupancy plotted as a function of distance to the nearest yeast ATG, using Danpos2 to define the most likely position of nucleosomes. Nucleosome occupancy at MORC-regulated genes in the context of MORC depletion are shown. **g,** Genome-wide panAc-H4 enrichment profiles at the cluster 1 genes, predefined as direct (Cluster 1a) and indirect (Cluster 1b) targets of both HDAC3 and of MORC. The ChIP-seq enrichment was carried in a Pru MORC KD strain left UT as well as following 30 hours treatment. The average signal profiles were plotted over a region spanning -2kb to + 6kb relatively to cluster 1 genes ATGs. The y-axis shows the average tag count of the enrichment.

III. Discussion

We functionally characterized the only MORC protein found in *T. gondii*. Our data provide solid evidence of the weight of this protein on the developmental gene regulation of the parasite life cycle and specifically its sexual differentiation.

A scarcity of knowledge exists around the molecular mechanisms guiding the developmental trajectory of the parasite towards its sexual stages. The involvement of MORC proteins in sexual development was already suggested in plants, when most of the orthologues were found to be mostly expressed in the reproductive tissues. Also, in mammals, MORC1 exhibited an exclusive expression in the germline and during early development, to inhibit aberrant transposable elements expression in such stages, consistent with the global relief of repression occurring as they acquire their pluripotency. The MORC of *C. elegans* was also acting on the level of the transgenerational fertility and germline siRNA signature inheritance.

By characterizing MORC protein in *T. gondii*, we bring additional knowledge not only to the developmental genetic regulation in the parasite and the phylum, but also to the MORC community by providing a new player for more evolutionary robust functional comparisons, and on the longer term for clearer mechanistic perceptions of this family's mode of action.

It cannot be denied that the development of a single-celled organism has unique features, and that its parasitic life style must have shaped greatly its regulatory pathways, including the ones that still hold great levels of conservation with other free-living eukaryotes. Considering these aspects, as well as all the years of evolution since their divergence, one must realize that any comparison made between these proteins, is merely speculative, unless defended and substantiated by concrete evidence.

However, as our work mainly provided functional characterization, and lacked exhaustive mechanistic depiction, I will allow myself, as the discussion flows, to philosophize on the possible mechanistic aspects as well as the putative similarities between our protein and the ones documented in other *phyla* that are delineated in the introduction, when these can fit to our distinct parasite and to what we know about its sophisticated development and molecular story.

1. Poised State and Chromatin Targeting

a. Poised State

In the quest of understanding the specificity of the targeting of the MORC-containing complexes, one must notice the abundance of MORC and HDAC3 binding near silenced genes embedded in a poised state chromatin, marked dually by H3K14ac and H3K9me3 (Sindikubwabo et al., 2017). Similar bimodal patterns were identified at developmental gene promoters in embryonic stem cells. However, as stated in the introduction, that in stem cells the reported combinatory consist of H3K14ac/H3K27me3 (Karmodiya et al., 2012), or of

H3K4me3 with H3K27me3 (Matsumura et al., 2015) which is a mark that is absent from *T. gondii* and *Plasmodium sp.*. In *P. falciparum*, more bivalent states were recognized, balancing activating marks as H3K9ac and H3K4me3, with repressive ones as H3K36me3 and H4K20me3 (Duraisingh and Skillman, 2018).

P. falciparum has been reported to carry most of its chromatin in an active state, with the clonally variant families in a poised state and employing the bivalent pattern in their advantage for ensuring its evasion of the immune system, with the aid of PfSET10 methylating the lysine 4 of histone H3 of the active gene for its transcription to be poised for reactivation in the following cycle (Volz et al., 2012).

The development-related genes of *T. gondii* being embedded in such chromatin states would seem only reasonable considering the speed at which the parasite must differentiate in order to ensure either its survival when facing the host immune response or its transmission by undergoing sexual conversion.

Histone tails acetylation is one of the most dynamic marks catalogued so far in the histone code, and over the years we have shown the impact of the acetylation state on gene regulation in *T. gondii*, and recently we provided MORC as a cornerstone for this process alongside its intimate partner HDAC3, with the complex probably recognizing these *loci* through the CW domain of MORC or its PHD-containing protein partner (detailed ahead).

Seeing MORC enriched at these sites raised the question about the mechanism of action of its complex and whether the bare absence of MORC and its partner would be enough for alleviating the repression through the mere disruption the balance between the activating and repressing marks duality at these *loci*.

But here we have to underline on the fact that what we observed *in vitro* following the engineering of a MORC-KD system might not be completely reflecting the real context, where it is still not clear what is upstream of MORC and regulating its activity, yet it is believed that MORC would be constitutively expressed and its activity or targeting would be regulated.

Therefore, we cannot conclude on its mechanism based solely on the aforementioned poised state *loci* enrichment. Yet, it is noteworthy to mention that there doesn't seem to be a considerable difference in the matter of which stage specific genes are embedded within this type of chromatin.

While searching for discrepancies that could explain the heterogeneity between the different vacuoles and the reasons behind their cell fate differences ([Figure 16](#) (Farhat et al., 2020)), one possibility arose that suggests that genes belonging to one and not the other stage would be in a poised state, especially considering that only one third of the bradyzoite-specific genes were MORC-regulated versus almost all of the merozoites ones, suggesting an extra layer of bradyzoite regulation that might require an activating signal instead of the mere poised and ready to be alleviated mode of repression.

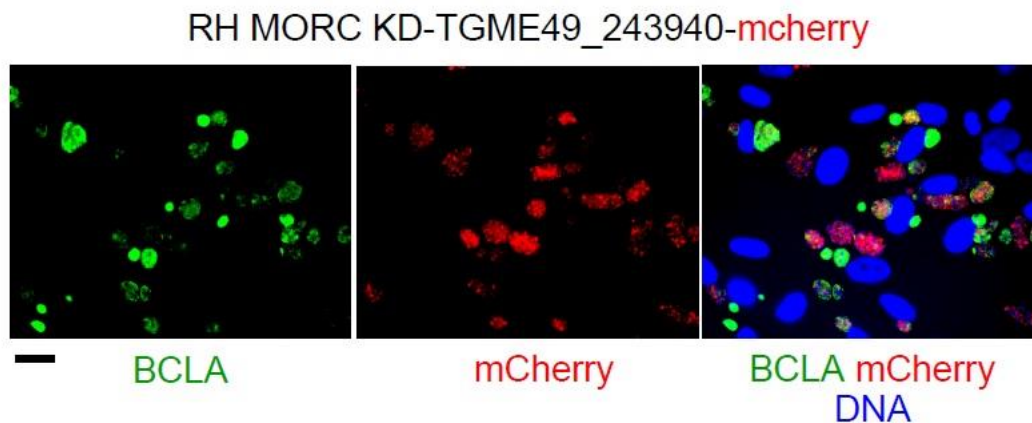


Figure 16. MORC KD results in a heterogenous protein expression.

IFA of HFFs that were infected with MORC–HA–mAID RH parasites expressing a targeted knock-in merozoite-specific gene (TGME49_243940). Immunofluorescence analysis of individual vacuoles after depletion of MORC, with a co-staining using a bradyzoite marker (BCLA, green) or a merozoite marker (red, TGME49_243940–mCherry) showing an expression pattern that is almost mutually exclusive between vacuoles. *Adapted from fig.5f in (Farhat et al., 2020)*

Many targeted genes belonging to the sexual stages intensely displayed this pattern including 7 paralogs of the SRS22 family and the SRS48 family (See **Extended Data Fig. 5** in (Farhat et al., 2020)), yet when looking more closely into some of the bradyzoite-specific targeted genes, a similar yet less pronounced pattern was observed, therefore, an explanation of the differences in the MORC-dependent regulatory events leading to one or the other stage, must not rely solely on interpreting their poised state.

Also to be taken in consideration, the fact that while conducting our ChIP-seq experiments, we didn't proceed to any particular mean of synchronization of the cells, deeming that all the parasites must be in a tachyzoite state and adopt a similar epigenetic signature, yet their cell cycle state, which is acknowledged to impact the histone marks enrichment (check cell cycle part for more details), was not homogeneous and this could lead to an apparent co-occupation of activating and repressing marks on the same *loci*, as reported in (Karmodiya et al., 2015).

Observing MORC at this state of chromatin would let one wonder about how dynamic is this repression action, herein how maintained would the action of MORC be and how easily reversible.

b. CW-mediated chromatin targeting of MORC

Our protein shares its structure with the one of human MORC3 and 4, both displaying an H3K4me3-reading CW domain, and carries no additional coiled coils domains separating the GHKL and S5 domains and potentially impeding the efficiency of this peptide binding, as was

seen for the aberrant CW domains of the human MORC1 and 2. This similarity pushes us to consider the possibility for the *T. gondii* MORC to rely on such chromatin reading ability, in order to recognize its targets. However, we did not verify in *T. gondii* the MORC enrichment at genes marked by methylation at H3K4, to allow ourselves to claim functional resemblance to the CW-dependent targeting observed in other MORCs.

It must be noted that *Chromera velia*, the most distant known ancestor of apicomplexans, possesses three orthologs of MORC, yet only one displays a CW-ZF, hinting to the functional relevance of such co-occurrence in *T. gondii* for this version to be conserved and not the other plant-like counterparts.

We tried to prove the reading abilities of TgMORC CW domain by conducting biochemical essays where we proceeded into producing a recombinant CW-ZF protein in bacteria. We were unable to purify the protein to a quality high enough for use in *in vitro* binding assays using modified histone peptides array.

In parallel, we also produced the PHD domain of the co-purified PHD-containing protein found in the MORC interactome (described ahead), in our attempts to identify the assistant of MORC in scanning its targets, assuming that this domain might be the reader of the acetylated H3K14 at which MORC is presumably enriched, especially after considering the reports about the ability of PHD to bind to this modified site (Zeng et al., 2010).

Unfortunately, the binding to the modified histone peptides array didn't succeed, and we were left with no conclusive answer regarding the TgMORC CW PTM reading abilities. It must be noted that phosphorylation sites were detected close to the PHD domain of this protein, as well as its additional EELM2 domain (described in HDAC3 section) which might be one of the reasons behind the failure of our histone binding assay when using bacterial recombinant proteins.

In fact, it is plausible that both these domains promote multivalent contacts to generate the hydrophobic pocket that is necessary for binding to the methylated lysine; however, detailed structural assays are needed to prove the reality of this assumption. In fact, it has been proposed that in yeast, the RSC nucleosome remodeling complex requires multiple bromodomains to act in concert, in order to increase the RSC-nucleosome interaction lifetime (Lorch et al., 2018), therefore one can imagine a similar mode of action that would grant MORC sufficient time to recruit partners or additional actors.

The CW is structurally seen as a variant of the PHD domain, as is the case for GATA-like zinc fingers, and together they can assemble, as the case for the LIM or ADD domains, potentially forming more potent or more diversified readers. One can imagine that such wider pockets might be thus evolved, enabling wider binding specificities other than the initial methylation of H3K4, similarly to the reported wider catalytic domain of TgHDAC3 (Bougdour et al., 2009) or the higher methylation catalyzing potential of TgSET8 (Sautel et al., 2007).

c. PHD-mediated chromatin targeting of MORC-containing complexes

In our attempts to uncover the actors behind the complex recognizing its targets and loci specifically, we stumbled upon one of the interactors of MORC, TGME49_230890, withholding a plant homeodomain (PHD) domain which is a C4HC3 zinc-finger-like motif. This well conserved plant homeodomain is one of the first identified alternatives for bromodomains in their ability to read acetylated histones such as H3K14ac (Zeng et al., 2010), a mark reported on poised *T. gondii* genes at which we found MORC to be enriched.

Originally, they were attributed a role in the recognition of the lysine methylated histone H3. Proteins harboring such domains were found within multi subunits complexes, mostly ones involved in gene regulation and development, precisely mediating repressive fates.

The BRAF35-histone deacetylase complex was isolated in human cells and included BHC180, a PHD zinc finger subunit and was shown to be involved in the repression of neuron-specific genes in non-neuronal cells (Hakimi et al., 2002). The corepressor complex NURD is also typified by its core component CHD4 harboring two PHD fingers able to bind H3 tails from distinct nucleosomes thereby acting on the compaction of chromatin (Mansfield et al., 2011).

This domain displays a stretched evolutionary range of conservation, and it is worth noting that the apicomplexans parasite must have acquired this plant-specific chromatin remodeling protein from their ancestor bearing an algal endosymbiont.

Homeodomain proteins are classified into different subfamilies according to the conservation of their sequence, but also to the presence of additional motifs besides the homeodomain with their functional differentiation seemingly being achieved through the acquisition of these different codomain architectures (Mukherjee et al., 2009).

One major group of plants homeodomain proteins is one characterized by the additional presence of the DDT domain which seems to have evolved early in green plant evolution since it was found even in green algae. However, this DDT motif was not identified on any of the six homeobox genes of the red algae, along with no detection of the additional HD domain on these proteins (Mukherjee et al., 2009). Moreover, having not detected any of these additional domains within the PHD containing proteins of apicomplexans renders it more plausible that the ancestor of phylum gained these domains during the endosymbiont event of this red algae.

The three-helix structure of the homeodomain shares similarities with families of prokaryotic transcription factors (Viola and Gonzalez, 2016), nevertheless the MORC-interacting PHD containing protein possess a stretch of 22 pentapeptide repeats (LSSSS) with the vast majority of proteins containing these domains being found in prokaryotes where they could be

mediating inhibitory interactions with gyrase proteins, and few examples being detected in both *A. thaliana* and *P. falciparum* (Vetting et al., 2006).

Aside from its possible implication in recognition of the MORC complex targets through its poised state chromatin reading abilities, this PHD-like protein could be participating in the intra-complex protein-protein interactions through the binding of proteins PTMs, noting that these latter were detected on some of the identified interactors, including the AP2s factors.

Additionally, we think that this PHD containing protein might be involved in the recruitment of TgHDAC3 to the nucleus mediated by its EELM2 additional domain, a concept described further in the TgHDAC3 section.

2. DNA-based targeting and Primary AP2s

The mechanism behind the chromatin targeting of MORC at specific *loci* remains cryptic. We elaborated on the possibility of this recognition being mediated by the chromatin state reading accomplished by either the CW domain of MORC or the PHD domain of TGME49_230890. The targeting of the MORC complex to its *loci* could also occur via the affinity of TFs for the chromatin embedded DNA. AP2s family of potential specific transcription factors are the first suspected actors here, especially seeing that the interactome of MORC included more than ten different AP2s (See **Extended Data Fig. 1a** in (Farhat et al., 2020)).

In our quest to unravel the mechanism(s) behind the cell fates heterogeneity emanating from the MORC-KD model, we pondered about the possible endogenous weight of these AP2s partners in establishing a stage-specific genes targeting.

It is acknowledged that the same TF can induce different cell fates owing to differences in its interaction partners, as the case of pMad/Brk or Smad (Spitz and Furlong, 2012) (X et al., 1998), thus we believe that similarly, MORC would induce distinct transcriptional programs based on its set of AP2 direct interactors. Our size exclusion chromatography of MORC-containing complexes revealed the association of MORC and HDAC3 within a high-molecular weight complex of around 800-KDa (predicted globular size); this number suggests that additionally to the constitutive partners, MORC is able to accommodate a (homo or hetero) dimer of AP2s at a time. However, we were unable to distinguish the different subcomplexes of MORC, as they all eluted in the same fraction corresponding to about 800-KDa, thus our understanding of the possible AP2s interactions between each other is as limited as our ability to provide the case about the binding motifs of these heterodimerizing AP2s.

MORC acts as a repressor of the sexual progression of the parasite, yet the mechanisms behind the regulation of its action remain unclear, especially considering that it seems to be expressed constitutively, at least on the mRNA level (Harb and Roos, 2020). mRNA of HDAC3 were also detected in all stages of the parasite life cycle, therefore the fact that the interacting AP2s are mostly not expressed in stages other than the tachyzoites stage. If we reason on the basis of MORC being a fully active repressor only in the tachyzoites, this interactors AP2s apparent exclusive expression to this stage might imply that the regulation of the complex

function is in part, held by the AP2s levels of expression and thus by their recognition of the targets becoming either absent or fine-tuned.

One would suspect that all AP2s belonging to the MORC complex would act as repressors too. In fact, we witnessed the action of the interactor AP2XI-4 as a repressor of our merozoite reporter gene *TgME49_243940*, as the CRISPR mediated disruption of this factor led to the induction of the *TgME49_243940* protein.

Similarly, we witnessed the implication of the interactor AP2XII-1 at the level of repression of our merozoite reporter gene, thus it seems to be acting in concert with MORC as a repressor of merozoite-specific genes and thus of the progression towards sexual development.

Also, to note, is the fact that the interactome of MORC in the virulent type I (RH) strain presented two additional AP2s that weren't detected in the one of the cystogenic type II (Pru) strain. We did not provide exhaustive evidence to whether this discrepancy emanates from technical limitations involving either the chromatography pull-down steps or the mass spectrometry sensitivity, or that these additional interactions are ones mediated by DNA and not a core complex protein-protein interaction, yet this last option is hardly plausible considering the highly stringent salinity used during the pull downs.

Taking in consideration the difficulty emanating from all the possible homo and heterodimers, it seems quite unattainable to aim for an in-vitro reconstitution of the order of assembly of these potential TFs and whether their targeting occurs in a random or hierarchical order.

Also, as in this proposed model, the AP2s are the ones to be most likely targeting the complex, yet it is unclear whether they establish the repression at genes that were active, or whether they maintain their poised state. As in, we are using tachyzoites parasites, but if we were to be following the parasite during their progression from sporozoites to tachyzoites for example we would witness the actual establishing mechanism of repression of sporozoite-specific genes, and not only their maintenance in that state, as is the case.

It remains to be determined whether the activity of these AP2 factors is merely regulated on the level of their transcription, or whether it can occur through post-translational modifications of their proteins. Other than the acetylated sites that we found to be regulated through the action of TgHDAC3, most of these proteins are phosphorylated at multiple sites of Serine and Threonine, yet not within their AP2 domains, with the exception of AP2XI-4 which displayed a calcium dependent motif in its AP2s domain, pointing at the parasite transcription regulation through indirect effects of the cellular metabolism and external environment (check metabolism part).

3. Secondary AP2s mediated regulation of targeting and action

During our analysis of the possible means of action of MORC, we briefly mentioned the weight of the secondary wave of AP2s in aiding MORC, for maintaining a stage appropriate epigenetic reprogramming. Following the engineered KD of MORC, we have observed a de-repression of a wide set of AP2 genes which we then arranged based on their acknowledged developmentally regulated expression ([Figure 17](#) (Farhat et al., 2020)). However, we reiterate

the non-exhaustive ability of our *in vitro* MORC KD model, to mimic the endogenous context. In fact, we haven't provided the full set of data concerning the secondary action of AP2s *in vitro*, which could have reinforced our theory about these latter holding a regulatory layer for unidirectional transcriptional programs.

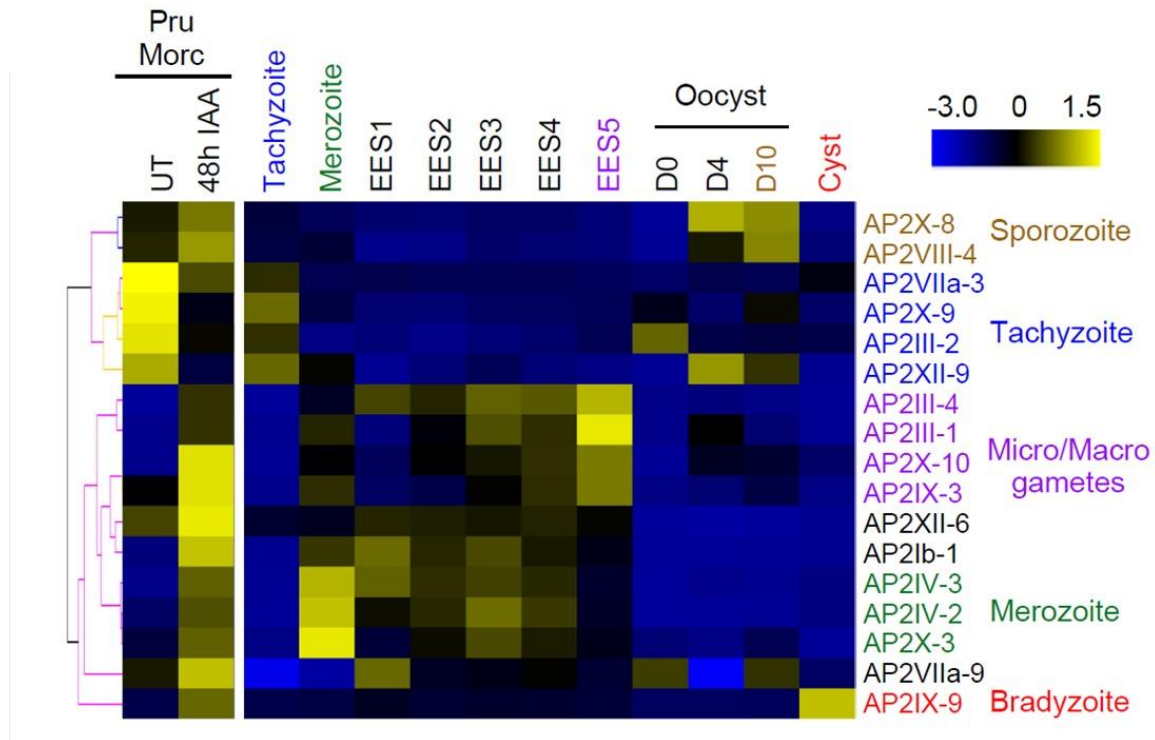


Figure 17. MORC guides developmental trajectories recruiting downstream regulating pathways.

Heat map showing mRNA hierarchical clustering analysis (Pearson correlation) of AP2 TFs shown to be regulated by MORC depletion. The abundance of their transcripts in the various stages of development, namely tachyzoite, bradyzoite/cyst, merozoite, EES and oocyst stages, is displayed. The colour scale indicates log₂-transformed fold changes. The AP2 factors were colour-coded according to the colour of the life stage during which they are most expressed. Adapted from figure 6.a. in (Farhat et al., 2020)

Some of these AP2s were already characterized, however we believe that the functions given to some of these factors were either not fully adequate, or that perceiving them requires further experiments, at least from our side. For instance, AP2IX-9 was reported as a repressor of bradyzoite gene expression, of which the deletion enhanced tissue cyst formation *in vitro*. In this study, they interpreted their observation basing on the tachyzoite-bradyzoite inter-conversion (Hong et al., 2017). However, we believe that this bradyzoite-resident repressor is acting instead on the level of the bradyzoite-merozoite conversion. Tachyzoite-bradyzoite transitions are bidirectional and highly dynamic, making their control more appropriate by specific activators at each stage. The unidirectionality of the bradyzoite to merozoite transition makes it more plausible that AP2IX-9 is acting on this level to prevent an aberrant bradyzoite genes expression once a development into merozoites is initiated.

If I reason now based on *our vitro* observations instead, I would mention that the KD of MORC, apart from derepressing this repressor, led to the upregulation of many bradyzoite-specific genes yet not all of them were touched by its action, herein the repressor AP2IX-9 became present in this context and kept these genes in their dormant state. In our model, we claim that MORC is a repressor of the sexual development trajectories, and based on this model, it would be repressing AP2IX-9 in the endogenous context in order to alleviate the action of this factor on the prevention of a merozoite to bradyzoite switch back, therefore maintaining the parasite in its tachyzoite-bradyzoite asexual cycle.

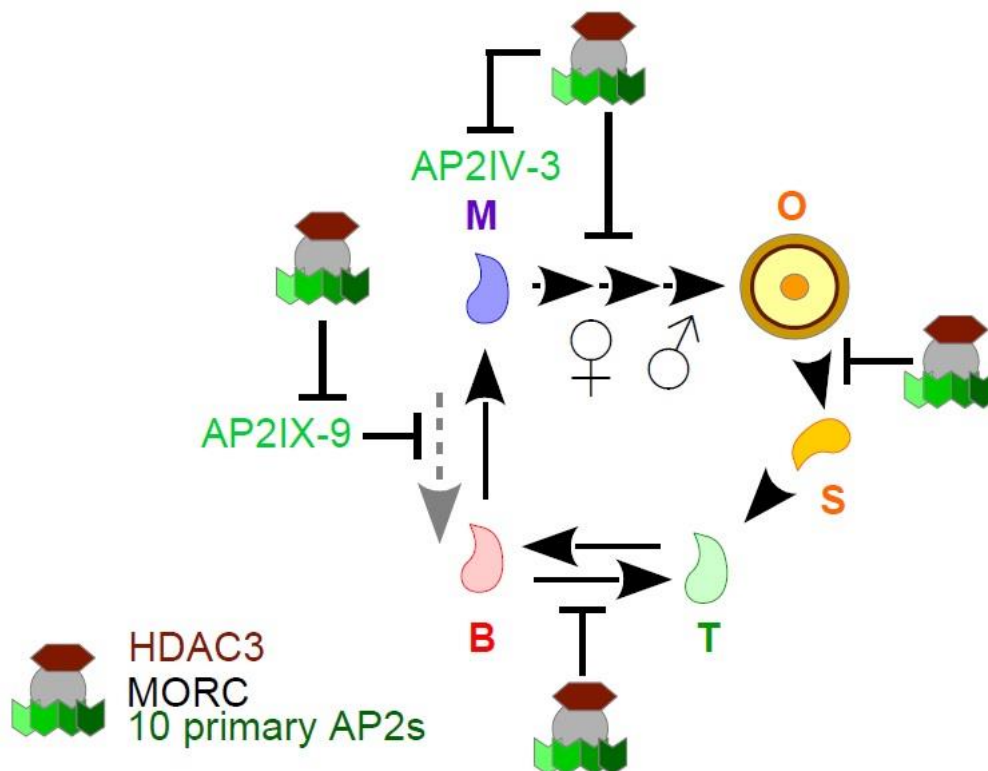


Figure 18. A proposed model.

A model describing the mechanism that underlies the MORC-complex-mediated modulations of the *T. gondii* life cycle. B, bradyzoite; M, merozoite; O, oocyst; S, sporozoite; T, tachyzoite. The genetic regulatory circuits would be mediated by the clusters of secondary AP2 TFs acting on the transcription of the stage they belong to. Adapted from (Farhat et al., 2020)

Going through the transcriptional profile of AP2IX-9, one would ponder about its semi-exclusive expression during the bradyzoites stage, with a slight mRNA detected during the beginning of EES, considering that reasonably if it were to be a true bradyzoite repressor it should be expressed in all other stages except the bradyzoite one, unless if its repressing action at this point is regulated at the posttranslational level rendering it unable to fulfill its potential repressor role.

It would seem reasonable that the AP2s would be fulfilling their roles in the stages during which they are normally expressed. Keeping up with the possibility that AP2IX-9 is as claimed a bradyzoite-specific repressor preventing a switch back from the merozoite stage, and basing on its exclusive bradyzoite expression profile, one can ponder about the upstream regulation of this factor, such as it would only repress its targets upon an external cue signaling the arrival into a feline's enterocyte and thus the necessity for repressing bradyzoite genes. This regulation could occur not only on the post-translational level of the protein, but also on the level of its proper transcription, herein this gene seems to harbor numerous different splicing possibilities, potentially stage dependent.

The same study (Hong et al., 2017) claimed finding an activator of bradyzoites and cyst formation, being AP2IV-3. We found this factor to be upregulated following the KD of MORC, and instead we believe that it is an activator of merozoite genes and thus of sexual progression. In fact, its transcriptional profile shows a semi exclusive expression during the merozoites stage and EES, herein going against a potential bradyzoites activator role.

We did not provide any solid evidence proving its activating or repressing role, but we reasoned that as in our model MORC is a sexual development repressor, then AP2IV-3 as one of its repressed targets would be acting directly or indirectly in the favor of going forward into the sexual stages.

As this factor is not essential, at least not in the tachyzoite stage, we attempt to generate its stable KO in the MORC-KD strain, in order to gain more insight on the role of this factor and to set straight the discrepancy. Moreover, transcriptomic assays would shed light on the potential functional resemblance between this factor and its *P. falciparum* architectural homologue PfAP2-G (See **Extended Data Fig. 10b,c** in (Farhat et al., 2020)).

Among the 17 detected AP2 domain-containing proteins that have their expression regulated by MORC, it is worth noting that three of them belong to the interactome of MORC comprised of about 10 AP2s proteins that we called primary AP2s. We discussed previously the idea of the weight of these primary AP2s on the stage-specific targeting of the MORC complex, however we overlooked the fact in some cases MORC seems to be behind their developmentally regulated expression

A thought loop regarding MORC's sexual repressing function, with a simplified view focusing on the crosstalk between MORC and AP2s:

If we consider that the MORC complexes act as repressors, due to the stable core complex interaction with HDAC3, then the outcome of the interactions that would occur with AP2 proteins must be of a similar nature, as they will be aiding in the targeting towards the genes to be repressed, keeping in mind that MORC doesn't have any DNA binding domain neither does HDAC3, and that their targets recognition can be mediated by the chromatin states, yet for their motif and promoter specificity they must rely on their AP2 partners. Two scenarios can be thought of, one with MORC always active and relying on AP2s for targeting (a repressor in all stages of all the genes that don't belong to the stage in question), and one with MORC only active in tachyzoite and its activity getting impaired by some external signal once the

sexual development gets initiated (a repressor of sexual development as our modest model proposes).

We claim in our paper that the function of MORC would be to repress the sexual development, and due to the lack of CHIP or Interactome data from other stages, we based many of our perspectives on this tachyzoite-based model. However, this view seems a bit limited, and MORC should be considered a repressor of sexual stages genes, when the parasite is in its tachyzoite stage, hence conversely MORC must also be considered or at least suspected to be, a repressor of the tachyzoite genes (and other aberrant genes) when the parasite is its sexual forms. The fact that MORC is an ATPase enzyme that lacks a DNA specificity, makes it that it is an enzyme that relies on the presence and the activity of interactors for it to be guided to its DNA targets.

For instance, MORC represses in tachyzoites the expression of a sporozoite gene, SporoSAG, (see **Extended data fig7.a.** in (Farhat et al., 2020)), directly at its promoter, and it also represses in tachyzoites the expression of a merozoite gene also in a direct manner (see **Fig4.e.** in (Farhat et al., 2020)), this targeting must be relying on a different set of AP2s that are acting in the tachyzoite stage as repressors, of sporozoite genes, and of bradyzoite genes, respectively. Thinking that MORC is by itself a sexual repressor that would become inactive at stages other than tachyzoite, would mean that both these genes would be de-repressed in the same way, which is not correct. Therefore, MORC is a master regulator, yet that would be at least partially active, this being throughout the developmental stages, at repressing the genes of other stages, herein relying on the states of expression as well as the activity of the differentially expressed interactors.

An unshown data-based example to illustrate the downstream effects of MORC, again if we keep reasoning on the basis of a gene regulation based only on the crosstalk between MORC and downstream TFs, is the following:

AP2IV-4 is reported as being a repressor of bradyzoite genes, and in the tachyzoite stage this AP2IV-4 is expressed at least at the mRNA level. Following the KD of MORC in this stage, this gene got repressed, yet considering that MORC proceeds by direct repression and not activation, MORC in the endogenous context, must have been repressing the repressor of this proteins, thus MORC in this case is acting as “ a repressor of a repressor of a repressor”!

However, this is a simplified view that focuses on the direct transcriptional targeting of the gene's promoters as a means of regulation, when the reality includes a lot more pathways and factors than only MORC protein and its interactors binding to their motifs. The fact that we cannot yet produce pertinent CHIP or MS-MS data from any stage other than tachyzoite, unless if we proceed by killing a huge number of cats, as well as that we lack a single cell approach for following a parasite going through its fate decision process, adds a layer of difficulty for interpreting what might be the endogenous case at other stages, and what complexes are forming then, and what promoters are targeted.

Also, some of the primary AP2s that were found to be interacting with MORC in the tachyzoite stage, were themselves found to be regulated by MORC. Hence, the interactome we identify in this stage is not a stable one throughout the life cycle, and what is seen as a secondary downstream AP2 can actually be a primary interactor in another stage.

4. Alternative means for developmental regulation of expression

We assigned to MORC the function of a sexual development master repressor; however, we wouldn't allow ourselves to claim that this is the sole factor guiding the tight regulation that this parasite acquired throughout its evolution and its adaptation to its different hosts environments.

As MORC was identified as an actor at the transcriptional level, it would be interesting to mention the other means of gene regulation that take place and could be related to the phenotypes we observed in our in vitro model, and that could also reflect the endogenous mechanisms of this developmentally regulated transcription.

One of the concerns that we raised after witnessing the upregulation of that many stage specific genes, was whether the effect of the KD of MORC stems from a direct transcriptional activation of its repressed targets, or whether its absence merely allowed the release of imprisoned stage specific mRNA, an event that in the endogenous case would have occurred only following an environmental signal.

We reason that both mechanisms can be taking place, wherein the time dependent increase of the upregulation of many of the targets of MORC following its KD, indicate that not all of the mRNAs of the genes in question were released simultaneously. This was a pattern for many genes that showed a higher level of mRNA after 24 hours of the KD of MORC than they did after 7hrs, taking in consideration that no MORC proteins were present after 3hrs of its posttranslational KD induction.

However, if we keep reasoning at the in-vitro level, we cannot forget that we haven't executed any single cell approach that could have allowed us to follow up on the transcriptional output of the step-by-step sexual development of the cells primed for this fate.

a. Translational repression and the pocket concept

Regarding the endogenous context, it would be interesting to investigate the possibility of a post-transcriptional regulation occurring at the beginning of the sexual trajectory and including all genes belonging to the merozoites as well as EES and oocysts, yet with each set of genes being released from an mRNA storage 'pocket' upon receiving an appropriate external signals, with this mechanism allowing for a faster progression through the stages rather than the energy and time consuming initiation of a transcription at the required time.

The pocket idea has been described in the literature, with mRNAs being stored either individually or in multiple non-translating mRNA aggregating together in cytoplasmic foci.

Cytoplasmic RNA granules can form under stressful conditions with the aid of factors that seem to be conserved. The duration of such translational blockade vary between mRNA and can be either transient or extend for long periods of time (Hooper and Hilliker, 2013).

In fact, the *P. berghei* female gamete has been seen to harbor such cytoplasmic bodies rich in translationally repressed mRNA, along with helicase proteins thought to establish the transition out of this state (Mair et al., 2006). Macrogametes stockpiling mRNA needed for the following oocyst stage was also described in *Eimeria* and *Cryptosporidium species*, as well as *T. gondii* (Ramakrishnan et al., 2019).

The sustained translational repression can be exemplified with the *T. gondii* Lactate dehydrogenase 1 (Holmes et al., 2014). The mRNA of this tachyzoite specific metabolic actor was reported to be present in both tachyzoite and bradyzoite stages, yet the corresponding protein was only detected in tachyzoite, thus it was suggested that this repression allows for the fast replenishment of this protein that could be assisting in the reversion to tachyzoites upon relief of stress. A sorting mechanism was also suggested, as to allow the transcripts destined for only transient repression to access to the limited supply of translation machinery.

Translational repression in *T. gondii* was also seen in the oocyst stages where an obvious upregulation of ribosomal proteins was detected especially at times considered to be representative of a mature state of the oocyst, prior to their development into infectious sporozoites. It was suggested that these ribosomes would allow the sporozoites-containing oocysts that are dormant for months, to rapidly restart the translation of secretory organelles needed for the next stage and the next intermediate host (Fritz et al., 2012). Translational repression waves were detected also in malaria parasites, during the maturation of their sporozoites, with one being relieved upon reaching the salivary gland, and another being relieved when reaching the liver (Lindner et al., 2019).

b. mRNA stability and maturity

Another mean to ensure rapidity in answering to the protein demand of the new environments, would be to act on the level of mRNA stability and maturity, herein the transcriptional repression would be alleviated, for instance at the loss of the action of MORC, yet the mRNA specific to a certain stage would go through maturation only when receiving the corresponding cues. We are currently conducting essays to unravel some of the mechanisms behind the maturation of mRNA involving the CPSF complex required for the poly-adenylation of RNA, and it would be of interest to check whether its action is timely regulated and stage dependent. (Farhat et al., in preparation)

While speaking of mRNA maturity, we can only mention the emerging evidence of the effect of the mRNA m6A mark on its stability. Studies are providing proofs about this mark acting on the stability of the mRNA, which was similarly reported in *P. falciparum* (Baumgarten et al., 2019). We are conducting attempts to characterize this mark in *T. gondii*, with an interest of examining their impact on the stage dependent regulation of the transcriptome.

We also believe that MORC might be acting somehow on this mark, but it is still unclear whether it is impacting its deposition or its reading. In fact, following the KD of MORC, one could easily perceive the decrease in the level of m6A staining during an IFA of intracellular parasites, with this decrease getting more robust with time. Although we cannot exclude the indirect means through which this outcome could be occurring, yet one could reflect on the possibility of this mark holding a layer of regulation of the cell fate maintenance, while noting the lack of any DNA methylation in this parasite, which is acknowledged to bear this substantial role in many species (Gissot et al., 2008).

c. RNA Polymerase II pausing

One additional aspect to be considered is the pausing of RNA polymerase II at the vicinity of promoters, a newly described concept that has been conserved from bacteria to mammals, and it is thought to be a mean for tuning the gene expression (Williams et al., 2015). It would be interesting to tackle the impact of the pausing of RNA pol-II on the maintenance of stage specific gene expression programs in *T. gondii*, especially considering that the release of the RNA pol-II can be guided by external cues. In fact, we witnessed an effect of MORC on the elongation of transcription (See **supplementary fig.6** in (Farhat et al., 2020)), as its deletion led to an increase, at its targets genes of the H3K41me mark which was previously shown to be related to the progression of pol II (Sindikubwabo et al., 2017). One can imagine an antagonist link existing between MORC and Pol-II, where this latter would be pausing and holding the active genes signature during the cell cycles and transitions, while MORC would aid in holding the imprint of the repressed ones.

5. HDAC3 involvement in post-transcriptional and post-translational regulation

The post-transcriptional regulation in *T. gondii* relies in part, on a homolog of Argonaute, Tg-AGO, which uniqueness is displayed through its ability to bind both plant-like trans-acting small interfering RNAs and metazoan-like microRNAs. This RNA silencing machinery seems to act mostly through translational repression, considering the loose pairing between the miRNA and the target, as well as the lack of endonucleolytic slicer residues in the Tg-AGO protein (Braun et al., 2010).

This protein was co-purified with the initial TgHDAC3 affinity chromatography pull down, along with MORC (TgCRC230 at the time), hinting to the possible involvement of Tg-AGO machinery in the post-transcriptional repression of the complex targets, or in some fine tuning of the extent of the transcription, or of it playing an indirect part in the recruitment of these actors, and others to some sites.

A Myb-domain-containing protein has recently been characterized in *T. gondii*, with evidence of its direct involvement in the ability of tachyzoites to differentiate into bradyzoites, as the

loss of this protein resulted in the loss of this ability even under in-vitro bradyzoites inducing conditions, as well as the loss of tissue-cyst formation during animal infections (Waldman et al., 2020), with this phenotype reaching extents that were not yet observed for any of the AP2s reported as putative bradyzoite development regulators.

What is intriguing is that this Myb protein, called BFD1, binds bradyzoite genes promoters, but also its own promoter, stressing here on the equal mRNA expression of this gene across stages, yet with the protein being only expressed under stressing conditions hinting to its expression being regulated at the post-transcriptional level. The relevance of this suggestion lies in the evidence of an unusually long 5' UTR as well as multiple open reading frames within this region (Waldman et al., 2020), recalling the translational regulation level of differentiation, described earlier.

Additionally, BFD1 has been detected as one of the non-histone targets of TgHDAC3, based on our unpublished acetylome sequencing following the drug-inactivation of HDAC3 by FR235222. In fact, two acetylation sites seem to be dependent on the activity of HDAC3, which adds a layer of complexity to the reprogramming pathways in this parasite, as well as it stresses once again on the indirect yet significant weight of HDAC3 on the bradyzoite development.

We previously conducted these acetylome-wide assays which proved that the targets of the deacetylating action of HDAC3 extend further than only histones to touch acetylation sites of many enzymes and proteins. Noting that most of these aforementioned HDAC3 deacetylated non-histone proteins were not directly targeted by its partner MORC, such that their RNA levels were not affected following its KD.

Of these non-histone targets, some were proteins related to the cell cycle, which explains the apparent arrest in the cystogenesis following the treatment of parasites with FR235222, others were enzymes with known histones PTM-reading or -writing functions, such as the acetylase GCN5b which was found to be holding a specific acetylation site targeted by HDAC3. This regulatory relationship between two enzymes that are recognized to be antagonist and to be contesting each other on many dynamic histones PTMs, held one layer of explanation for the changes in acetylation levels extending beyond the promoter regions of the targets of HDAC3 following its chromatin dissociation and the KD of MORC.

6. HDAC3 and MORC: a non-exclusive relationship

In fact, one question regarding the activity of HDAC3 is the extent to which this protein is carrying an exclusive interaction within the MORC complex. In fact, other than its non-histone targets and the possible indirect impact of such protein activity post-translational regulation, it cannot be excluded that HDAC3 might be a part of other complexes, noting that the assumption of it acting within complexes and not independently, stems mostly of the fact that this nuclear based protein lacks any Nuclear Localization Signal (NLS) and thus requires targeting assistance from other proteins. Here, it must be noted that the KD of MORC led to

the disruption of the targeting of HDAC3 to the *loci*, yet it didn't affect its nuclear localization (See **Fig. 2e** in (Farhat et al., 2020)), thus MORC is not the one directly carrying HDAC3 to the nucleus, and we also don't have proof for the dissociation of the whole complex once MORC is gone, thus many of the intra-complex interactions could be maintained.

Proteins harboring ELM2 (EGL-27 and MTA1) domains were reported to fulfill similar mission for the class I of human HDACs to be recruited into corepressor complexes. These domains undergo major folding transitions on binding to the HDAC, as in the NURD complex where this folding would mediate the dimerization of the whole complex containing a homo or heterodimer of HDAC1 and/or HDAC2, herein resulting in chromatin condensation and transcriptional silencing (Millard et al., 2013).

Many of the 20 PHD domain containing proteins identified in *T. gondii* harbor specific co-domains, however only the one interacting with MORC (described earlier) possess an additional EELM2 domain. This extended ELM2 homology domain occurs exclusively in proteins from apicomplexan parasites, with this extension stemming from the longer boundaries it harbors compared to the orthodox ELM2 domain (Oehring et al., 2012).

Therefore, it could be that the recruitment of TgHDAC3 to the MORC corepressor complex is mediated by one of the 3 EELM2 domains-containing proteins identified in the interactome. These domains are found on the aforementioned protein co-displaying the PHD, as well as on two other proteins. However, only the protein harboring both the PHD and the ELM2 is reported to being essential for the fitness of the parasite, as well as the only one carrying an NLS, both aspects missing from the 2 other ELM2 proteins.

To note that no other protein than these three MORC interactors possess the EELM2 domain, hence if we were to consider that they represent the only recruiters of HDAC3, then either HDAC3 is exclusive to the MORC corepressor complex, or most likely that these factors are also part of other HDAC3-containing complexes. However, other actors might be fulfilling this mission.

7. MORC in *Plasmodium falciparum*

Recently, a study was published describing *in silico* attempts to identify the complexome in three species of *Plasmodium* (*P. falciparum*, *P. knowlesi*, and *P. berghei*) where they managed to determine a complex assembling the analogue of TgMORC in *P. falciparum* with PfHDAC1 an EELM2 containing protein, and many multigene regulatory ApiAP2s (Hillier et al., 2019). They validated these predictions *in vitro* and considered MORC as the potential scaffolding protein connecting transcription factors with epigenetic regulators. It must be noted that the PfHDAC1 harbors a domain that shares the most homology with TgHDAC3. PfHDAC1 also displays a nuclear localization, one that wasn't attributed to the other HDACs of this parasite for lack of antibodies.

No studies have tackled directly the function of MORC in *Plasmodium* sp., yet a freshly out paper detected an abundance of the PfMORC on *var* gene promoters (Bryant et al., 2020). The study relied on a catalytically inactivated dCas9 for identifying chromatin complexes on these genes regulatory elements. In parallel, they relied on the identified interactome of the ISWI protein also found to be enriched at the promoters of these elements. Among others, PfMORC and PfHDAC1 were co-purified with ISWI, along with GCN5, and an ApiAP2 that was already identified in the previous complexome to be associating with the MORC homologue (Hillier et al., 2019). The putative *P. falciparum* MORC displays the apicomplexan KELCH repeats signature for MORC proteins, yet intriguingly lacks the S5 fold domain, characteristic of the MORC ATPase module. Thus, it was suggested that this newly identified *var* elements-associated complex would be acting on the level of subtelomeric chromatin alteration through mechanisms not yet elucidated.

Moreover, a yeast two hybrid screen identified an association existing between PfMORC and a telomere repeat-binding zinc finger protein (TRZ) (Bertschi et al., 2017), hinting to a possible role for MORC at such regions.

8. MORC and HDAC3 at telomeric repeats?

In fact, we noted but did not report in (Farhat et al., 2020) the enrichment of both MORC and HDAC3 on the ends of the chromosomes of *T. gondii*, yet the repetitive nature of these elements led us to ponder whether this is merely due to the known technical limitations of ChIP-sequencing approaches when it comes to satellites and telomeres. However, these doubts were appeased when the enrichment of MORC was lost on these elements following its KD but not the one of HDAC3. This suggested that on these regions in particular, HDAC3 would be fulfilling some other MORC-independent roles, as its targeting here was not dependent on its presence, as the case for their developmentally regulated targets (Figure 19).

It is plausible that on these regions, HDAC3 is part of another complex, one that would be acting on the regulation of telomeres. In order to identify this interactome, we ran pull-down assays in very stringent salt conditions to release the chromatin-bound proteins. The data from the mass-spectrometry(MS)-based proteomic analysis didn't provide compelling evidence of any clear-cut different core complex.

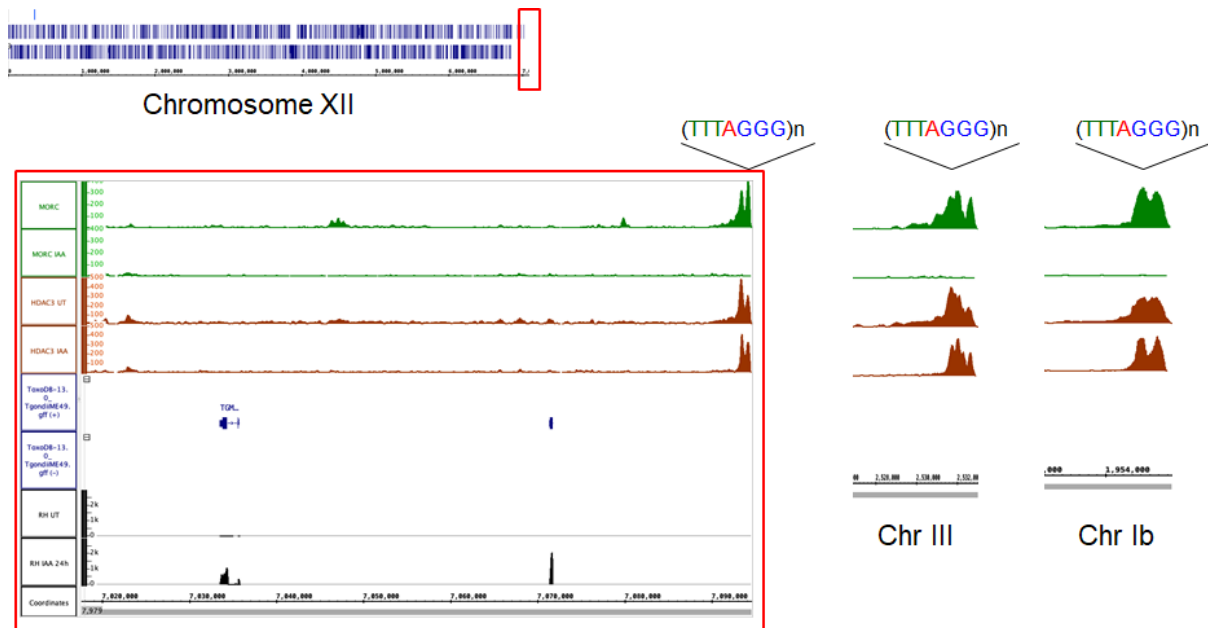


Figure 19. *MORC and HDAC3 are enriched on telomeric regions.*

Density profiles for representative regions of chromosomes XII, as well as III and Ib. Chromosomal positions are indicated on the x axis. The ChIP-seq profiles were obtained using antibodies directed against HDAC3 (here shown in brown) and HA for MORC detection (here shown in green), and chromatin sampled from a Pru MORC KD strain that was either untreated or treated with IAA for 24 h. RNA-seq data for both treated and untreated for induction are shown in black. Read density from ChIP-seq data and RPKM values for RNA-seq data are shown on the y axis.

Hardly any studies were conducted in attempt to unravel the secrets of the telomeric regions mainly because of technical challenges, for example it holds great difficulty to sequence their length with high precision in order to evaluate regulatory phenotypes of a suspect actor protein. However, few aspects can be easily observed as the plant-like nature of the telomeres of *T. gondii*, with their repeated sequence consisting of TTTAGGG instead of the TTAGGG sequence repeated at the ends of human chromosomes (Ichikawa et al., 2015); It must be noted that unlike yeast, *T. gondii* harbor their telomeric regions within nucleosomes as do mammals and plants (Figure 20). The short cell cycle of these parasites would let one doubt the existence of a telomere shortening process occurring in these parasites, keeping in mind that this process doesn't exist in plants.

Organism	<i>In vivo</i>	
	Repeat unit	Telomeric chromatin
<i>Homo sapiens</i>	TTAGGG	Nucleosomal
<i>Mus musculus</i>	TTAGGG	Nucleosomal
<i>Arabidopsis thaliana</i>	TTTAGGG	Nucleosomal
<i>Tetrahymena thermophila</i>	TTGGGG	Nonnucleosomal ^b
<i>Schizosaccharomyces pombe</i>	T ₁₋₂ ACA ₀₋₁ C ₀₋₁ G ₁₋₆	Nonnucleosomal
<i>Saccaromyces cerevisiae</i>	TG ₁₋₃	Nonnucleosomal

Figure 20. Nucleosome organization in telomeric chromatin.

Table displaying the sequences of the telomeric repeats of some organisms and their nucleosomal nature. Highlighted is the one from *A.thaliana* for resemblance of features with *T. gondii*.

We also managed to identify some proteins that carry domains with homology to telomere binding domains TBDs, and it is probable that one of them could aid HDAC3 in recognizing its targets, as HDAC enzymes are believed to rely on their partners for a specific identification of their targets. Spotting HDAC3 would suggest that in *T. gondii* it is the type I HDACs instead of the Sirtuins that are required for the deacetylation step that must precede the methylation of histones on these regions and which leads to a more constitutive heterochromatic state.

The apparent enrichment of MORC at the ends of chromosomes is worth being speculated. The accumulation of the protein at the ends of the chromosome recalls the model of the *C. elegans* MORC-1 which was seen *in vitro* to form static multimers at the end of linear DNA thus preventing the release of the mobile proteins that were encircling DNA. It would be interesting to further tackle the nature of the MORC enrichment at telomeric regions and to check whether its nuclear *puncta* pattern stems from DNA entrapping multimeric *foci*.

9. Host impact

a. Host-parasite interactions

In order to establish an adapted epigenetic and transcriptional reprogramming, the parasite must respond to external cues arising mainly from its multiple hosts, here considering that most of the stages this parasite passes through are intracellular, with the exception of the parasites residing within oocysts shed in the environment by felids, which nevertheless are responding to the environment, as in the oxygen dependent maturation of the sporozoites but not in a direct interactive relationship as the one linking the parasite to its various hosts.

However, the study of this interactive host-parasite link is not always attainable, namely for the parasites in their sexual forms (merozoites and EES) at the encounter of their definitive Felidae host. It is deemed challenging and clearly unethical, to investigate the signaling pathways occurring at the level of the cat's intestines.

The most researched area here is based on the tachyzoite stage. An infection of the intermediate host by tachyzoites, is responsible for the expansion of the parasite population that triggers the acute phase of infection that is usually restrained by the host immune system. It is believed that the parasite would also be required to establish a protection against the oxidative stresses emanating from the infected host.

This would be likely achieved by factors acting to maintain cellular homeostasis, as the reported KMTox, a histone lysine methyltransferase, however no definitive *in vivo* proof has been provided (Sautel et al.). This anti-oxidant defense seems to be needed also by the bradyzoite stage, where an upregulation of a number of oxygen radicals metabolizing enzymes, was stated. These encysted forms seem to be especially armed by various DNA repair enzymes to cope with the long-term exposure to reactive metabolites (Manger et al., 1998).

For a long time, it was believed that no host-pathogen interplay is present during the chronic infection, until a study produced RNA-seq based data to advocate the bradyzoites as metabolically and transcriptionally active forms and not as dormant as was thought (Pittman et al., 2014). For instance, bradyzoites were shown to hijack the host brain microRNome by stimulating miR-146a and miR-155 expression, two key inflammatory response regulators (Cannella et al., 2014).

The influence of the host environment on the parasite gene transcription and differentiation can be observed at many levels, as the external host derived metabolic cues (check metabolism state section), or the different biophysical properties (temperature, consistency,

density and internal cytoskeletal processes) and mechanical load of the host cell (check actin section).

However, this influence is met by the one exerted by the parasite development upon its host, which as mentioned upwards, would be set differently depending on the stage of the parasite. In fact, Laura Knoll's lab reported the detection of many host immune regulatory genes were more abundant during chronic infection compared to the acute one (Pittman et al., 2014).

In *Leishmania*, intracellular infection of macrophages induced epigenetic modifications of the host cell genome resulting in the inhibition of their activation and hence of the host defense mechanisms (Marr et al., 2014; McMaster et al., 2016). Host cell remodeling was observed during gametogenesis in *Plasmodium falciparum*, as many of the early gametocyte genes were employed for that purpose (Brancucci et al., 2014b). In regards to *T. gondii*, we believe that we have brought a new layer to the parasite influence picture consisting of the MORC protein. We have observed a great number of host cells to be affected at their mRNA level following the KD of MORC, raising intriguing questions about its implication in the remodeling of the host for the benefit of the parasitism.

The signaling cascade leading to this parasite-external outcome is yet to be unraveled, but we believe that MORC might be acting through the secretome of *Toxoplasma* by indirectly regulating the action of effectors at the level of their expression, their activity or their secretion.

b. Rhoptry- and Dense granules-resident Effectors

It is now acknowledged that the parasite delivers a set of its organelles-resident proteins, so-called effectors, to the host cell cytosol and nucleus in order to manipulate the signaling pathways and responses of this latter, in the benefit of both the host and the parasite. These various effectors are released at distinct moments of the infection, with rhoptries and micronemes getting discharged at the moment of invasion, and the dense granules exocytosed later on out of the nascent parasitophorous vacuole (Carruthers and Sibley, 1997).

Many effectors were reported to target the immune defenses of the host, here mentioning some of the ones employing the host transcriptional programs in order to achieve their survival and immune evasion. It must be noted that the infection effect on the transcriptional patterns of the host would depend on the host cell type as well as on the parasite strains, as is the case of the NF- κ B pathway which was seen to be modulated exclusively by type II strain background (Rosowski et al., 2011).

The mechanisms through which the parasite is modulating the immune responses were unknown, until a few years ago when many effectors got characterized. TgIST was shown to inhibit the STAT1-dependent interferon-gamma-induced pro-inflammatory gene expression by altering the host chromatin through the recruitment of NuRD transcriptional co-repressor complex (Gay et al., 2016). More recently, another effector targeting the host chromatin state, was reported. TEEGR counteracts the nuclear factor NF- κ B pathway thereby restraining the host inflammatory responses. This effector is able to interact with host E2F3 and E2F4 transcription factors to promote the induction of the expression of EZH2 (a PRC2 subunit) which would target and downregulate the NF- κ B target genes. EZH2 would also repress the expression of some pro-inflammatory cytokines, as IL-6 and IL-8 through promoting a non-permissive chromatin state (Braun et al., 2019).

However, an under-active immune defense would lead to an excessive parasite burden resulting in killing the host. Bearing in mind that the parasite must undergo a transition into its bradyzoite form, the holder of the chronic infection holder, in order to proceed towards its sexual development and its transmission, it would seem only reasonable that it is not in the advantage of the parasite to kill its own host through high virulence during the acute phase.

Therefore, it is crucial for the parasite to establish a balanced host immune response, thus counterbalancing the parasite protecting effect through cytokines establishing pro-inflammatory responses thus ensuring the homeostasis required for preserving both the host and the parasite. One example of the pro-inflammation inducing effector would be GRA15 which would trigger a pro-inflammatory response mediated by NF- κ B, an action recently reported to depend on the interaction of GRA15 with tumor necrosis factors acting upstream of NF- κ B (Sangaré et al., 2019).

Another example is the GRA24 which modulated positively the host p38-alpha MAPK pathway, thus triggering the expression of an array of pro-inflammatory cytokines and chemokines, in turn contributing to the stimulation of a decisive interferon-gamma mediated parasite burden reduction (Braun et al., 2013).

Nevertheless, Inflammatory responses are not all what the parasite effectors modulate, with the example of GRA16 shown to positively regulate the host cell cycle progression and the p53 tumor suppressor pathway (Bougdour et al., 2013). *Theileria* parasites would employ a different strategy to trigger this same signaling pathway, involving the cytoplasmic sequestration of the host p53 to the schizont membrane (Haller et al., 2010). Herein to say that the effectors of *T. gondii* have co-evolved with their different hosts, and that not all parasitism approaches are common.

For instance, one of the closest relatives of *T. gondii*, *Hammondia hammondi* seem to share more than 95% of the host pathways which the infection of both would alter, yet the

magnitude was not as similar, especially regarding their pro-inflammatory potential. While *T. gondii* would create a homeostatic balanced immune response mostly suited for its own survival but also for its ability for dissemination and encystment within the host brain cells, *H. hammondi* seem to lack such trait, potentially due to its divergent IST effector, one which seem to hold a *T. gondii* innovation regarding the immune modulation, as this factor seems aberrantly inactive in *H. hammondi*, as well as being absent in the close relative *Neospora caninum* (Wong et al., 2020).

It remains uncovered how *T. gondii* establishes similar persisting mechanisms in its definitive host, and whether a distinct set of effectors would be involved in setting the appropriate host environments for the parasite to successfully undergo its sexual development. For instance, it was reported that GRA15, GRA16, GRA24 as well as the rhoptry-resident kinase ROP16 were not expressed in merozoites, so either the host pathways that these effectors are targeting in tachyzoite are not targeted in merozoite, or that other unidentified merozoite-specific paralogs would replace and perform their functions (Hehl et al., 2015).

However, the felines must have developed a sort of resistance against *T. gondii*, yet it seems different than the innate resistance adapted in rodents, which is at the base of *T. gondii*'s completion of its life cycle. The genes encoding homologues of the mice IFN γ induced IRG immune actors are in majority lacking in cats (Howard et al., 2011). This type of innate immune response is lacking also in higher mammals, which seems consistent with the absence of a co-adaptation between these species and the parasite, and their status as accidental intermediate hosts.

We mentioned earlier our finding on MORC depletion upregulating a set of the host mRNAs and our theory of a possible role of the secretome in this process. However, we are now seeking beyond the tachyzoite stage effectors, in attempt to discover sexual stages-specific effectors, Myr-1-independent (likely merozoite) or Myr-1-dependent (i.e. sporozoite or/and bradyzoite). The MORC designed KD tool would hinder the aforementioned limitations regarding studying the definitive host interactions with the parasite.

It also remains unknown whether the effects on the host on the chromatin and transcription levels emanating from the actions of the parasite effectors, would persist beyond the clearance of the parasite. Some of the effectors, discovered to date, were found to be either interacting with or regulating the activity of epigenetic factors of the host, recalling TgIST with NuRD, and TEEGR with EZH2.

It had been reported that the intracellular infection of macrophages by *Leishmania* would induce modifications in the host DNA methylation, letting one suspect a static effect on the host chromatin (Marr et al., 2014). Bacterial infections were also seen to affect the DNA methylome of the host cells (Bierne et al., 2012).

Therefore, it would be interesting to investigate whether an infection by *T. gondii* would induce lasting influence on the host epigenome. However, we must be also able to distinguish whether the epigenetic changes are induced by the parasite or whether they merely emanate from the infected host response and in this case the parasite turning these events into its advantage.

c. Developmental influence

The host immune response induced-conversion of the tachyzoite into the chronic bradyzoite forms, is required for progressing into the parasite's sexual development, therefore it would be reasonable to assess the impact of the host cell on this transition. In fact, a distinct level of permissiveness between different host cell regarding the development of tissue cysts has been described, in part due to their level of lactate and glycolysis (Lüder and Rahman, 2017). (Goerner et al., 2020).

Many features of the host cell could be priming the parasites into one or the other cell fate. For instance, the host nutrient level could vary this outcome as lipid contents, as well as oxygen levels and internal temperatures, such as the cat's gut would differ in its temperature than the neural or muscle tissues cells. These aspects represent some of many external signals that the parasite senses and responds to during its development, and the specificity of this response can be seen in, for example, the inability of oocysts to infect cats whereas they could excrete millions of oocysts following the ingestion of one bradyzoite (Dubey, 2006).

It has been seen that host-derived enzymes were responsible for post-translationally modifying the proteins of the tissue cyst wall (Tomita et al., 2017), herein the host cell granting the parasite with a shield against its own immune defense.

Another observation illustrating the host impact on the parasite differentiation lies in the particularity of the cat guts, where it would seem that felids are the only mammals known to lack the enzyme that catalyzes linoleic acid rendering their intestines charged by this particular fatty acid, of which the addition was reported to lead to an increase in the expression of merozoite stage-specific markers (Martorelli Di Genova et al., 2019).

Martorelli Di Genova B and colleagues sought to identify the potential factors behind the species specificity of *T. gondii* sexual development. They reported that adding linoleic acid to *in vitro* generated cat intestinal organoids, caused 35% of the parasite populations to express merozoite-specific markers. In fact, felids are the only mammals known to lack the enzyme that catalyzes linoleic acid rendering their intestines charged by this particular fatty acid. Their

attempts were carried on towards trying to mimic the cat gut environment within mice , but they couldn't succeed in producing efficiently competent oocysts.

In our attempts to assess the infectivity of the possible sexual forms arising following the MORC KD, we infected mice models with the treated parasites. Although the depletion of MORC protein resulted in the induction *in vitro* of the expression of cyst wall and sporozoite proteins (e.g. SporoSAG, **ED Fig. 7a**, Farhat *et al.*, 2020) recognized as hallmarks of mature oocysts, yet, unfortunately, few round shaped structures were obtained, and they were not infectious in mice when delivered *per os*. We speculate that the infectivity would require complementary subtle regulatory factors other than MORC's downregulation, ones that are normally triggered *in vivo* in cat intestinal milieu. We expect that artificially applying these factors would assist our MORC mediated genetic reprogramming into recapitulating the in-tissue culture of the complete sexual life cycle.

We aimed to assess the impact of the linoleic fatty acid on the activity of MORC, and following its addition to our infected HFF, we noticed a slight decrease in the expression of the protein of MORC, along with a modest increase of our merozoite reporter gene *TgME49_243940*. The aforementioned study reinforced our findings by demonstrating that employing metabolic components by themselves is not sufficient for recreating the felids gut environment required for the parasite to go through its full sexual development.

In order for the parasite to progress through its development, the activity of MORC must be dynamic as well as adjusted and complemented by unknown factors to date, yet It remains to be unraveled how the regulatory interactions are being translated into highly specific and adapted cell fate outcomes.

10. Cell cycle impact

Before addressing the impact of the parasite cell cycle on its own development, it is noteworthy to point at the ability of this parasite to alter the cell cycle progression of the host. Although the weight of intervening with such host pathways on the parasite persistence remains unclear, yet the fact that many parasite effectors target this path in an overlapping way suggests that the induced host G2/M arrest might involve a mean for preventing the host cell senescence (Brunet *et al.*, 2008; Molestina *et al.*, 2008).

Many cell cycle regulators were spotted to be regulated by the action of MORC and HDAC3, with many cell cycle inhibitors getting derepressed following the KD of MORC and the parasite being forced into deviating from its normal cycle progression. This was observed also on the

phenotypic cellular level where many parasites displayed round forms reminiscing dead cells with great damages on the nuclear level.

Many discussions suggested that cells respond differently to the differentiation signals depending on their cell cycle status (Sela et al., 2012) (Pauklin and Vallier, 2013). It has been reported that a permissive window for cell fate reprogramming occurs at mitosis (Halley-Stott et al., 2014).

Mitosis of *T. gondii* seems to be a semi-closed one, as adopted by *P. falciparum* during which, the nuclear membrane remains largely intact, and chromosomes don't seem to condense at any step of mitosis (Hager et al., 1999; White and Suvorova, 2018). To note that the replication of the parasite seems to involve an induction of the host cells to produce soluble factors to increase the replication of parasites in the neighboring host cells (Wong et al., 2020).

The division process must be finely tuned for an optimal speed balancing between a too fast host destructive division and a too slow virulence loosening one. The multitude of hosts environments and of life stages these parasites go through, acquired them with distinct division mechanisms suitable for the requirements of each stage, such as in the lack, in bradyzoite, of the cytokinesis step following mitosis, therefore reasonably they would also harbor developmentally controlled cell cycle progression checkpoints.

In higher Eukaryotes, the replication of DNA is accompanied by a passive dilution of DNA methylation, a process suggested to be facilitating cells reprogramming (Tsubouchi et al., 2013) (Ma et al., 2015), with others suggesting a mitotic advantage for reprogramming, based instead on cell cycle regulated active de-methylation (Halley-Stott et al., 2014).

In *T. gondii* no DNA methylation has been reported, yet the chromatin acetylation state is proven to be of great impact on the cell cycle, as well as on its differentiation outcomes, therefore a passive dilution of the histone marks during the S phase could maybe have similar fallout such as the ones described above, noting the impact that has the S/G2 checkpoint on bradyzoite differentiation (Radke et al., 2003).

In theory, if the KD effect of MORC hits the parasite when this latter is going through its mitosis, it would be much more responsive to getting its epigenetic programs reprogrammed, and this might explain some of the discrepancies we observed in the resulting cell fates of the different vacuoles following the KD.

Also, to be taken in consideration, that at no point during our experiments did we employ any cell synchronization tools, adding complexity to the gene regulatory aspects interpretations, as the used populations would consist of parasites at different points of their cell cycle.

The level of responsiveness would still be a matter of debate especially considering that long-distance interactions across the genome are lost during mitosis (Palozola et al., 2017) , a mechanism suggested to be adopted by MORC for its differentiation regulatory functions.

Here, an additional aspect might be highlighted which represents another *in vitro* limitation. In fact, while passing the parasites, they go through a period of time where they have no more host cells to feed on, therefore we are not recapitulating the real cell cycle of the parasite *in vivo* (oocysts stage considered as a non *in vivo* stage), where no real extracellular G0 stage takes place. Such extension of a phase that the parasite is not adapted to in normal conditions, adds another layer to be considered while interpreting the resulting differentiation outcomes.

11. Metabolism impact

Many enzymes rely on ions for an optimal activity, if not for any, and HDAC3 is no exception, therefore the metabolic balance is a crucial aspect for establishing the needed environment for these proteins to accomplish their roles. To name a few examples, many demethylases require α -ketoglutarate, a metabolic intermediate of the Krebs cycle, the Sirtuin family of histone deacetylases requires NAD⁺ as a cofactor, and Acetyl-CoA is the acetyl donor used by histone acetyltransferases.

Therefore, it would be of interest to dig deeper into the impact of this metabolic balance on the functioning of these enzymes, on the chromatin state and in that, on the differentiation from one stage to another. The crosstalk between metabolism homeostasis and gene regulation is an acknowledged concept as many metabolites represent corner stones of the nuclear enzymatic activities. For instance, the acetyl-CoA synthetase ACS2, was reported to undergo, upon metabolic stress, a translocation to the nucleus mediated by its phosphorylation, where it would hijack the HDAC generated acetate, and utilize it in favor of inducing the expression of autophagy genes (Li et al., 2017).

It has been reported that different stages adapted distinct modes of energy resourcing, as in bradyzoite relying on anaerobic glycolysis instead of aerobic respiration for energy production (Shukla et al., 2018) (Denton et al., 1996), meaning the Krebs cycle would be differed and with it the intermediate metabolites and in that the effect on chromatin regulatory enzymes activity. It is also known that each stage would express one gene isoform of the enolase protein (ENO) (Dzierszinski et al., 2001) and of the lactate dehydrogenase (LDH) with LDH being utilized during anaerobic growth conditions (Yang and Parmley, 1997), and even being described as having a role during tachyzoite conversion into bradyzoite (Abdelbaset et al., 2017).

It is worth noting that, following the KD of MORC and the upregulation of bradyzoite- and sexual stages-specific genes, we noticed a clean shift in the metabolic isoforms, herein the tachyzoite ENO2 was downregulated where the bradyzoite ENO1 got upregulated, keeping in mind the tight chromosomal proximity of these 2 genes, emphasizing more on the specific action of MORC and the lack of a leaky spreading transcription. The same pattern was observed in the stage specific LDH enzymes. It must be noted that in addition to their function

in glycolysis, the nuclear-targeted ENO isoenzymes were shown to bind chromosomal DNA, potentially acting on gene expression (Mouveaux et al., 2014).

The metabolic state of the parasite seems crucial for its fate determination, and it is potentially being sensed and translated with the aid of the aforementioned chromatin regulators, providing a new layer of investigation of the upstream events coordinating the activity of HDAC3, MORC and all the other yet unknown players involved in parasite differentiation. In *P. falciparum*, pfsIR2 has been accredited the role of metabolic sensor through its dependency to NAD⁺ for its deacetylating activity, and it is believed to play at adapting the rRNA synthesis and parasite proliferation according to the parasite's energy status (Mancio-Silva et al., 2013) (Imai and Guarente, 2010).

Here, light must be shed on the impact of the infected host cell and its metabolic status on the parasite proliferation and differentiation abilities. This concept has been reported in (Ferreira-da-Silva et al., 2009) (Lüder and Rahman) where they describe their *in vitro* trials of inducing tachyzoite differentiation using different host cells and claimed to observe significant differences in the conversion rates. Also, astrocytes seem to be a lot more permissive for the *in-vitro* auto conversion of tachyzoite-to-bradyzoite, without the need to stress the parasites with alkaline media (Goerner et al., 2020).

Also, with each host cell type, the parasite is faced with a different flux of available metabolites. For instance, the merozoite which infect only the *Felidae* enterocyte monolayer, encounter less variability of bioavailable metabolites, than the tachyzoite which infects different cell types. In addition, the position of merozoite close to the nutrients stream of the host enterocyte, as well as the low oxygen conditions in the gut, might explain the few discrepancies between these two forms regarding the differential metabolic gene expression.

The glycerol 3-phosphate dehydrogenase (NAD⁺), for instance, was seen to be highly expressed in merozoite compared to the tachyzoite, noting that its expression was detected to be induced following the KD of MORC. Also, the ENO2 protein was slightly higher in merozoite, hinting to the adaptive response towards a higher glycolytic throughput than the one in tachyzoite (Hehl et al., 2015).

It seems only reasonable to consider that the event of host cell infection would acquire both of them with metabolic changes that could be translated at the level of immunological responses of the host. For instance, the oxidative stress of the infected cell leading to proteins having their activities be altered post-translationally, and to genetic signatures from both players being epigenetically modified, as was seen in (Metheni et al., 2014) where they considered this stress as a metabolic epigenetic initiator (Berger et al., 2009).

12. Contractility and KELCH

The difference in the signaling environments of each of the host cell types that the parasite invades, might hold a layer of control for shaping the parasite into its most suited form for this or that life stage in particular. While studying the host cell effect on the parasite, one cannot disregard the biophysical properties of the host cell, and how the parasite might be sensing and translating these features, thereby adapting towards the fittest convenient.

It has been proposed that the fine tuning of the contractile forces exerted by actin would affect gene expression through determining the morphology of the nucleus (Uhler and Shivashankar, 2017). This can be of importance at the moment of invasion, yet whether the parasite invading distinct host cells harboring different contractility and density properties, would impact its gene regulation differently is still under studied.

We have brought solid evidence that links the MORC-dependent gene regulation to the cell fate of the parasite, therefore it would seem reasonable to philosophize the link between the actin-dependent contractility of the host cell, and the corresponding epigenetic signature the parasite would adopt in order to switch to its other form.

Moreover, the apicomplexan MORC protein, and unlike orthologues in other phyla, holds on its N-terminal repetitive stretches forming a KELCH domain, which we mentioned in Farhat *et al.*, 2020.

It is still unknown how is MORC's activity regulated, and in this regard, the protein presents many putative phosphorylation sites across its sequence, along with a monoubiquitination at the residue K472 in the vicinity of the Kelch motifs (See **Extended Data Fig. 1** in (Farhat *et al.*, 2020)), which are now acknowledged to act as substrate adaptors for the Cullin-RING E3 ubiquitin ligases (Shi *et al.*, 2019). Beyond proteasomal degradation, K472 mono-ubiquitination may serve as a signaling hub for others modifications (e.g. acetylation) in this so-called ubiquitin code (Swatek and Komander, 2016). Hence these repetitions might aid MORC in sensing the external environments, this leading potentially to a more adapted genetical reprogramming of the different stage-specific signatures.

Structural studies reported detecting these domains within some actin-binding and actin-responsive proteins (Hara *et al.*, 2004; Peng *et al.*, 2018; Soltysik-Espanola *et al.*, 1999). Therefore, there is a possibility that they might be able to sense the differential host contractile properties, creating a cascade of events starting with regulating MORC's function and ending in shaping the parasite's adopted form.

Another aspect to consider regarding actin's effect on the gene regulation, is underlined by its effect on the assembly of protein complexes and their translocation to the nucleus. We didn't provide any proof of the order of assembly of the MORC corepressor complex, nor did we

determine whether the interaction happens in the cytoplasm then gets translocated or whether it is all nuclear-based.

Actin was reported to aid proteins in coming together as well as in carrying them towards their targets, as well as the repositioning of genes from one nuclear compartment to the other (Zhang et al., 2011). Recalling the 800-KDa of globular weight of the different complexes MORC is embedded within as well as the many MORC-MORC dimers that could be formed, it would seem of much help for the nuclear actin to provide support of the nuclear bodies that could otherwise sediment and aggregate (Feric and Brangwynne, 2013; Miroshnikova et al., 2017).

13. Last words...

Taking all these aspects into account, I could not call this section a conclusion, because I believe we're far from revealing the developmental secrets of this parasite. I put the discussion ideas into distinct chapters, but in reality, they are all part of each other, all gathered, synchronized, ordered in their hierarchy, co-operating in order to bring out such a successful parasite. This is why I won't be able to put big open questions in these last few sentences, because the complexity of this system, and the lack of knowledge and lack of answers that we are facing, make it so that the open question is still the one main question:

What mechanisms are guiding the development of *T. gondii* in each life cycle stage?

And the answer is all of the aspects above!

Such a sophisticated parasite in need of tightly controlled responses to its various environments, would not possibly rely on one factor! This is why I am convinced that we merely scratched the surface by providing one additional piece of the puzzle to the community, that being the crucial weight of the MORC protein in the developmental pathways.

IV. Bibliography

- Abdelbaset, A.E., Fox, B.A., Karram, M.H., Abd Ellah, M.R., Bzik, D.J., and Igarashi, M. (2017). Lactate dehydrogenase in *Toxoplasma gondii* controls virulence, bradyzoite differentiation, and chronic infection. *PLoS ONE* *12*, e0173745.
- Abrahamsen, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., et al. (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* *304*, 441–445.
- Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nussaume, L., Noh, Y.-S., Amasino, R., and Scheres, B. (2004). The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell* *119*, 109–120.
- Akkaya, M., Bansal, A., Sheehan, P.W., Pena, M., Molina-Cruz, A., Orchard, L.M., Cimperman, C.K., Qi, C.-F., Ross, P., Yazew, T., et al. (2020). A single-nucleotide polymorphism in a *Plasmodium berghei* ApiAP2 transcription factor alters the development of host immunity. *Sci Adv* *6*, eaaw6957.
- Alizon, S., Hurford, A., Mideo, N., and Van Baalen, M. (2009). Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol.* *22*, 245–259.
- Allen, M.D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., and Suzuki, M. (1998). A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J* *17*, 5484–5496.
- Amit-Avraham, I., Pozner, G., Eshar, S., Fastman, Y., Kolevzon, N., Yavin, E., and Dzikowski, R. (2015). Antisense long noncoding RNAs regulate var gene activation in the malaria parasite *Plasmodium falciparum*. *PNAS* *112*, E982–E991.
- Anbar, A.D. (2008). Oceans. Elements and evolution. *Science* *322*, 1481–1483.
- Andrabi, S.B.A., Tahara, M., Matsubara, R., Toyama, T., Aonuma, H., Sakakibara, H., Suematsu, M., Tanabe, K., Nozaki, T., and Nagamune, K. (2018). Plant hormone cytokinins control cell cycle progression and plastid replication in apicomplexan parasites. *Parasitology International* *67*, 47–58.
- Andrews, F.H., Tong, Q., Sullivan, K.D., Cornett, E.M., Zhang, Y., Ali, M., Ahn, J., Pandey, A., Guo, A.H., Strahl, B.D., et al. (2016). Multivalent Chromatin Engagement and Inter-domain Crosstalk Regulate MORC3 ATPase. *Cell Reports* *16*, 3195–3207.
- Antonovics, J., Wilson, A.J., Forbes, M.R., Hauffe, H.C., Kallio, E.R., Leggett, H.C., Longdon, B., Okamura, B., Sait, S.M., and Webster, J.P. (2017). The evolution of transmission mode. *Philosophical Transactions of the Royal Society B: Biological Sciences* *372*, 20160083.
- Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Fejes Toth, K., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* *31*, 785–799.
- Aravind, L., and Landsman, D. (1998). AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* *26*, 4413–4421.
- Aravind, L., Iyer, L.M., Wellems, T.E., and Miller, L.H. (2003). *Plasmodium* biology: genomic gleanings. *Cell* *115*, 771–785.
- Archibald, J.M. (2009). The Puzzle of Plastid Evolution. *Current Biology* *19*, R81–R88.

- Argentaro, A., Yang, J.-C., Chapman, L., Kowalczyk, M.S., Gibbons, R.J., Higgs, D.R., Neuhaus, D., and Rhodes, D. (2007). Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc. Natl. Acad. Sci. U.S.A.* *104*, 11939–11944.
- Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., Bechner, M., et al. (2004). The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism. *Science* *306*, 79–86.
- Attar, N., Campos, O.A., Vogelauer, M., Cheng, C., Xue, Y., Schmollinger, S., Salwinski, L., Mallipeddi, N.V., Boone, B.A., Yen, L., et al. (2020). The histone H3-H4 tetramer is a copper reductase enzyme. *Science* *369*, 59–64.
- Ay, F., Bunnik, E.M., Varoquaux, N., Vert, J.-P., Noble, W.S., and Le Roch, K.G. (2015). Multiple dimensions of epigenetic gene regulation in the malaria parasite *Plasmodium falciparum*. *BioEssays* *37*, 182–194.
- Babu, M.M., Iyer, L.M., Balaji, S., and Aravind, L. (2006). The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* *34*, 6505–6520.
- Balaji, S., Babu, M.M., Iyer, L.M., and Aravind, L. (2005). Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res* *33*, 3994–4006.
- Bannister, L.H., Hopkins, J.M., Fowler, R.E., Krishna, S., and Mitchell, G.H. (2000). A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. *Parasitol. Today (Regul. Ed.)* *16*, 427–433.
- Baumgarten, S., Bryant, J.M., Sinha, A., Reyser, T., Preiser, P.R., Dedon, P.C., and Scherf, A. (2019). Transcriptome-wide dynamics of extensive m⁶A mRNA methylation during *Plasmodium falciparum* blood-stage development. *Nat Microbiol* 1–14.
- Behnke, M.S., Wootton, J.C., Lehmann, M.M., Radke, J.B., Lucas, O., Nawas, J., Sibley, L.D., and White, M.W. (2010). Coordinated Progression through Two Subtranscriptomes Underlies the Tachyzoite Cycle of *Toxoplasma gondii*. *PLOS ONE* *5*, e12354.
- Berg, A., Meza, T.J., Mahić, M., Thorstensen, T., Kristiansen, K., and Aalen, R.B. (2003). Ten members of the Arabidopsis gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development. *Nucleic Acids Res* *31*, 5291–5304.
- Berger, S.L., Kouzarides, T., Shiekhhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev* *23*, 781–783.
- Berna, L., Marquez, P., Cabrera, A., Greif, G., Francia, M.E., and Robello, C. (2020). Reevaluation of the *Toxoplasma gondii* and *Neospora caninum* genomes reveals misassembly, karyotype differences and chromosomal rearrangements. *BioRxiv* 2020.05.22.111195.
- Berney, C., and Pawlowski, J. (2006). A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proceedings of the Royal Society B: Biological Sciences* *273*, 1867–1872.

Bertschi, N.L., Toenhake, C.G., Zou, A., Niederwieser, I., Henderson, R., Moes, S., Jenoe, P., Parkinson, J., Bartfai, R., and Voss, T.S. (2017). Malaria parasites possess a telomere repeat-binding protein that shares ancestry with transcription factor IIIA. *Nature Microbiology* 2, 17033.

Bhattacharya, D., Yoon, H.S., and Hackett, J.D. (2004). Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* 26, 50–60.

Bhatti, M.M., Livingston, M., Mullapudi, N., and Sullivan, W.J. (2006). Pair of unusual GCN5 histone acetyltransferases and ADA2 homologues in the protozoan parasite *Toxoplasma gondii*. *Eukaryotic Cell* 5, 62–76.

Bierne, H., Hamon, M., and Cossart, P. (2012). Epigenetics and bacterial infections. *Cold Spring Harb Perspect Med* 2, a010272.

Bolden, J.E., Peart, M.J., and Johnstone, R.W. (2006). Anticancer activities of histone deacetylase inhibitors. *Nat Rev Drug Discov* 5, 769–784.

Botté, C.Y., and Yamaryo-Botté, Y. (2018). Complex Endosymbioses II: The Nonphotosynthetic Plastid of Apicomplexa Parasites (The Apicoplast) and Its Integrated Metabolism. *Methods Mol. Biol.* 1829, 37–54.

Bougdour, A., Maubon, D., Baldacci, P., Ortet, P., Bastien, O., Bouillon, A., Barale, J.-C., Pelloux, H., Ménard, R., and Hakimi, M.-A. (2009). Drug inhibition of HDAC3 and epigenetic control of differentiation in Apicomplexa parasites. *Journal of Experimental Medicine* 206, 953–966.

Bougdour, A., Durandau, E., Brenier-Pinchart, M.-P., Ortet, P., Barakat, M., Kieffer, S., Curt-Varesano, A., Curt-Bertini, R.-L., Bastien, O., Coute, Y., et al. (2013). 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.

Boyle, J.P., Rajasekar, B., Saeij, J.P.J., Ajioka, J.W., Berriman, M., Paulsen, I., Roos, D.S., Sibley, L.D., White, M.W., and Boothroyd, J.C. (2006). Just one cross appears capable of dramatically altering the population biology of a eukaryotic pathogen like *Toxoplasma gondii*. *PNAS* 103, 10514–10519.

Brabbs, T.R., He, Z., Hogg, K., Kamenski, A., Li, Y., Paszkiewicz, K.H., Moore, K.A., O'Toole, P., Graham, I.A., and Jones, L. (2013). The stochastic silencing phenotype of *Arabidopsis* *morc6* mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J.* 75, 836–846.

Brancucci, N.M.B., Bertschi, N.L., Zhu, L., Niederwieser, I., Chin, W.H., Wampfler, R., Freymond, C., Rottmann, M., Felger, I., Bozdech, Z., et al. (2014a). Heterochromatin Protein 1 Secures Survival and Transmission of Malaria Parasites. *Cell Host & Microbe* 16, 165–176.

Brancucci, N.M.B., Bertschi, N.L., Zhu, L., Niederwieser, I., Chin, W.H., Wampfler, R., Freymond, C., Rottmann, M., Felger, I., Bozdech, Z., et al. (2014b). Heterochromatin Protein 1 Secures Survival and Transmission of Malaria Parasites. *Cell Host & Microbe* 16, 165–176.

Braun, L., Cannella, D., Ortet, P., Barakat, M., Sautel, C.F., Kieffer, S., Garin, J., Bastien, O., Voinnet, O., and Hakimi, M.-A. (2010). A Complex Small RNA Repertoire Is Generated by a Plant/Fungal-Like Machinery and Effected by a Metazoan-Like Argonaute in the Single-Cell Human Parasite *Toxoplasma gondii*. *PLOS Pathogens* 6, e1000920.

Braun, L., Brenier-Pinchart, M.-P., Yogavel, M., Curt-Varesano, A., Curt-Bertini, R.-L., Hussain, T., Kieffer-Jaquinod, S., Coute, Y., Pelloux, H., Tardieux, I., et al. (2013). A *Toxoplasma* dense granule

protein, GRA24, modulates the early immune response to infection by promoting a direct and sustained host p38 MAPK activation. *The Journal of Experimental Medicine* 210, 2071–2086.

Braun, L., Brenier-Pinchart, M.-P., Hammoudi, P.-M., Cannella, D., Kieffer-Jaquinod, S., Voltaire, J., Josserand, V., Touquet, B., Couté, Y., Tardieux, I., et al. (2019). The *Toxoplasma* effector TEEGR promotes parasite persistence by modulating NF- κ B signalling via EZH2. *Nature Microbiology* 1.

Brogliè, K.E., Biddle, P., Cressman, R., and Brogliè, R. (1989). Functional analysis of DNA sequences responsible for ethylene regulation of a bean chitinase gene in transgenic tobacco. *Plant Cell* 1, 599–607.

Brooks, C.F., Francia, M.E., Gissot, M., Croken, M.M., Kim, K., and Striëpen, B. (2011). *Toxoplasma gondii* sequesters centromeres to a specific nuclear region throughout the cell cycle. *Proc Natl Acad Sci U S A* 108, 3767–3772.

Brunet, J., Pfaff, A.W., Abidi, A., Unoki, M., Nakamura, Y., Guinard, M., Klein, J.-P., Candolfi, E., and Mousli, M. (2008). *Toxoplasma gondii* exploits UHRF1 and induces host cell cycle arrest at G2 to enable its proliferation. *Cell. Microbiol.* 10, 908–920.

Bryant, J.M., Baumgarten, S., Dingli, F., Loew, D., Sinha, A., Claës, A., Preiser, P.R., Dedon, P.C., and Scherf, A. (2020). Exploring the virulence gene interactome with CRISPR/dCas9 in the human malaria parasite. *Molecular Systems Biology* 16, e9569.

Bunnik, E.M., Venkat, A., Shao, J., McGovern, K.E., Batugedara, G., Worth, D., Prudhomme, J., Lapp, S.A., Andolina, C., Ross, L.S., et al. (2019). Comparative 3D genome organization in apicomplexan parasites. *PNAS* 116, 3183–3192.

Burroughs, A.M., Allen, K.N., Dunaway-Mariano, D., and Aravind, L. (2006). Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol.* 361, 1003–1034.

Bushell, E., Gomes, A.R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modrzynska, K., Schwach, F., Martin, R.E., et al. (2017). Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell* 170, 260-272.e8.

Ca, S., and Jp, D. (1999). Ultrastructure of shizonts and merozoites of *Sarcocystis falcatula* in the lungs of budgerigars (*Melopsittacus undulatus*). *J Parasitol* 85, 630–637.

Campbell, T.L., De Silva, E.K., Olszewski, K.L., Elemento, O., and Llinás, M. (2010). Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* 6, e1001165.

Cannella, D., Brenier-Pinchart, M.-P., Braun, L., van Rooyen, J.M., Bougdour, A., Bastien, O., Behnke, M.S., Curt, R.-L., Curt, A., Saeij, J.P.J., et al. (2014). miR-146a and miR-155 delineate a MicroRNA fingerprint associated with *Toxoplasma* persistence in the host brain. *Cell Rep* 6, 928–937.

Carruthers, V.B., and Sibley, L.D. (1997). Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *Eur. J. Cell Biol.* 73, 114–123.

Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* 46, 347–366.

- Cavalier-smith, T. (2002). Origins of the machinery of recombination and sex. *Heredity* *88*, 125–141.
- Cavalier-Smith, T. (2009). Predation and eukaryote cell origins: A coevolutionary perspective. *The International Journal of Biochemistry & Cell Biology* *41*, 307–322.
- Cavalier-smith, T. (2010). Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biology Direct* *5*, 7.
- Cobbold, S.A., Santos, J.M., Ochoa, A., Perlman, D.H., and Llinás, M. (2016). Proteome-wide analysis reveals widespread lysine acetylation of major protein complexes in the malaria parasite. *Sci Rep* *6*, 19722.
- Coleman, B.I., Skillman, K.M., Jiang, R.H.Y., Childs, L.M., Altenhofen, L.M., Ganter, M., Leung, Y., Goldowitz, I., Kafsack, B.F.C., Marti, M., et al. (2014). A *Plasmodium falciparum* Histone Deacetylase Regulates Antigenic Variation and Gametocyte Conversion. *Cell Host & Microbe* *16*, 177–186.
- Corbett, K.D., and Berger, J.M. (2003). Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *EMBO J* *22*, 151–163.
- Coulson, R.M.R., Hall, N., and Ouzounis, C.A. (2004). Comparative Genomics of Transcriptional Control in the Human Malaria Parasite *Plasmodium falciparum*. *Genome Res* *14*, 1548–1554.
- Curt-Varesano, A., Braun, L., Ranquet, C., Hakimi, M.-A., and Bougdour, A. (2016). The aspartyl protease TgASP5 mediates the export of the *Toxoplasma* GRA16 and GRA24 effectors into host cells. *Cell. Microbiol.* *18*, 151–167.
- Darkin-Rattray, S.J., Gurnett, A.M., Myers, R.W., Dulski, P.M., Crumley, T.M., Allocco, J.J., Cannova, C., Meinke, P.T., Colletti, S.L., Bednarek, M.A., et al. (1996). Apicidin: a novel antiprotozoal agent that inhibits parasite histone deacetylase. *Proc. Natl. Acad. Sci. U.S.A.* *93*, 13143–13147.
- Daujat, S., Bauer, U.-M., Shah, V., Turner, B., Berger, S., and Kouzarides, T. (2002). Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Curr. Biol.* *12*, 2090–2097.
- Dave, A.J., and Godward, M.B.E. (1982). ULTRASTRUCTURAL STUDIES IN THE RHODOPHYTA. I. DEVELOPMENT OF MITOTIC SPINDLE POLES IN APOGLOSSUM RUSCIFOLIUM, KYLIN. *18*.
- DeBarry, J.D., and Kissinger, J.C. (2011). Jumbled Genomes: Missing Apicomplexan Synteny. *Mol Biol Evol* *28*, 2855–2871.
- Deitsch, K.W., Calderwood, M.S., and Wellems, T.E. (2001). Malaria. Cooperative silencing elements in var genes. *Nature* *412*, 875–876.
- Denton, H., Roberts, C.W., Alexander, J., Thong, K.W., and Coombs, G.H. (1996). Enzymes of energy metabolism in the bradyzoites and tachyzoites of *Toxoplasma gondii*. *FEMS Microbiol. Lett.* *137*, 103–108.
- Ding, Z., Gillespie, L.L., and Paterno, G.D. (2003). Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. *Mol. Cell. Biol.* *23*, 250–258.
- Dixon, S.E., Bhatti, M.M., Uversky, V.N., Dunker, A.K., and Sullivan, W.J. (2011). Regions of intrinsic disorder help identify a novel nuclear localization signal in *Toxoplasma gondii* histone acetyltransferase TgGCN5-B. *Molecular and Biochemical Parasitology* *175*, 192–195.

- Dong, W., Vannozzi, A., Chen, F., Hu, Y., Chen, Z., and Zhang, L. (2018). MORC Domain Definition and Evolutionary Analysis of the MORC Gene Family in Green Plants. *Genome Biol Evol* *10*, 1730–1744.
- van Dooren, G.G., Marti, M., Tonkin, C.J., Stimmler, L.M., Cowman, A.F., and McFadden, G.I. (2005). Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. *Mol. Microbiol.* *57*, 405–419.
- van Dooren, G.G., Stimmler, L.M., and McFadden, G.I. (2006). Metabolic maps and functions of the *Plasmodium* mitochondrion. *FEMS Microbiol. Rev.* *30*, 596–630.
- Dou, Z., McGovern, O.L., Cristina, M.D., and Carruthers, V.B. (2014). *Toxoplasma gondii* Ingests and Digests Host Cytosolic Proteins. *MBio* *5*.
- Douglas, A.E., and Raven, J.A. (2003). Genomes at the interface between bacteria and organelles. *Philos Trans R Soc Lond B Biol Sci* *358*, 5–518.
- Douse, C.H., Bloor, S., Liu, Y., Shamin, M., Tchasovnikarova, I.A., Timms, R.T., Lehner, P.J., and Modis, Y. (2018). Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. *Nature Communications* *9*, 651.
- Dubey, J.P. (2005). Unexpected oocyst shedding by cats fed *Toxoplasma gondii* tachyzoites: in vivo stage conversion and strain variation. *Vet. Parasitol.* *133*, 289–298.
- Dubey, J.P. (2006). Comparative infectivity of oocysts and bradyzoites of *Toxoplasma gondii* for intermediate (mice) and definitive (cats) hosts. *Vet. Parasitol.* *140*, 69–75.
- Dubey, J.P. (2009). History of the discovery of the life cycle of *Toxoplasma gondii*. *International Journal for Parasitology* *39*, 877–882.
- Dubey, J.P., and Frenkel, J.K. (1972). Cyst-Induced Toxoplasmosis in Cats*. *The Journal of Protozoology* *19*, 155–177.
- Duraisingh, M.T., and Skillman, K.M. (2018). Epigenetic Variation and Regulation in Malaria Parasites. *Annu. Rev. Microbiol.* *72*, 355–375.
- Dutta, R., and Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences* *25*, 24–28.
- Dzierszynski, F., Mortuaire, M., Dendouga, N., Popescu, O., and Tomavo, S. (2001). Differential expression of two plant-like enolases with distinct enzymatic and antigenic properties during stage conversion of the protozoan parasite *Toxoplasma gondii* 1 Edited by R. Huber. *Journal of Molecular Biology* *309*, 1017–1027.
- Eeftens, J., and Dekker, C. (2017). Catching DNA with hoops—biophysical approaches to clarify the mechanism of SMC proteins. *Nature Structural & Molecular Biology* *24*, 1012–1020.
- El Bissati, K., Suvorova, E.S., Xiao, H., Lucas, O., Upadhyya, R., Ma, Y., Angeletti, R.H., White, M.W., Weiss, L.M., and Kim, K. (2016). *Toxoplasma gondii* Arginine Methyltransferase 1 (PRMT1) Is Necessary for Centrosome Dynamics during Tachyzoite Cell Division. *MBio* *7*, e02094-02015.
- Epp, C., Li, F., Howitt, C.A., Chookajorn, T., and Deitsch, K.W. (2009). Chromatin associated sense and antisense noncoding RNAs are transcribed from the var gene family of virulence genes of the malaria parasite *Plasmodium falciparum*. *RNA* *15*, 116–127.

- Esseiva, A.C., Naguleswaran, A., Hemphill, A., and Schneider, A. (2004). Mitochondrial tRNA import in *Toxoplasma gondii*. *J. Biol. Chem.* 279, 42363–42368.
- Fan, Q., An, L., and Cui, L. (2004). *Plasmodium falciparum* Histone Acetyltransferase, a Yeast GCN5 Homologue Involved in Chromatin Remodeling. *Eukaryotic Cell* 3, 264–276.
- Farhat, D.C., Bowler M, M., Pontier, D., Belmudes, L., Mas, C., Couté, Y., Bougdour, A., Lagrange, T., Hakimi, M.A., and Swale, C. (in preparation). Structural Basis of m6A recognition by the CPSF complex in *Toxoplasma gondii*.
- Farhat, D.C., Swale, C., Dard, C., Cannella, D., Ortet, P., Barakat, M., Belmudes, L., De Bock, P.-J., Couté, Y., Bougdour, A., et al. (2020). A MORC-driven transcriptional switch controls *Toxoplasma* developmental trajectories and sexual commitment. *Nature Microbiology* 5, 570–583.
- Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C analyses in wild type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Mol Cell* 55, 694–707.
- Ferguson, D.J., Hutchison, W.M., Dunachie, J.F., and Siim, J.C. (1974). Ultrastructural study of early stages of asexual multiplication and microgametogony of *Toxoplasma gondii* in the small intestine of the cat. *Acta Pathol Microbiol Scand B Microbiol Immunol* 82, 167–181.
- Ferguson, D.J., Hutchison, W.M., and Siim, J.C. (1975). The ultrastructural development of the macrogamete and formation of the oocyst wall of *Toxoplasma gondii*. *Acta Pathol Microbiol Scand B* 83, 491–505.
- Ferguson, D.J.P., Sahoo, N., Pinches, R.A., Bumstead, J.M., Tomley, F.M., and Gubbels, M.-J. (2008). MORN1 Has a Conserved Role in Asexual and Sexual Development across the Apicomplexa. *Eukaryot Cell* 7, 698–711.
- Feric, M., and Brangwynne, C.P. (2013). A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells. *Nat. Cell Biol.* 15, 1253–1259.
- Ferreira-da-Silva, M. da F., Takács, A.C., Barbosa, H.S., Gross, U., and Lüder, C.G.K. (2009). Primary skeletal muscle cells trigger spontaneous *Toxoplasma gondii* tachyzoite-to-bradyzoite conversion at higher rates than fibroblasts. *Int. J. Med. Microbiol.* 299, 381–388.
- Filarsky, M., Frasnka, S.A., Niederwieser, I., Brancucci, N.M.B., Carrington, E., Carrió, E., Moes, S., Jenoe, P., Bartfai, R., and Voss, T.S. (2018). GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science* 359, 1259–1263.
- Flueck, C., Bartfai, R., Volz, J., Niederwieser, I., Salcedo-Amaya, A.M., Alako, B.T.F., Ehlgén, F., Ralph, S.A., Cowman, A.F., Bozdech, Z., et al. (2009). *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog.* 5, e1000569.
- Flueck, C., Bartfai, R., Niederwieser, I., Witmer, K., Alako, B.T.F., Moes, S., Bozdech, Z., Jenoe, P., Stunnenberg, H.G., and Voss, T.S. (2010). A Major Role for the *Plasmodium falciparum* ApiAP2 Protein PfSIP2 in Chromosome End Biology. *PLoS Pathog* 6, e1000784.
- Franco, M., Panas, M.W., Marino, N.D., Lee, M.-C.W., Buchholz, K.R., Kelly, F.D., Bednarski, J.J., Sleckman, B.P., Pourmand, N., and Boothroyd, J.C. (2016). A Novel Secreted Protein, MYR1, Is Central to *Toxoplasma*'s Manipulation of Host Cells. *MBio* 7, e02231-15.

- Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U.S.A.* *99*, 14584–14589.
- Frenkel, J.K. (1970). Pursuing *Toxoplasma*. *The Journal of Infectious Diseases* *122*, 553–559.
- Frenkel, J.K., and Smith, D.D. (2003). Determination of the genera of cyst-forming coccidia. *Parasitol Res* *91*, 384–389.
- Fritz, H.M., Buchholz, K.R., Chen, X., Durbin-Johnson, B., Rocke, D.M., Conrad, P.A., and Boothroyd, J.C. (2012). Transcriptomic analysis of *Toxoplasma* development reveals many novel functions and structures specific to sporozoites and oocysts. *PLoS ONE* *7*, e29998.
- Fujimoto, S.Y., Ohta, M., Usui, A., Shinshi, H., and Ohme-Takagi, M. (2000). *Arabidopsis* Ethylene-Responsive Element Binding Factors Act as Transcriptional Activators or Repressors of GCC Box-Mediated Gene Expression. *Plant Cell* *12*, 393–405.
- Ganji, M., Shaltiel, I., Bisht, S., Kim, E., Kalichava, A., Haering, C., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. *Science* *360*, eaar7831.
- Gardner, M.J., Williamson, D.H., and Wilson, R.J.M. (1991). A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Molecular and Biochemical Parasitology* *44*, 115–123.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* *419*, 498–511.
- Gay, G., Braun, L., Brenier-Pinchart, M.-P., Vollaire, J., Josserand, V., Bertini, R.-L., Varesano, A., Touquet, B., Bock, P.-J.D., Coute, Y., et al. (2016). *Toxoplasma gondii* TgIST co-opts host chromatin repressors dampening STAT1-dependent gene regulation and IFN- γ -mediated host defenses. *Journal of Experimental Medicine* jem.20160340.
- Gazzinelli, R.T., Mendonça-Neto, R., Lilue, J., Howard, J., and Sher, A. (2014). Innate resistance against *Toxoplasma gondii*: an evolutionary tale of mice, cats, and men. *Cell Host Microbe* *15*, 132–138.
- Gisselberg, J.E., Dellibovi-Ragheb, T.A., Matthews, K.A., Bosch, G., and Prigge, S.T. (2013). The Suf Iron-Sulfur Cluster Synthesis Pathway Is Required for Apicoplast Maintenance in Malaria Parasites. *PLOS Pathogens* *9*, e1003655.
- Gissot, M., Choi, S.-W., Thompson, R.F., Greally, J.M., and Kim, K. (2008). *Toxoplasma gondii* and *Cryptosporidium parvum* Lack Detectable DNA Cytosine Methylation. *Eukaryot Cell* *7*, 537–540.
- Goerner, A.L., Vizcarra, E.A., Hong, D.D., Bergersen, K.V., Alvarez, C.A., Talavera, M.A., Wilson, E.H., and White, M.W. (2020). An ex vivo model of *Toxoplasma* recrudescence. *BioRxiv* 2020.05.18.101931.
- Gould, S.B., Waller, R.F., and McFadden, G.I. (2008). Plastid Evolution. *Annual Review of Plant Biology* *59*, 491–517.
- Gray, M.W., Lang, B.F., and Burger, G. (2004). Mitochondria of protists. *Annu. Rev. Genet.* *38*, 477–524.

- Green, B.R. (2011). After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl *c*. *Photosynth Res* *107*, 103–115.
- Guenther, M.G., Lane, W.S., Fischle, W., Verdin, E., Lazar, M.A., and Shiekhhattar, R. (2000). A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-repeat protein linked to deafness. *Genes Dev* *14*, 1048–1057.
- Guérardel, Y., Leleu, D., Coppin, A., Liénard, L., Slomianny, C., Strecker, G., Ball, S., and Tomavo, S. (2005). Amylopectin biogenesis and characterization in the protozoan parasite *Toxoplasma gondii*, the intracellular development of which is restricted in the HepG2 cell line. *Microbes Infect.* *7*, 41–48.
- Haberland, M., Montgomery, R.L., and Olson, E.N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat. Rev. Genet.* *10*, 32–42.
- Hager, K.M., Striepen, B., Tilney, L.G., and Roos, D.S. (1999). The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*. *Journal of Cell Science* *112*, 2631–2638.
- Hailu, G.S., Robaa, D., Forgione, M., Sippl, W., Rotili, D., and Mai, A. (2017). Lysine Deacetylase Inhibitors in Parasites: Past, Present, and Future Perspectives. *J. Med. Chem.* *60*, 4780–4804.
- Hakimi, M.-A., Bochar, D.A., Chenoweth, J., Lane, W.S., Mandel, G., and Shiekhhattar, R. (2002). A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci U S A* *99*, 7420–7425.
- Hakimi, M.-A., Olias, P., and Sibley, L.D. (2017). review ali sibley. *Clin. Microbiol. Rev.* *30*, 615–645.
- Haller, D., Mackiewicz, M., Gerber, S., Beyer, D., Kullmann, B., Schneider, I., Ahmed, J.S., and Seitzer, U. (2010). Cytoplasmic sequestration of p53 promotes survival in leukocytes transformed by *Theileria*. *Oncogene* *29*, 3079–3086.
- Halley-Stott, R.P., Jullien, J., Pasque, V., and Gurdon, J. (2014). Mitosis Gives a Brief Window of Opportunity for a Change in Gene Transcription. *PLOS Biology* *12*, e1001914.
- Hammoudi, P.-M., Jacot, D., Mueller, C., Di Cristina, M., Dogga, S.K., Marq, J.-B., Romano, J., Tosetti, N., Dubrot, J., Emre, Y., et al. (2015). Fundamental Roles of the Golgi-Associated *Toxoplasma* Aspartyl Protease, ASP5, at the Host-Parasite Interface. *PLoS Pathog.* *11*, e1005211.
- Hara, T., Ishida, H., Raziuddin, R., Dorkhom, S., Kamijo, K., and Miki, T. (2004). Novel kelch-like protein, KLEIP, is involved in actin assembly at cell-cell contact sites of Madin-Darby canine kidney cells. *Mol. Biol. Cell* *15*, 1172–1184.
- Harb, O.S., and Roos, D.S. (2020). ToxoDB: Functional Genomics Resource for *Toxoplasma* and Related Organisms. In *Toxoplasma Gondii: Methods and Protocols*, C.J. Tonkin, ed. (New York, NY: Springer US), pp. 27–47.
- Harris, C.J., Husmann, D., Liu, W., Kasmi, F.E., Wang, H., Papikian, A., Pastor, W.A., Moissiard, G., Vashisht, A.A., Dangl, J.L., et al. (2016). Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. *PLOS Genetics* *12*, e1005998.
- Haumaitre, C., Lenoir, O., and Scharfmann, R. (2009). Directing cell differentiation with small-molecule histone deacetylase inhibitors: the example of promoting pancreatic endocrine cells. *Cell Cycle* *8*, 536–544.

- Hehl, A.B., Basso, W.U., Lippuner, C., Ramakrishnan, C., Okoniewski, M., Walker, R.A., Grigg, M.E., Smith, N.C., and Deplazes, P. (2015). Asexual expansion of *Toxoplasma gondii* merozoites is distinct from tachyzoites and entails expression of non-overlapping gene families to attach, invade, and replicate within feline enterocytes. *BMC Genomics* 16.
- Herrmann, D.C., Bärwald, A., Maksimov, A., Pantchev, N., Vrhovec, M.G., Conraths, F.J., and Schares, G. (2012). *Toxoplasma gondii* sexual cross in a single naturally infected feline host: Generation of highly mouse-virulent and avirulent clones, genotypically different from clonal types I, II and III. *Vet Res* 43, 39.
- Hillier, C., Pardo, M., Yu, L., Bushell, E., Sanderson, T., Metcalf, T., Herd, C., Anar, B., Rayner, J.C., Billker, O., et al. (2019). Landscape of the Plasmodium Interactome Reveals Both Conserved and Species-Specific Functionality. *Cell Reports* 28, 1635-1647.e5.
- Holmes, M., Itaas, V., and Ananvoranich, S. (2014). Sustained translational repression of lactate dehydrogenase 1 in *Toxoplasma gondii* bradyzoites is conferred by a small regulatory RNA hairpin. *The FEBS Journal* 281, 5077–5091.
- Hong, Y. (2010). Medaka Haploid Embryonic Stem Cells. In *Methods in Cell Biology*, (Elsevier), pp. 55–69.
- Hong, D.-P., Radke, J.B., and White, M.W. (2017). Opposing Transcriptional Mechanisms Regulate *Toxoplasma* Development. *MSphere* 2.
- Hooper, C., and Hilliker, A. (2013). Packing them up and dusting them off: RNA helicases and mRNA storage. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1829, 824–834.
- Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S.V., Rahman, M.A., Finne, K., Aalen, R.B., and Aasland, R. (2011). The CW domain, a new histone recognition module in chromatin proteins. *EMBO J* 30, 1939–1952.
- Howard, J.C., Hunn, J.P., and Steinfeldt, T. (2011). The IRG protein-based resistance mechanism in mice and its relation to virulence in *Toxoplasma gondii*. *Current Opinion in Microbiology* 14, 414–421.
- Hsiao, C.-H.C., Luisa Hiller, N., Haldar, K., and Knoll, L.J. (2013). A HT/PEXEL motif in *Toxoplasma* dense granule proteins is a signal for protein cleavage but not export into the host cell. *Traffic* 14, 519–531.
- Huang, S., Holmes, M.J., Radke, J.B., Hong, D.-P., Liu, T.-K., White, M.W., and Sullivan, W.J. (2017). *Toxoplasma gondii* AP2IX-4 Regulates Gene Expression during Bradyzoite Development. *MSphere* 2.
- Hunter, C.A., and Sibley, L.D. (2012). Modulation of innate immunity by *Toxoplasma gondii* virulence effectors. *Nat. Rev. Microbiol.* 10, 766–778.
- Ichikawa, Y., Nishimura, Y., Kurumizaka, H., and Shimizu, M. (2015). Nucleosome organization and chromatin dynamics in telomeres. *Biomolecular Concepts* 6, 67–75.
- Imai, S., and Guarente, L. (2010). Ten years of NAD-dependent SIR2 family deacetylases: implications for metabolic diseases. *Trends Pharmacol. Sci.* 31, 212–220.

- Inoue, N., Hess, K.D., Moreadith, R.W., Richardson, L.L., Handel, M.A., Watson, M.L., and Zinn, A.R. (1999). New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum. Mol. Genet.* *8*, 1201–1207.
- Iwanaga, S., Kaneko, I., Kato, T., and Yuda, M. (2012). Identification of an AP2-family Protein That Is Critical for Malaria Liver Stage Development. *PLOS ONE* *7*, e47557.
- Iyer, L.M., and Aravind, L. (2012). Insights from the architecture of the bacterial transcription apparatus. *J Struct Biol* *179*, 299–319.
- Iyer, L.M., Anantharaman, V., Wolf, M.Y., and Aravind, L. (2008a). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* *38*, 1–31.
- Iyer, L.M., Abhiman, S., and Aravind, L. (2008b). MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol Direct* *3*, 8.
- Jackson, A.P. (2015). The evolution of parasite genomes and the origins of parasitism. *Parasitology* *142*, S1–S5.
- Janouskovec, J., and Keeling, P.J. (2016). Evolution: Causality and the Origin of Parasitism. *Current Biology* *26*, R174–R177.
- Janouškovec, J., Horák, A., Oborník, M., Lukeš, J., and Keeling, P.J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *PNAS* *107*, 10949–10954.
- Janouškovec, J., Tikhonenkov, D.V., Burki, F., Howe, A.T., Kolísko, M., Mylnikov, A.P., and Keeling, P.J. (2015). Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *PNAS* *112*, 10200–10207.
- Jeffers, V., and Sullivan, W.J. (2012). Lysine Acetylation Is Widespread on Proteins of Diverse Function and Localization in the Protozoan Parasite *Toxoplasma gondii*. *Eukaryot Cell* *11*, 735–742.
- Jeffers, V., Yang, C., Huang, S., and Sullivan, W.J. (2017). Bromodomains in Protozoan Parasites: Evolution, Function, and Opportunities for Drug Development. *Microbiol. Mol. Biol. Rev.* *81*, e00047-16.
- Jeninga, M.D., Quinn, J.E., and Petter, M. (2019). ApiAP2 Transcription Factors in Apicomplexan Parasites. *Pathogens* *8*, 47.
- Jenkins, M., Trout, J.M., Higgins, J., Dorsch, M., Veal, D., and Fayer, R. (2003). Comparison of tests for viable and infectious *Cryptosporidium parvum* oocysts. *Parasitol. Res.* *89*, 1–5.
- Jones-Brando, L., Torrey, E.F., and Yolken, R. (2003). Drugs used in the treatment of schizophrenia and bipolar disorder inhibit the replication of *Toxoplasma gondii*. *Schizophr. Res.* *62*, 237–244.
- Josling, G.A., and Llinás, M. (2015). Sexual development in Plasmodium parasites: knowing when it's time to commit. *Nat. Rev. Microbiol.* *13*, 573–587.
- Josling, G.A., Russell, T.J., Venezia, J., Orchard, L., van Biljon, R., Painter, H.J., and Llinás, M. (2020). Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nature Communications* *11*, 1–13.

Jung, C., Lee, C.Y.-F., and Grigg, M.E. (2004). The SRS superfamily of Toxoplasma surface proteins. *Int. J. Parasitol.* *34*, 285–296.

Kafsack, B.F.C., Rovira-Graells, N., Clark, T.G., Bancells, C., Crowley, V.M., Campino, S.G., Williams, A.E., Drought, L.G., Kwiatkowski, D.P., Baker, D.A., et al. (2014). A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* *507*, 248–252.

Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* *27*, 470–478.

Kang, H.-G., Kuhl, J.C., Kachroo, P., and Klessig, D.F. (2008). CRT1, an Arabidopsis ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. *Cell Host Microbe* *3*, 48–57.

Kanno, T., Berta, D.G., and Sjögren, C. (2015). The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. *Cell Reports* *12*, 1471–1482.

Kanyal, A., Rawat, M., Gurung, P., Choubey, D., Anamika, K., and Karmodiya, K. (2018). Genome-wide survey and phylogenetic analysis of histone acetyltransferases and histone deacetylases of *Plasmodium falciparum*. *FEBS J.* *285*, 1767–1782.

Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* *13*, 424.

Karmodiya, K., Pradhan, S.J., Joshi, B., Jangid, R., Reddy, P.C., and Galande, S. (2015). A comprehensive epigenome map of *Plasmodium falciparum* reveals unique mechanisms of transcriptional regulation and identifies H3K36me2 as a global mark of gene suppression. *Epigenetics & Chromatin* *8*, 32.

Keeling, P.J. (2008). Bridge over troublesome plastids. *Nature* *451*, 896–897.

Keeling, P.J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* *365*, 729–748.

Keithly, J.S. (2008). The Mitochondrion-Related Organelle of *Cryptosporidium parvum*. In *Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes*, J. Tachezy, ed. (Berlin, Heidelberg: Springer), pp. 231–253.

Khan, A., Fux, B., Su, C., Dubey, J.P., Darde, M.L., Ajioka, J.W., Rosenthal, B.M., and Sibley, L.D. (2007). Recent transcontinental sweep of *Toxoplasma gondii* driven by a single monomorphic chromosome. *Proceedings of the National Academy of Sciences* *104*, 14872–14877.

Kilejian, A. (1975). Circular mitochondrial DNA from the avian malarial parasite *Plasmodium lophurae*. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* *390*, 276–284.

Kim, H., Yen, L., Wongpalee, S.P., Kirshner, J.A., Mehta, N., Xue, Y., Johnston, J.B., Burlingame, A.L., Kim, J.K., Loparo, J.J., et al. (2019). The Gene-Silencing Protein MORC-1 Topologically Entraps DNA and Forms Multimeric Assemblies to Cause DNA Compaction. *Molecular Cell* *75*, 700-710.e6.

- Kishore, S.P., Stiller, J.W., and Deitsch, K.W. (2013). Horizontal gene transfer of epigenetic machinery and evolution of parasitism in the malaria parasite *Plasmodium falciparum* and other apicomplexans. *BMC Evolutionary Biology* 13, 37.
- Kissinger, J.C., and DeBarry, J. (2011). Genome cartography: charting the apicomplexan genome. *Trends Parasitol* 27, 345–354.
- Kloehn, J., Oppenheim, R.D., Siddiqui, G., De Bock, P.-J., Kumar Dogga, S., Coute, Y., Hakimi, M.-A., Creek, D.J., and Soldati-Favre, D. (2020). Multi-omics analysis delineates the distinct functions of sub-cellular acetyl-CoA pools in *Toxoplasma gondii*. *BMC Biol* 18, 67.
- Koch, A., Kang, H.-G., Steinbrenner, J., Dempsey, D.A., Klessig, D.F., and Kogel, K.-H. (2017). MORC Proteins: Novel Players in Plant and Animal Health. *Front Plant Sci* 8.
- Köhler, S., Delwiche, C.F., Denny, P.W., Tilney, L.G., Webster, P., Wilson, R.J., Palmer, J.D., and Roos, D.S. (1997). A plastid of probable green algal origin in Apicomplexan parasites. *Science* 275, 1485–1489.
- Komaki-Yasuda, K., Okuwaki, M., Nagata, K., Kawazu, S., and Kano, S. (2013). Identification of a Novel and Unique Transcription Factor in the Intraerythrocytic Stage of *Plasmodium falciparum*. *PLOS ONE* 8, e74701.
- Koreny, L., and Field, M.C. (2016). Ancient Eukaryotic Origin and Evolutionary Plasticity of Nuclear Lamina. *Genome Biol Evol* 8, 2663–2671.
- Kozikowski, A.P., Chen, Y., Gaysin, A., Chen, B., D’Annibale, M.A., Suto, C.M., and Langley, B.C. (2007). Functional differences in epigenetic modulators—superiority of mercaptoacetamide-based histone deacetylase inhibitors relative to hydroxamates in cortical neuron neuroprotection studies. *J. Med. Chem.* 50, 3054–3061.
- Kuma, K., Iwabe, N., and Miyata, T. (1995). Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. *Mol Biol Evol* 12, 123–130.
- LeBoeuf, M., Terrell, A., Trivedi, S., Sinha, S., Epstein, J.A., Olson, E.N., Morrisey, E.E., and Millar, S.E. (2010). Hdac1 and Hdac2 act redundantly to control p63 and p53 functions in epidermal progenitor cells. *Dev. Cell* 19, 807–818.
- Lee, M.G., Wynder, C., Cooch, N., and Shiekhatar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* 437, 432–435.
- Lee, M.G., Wynder, C., Bochar, D.A., Hakimi, M.-A., Cooch, N., and Shiekhatar, R. (2006a). Functional Interplay between Histone Demethylase and Deacetylase Enzymes. *Mol. Cell. Biol.* 26, 6395–6402.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006b). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313.
- Leipe, D.D., and Landsman, D. (1997). Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily. *Nucleic Acids Res* 25, 3693–3697.

- Lesage, K.M., Huot, L., Mouveaux, T., Courjol, F., Saliou, J.-M., and Gissot, M. (2018). Cooperative binding of ApiAP2 transcription factors is crucial for the expression of virulence genes in *Toxoplasma gondii*. *Nucleic Acids Res.* *46*, 6057–6068.
- Levin, N.D. (1989). N.D. Levine, *The Protozoan Phylum Apicomplexa*, Vol. I and II, CRC Press Inc., Boca Raton, Florida (1988), £ 168.00 (set). *European Journal of Protistology* *24*, 291.
- Li, S., Yen, L., Pastor, W.A., Johnston, J.B., Du, J., Shew, C.J., Liu, W., Ho, J., Stender, B., Clark, A.T., et al. (2016). Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. *PNAS* *113*, E5108–E5116.
- Li, X., Yu, W., Qian, X., Xia, Y., Zheng, Y., Lee, J.-H., Li, W., Lyu, J., Rao, G., Zhang, X., et al. (2017). Nucleus-Translocated ACS2 Promotes Gene Transcription for Lysosomal Biogenesis and Autophagy. *Molecular Cell* *66*, 684-697.e9.
- Liggins, A.P., Cooper, C.D.O., Lawrie, C.H., Brown, P.J., Collins, G.P., Hatton, C.S., Pulford, K., and Banham, A.H. (2007). MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas. *Br. J. Haematol.* *138*, 479–486.
- Lilue, J., Müller, U.B., Steinfeldt, T., and Howard, J.C. (2013). Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. *ELife* *2*.
- Lindner, S.E., De Silva, E.K., Keck, J.L., and Llinás, M. (2010). Structural determinants of DNA binding by a *P. falciparum* ApiAP2 transcriptional regulator. *J. Mol. Biol.* *395*, 558–567.
- Lindner, S.E., Swearingen, K.E., Shears, M.J., Walker, M.P., Vrana, E.N., Hart, K.J., Minns, A.M., Sinnis, P., Moritz, R.L., and Kappe, S.H.I. (2019). Transcriptomics and proteomics reveal two waves of translational repression during the maturation of malaria parasite sporozoites. *Nat Commun* *10*, 4964.
- Liu, C., Schroeder, A.A., Kapur, V., and Abrahamsen, M.S. (1998). Telomeric sequences of *Cryptosporidium parvum*. *Mol. Biochem. Parasitol.* *94*, 291–296.
- Liu, N., Lee, C.H., Swigut, T., Grow, E., Gu, B., Bassik, M.C., and Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* *553*, 228–232.
- Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J., et al. (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in *Chlamydomonas* and *Plasmodium* gametes. *Genes Dev.* *22*, 1051–1068.
- Liu, Y., Tempel, W., Zhang, Q., Liang, X., Loppnau, P., Qin, S., and Min, J. (2016). Family-wide Characterization of Histone Binding Abilities of Human CW Domain-containing Proteins. *J. Biol. Chem.* *291*, 9000–9013.
- Lively, C.M., Craddock, C., and Vrijenhoek, R.C. (1990). Red Queen hypothesis supported by parasitism in sexual and clonal fish. *Nature* *344*, 864–866.
- Llinas, M., Kw, D., and Ts, V. (2008). *Plasmodium* gene regulation: far more to factor in. *Trends Parasitol* *24*, 551–556.
- Lopez-Rubio, J.-J., Mancio-Silva, L., and Scherf, A. (2009). Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* *5*, 179–190.

- Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2018). Histone Acetylation Inhibits RSC and Stabilizes the +1 Nucleosome. *Molecular Cell* 72, 594-600.e2.
- Lorenzi, H., Khan, A., Behnke, M.S., Namasivayam, S., Swapna, L.S., Hadjithomas, M., Karamycheva, S., Pinney, D., Brunk, B.P., Ajioka, J.W., et al. (2016). Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic *Toxoplasma gondii* genomes. *Nature Communications* 7, 10147.
- Lorković, Z.J., Naumann, U., Matzke, A.J.M., and Matzke, M. (2012). Involvement of a GHKL ATPase in RNA-Directed DNA Methylation in *Arabidopsis thaliana*. *Current Biology* 22, 933–938.
- Lüder, C.G.K., and Rahman, T. (2017). Impact of the host on *Toxoplasma* stage differentiation. *Microb Cell* 4, 203–211.
- Lüder, C.G.K., and Rahman, T. Impact of the host on *Toxoplasma* stage differentiation. *Microb Cell* 4, 203–211.
- Lyons, R.E., McLeod, R., and Roberts, C.W. (2002). *Toxoplasma gondii* tachyzoite–bradyzoite interconversion. *Trends in Parasitology* 18, 198–201.
- Ma, J.S., Sasai, M., Ohshima, J., Lee, Y., Bando, H., Takeda, K., and Yamamoto, M. (2014). Selective and strain-specific NFAT4 activation by the *Toxoplasma gondii* polymorphic dense granule protein GRA6. *J. Exp. Med.* 211, 2013–2032.
- Ma, Y., Kanakousaki, K., and Buttitta, L. (2015). How the cell cycle impacts chromatin architecture and influences cell fate. *Front. Genet.* 6.
- Magnani, E., Sjölander, K., and Hake, S. (2004). From Endonucleases to Transcription Factors: Evolution of the AP2 DNA Binding Domain in Plants. *Plant Cell* 16, 2265–2277.
- Mair, G.R., Braks, J.A.M., Garver, L.S., Dimopoulos, G., Hall, N., Wiegant, J.C.A.G., Dirks, R.W., Khan, S.M., Janse, C.J., and Waters, A.P. (2006). Translational Repression is essential for *Plasmodium* sexual development and mediated by a DDX6-type RNA helicase. *Science* 313, 667–669.
- Makarova, M., and Oliferenko, S. (2016). Mixing and matching nuclear envelope remodeling and spindle assembly strategies in the evolution of mitosis. *Current Opinion in Cell Biology* 41, 43–50.
- Mallo, N., Fellows, J., Johnson, C., and Sheiner, L. (2018). Protein Import into the Endosymbiotic Organelles of Apicomplexan Parasites. *Genes (Basel)* 9.
- Mancio-Silva, L., Lopez-Rubio, J.J., Claes, A., and Scherf, A. (2013). Sir2a regulates rDNA transcription and multiplication rate in the human malaria parasite *Plasmodium falciparum*. *Nat Commun* 4, 1530.
- Manger, I.D., Hehl, A., Parmley, S., Sibley, L.D., Marra, M., Hillier, L., Waterston, R., and Boothroyd, J.C. (1998). Expressed Sequence Tag Analysis of the Bradyzoite Stage of *Toxoplasma gondii*: Identification of Developmentally Regulated Genes. *Infect Immun* 66, 1632–1637.
- Manosalva, P., Manohar, M., Kogel, K.-H., Kang, H.-G., and Klessig, D.F. (2015). The GHKL ATPase MORC1 Modulates Species-Specific Plant Immunity in Solanaceae. *Mol. Plant Microbe Interact.* 28, 927–942.
- Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., and Mackay, J.P. (2011). Plant Homeodomain (PHD) Fingers of CHD4 Are Histone

- H3-binding Modules with Preference for Unmodified H3K4 and Methylated H3K9. *J Biol Chem* *286*, 11779–11791.
- Marr, A.K., Maclsaac, J.L., Jiang, R., Airo, A.M., Kobor, M.S., and McMaster, W.R. (2014). *Leishmania donovani* Infection Causes Distinct Epigenetic DNA Methylation Changes in Host Macrophages. *PLOS Pathogens* *10*, e1004419.
- Martins, R.M., Macpherson, C.R., Claes, A., Scheidig-Benatar, C., Sakamoto, H., Yam, X.Y., Preiser, P., Goel, S., Wahlgren, M., Sismeiro, O., et al. (2017). An ApiAP2 member regulates expression of clonally variant genes of the human malaria parasite *Plasmodium falciparum*. *Sci Rep* *7*, 14042.
- Martorelli Di Genova, B., Wilson, S.K., Dubey, J.P., and Knoll, L.J. (2019). Intestinal delta-6-desaturase activity determines host range for *Toxoplasma* sexual reproduction. *PLoS Biol* *17*.
- Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., et al. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Mol. Cell* *60*, 584–596.
- McFadden, G.I., Reith, M.E., Munholland, J., and Lang-Unnasch, N. (1996). Plastid in human parasites. *Nature* *381*, 482.
- McMaster, W.R., Morrison, C.J., and Kobor, M.S. (2016). Epigenetics: A New Model for Intracellular Parasite–Host Cell Regulation. *Trends in Parasitology* *32*, 515–521.
- Metheni, M., Echebli, N., Chaussepied, M., Ransy, C., Chéreau, C., Jensen, K., Glass, E., Batteux, F., Bouillaud, F., and Langsley, G. (2014). The level of H₂O₂ type oxidative stress regulates virulence of *Theileria*-transformed leukocytes. *Cell Microbiol* *16*, 269–279.
- Miao, J., Fan, Q., Cui, L., Li, X., Wang, H., Ning, G., Reese, J.C., and Cui, L. (2010). The MYST Family Histone Acetyltransferase Regulates Gene Expression and Cell Cycle in Malaria Parasite *Plasmodium falciparum*. *Mol Microbiol* *78*, 883–902.
- Millard, C.J., Watson, P.J., Celardo, I., Gordiyenko, Y., Cowley, S.M., Robinson, C.V., Fairall, L., and Schwabe, J.W.R. (2013). Class I HDACs Share a Common Mechanism of Regulation by Inositol Phosphates. *Mol Cell* *51*, 57–67.
- Mimura, Y., Takahashi, K., Kawata, K., Akazawa, T., and Inoue, N. (2010). Two-step colocalization of MORC3 with PML nuclear bodies. *Journal of Cell Science* *123*, 2014–2024.
- Miroshnikova, Y.A., Nava, M.M., and Wickström, S.A. (2017). Emerging roles of mechanical forces in chromatin regulation. *J Cell Sci* *130*, 2243–2250.
- Mizoi, J., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim. Biophys. Acta* *1819*, 86–96.
- Modrzynska, K., Pfander, C., Chappell, L., Yu, L., Suarez, C., Dundas, K., Gomes, A.R., Goulding, D., Rayner, J.C., Choudhary, J., et al. (2017). A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting Transcriptional Regulators Controlling the *Plasmodium* Life Cycle. *Cell Host Microbe* *21*, 11–22.
- Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., et al. (2012). MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science* *336*, 1448–1451.

- Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., Yen, L., Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A., et al. (2014). Transcriptional gene silencing by *Arabidopsis* microRNA homologues involves the formation of heteromers. *Proc. Natl. Acad. Sci. U.S.A.* *111*, 7474–7479.
- Molestina, R.E., El-Guendy, N., and Sinai, A.P. (2008). Infection with *Toxoplasma gondii* results in dysregulation of the host cell cycle. *Cell Microbiol* *10*, 1153–1165.
- Moore, R.B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D.H., Wright, S.W., Davies, N.W., Bolch, C.J.S., Heimann, K., et al. (2008). A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* *451*, 959–963.
- Mori, H., Urano, Y., Abe, F., Furukawa, S., Furukawa, S., Tsurumi, Y., Sakamoto, K., Hashimoto, M., Takase, S., Hino, M., et al. (2003). FR235222, a Fungal Metabolite, is a Novel Immunosuppressant that Inhibits Mammalian Histone Deacetylase (HDAC) I. Taxonomy, Fermentation, Isolation and Biological Activities. *J. Antibiot.* *56*, 72–79.
- Mouveaux, T., Oria, G., Werkmeister, E., Slomianny, C., Fox, B.A., Bzik, D.J., and Tomavo, S. (2014). Nuclear Glycolytic Enzyme Enolase of *Toxoplasma gondii* Functions as a Transcriptional Regulator. *PLOS ONE* *9*, e105820.
- Mukherjee, K., Brocchieri, L., and Bürglin, T.R. (2009). A Comprehensive Classification and Evolutionary Analysis of Plant Homeobox Genes. *Mol Biol Evol* *26*, 2775–2794.
- Nagaraj, V.A., Sundaram, B., Varadarajan, N.M., Subramani, P.A., Kalappa, D.M., Ghosh, S.K., and Padmanaban, G. (2013). Malaria parasite-synthesized heme is essential in the mosquito and liver stages and complements host heme in the blood stages of infection. *PLoS Pathog.* *9*, e1003522.
- Naguleswaran, A., Elias, E.V., McClintick, J., Edenberg, H.J., and Jr, W.J.S. (2010). *Toxoplasma gondii* Lysine Acetyltransferase GCN5-A Functions in the Cellular Response to Alkaline Stress and Expression of Cyst Genes. *PLOS Pathogens* *6*, e1001232.
- Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S. (2001). Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science* *292*, 110–113.
- Nardelli, S.C., Che, F.-Y., Silmon de Monerri, N.C., Xiao, H., Nieves, E., Madrid-Aliste, C., Angel, S.O., Sullivan, W.J., Angeletti, R.H., Kim, K., et al. (2013). The Histone Code of *Toxoplasma gondii* Comprises Conserved and Unique Posttranslational Modifications. *MBio* *4*, e00922-13.
- Nishi, M., Hu, K., Murray, J.M., and Roos, D.S. (2008). Organellar dynamics during the cell cycle of *Toxoplasma gondii*. *J. Cell. Sci.* *121*, 1559–1568.
- Nole-Wilson, S., and Krizek, B.A. (2000). DNA binding properties of the *Arabidopsis* floral development protein AINTEGUMENTA. *Nucleic Acids Res* *28*, 4076–4082.
- Normark, B.B. (2003). The evolution of alternative genetic systems in insects. *Annu. Rev. Entomol.* *48*, 397–423.
- Oberstaller, J., Pumpalova, Y., Schieler, A., Llinás, M., and Kissinger, J.C. (2014). The *Cryptosporidium parvum* ApiAP2 gene family: insights into the evolution of apicomplexan AP2 regulatory systems. *Nucleic Acids Res.* *42*, 8271–8284.

Oehring, S.C., Woodcroft, B.J., Moes, S., Wetzel, J., Dietz, O., Pulfer, A., Dekiwadia, C., Maeser, P., Flueck, C., Witmer, K., et al. (2012). Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum*. *Genome Biol* *13*, R108.

Ohki, I., Shimotake, N., Fujita, N., Nakao, M., and Shirakawa, M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J* *18*, 6653–6661.

Omodeo, O.P. (2010). The biggest evolutionary jump: restructuring of the genome and some consequences. *Tsitologiya* *52*, 797–816.

Otto, S.P., and Gerstein, A.C. (2008). The evolution of haploidy and diploidy. *Current Biology* *18*, R1121–R1124.

Otto, T.D., Wilinski, D., Assefa, S., Keane, T.M., Sarry, L.R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., et al. (2010). New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Molecular Microbiology* *76*, 12–24.

Otto, T.D., Böhme, U., Jackson, A.P., Hunt, M., Franke-Fayard, B., Hoeijmakers, W.A.M., Religa, A.A., Robertson, L., Sanders, M., Ogun, S.A., et al. (2014). A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol* *12*.

Palozola, K.C., Donahue, G., Liu, H., Grant, G.R., Becker, J.S., Cote, A., Yu, H., Raj, A., and Zaret, K.S. (2017). Mitotic Transcription and Waves of Gene Reactivation During Mitotic Exit. *Science* *358*, 119–122.

Parker, G.A., Ball, M.A., and Chubb, J.C. (2015). Evolution of complex life cycles in trophically transmitted helminths. I. Host incorporation and trophic ascent. *J. Evol. Biol.* *28*, 267–291.

Pastor, W.A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S.A., Moissiard, G., Zamudio, N., Bourc'his, D., et al. (2014). MORC1 represses transposable elements in the mouse male germline. *Nat Commun* *5*.

Patron, N.J., and Waller, R.F. (2007). Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *Bioessays* *29*, 1048–1058.

Pauklin, S., and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. *Cell* *155*, 135–147.

Peng, J.-M., Bera, R., Chiou, C.-Y., Yu, M.-C., Chen, T.-C., Chen, C.-W., Wang, T.-R., Chiang, W.-L., Chai, S.-P., Wei, Y., et al. (2018). Actin cytoskeleton remodeling drives epithelial-mesenchymal transition for hepatoma invasion and metastasis in mice. *Hepatology* *67*, 2226–2243.

Pieszko, M., Weir, W., Goodhead, I., Kinnaird, J., and Shiels, B. (2015). ApiAP2 Factors as Candidate Regulators of Stochastic Commitment to Merozoite Production in *Theileria annulata*. *PLoS Negl Trop Dis* *9*, e0003933.

Pittman, K.J., Aliota, M.T., and Knoll, L.J. (2014). Dual transcriptional profiling of mice and *Toxoplasma gondii* during acute and chronic infection. *BMC Genomics* *15*.

Preiser, P.R., Wilson, R.J., Moore, P.W., McCready, S., Hajibagheri, M.A., Blight, K.J., Strath, M., and Williamson, D.H. (1996). Recombination associated with replication of malarial mitochondrial DNA. *The EMBO Journal* *15*, 684–693.

- Radke, J.B., Lucas, O., Silva, E.K.D., Ma, Y., Sullivan, W.J., Weiss, L.M., Llinas, M., and White, M.W. (2013). ApiAP2 transcription factor restricts development of the *Toxoplasma* tissue cyst. *PNAS* *110*, 6871–6876.
- Radke, J.R., Striepen, B., Guerini, M.N., Jerome, M.E., Roos, D.S., and White, M.W. (2001). Defining the cell cycle for the tachyzoite stage of *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* *115*, 165–175.
- Radke, J.R., Guerini, M.N., Jerome, M., and White, M.W. (2003). A change in the premitotic period of the cell cycle is associated with bradyzoite differentiation in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* *131*, 119–127.
- Radke, J.R., Gubbels, M.-J., Jerome, M.E., Radke, J.B., Striepen, B., and White, M.W. (2004). Identification of a sporozoite-specific member of the *Toxoplasma* SAG superfamily via genetic complementation. *Mol. Microbiol.* *52*, 93–105.
- Ramakrishnan, C., Maier, S., Walker, R.A., Rehrauer, H., Joekel, D.E., Winiger, R.R., Basso, W.U., Grigg, M.E., Hehl, A.B., Deplazes, P., et al. (2019). An experimental genetically attenuated live vaccine to prevent transmission of *Toxoplasma gondii* by cats. *Sci Rep* *9*.
- Rauch, G., Kalbe, M., and Reusch, T.B.H. (2005). How a complex life cycle can improve a parasite's sex life. *J. Evol. Biol.* *18*, 1069–1075.
- Rider, S.D., and Zhu, G. (2009). An apicomplexan ankyrin-repeat histone deacetylase with relatives in photosynthetic eukaryotes. *Int. J. Parasitol.* *39*, 747–754.
- Riechmann, J.L., and Meyerowitz, E.M. (1998). The AP2/EREBP family of plant transcription factors. *Biol. Chem.* *379*, 633–646.
- Riester, D., Hildmann, C., Grünewald, S., Beckers, T., and Schwienhorst, A. (2007). Factors affecting the substrate specificity of histone deacetylases. *Biochem. Biophys. Res. Commun.* *357*, 439–445.
- Rosowski, E.E., Lu, D., Julien, L., Rodda, L., Gaiser, R.A., Jensen, K.D.C., and Saeij, J.P.J. (2011). Strain-specific activation of the NF- κ B pathway by GRA15, a novel *Toxoplasma gondii* dense granule protein. *J Exp Med* *208*, 195–212.
- Roy, S.W., and Penny, D. (2007). Widespread Intron Loss Suggests Retrotransposon Activity in Ancient Apicomplexans. *Mol Biol Evol* *24*, 1926–1933.
- Saksouk, N., Bhatti, M.M., Kieffer, S., Smith, A.T., Musset, K., Garin, J., Sullivan, W.J., Cesbron-Delauw, M.-F., and Hakimi, M.-A. (2005). Histone-Modifying Complexes Regulate Gene Expression Pertinent to the Differentiation of the Protozoan Parasite *Toxoplasma gondii*. *Mol. Cell. Biol.* *25*, 10301–10314.
- Sakuma, Y., Liu, Q., Dubouzet, J.G., Abe, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2002). DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem. Biophys. Res. Commun.* *290*, 998–1009.
- Sánchez-Solana, B., Li, D.-Q., and Kumar, R. (2014). Cytosolic Functions of MORC2 in Lipogenesis and Adipogenesis. *Biochim Biophys Acta* *1843*, 316–326.
- Sangaré, L.O., Yang, N., Konstantinou, E.K., Lu, D., Mukhopadhyay, D., Young, L.H., and Saeij, J.P.J. (2019). *Toxoplasma* GRA15 Activates the NF- κ B Pathway through Interactions with TNF Receptor-Associated Factors. *MBio* *10*.

- Sautel, C.F., Cannella, D., Bastien, O., Kieffer, S., Aldebert, D., Garin, J., Tardieux, I., Belrhali, H., and Hakimi, M.-A. (2007). SET8-Mediated Methylations of Histone H4 Lysine 20 Mark Silent Heterochromatic Domains in Apicomplexan Genomes. *MCB* 27, 5711–5724.
- Sautel, C.F., Ortet, P., Saksouk, N., Kieffer, S., Garin, J., Bastien, O., and Hakimi, M.-A. The histone methylase KMTox interacts with the redox-sensor peroxiredoxin-1 and targets genes involved in *Toxoplasma gondii* antioxidant defences. *Molecular Microbiology* 71, 212–226.
- Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J., and Lanzer, M. (1998). Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J* 17, 5418–5426.
- Scherf, A., Figueiredo, L.M., and Freitas-Junior, L.H. (2001). *Plasmodium* telomeres: a pathogen's perspective. *Curr. Opin. Microbiol.* 4, 409–414.
- Scherf, A., Lopez-Rubio, J.J., and Riviere, L. (2008). Antigenic Variation in *Plasmodium falciparum*. *Annual Review of Microbiology* 62, 445–470.
- Scholtyssek, E., Mehlhorn, H., and Hammond, D.M. (1972). Electron microscope studies of microgametogenesis in coccidia and related groups. *Z. Parasitenk.* 38, 95–131.
- Seeber, F., Limenitakis, J., and Soldati-Favre, D. (2008). Apicomplexan mitochondrial metabolism: a story of gains, losses and retentions. *Trends in Parasitology* 24, 468–478.
- Sela, Y., Molotski, N., Golan, S., Itskovitz-Eldor, J., and Soen, Y. (2012). Human embryonic stem cells exhibit increased propensity to differentiate during the G1 phase prior to phosphorylation of retinoblastoma protein. *Stem Cells* 30, 1097–1108.
- Sheffield, H.G., and Melton, M.L. (1968). The Fine Structure and Reproduction of *Toxoplasma gondii*. *The Journal of Parasitology* 54, 209–226.
- Shi, X., Xiang, S., Cao, J., Zhu, H., Yang, B., He, Q., and Ying, M. (2019). Kelch-like proteins: Physiological functions and relationships with diseases. *Pharmacol. Res.* 148, 104404.
- Shukla, A., Olszewski, K.L., Llinás, M., Rommereim, L.M., Fox, B.A., Bzik, D.J., Xia, D., Wastling, J., Beiting, D., Roos, D.S., et al. (2018). Glycolysis is important for optimal asexual growth and formation of mature tissue cysts by *Toxoplasma gondii*. *Int. J. Parasitol.* 48, 955–968.
- Sierra-Miranda, M., Vembar, S.-S., Delgadillo, D.M., Ávila-López, P.A., Herrera-Solorio, A.-M., Amado, D.L., Vargas, M., and Hernandez-Rivas, R. (2017). PfAP2Tel, harbouring a non-canonical DNA-binding AP2 domain, binds to *Plasmodium falciparum* telomeres. *Cellular Microbiology* 19, e12742.
- Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* 14, 6711–6719.
- Sindikubwabo, F., Ding, S., Hussain, T., Ortet, P., Barakat, M., Baumgarten, S., Cannella, D., Palencia, A., Bougdour, A., Belmudes, L., et al. (2017). Modifications at K31 on the lateral surface of histone H4 contribute to genome structure and expression in apicomplexan parasites. *ELife* 6.
- Sinha, A., Hughes, K.R., Modrzynska, K.K., Otto, T.D., Pfander, C., Dickens, N.J., Religa, A.A., Bushell, E., Graham, A.L., Cameron, R., et al. (2014). A cascade of DNA binding proteins for sexual commitment and development in *Plasmodium*. *Nature* 507, 253–257.

- Smith, D.R., and Keeling, P.J. (2015). Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci U S A* *112*, 10177–10184.
- Soltysik-Espanola, M., Rogers, R.A., Jiang, S., Kim, T.-A., Gaedigk, R., White, R.A., Avraham, H., and Avraham, S. (1999). Characterization of Mayven, a Novel Actin-binding Protein Predominantly Expressed in Brain. *MBoC* *10*, 2361–2375.
- Speer, C.A., and Dubey, J.P. (2005). Ultrastructural differentiation of *Toxoplasma gondii* schizonts (types B to E) and gamonts in the intestines of cats fed bradyzoites. *Int. J. Parasitol.* *35*, 193–206.
- Spillman, N.J., Beck, J.R., and Goldberg, D.E. (2015). Protein export into malaria parasite-infected erythrocytes: mechanisms and functional consequences. *Annu. Rev. Biochem.* *84*, 813–841.
- Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics* *13*, 613–626.
- Srivastava, S., White, M.W., and Sullivan, W.J. (2020). *Toxoplasma gondii* AP2XII-2 contributes to proper progression through S-phase of the cell cycle. *BioRxiv* 2020.06.09.143586.
- Striepen, B., Crawford, M.J., Shaw, M.K., Tilney, L.G., Seeber, F., and Roos, D.S. (2000). The Plastid of *Toxoplasma gondii* Is Divided by Association with the Centrosomes. *J Cell Biol* *151*, 1423–1434.
- Strobl, J.S., Cassell, M., Mitchell, S.M., Reilly, C.M., and Lindsay, D.S. (2007). Scriptaid and suberoylanilide hydroxamic acid are histone deacetylase inhibitors with potent anti-*Toxoplasma gondii* activity in vitro. *J. Parasitol.* *93*, 694–700.
- Su, C., Evans, D., Cole, R.H., Kissinger, J.C., Ajioka, J.W., and Sibley, L.D. (2003). Recent expansion of *Toxoplasma* through enhanced oral transmission. *Science* *299*, 414–416.
- Sullivan, W.J., Naguleswaran, A., and Angel, S.O. (2006). Histones and histone modifications in protozoan parasites. *Cell. Microbiol.* *8*, 1850–1861.
- Swatek, K.N., and Komander, D. (2016). Ubiquitin modifications. *Cell Res* *26*, 399–422.
- Takahashi, K., Yoshida, N., Murakami, N., Kawata, K., Ishizaki, H., Tanaka-Okamoto, M., Miyoshi, J., Zinn, A.R., Shime, H., and Inoue, N. (2007). Dynamic regulation of p53 subnuclear localization and senescence by MORC3. *Mol. Biol. Cell* *18*, 1701–1709.
- Tamkun, J.W., Deuring, R., Scott, M.P., Kissinger, M., Pattatucci, A.M., Kaufman, T.C., and Kennison, J.A. (1992). brahma: A regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF 2 SWI 2. *Cell* *68*, 561–572.
- Tandel, J., English, E.D., Sateriale, A., Gullicksrud, J.A., Beiting, D.P., Sullivan, M.C., Pinkston, B., and Striepen, B. (2019). Life cycle progression and sexual development of the apicomplexan parasite *Cryptosporidium parvum*. *Nat Microbiol* *4*, 2226–2236.
- Templeton, T.J., Iyer, L.M., Anantharaman, V., Enomoto, S., Abrahante, J.E., Subramanian, G.M., Hoffman, S.L., Abrahamsen, M.S., and Aravind, L. (2004). Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res.* *14*, 1686–1695.
- Thiriet-Rupert, S., Carrier, G., Chénais, B., Trottier, C., Bougaran, G., Cadoret, J.-P., Schoefs, B., and Saint-Jean, B. (2016). Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC Genomics* *17*.

- Thomas, S., Green, A., Sturm, N.R., Campbell, D.A., and Myler, P.J. (2009). Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics* *10*, 152.
- Toenhake, C.G., Frasncka, S.A.-K., Vijayabaskar, M.S., Westhead, D.R., van Heeringen, S.J., and Bártfai, R. (2018). Chromatin Accessibility-Based Characterization of the Gene Regulatory Network Underlying *Plasmodium falciparum* Blood-Stage Development. *Cell Host & Microbe* *23*, 557-569.e9.
- Tomavo, S., Slomianny, C., Meissner, M., and Carruthers, V.B. (2013). Protein trafficking through the endosomal system prepares intracellular parasites for a home invasion. *PLoS Pathog.* *9*, e1003629.
- Tomita, T., Sugi, T., Yakubu, R., Tu, V., Ma, Y., and Weiss, L.M. (2017). Making Home Sweet and Sturdy: *Toxoplasma gondii* ppGalNAc-Ts Glycosylate in Hierarchical Order and Confer Cyst Wall Rigidity. *MBio* *8*.
- Tsubouchi, T., Soza-Ried, J., Brown, K., Piccolo, F.M., Cantone, I., Landeira, D., Bagci, H., Hohegger, H., Merkschlager, M., and Fisher, A.G. (2013). DNA synthesis is required for reprogramming mediated by stem cell fusion. *Cell* *152*, 873–883.
- Uhler, C., and Shivashankar, G.V. (2017). Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nature Reviews Molecular Cell Biology* *18*, 717–727.
- Ukaegbu, U.E., Kishore, S.P., Kwiatkowski, D.L., Pandarinath, C., Dahan-Pasternak, N., Dzikowski, R., and Deitsch, K.W. (2014). Recruitment of PfSET2 by RNA Polymerase II to Variant Antigen Encoding Loci Contributes to Antigenic Variation in *P. falciparum*. *PLOS Pathogens* *10*, e1003854.
- Vaidya, A.B., Akella, R., and Suplick, K. (1989). Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol. Biochem. Parasitol.* *35*, 97–107.
- Vetting, M.W., Hegde, S.S., Fajardo, J.E., Fiser, A., Roderick, S.L., Takiff, H.E., and Blanchard, J.S. (2006). The Pentapeptide Repeat Proteins. *Biochemistry* *45*, 1–10.
- Viola, I.L., and Gonzalez, D.H. (2016). Chapter 6 - Structure and Evolution of Plant Homeobox Genes. In *Plant Transcription Factors*, D.H. Gonzalez, ed. (Boston: Academic Press), pp. 101–112.
- Vleugel, M., Hoogendoorn, E., Snel, B., and Kops, G.J.P.L. (2012). Evolution and Function of the Mitotic Checkpoint. *Developmental Cell* *23*, 239–250.
- Volz, J.C., Bártfai, R., Petter, M., Langer, C., Josling, G.A., Tsuboi, T., Schwach, F., Baum, J., Rayner, J.C., Stunnenberg, H.G., et al. (2012). PfSET10, a *Plasmodium falciparum* Methyltransferase, Maintains the Active var Gene in a Poised State during Parasite Division. *Cell Host & Microbe* *11*, 7–18.
- Waldman, B.S., Schwarz, D., Wadsworth, M.H., Saeij, J.P., Shalek, A.K., and Lourido, S. (2020). Identification of a Master Regulator of Differentiation in *Toxoplasma*. *Cell* *180*, 359-372.e16.
- Walker, D., Chaddock, A.M., Chaddock, J.A., Roberts, L.M., Lord, J.M., and Robinson, C. (1996). Ricin A chain fused to a chloroplast-targeting signal is unfolded on the chloroplast surface prior to import across the envelope membranes. *J. Biol. Chem.* *271*, 4082–4085.
- Walker, R., Gissot, M., Huot, L., Alayi, T.D., Hot, D., Marot, G., Schaeffer-Reiss, C., Van Dorsselaer, A., Kim, K., and Tomavo, S. (2013a). *Toxoplasma* transcription factor TgAP2XI-5 regulates the expression of genes involved in parasite virulence and host invasion. *J. Biol. Chem.* *288*, 31127–31138.

- Walker, R., Gissot, M., Croken, M.M., Huot, L., Hot, D., Kim, K., and Tomavo, S. (2013b). The Toxoplasma nuclear factor TgAP2XI-4 controls bradyzoite gene expression and cyst formation. *Mol Microbiol* 87, 641–655.
- Walker, R.A., Sharman, P.A., Miller, C.M., Lippuner, C., Okoniewski, M., Eichenberger, R.M., Ramakrishnan, C., Brossier, F., Deplazes, P., Hehl, A.B., et al. (2015). RNA Seq analysis of the Eimeria tenella gametocyte transcriptome reveals clues about the molecular basis for sexual reproduction and oocyst biogenesis. *BMC Genomics* 16, 94.
- Wang, J., Dixon, S.E., Ting, L.-M., Liu, T.-K., Jeffers, V., Croken, M.M., Calloway, M., Cannella, D., Hakimi, M.A., Kim, K., et al. (2014). Lysine Acetyltransferase GCN5b Interacts with AP2 Factors and Is Required for Toxoplasma gondii Proliferation. *PLOS Pathogens* 10, e1003830.
- Waters, A.P. (2016). Epigenetic Roulette in Blood Stream Plasmodium: Gambling on Sex. *PLOS Pathogens* 12, e1005353.
- Watson, M.L., Zinn, A.R., Inoue, N., Hess, K.D., Cobb, J., Handel, M.A., Halaban, R., Duchene, C.C., Albright, G.M., and Moreadith, R.W. (1998). Identification of morc (microrchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14361–14366.
- Watve, M. (1998). The Red Queen: Sex and the evolution of human nature. *Resonance* 3, 79–80.
- Webster, J.P. (1994). Prevalence and transmission of Toxoplasma gondii in wild brown rats, Rattus norvegicus. *Parasitology* 108 (Pt 4), 407–411.
- Webster, J.P. (2010). Review of “Toxoplasmosis of Animals and Humans (Second Edition)” by J.P. Dubey. *Parasit Vectors* 3, 112.
- Weedall, G.D., and Hall, N. (2015). Sexual reproduction and genetic exchange in parasitic protists. *Parasitology* 142, S120–S127.
- Weiser, N.E., Yang, D.X., Feng, S., Kalinava, N., Brown, K.C., Khanikar, J., Freeberg, M.A., Snyder, M.J., Csankovszki, G., Chan, R.C., et al. (2017). MORC-1 integrates nuclear RNAi and transgenerational chromatin architecture to promote germline immortality. *Dev Cell* 41, 408-423.e7.
- Weiss, L.M., and Dubey, Jitender.P. (2009). Toxoplasmosis: a history of clinical observations. *Int J Parasitol* 39, 895–901.
- White, M.W., and Suvorova, E.S. (2018). Apicomplexa Cell Cycles: Something Old, Borrowed, Lost, and New. *Trends in Parasitology* 34, 759–771.
- Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA polymerase II Regulates Mammalian Developmental Potential Through Control of Signaling Networks. *Mol Cell* 58, 311–322.
- Wojciak, J.M., Connolly, K.M., and Clubb, R.T. (1999). NMR structure of the Tn916 integrase-DNA complex. *Nat. Struct. Biol.* 6, 366–373.
- Wong, Z.S., Sokol-Borrelli, S.L., Olias, P., Dubey, J.P., and Boyle, J.P. (2020). Head-to-head comparisons of Toxoplasma gondii and its near relative Hammondia hammondi reveal dramatic differences in the host response and effectors with species-specific functions. *PLOS Pathogens* 16, e1008528.

- Woo, Y.H., Ansari, H., Otto, T.D., Klinger, C.M., Kolisko, M., Michálek, J., Saxena, A., Shanmugam, D., Tayyrov, A., Veluchamy, A., et al. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *ELife* 4, e06974.
- X, X., Z, Y., Jb, H., El, F., and M, F. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the *Drosophila* mesoderm (*Genes Dev*).
- Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell* 2, 851–861.
- Yahiaoui, B., Dzierszynski, F., Bernigaud, A., Slomianny, C., Camus, D., and Tomavo, S. (1999). Isolation and characterization of a subtractive library enriched for developmentally regulated transcripts expressed during encystation of *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* 99, 223–235.
- Yamamoto, M., Standley, D.M., Takashima, S., Saiga, H., Okuyama, M., Kayama, H., Kubo, E., Ito, H., Takaura, M., Matsuda, T., et al. (2009). A single polymorphic amino acid on *Toxoplasma gondii* kinase ROP16 determines the direct and strain-specific activation of Stat3. *Journal of Experimental Medicine* 206, 2747–2760.
- Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., and Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends in Plant Science* 18, 267–276.
- Yang, S., and Parmley, S.F. (1997). *Toxoplasma gondii* expresses two distinct lactate dehydrogenase homologous genes during its life cycle in intermediate hosts. *Gene* 184, 1–12.
- Yant, L., Mathieu, J., Dinh, T.T., Ott, F., Lanz, C., Wollmann, H., Chen, X., and Schmid, M. (2010). Orchestration of the Floral Transition and Floral Development in *Arabidopsis* by the Bifunctional Transcription Factor APETALA2[W][OA]. *Plant Cell* 22, 2156–2170.
- Yuda, M., Iwanaga, S., Shigenobu, S., Mair, G.R., Janse, C.J., Waters, A.P., Kato, T., and Kaneko, I. (2009). Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Molecular Microbiology* 71, 1402–1414.
- Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T., and Kaneko, I. (2010). Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Mol. Microbiol.* 75, 854–863.
- Zeng, L., Zhang, Q., Li, S., Plotnikov, A.N., Walsh, M.J., and Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* 466, 258–262.
- Zeyl, C., Vanderford, T., and Carter, M. (2003). An Evolutionary Advantage of Haploidy in Large Yeast Populations. *Science* 299, 555–558.
- Zhang, C., Li, Z., Cui, H., Jiang, Y., Yang, Z., Wang, X., Gao, H., Liu, C., Zhang, S., Su, X., et al. (2017). Systematic CRISPR-Cas9-Mediated Modifications of *Plasmodium yoelii* ApiAP2 Genes Reveal Functional Insights into Parasite Development. *MBio* 8.
- Zhang, Q., Huang, Y., Zhang, Y., Fang, X., Claes, A., Duchateau, M., Namane, A., Lopez-Rubio, J.-J., Pan, W., and Scherf, A. (2011). A Critical Role of Perinuclear Filamentous Actin in Spatial Repositioning and Mutually Exclusive Expression of Virulence Genes in Malaria Parasites. *Cell Host & Microbe* 10, 451–463.

Zhang, Y.W., Halonen, S.K., Ma, Y.F., Wittner, M., and Weiss, L.M. (2001). Initial Characterization of CST1, a *Toxoplasma gondii* Cyst Wall Glycoprotein. *Infect Immun* 69, 501–507.

Zhu, A.Y., Zhou, Y., Khan, S., Deitsch, K.W., Hao, Q., and Lin, H. (2012). *Plasmodium falciparum* Sir2A preferentially hydrolyzes medium and long chain fatty acyl lysine. *ACS Chem Biol* 7, 155–159.