Understanding DNA replication of Plasmodium falciparum

Casilda Muñoz Castellano

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale Sciencies Chimiques et Biologuiques pour la Santé (CBS2)

Unité de recherche Laboratory of Pathogen and Host Immunity (LPHI) UMR5294 CNRS

Understanding DNA replication of *Plasmodium falciparum*

Présentée par Casilda MUÑOZ CASTELLANO

Le 03 Octobre 2023

Sous la direction de Ana Rita Gomes

Devant le jury composé de

Elena GOMEZ DIAZ, Professor, Consejo Superior de Investigaciones Científicas (CSIC)	**Rapportrice**
Benoit MIOTTO, DR, Institut Cochin	**Rapporteur**
Mathieu BROCHET, Professor, Université de Genève	**Examinateur**
Jérôme POLI, MCU, Institut de Génétique Humaine (IGH), Université de Montpellier	**Examinateur**
Jenny WU, DR, Institute of Biochemistry and Cellular Genetics (IBGC)	**Examinatrice**
Ana Rita GOMES, CRCN, Laboratory of Pathogen and Host Immunity (LPHI)	**Directrice de thèse**

UNIVERSITÉ DE MONTPELLIER

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Ana Rita Gomes. For giving me the opportunity to work in a laboratory in the vanguard of malaria research. Your continuous guidance and boundless patience throughout this research journey, which has been often challenging, were pivotal for the development of this project. Your mentorship has had a profound impact on my academic and personal growth.

I also sincerely appreciate the members of the jury for their invaluable insights regarding my manuscripts and the enlightening discussions shared during my thesis defense. Elena, Benoit, Jenny, Jérôme and Mathieu; your constructive feedback has been essential in shaping the quality and rigor of this research.

I would also like to extend my thanks to Kai, Mauld, JJ, Rachel, Sharon and the rest of the dedicated members of the LPHI lab. Your support and insightful scientific discussions were invaluable. The collaborative spirit within the lab significantly enriched the quality of my work.

I would also like to acknowledge the funding provided by the CNRS, and the Université de Montpellier for the academic training received, for giving me a solid knowledge base without which this research would not have been possible. I would also like to thank Laurent Lacroix for the invaluable help with NanoForkSpeed; and to Vicky and Elodie from the MRI for your training and help with the imaging.

The biggest thank you to my fellow PhD army. Where to start? Marie, my chaotic sister but with great taste in music; you have been next to me throughout these three years and your company during those late nights in the lab has made the struggle worthwhile. Syrian, thank you and your gang of singes for adopting me from the start and introducing me to the wonders of Babylon. Jason, Elea, Maria, Inayat, who have made me feel at home from the beginning, and to the new acquisitions Aytac and Rea, for the fun tea times and discussions that have made these three years unforgettable. Meeting you during this process has been a true gift. Yann, thank you for being the best office buddy possible, thank you for your infinite support and the fun discussions we shared. I hope sharing the office with three crazy PhD students was not too unbearable and we made your days a bit more fun.

A mi familia de españoles expatriados en Montpellier, habéis sido un soplo de aire fresco que hacía que desapareciera el estrés y el agobio de la tesis. Rafa, Manuel, Elena, Irene, Javi y todos los demás; gracias por acompañarme y alegrarme las noches y días de fiestas, risas y planes durante estos tres años.

Y lo mejor para el final. Millones de gracias a mi familia, amigos, y a mi compañero de aventuras, por estar siempre ahí dándome fuerza desde la distancia. Vuestro apoyo incondicional ha sido imprescindible estos tres años y nunca podré agradecéroslo suficiente. Este logro es tanto mío como vuestro. Os quiero.

Abstract (English)

Plasmodium falciparum parasites have a complex life cycle and employ atypical multiplication strategies, tailored for fast population expansion during host colonisation and transmission, that are not fully understood. In an effort to unravel the initiation of DNA replication of these human pathogens my project focused on two main aspects of this process: (i) identifying and characterising the origins of replication and (ii) identifying and characterising the replication machinery.

Regarding the first part, I combined three different approaches to map the origins of replication in *P. falciparum*. These were ChIP-seq, SNS-Seq and NFS. I first performed ChIP-seq on two subunits of the origin recognition complex (of *Pf*ORC1 and *Pf*ORC2) at the beginning of schizogony to obtain a cartography of all potential replication initiation sites. Next, I mapped sites of active replication using two strategies: sequencing DNA nascent strands (SNS-seq); and mapping the incorporation of the thymidine analogue BrdU into replicating DNA, using nanopore sequencing combined with NanoForkSpeed (NFS). By combining data from these different methods, I have obtained a robust set of origins of replication that display some characteristics similar to those of mammalian origins, such as the non-random distribution in initiation zones or clusters and the association with G-quadruplex forming sequences. Strikingly, they also display unique characteristics, since they are associated with highly transcribed genes but depleted from transcription start sites (TSS). Additionally, the results showed a uniform fork speed across the genome with a significant decrease in centromeres and telomeres. Single molecule information, using reads containing multiple initiation events which could have only come from individual cells, revealed a relationship between the pace at which replication forks travel and the distance to the nearest origin. This multifaceted approach provided the first comprehensive analysis of the genetic landscape of the origins of replication in *P. falciparum*.

The second part of my project focused on the characterization of the replicative complex of *P. falciparum* by isolating proteins bound to nascent DNA at active replication forks.

Overall, this work contributes to the growing field of studying how *Plasmodium falciparum* parasites replicate their genome within the human host, and I am confident that it will serve as a solid foundation for further investigations.

Résumé (Français)

Le parasite *Plasmodium falciparum*, l'agent responsable du paludisme, a un cycle de vie complexe et emploie des stratégies de multiplication atypiques, adaptées à l'expansion rapide de la population pendant la colonisation de l'hôte et la transmission, qui ne sont pas entièrement comprises. Afin de comprendre l'initiation de la réplication de l'ADN de ce pathogène humain, mon projet s'est concentré sur deux aspects principaux de ce processus : l'identification et la caractérisation des origines de la réplication et de la machinerie de réplication.

Dans la première partie de ce projet, j'ai combiné trois approches différentes pour cartographier les origines de réplication chez *P. falciparum* : ChIP-seq, le séquençage des brins naissants d'ADN (SNS-Seq) et la méthode NanoForkSpeed (NFS). J'ai d'abord réalisé les expériences ChIP-seq sur deux sous-unités du complexe de reconnaissance des origines de réplication (*Pf*ORC1 et *Pf*ORC2) au début de la schizogonie afin d'obtenir une cartographie de tous les sites d'initiation potentiels. Ensuite, j'ai identifié les sites de réplication actifs en utilisant deux stratégies : SNS-seq et la cartographie de l'incorporation de l'analogue de la thymidine BrdU dans l'ADN en réplication, en utilisant le séquençage nanopore et l'algorithme NFS. En combinant les données issues de ces différentes méthodes, j'ai obtenu un ensemble robuste d'origines de réplication qui présentent des caractéristiques similaires à celles des origines des mammifères, comme la distribution non aléatoire dans des zones d'initiation ou 'clusters' et l'association avec des séquences formant des G-quadruplex. Les origines de réplication montrent également des caractéristiques uniques. Elles sont associées à des gènes fortement transcrits, mais dépourvus de sites de début de transcription (TSS). En outre, les résultats montrent que la vitesse de la fourche de réplication est uniforme sur l'ensemble du génome, à l'exception d'une diminution significative au niveau des centromères et des télomères. Des informations obtenues sur molécules uniques, utilisant des lectures contenant de multiples événements d'initiation qui n'auraient pu provenir que de cellules individuelles, ont révélé une relation entre la vitesse à laquelle les fourches de réplication se déplacent et la distance par rapport à l'origine la plus proche. Cette approche à multiples facettes a fourni la première analyse complète du paysage génétique des origines de réplication chez *P. falciparum*.

La deuxième partie de mon projet s'est concentrée sur la caractérisation du complexe réplicatif de *P. falciparum* en isolant des protéines sur l'ADN naissant, qui font potentiellement partie de la machinerie du réplisome. Dans l'ensemble, ce travail contribue aux études en plein essor, sur la réplication de l'ADN des parasites *Plasmodium falciparum,* dans l'hôte humain, et fournira une base solide pour les futures recherches dans ce domaine.

Résumé de la thèse en français

Contexte

Tout au long du développement de mon projet de doctorat, j'ai exploré divers aspects de l'initiation de la réplication de l'ADN chez *Plasmodium falciparum* (*P. falciparum*), l'agent responsable du paludisme. Ces recherches ont porté notamment sur l'identification des origines de réplication et la caractérisation de leur environnement génétique, l'exploration de la dynamique de progression des fourches de réplication, ainsi que l'étude approfondie des composants du réplisome. Les résultats de ces travaux m'a permis d'acquérir une meilleure compréhension des mécanismes sous-jacents à l'initiation de la réplication de l'ADN chez cet agent pathogène humain.

Les connaissances acquises sur les organismes modèles du groupe des Opisthokonta, tels que les métazoaires et les levures, ont servi de base aux efforts visant à étudier la réplication de l'ADN chez les eucaryotes. Bien que quelques études récentes aient commencé à explorer la dynamique de la réplication de l'ADN chez les parasites asexués intraérythrocytaires de *P. falciparum* [1]–[4], les mécanismes moléculaires régissant l'initiation de la réplication et les déterminants génomiques de la spécification des origines de réplication restent à comprendre. En effet, les parasites apicomplexes ont dévié au cours de l'évolution et présentent des différences significatives dans leur biologie et n'adhèrent pas nécessairement aux principes appliqués à la réplication canonique de l'ADN eucaryote. Au cours de leur développement à l'intérieur des érythrocytes de l'hôte, les parasites répliquent leur génome à plusieurs reprises, ce qui leur permet d'augmenter leur population de plusieurs ordres de grandeur, et leur taux de multiplication élevé est lié à leur pathogénicité. Par conséquent, les traitements ciblant ce processus de réplication ont le potentiel d'être très efficaces. L'émergence croissante d'une résistance répandue aux antipaludéens couramment utilisés souligne la nécessité de développer de nouveaux médicaments ciblant le stade prolifératif de ce pathogène.

Objectifs de l'étude

L'objectif principal de cette étude était de démêler les complexités de la réplication de l'ADN chez *P. falciparum*, un processus vital pour la survie et la prolifération des parasites dans les cellules humaines. Afin de comprendre comment ces parasites se multiplient grâce à des cycles consécutifs de réplication de l'ADN et de division nucléaire, j'ai combiné des approches différentes pour cartographier les origines plasmodiales de la réplication de l'ADN et analyser la composition du complexe de réplication en isolant les protéines liées aux fourches de réplication actives. En adoptant une approche multidisciplinaire englobant des techniques telles que l'analyse génomique, la biologie moléculaire et la protéomique, j'ai cherché à contribuer au corpus croissant de connaissances

entourant ce processus complexe. Mes recherches ont permis de mettre en lumière les événements moléculaires fondamentaux qui sous-tendent l'initiation de la réplication de l'ADN, d'élucider les mécanismes par lesquels *P. falciparum* réplique son génome et de fournir des informations cruciales sur sa dynamique de réplication.

Résultats et conclusions

Ma première approche a été de combiner trois méthodologies différentes pour cartographier les origines plasmodiales de réplication (ORIs). J'ai réalisé en premier temps ChIP-seq sur deux sous-unités du complexe de reconnaissance de l'origine *Pf*ORC1 et *Pf*ORC2, marquées par une étiquette HA, au début de la schizogonie afin d'obtenir une cartographie de tous les sites d'initiation. Ensuite, j'ai cartographié les sites de réplication active en utilisant deux stratégies: en séquençant les petits brins naissants d'ADN (SNS-seq); et en cartographiant l'incorporation de l'analogue de la thymidine BrdU dans l'ADN des parasites en réplication, en utilisant le séquençage par nanopores combiné à NanoForkSpeed [5]. En combinant les données de ces différentes stratégies, qui ont donné des résultats très concordants, j'ai obtenu un ensemble robuste d'origines putatives de réplication.

On a commencé par mesurer l'augmentation du contenu en ADN par cytométrie de flux et en chronométrant l'incorporation d'un analogue de la thymidine dans l'ADN, cela m'a permis de définir des fenêtres pour cartographier les origines de la réplication de l'ADN de *P. falciparum*: 25 hpi (± 1h) comme fenêtre pour étudier l'occupation du *Pf*ORC dans l'ensemble du génome avant l'initiation de la synthèse de l'ADN; et 29 hpi (± 1h) comme fenêtre pour étudier les événements d'initiation de la réplication pendant la première phase S du cycle intra-érythrocytaire.

L'expérience ChIP-seq m'a permis de cartographier les sites de liaison des protéines *Pf*ORC1 et *Pf*ORC2 avant le début de la réplication (25 hpi). Cette expérience a montré que les origines de réplication manquent de spécificité de séquence (de manière similaire au système mammifère mais différente de la levure *S. cerevisiae*, où les origines sont définies par la présence d'une séquence répliquant de manière autonome ou ARS [6]). Compte tenu de la teneur extrêmement élevée en AT du génome de *P. falciparum* (>80 %), les sites de liaison de *Pf*ORC$_{1\text{-}2}$ ont été étonnamment trouvés dans des régions ayant une teneur relativement élevée en GC (27 %) et associées à la présence de séquences formant des G-quadruplexes (G4FS). Il a été déjà suggéré que les G4FS jouent un rôle dans la spécification de origines dans les cellules de mammifères, ainsi que dans les parasites de la drosophile et des *Leishmania major* [7]–[9], en excluant l'occupation des nucléosomes et, par conséquent, en favorisant la liaison de l'ORC.

La transcription active s'est avérée être un déterminant important de la spécification de l'origine, bien que les TSS et TTS soient appauvris dans les sites de *Pf*ORC$_{1-2}$. Il est intéressant de noter que les sites de liaison de *Pf*ORC$_{1-2}$ sont enrichis sur les corps des gènes transcriptionnellement actifs. Une autre caractéristique surprenante a été trouvée dans les gènes *var*, où les promoteurs étaient significativement enrichis en sites de liaison de *Pf*ORC$_{1-2}$, montrant un modèle d'enrichissement complètement différent de celui des gènes du core génome ou même d'autres familles de multigènes à hétérochromatine comme les gènes *rifin* et *stevor*, qui montrent également un enrichissement en sites de liaison de *Pf*ORC$_{1-2}$ à l'intérieur de leurs corps génétiques.

En outre, j'ai identifié un ensemble d'origines actives de réplication chez *P. falciparum* en séquençant de courts brins naissants (SNS-seq) au début de la schizogonie, qui présentaient une distribution à l'échelle du génome très similaire à celle des sites de liaison de *Pf*ORC$_{1-2}$, la plupart étant situés à moins de 2 kb d'un site actif d'initiation de la réplication de l'ADN. Les SNS-seq ORIs actives sont également associées à G4FS, de manière encore plus prononcée, et présentent une déplétion similaire en TSS et un enrichissement dans les régions internes des gènes. Il est intéressant de noter que les données de cette thèse suggèrent une forte corrélation entre la transcription active chez *P. falciparum* et la mise à feu de l'origine, étant donné que les ORIs actives étaient fortement enrichies en gènes du quartile supérieur d'expression.

Enfin, le séquençage par nanopore que j'ai utilisé pour cartographier les origines actives de réplication s'est avéré très efficace et précis pour un organisme doté d'un génome relativement petit (23 Mb) comme *P. falciparum*. Grâce à cette technologie de séquençage d'ONT, nous pouvons cartographier l'incorporation d'analogues nucléosidiques dans l'ADN à partir des signatures de profil électrique spécifiques qu'ils génèrent lorsque le brin d'ADN transite à travers le pore, ce qui permet de détecter les fourches de réplication actives et de déduire les coordonnées des sites d'initiation de la réplication à l'aide de l'algorithme NanoForkSpeed [5], [10]. Le jeu de données NFS ORIs obtenu est hautement concordant avec les sites de liaison de *Pf*ORC$_{1-2}$ cartographiés et les origines SNS-seq, illustrant la robustesse de la méthode NFS et validant son efficacité pour identifier les origines de réplication dans l'ensemble du génome de *P. falciparum*.

J'ai également observé que les origines de réplication de *P. falciparum* ne sont pas distribuées au hasard, et que la plupart des origines identifiées (~75%) par les trois différentes méthodes semblent regroupées dans 523 "zones d'initiation". L'octroi de licences est biaisé vers des régions à forte teneur en GC et associées à des séquences formant des G4FS et à des gènes activement transcrits. Alors qu'une forte transcription peut favoriser la mise à feu des origines, les origines actives ne sont pas

présentes sur les sites de départ de la transcription. Au lieu de cela, la plupart s'accumulent dans les corps de gènes transcriptionnellement actifs.

L'utilisation du séquençage nanopore et de NanoForkSpeed m'a permis aussi de mesurer la vitesse des fourches de réplication. J'ai constaté que, à l'exception des centromères et des télomères, la vitesse des fourches est similaire sur l'ensemble du génome et à différents moments de la schizogonie. Des informations au niveau des molécules uniques, contenant de multiples événements d'initiation, qui n'auraient pu provenir que de noyaux individuels en cours de réplication, a montré une relation entre la vitesse à laquelle les fourches de réplication se déplacent et la distance par rapport à l'origine la plus proche (inter-origin distance, IOD). Des IOD plus longues sont corrélées avec une progression plus rapide des fourches, tandis que des IOD plus courtes sont corrélées avec une progression plus lente des fourches, indépendamment de l'état avancé de la schizogonie.

Ces résultats indiquent que malgré les changements modestes mais significatifs de la vitesse et des IOD, les parasites asexués de *P. falciparum* peuvent compenser le fait d'avoir des origines de réplication plus espacées (reflétées par des IOD plus élevés) par une vitesse de fourche de réplication plus élevée, et *vice versa*, pour achever chaque cycle de réplication de l'ADN en 15 minutes, à la fois au début et au milieu de la schizogonie.

Dans l'ensemble, cette première partie permet de mieux comprendre le processus de réplication qui permet à ces parasites de proliférer massivement au sein de l'hôte humain, de plusieurs ordres de grandeur, en l'espace de quelques jours. Ça ouvrira la voie à une meilleure compréhension des mécanismes moléculaires qui sous-tendent la réplication de l'ADN des parasites, toujours dans le but ultime de développer de nouvelles stratégies thérapeutiques contre le paludisme.

Le deuxième objectif principal de mon projet de doctorat était de démêler les interactions complexes entre les protéines au sein du complexe réplicatif de *P. falciparum*, afin de mettre en lumière les processus sous-jacents qui entraînent la réplication de l'ADN pendant la schizogonie au stade sanguin. J'ai utilisé une gamme complète de méthodologies pour tenter de caractériser les composants clés de la machinerie du réplisome, ainsi que pour identifier leurs partenaires d'interaction.

L'approche consistait à utiliser des protéines comme molécules-appâts dans des expériences d'immunoprécipitation et de marquage de proximité, afin de capturer sélectivement les protéines qui interagissent au sein du complexe du réplisome. Bien que ces premières tentatives n'aient pas donné les résultats escomptés, j'ai modifié l'approche pour explorer des méthodologies alternatives qui permettraient d'isoler et de caractériser les protéines spécifiquement associées à l'ADN naissant dans

les fourches de réplication actives. J'ai utilisé avec succès l'isolation des protéines sur l'ADN naissant (*isolation of proteins on nascent DNA*, iPOND), qui utilise la chimie click pour purifier sélectivement l'EdU (5-éthynyl-2'-désoxyuridine) incorporé dans l'ADN naissant ainsi que les protéines qui y sont liées et qui font potentiellement partie de la machinerie du réplisome.

En tirant parti de la méthodologie iPOND, j'ai pu isoler les protéines directement associées au complexe réplicatif pendant la réplication active de l'ADN et identifier un ensemble de protéines enrichies dans l'ADN naissant. Parmi ces protéines figurent des composants clés connus du réplisome, tels que *Pf*PCNA1, mais d'autres membres connus n'ont pas été détectés, tels que des sous-unités du complexe MCM ou ORC et des ADN polymérases, ce qui renforce la nécessité d'optimiser la purification des protéines associées aux brins d'ADN naissants. Néanmoins, une analyse des interactions protéine-protéine entre les protéines enrichies dans les fourches de réplication actives, donc potentiellement impliquées dans la réplication de l'ADN, a révélé un grand nombre d'interactions fonctionnelles au sein de la plupart des membres. En outre, cette notion est renforcée par les résultats de l'analyse de l'ontologie des gènes, où les processus biologiques enrichis comprenaient des termes clairement liés à la réplication de l'ADN, tels que l'élongation du brin principal, le processus métabolique de l'ADN ou la régulation de la réplication de l'ADN.

Grâce à iPOND, bien qu'il s'agisse de résultats préliminaires et que le protocole doit encore être optimisé, j'ai obtenu des informations précieuses sur la composition et la dynamique de la machinerie du réplisome, fournissant un aperçu détaillé des interactions protéiques qui se produisent pendant la réplication de l'ADN chez *P. falciparum*. Les connaissances acquises grâce à ces études sur les interactions protéiques au sein du complexe réplicatif de ce parasite constituent une base solide pour la poursuite des recherches. Les interacteurs protéiques identifiés peuvent maintenant être soumis à des analyses fonctionnelles afin d'élucider leurs rôles spécifiques dans la réplication de l'ADN et d'évaluer leur potentiel en tant que cibles pour des interventions antipaludiques.

En ce qui concerne l'obtention de lignées de parasites transgéniques, bien que les résultats de certains de mes efforts n'aient pas répondu à mes attentes initiales, je reste déterminé à explorer des stratégies alternatives pour obtenir des mutants avec certains des membres du réplisome marqués de manière endogène, et à affiner nos méthodologies dans les recherches en cours et à venir. Les connaissances acquises grâce à ces expériences servent de base aux perspectives futures, soulignant la nécessité de poursuivre les efforts pour élucider les mécanismes complexes qui sous-tendent la réplication de l'ADN de cet agent pathogène. Plusieurs membres du réplisome n'ont pas encore été étudiés, et l'équipe s'efforcera de caractériser leur contribution à la réplication de l'ADN et d'identifier d'autres protéines qui interagissent avec eux.

Les multiples approches adoptées pour approfondir notre compréhension de la machinerie complexe et des interactions qui régissent la réplication de l'ADN chez les parasites du paludisme ont été difficiles et illustrent la complexité du travail avec *P. falciparum*. Cependant, elles m'ont permis d'explorer de nouvelles stratégies et d'acquérir une expertise dans la conception de plasmides pour l'obtention de lignées de parasites transgéniques, ainsi que dans la planification d'expériences et l'optimisation de protocoles pour la caractérisation des protéines.

Table of contents

List of figures

List of tables

List of abbreviations

AAA+	ATPases Associated with various cellular Activities
ACT	Artemisinin based Combination Therapy
ACS	ARS consensus sequence
ApiAP2	Apicomplexa specific AP2
ARS	Autonomously replicating sequence
AT	Adenine and Thymine
BirA	Promiscuous biotin ligase
BrdU	5-Bromo-2-deoxyuridine
Cdc6	Cell division cycle 6 protein
CDK	Cyclin-dependent kinase
CDS	Coding sequence
Cdt1	Cdc10-dependent transcript 1
GC	Guanine and Cytosine
ChIP	Chromatin Immunoprecipitation
CM	Cerebral malaria
CMG	Cdc45/MCM2-7/GINS replication complex
CRK	Cdk/Cdc2-related protein kinase
CSP	Circumsporozoite protein
CuAAC	Copper(I)-catalyzed azide-alkyne cycloaddition reaction
DDK	Cdc7-Dbf4 protein kinase
hDHFR	Human dihydrofolate reductase
DNA	Deoxyribonucleic acid
DNA Pol	DNA Polymerase
dsDNA	double stranded DNA
EdU	5-ethynyl-2-deoxyuridine
EMP1	Erythrocyte membrane protein 1

FC	Fold Change
FL	Full length
G4	G-quadruplex
G4FS	G-Quadruplex Forming Sequence
GINS	go-ichi-ni-san complex, composed of the proteins Sld5, Psf1, Psf2 and Psf3
GO	Gene ontology
gRNA	Guide RNA
H3	Histone 3
HA	Hemagglutinin
HFF	Human Foreskin Fibroblast
hpi	hours post invasion
HR	Homology region
HRP	Horseradish peroxidase
IDC	Intraerythrocytic Developmental Cycle
IFA	Immunofluorescence assay
IOD	Inter Origin Distance
IP	Immunoprecipitation
iRBC	Infected Red Blood Cell
MCM	Mini-chromosome maintenance protein
MJ	Moving junction
MNG	MNeonGreen
MSP	Merozoite surface protein
MO	MCM-ORC complex
MS	Mass Spectrometry
NCC	No Click Control
NFS	NanoForkSpeed
NGS	Next Generation Sequencing

OCCM	ORC-Cdc6-Cdt1-MCM complex
OGRE	Origin G-rich Repeated Elements
OM	ORC-MCM complex
ONT	Oxford Nanopore Technologies
ORC	Origin Recognition Complex
ORI	Origin of replication
PBS	Phosphaste buffered saline
PCNA1	Proliferative Cell Nuclear Antigen 1 protein
PCR	Polymerase Chain Reaction
Pf	*Plasmodium falciparum*
PIC	Protease inhibitor cocktail
POI	Protein of interest
PPI	Protein-protein Interaction
PPM	Parasite's plasma membrane
pre-RC	pre Replicative Complex
PTMs	Post translational modifications
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
RBC	Red Blood Cell
RBCM	Red Blood Cell Membrane
RC	Replicative Complex
RDT	Rapid Diagnostic Test
RFB	replication fork barriers
RFC	Replication factor C
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPA	Replication Protein A

rRNA	ribosomal RNA
RT	Room temperature
Sc	Single cell
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide gel electrophoresis
Seq	Sequencing
SLI	Selection Linked Integration
SNS	Short Nascent Strands
ssDNA	single stranded DNA
TARE	Telomere Associated Repetitive Element
TAS	Telomere Associated Sequences
TF	Transcription factor
TK	Thymidine kinase
tRNA	transference RNA
TSS	Transcription Start Site
TTS	Transcription Termination Site
UTR	Untranslated region
VSA	Variant Surface Antigen
WB	Western blot
WHO	World Health Organization
WT	Wild type

Introduction

Chapter 1. Malaria: a global burden

Malaria is a significant global health challenge caused by *Plasmodium* parasites and transmitted by female *Anopheles* mosquitoes. This vector-borne disease afflicts nearly half of the world's population, particularly those living in tropical and subtropical areas. Despite decades of research and control efforts, malaria remains one of the deadliest infectious diseases worldwide, claiming hundreds of thousands of lives every year. Limited funding and widespread drug resistance continue to impede progress in the fight against malaria, making the development of new and effective treatments urgent. However, the complexity of the malaria parasite's life cycle, which involves multiple stages and requires a mosquito vector, presents unique challenges for drug development. Therefore, a comprehensive understanding of *Plasmodium* parasites' biology is essential to identify new targets for drug development and to design effective strategies for disease control. To address these challenges, this thesis aims to deepen our understanding of the biology of *Plasmodium falciparum*, with a particular focus on the molecular mechanisms underlying their multiplication within the human host. By leveraging state-of-the-art technologies, such as genomics and proteomics, we aim to uncover new insights into the fundamental biology of the parasite. We hope these insights may contribute to the global effort to reduce the burden of malaria and help to identify new targets for potential drug development.

1. History

Malaria is one of the most ancient diseases, predating even the history of man. From Mesopotamian clay tablets from 2000 B.C. to the early Greeks in 800 B.C., references to recurring fevers indicative of malaria illness have been found throughout history [11], [12]. The first theory put forth was that an unknown substance in the air arising from marshes was the origin of the disease, giving name to the word malaria, which literally means "bad air" coming from the Italian "mal aria". It was not until the year 1880 that the parasites were discovered in blood by Alphonse Laveran and 1898 that Ronald Ross showed that malaria can be transmitted to healthy individuals from infected mosquito bites, leading to the implementation of the first vector-control measures [11], [13].

1.1. Epidemiology

The impact of this disease on human health across evolution has probably been greater than that of any other infectious agent, and its burden remains one of the most devastating globally. Despite significant global efforts and partial success in combating malaria, more than 3.3 billion people still live at risk of infection, mainly in tropical and subtropical regions where the disease is endemic.

According to the most recent World Malaria Report of the World Health Organization [14], there were an estimated 247 million cases of malaria and 619 000 malaria-related fatalities in 2021, which represented a 10% increase compared to 2019. Besides, the COVID-19 pandemic has caused significant delays in the delivery of malaria prevention, diagnosis and treatment, that led to the increase in cases and fatalities compared to those of 2019.

As shown in Fig. 1, most of the cases affect Sub-Saharan African countries (82% of cases and 95% of deaths), and South-East Asia (10% of cases and 3% deaths). More specifically, 96% of malaria cases worldwide were reported from 29 countries, and almost half of the cases from just four countries: Nigeria (27%), the Democratic Republic of the Congo (12%), Uganda (5%) and Mozambique (4%) [14]. Fortunately, global efforts have permitted a decline in cases, deaths and in the number of countries classified as endemic since the year 2000. Currently, 84 countries are considered to be malaria endemic in contrast to the year 2000 when there were 108 countries.

Epidemiological data reveals that malaria disproportionately affects young children and pregnant women. Among these groups, children under five years of age are the most vulnerable to the disease, accounting for over 77% of malaria-related fatalities.

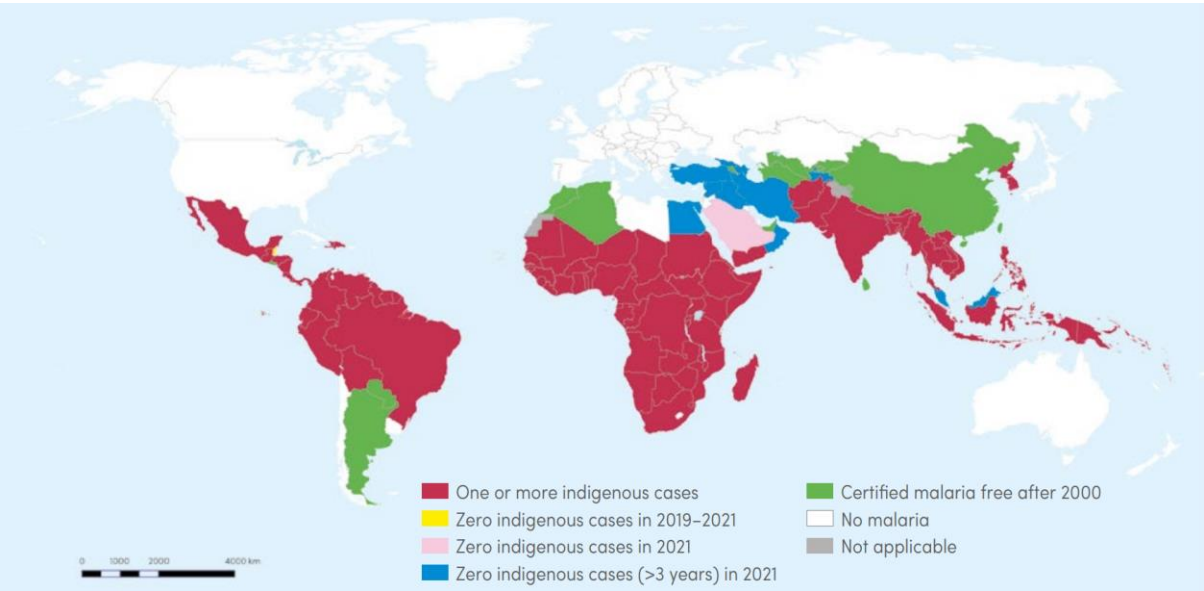


Figure 1. Countries with indigenous malaria cases in 2000 and their status by 2021. *Countries without any indigenous cases for at least 3 consecutive years are considered to have eliminated malaria. Source: WHO world malaria report 2022* [14].

1.2. Infection and transmission

Malaria is caused by protozoa of the genus *Plasmodium*, phylum Apicomplexa, order Haemosporida and family Plasmodiidae [15]. More than 200 species of *Plasmodium* have been described and all of them exhibit an heteroxenous life cycle, meaning that they parasitize more than one host. *Plasmodium* parasites require both a vertebrate host (including reptiles, birds and mammals); and a female mosquito of a susceptible *Anopheles* species as vector for sexual reproduction and transmission [16], [17]. Five *Plasmodium* species are known to infect humans: *P. falciparum*, *P. vivax*, *P. ovale, P. malariae* and *P. knowlesi*, which primarily infects macaques but is also a zoonotic parasite of humans in Malaysia and Borneo [18]. All these species are found naturally in tropical and subtropical areas worldwide, as that is the ecosystem of their mosquito vector. In particular, *P. falciparum* has the highest mortality and morbidity rates and is the most prevalent malaria parasite in sub-Saharan Africa, accounting for over 99% of the reported cases [14].

2. *Plasmodium falciparum* life cycle

The heteroxenous life cycle of *P. falciparum* is an intricate process where the parasite undergoes more than ten stages of cellular differentiation and invades at least four types of cells within two different hosts to establish infection, proliferation, and secure transmission (Fig. 2).

2.1. Mosquito stage

The transmission of malaria starts with the blood meal of a female mosquito from a susceptible *Anopheles* species in an infected human. There are approximately 430 *Anopheles* species, of which 30-40 transmit malaria. They are widely distributed around the world, with different environments supporting the presence of different species. In Africa, the main *Anopheles* mosquito species transmitting malaria is *A. gambiae* [19].

In the infective bloodmeal, the sexual parasite forms named gametocytes are the only ones able to carry on the infection to the mosquito. Once in the mosquito gut, these initiate a process to produce gametes termed gametogenesis which is triggered by several stimuli in the mosquito midgut, including a drop in temperature inside the mosquito, as compared to the human host, and the presence of xanthurenic acid. The female macrogametocyte becomes a non-motile macrogamete within 5 minutes, while the male microgametocyte takes longer (< 20 minutes). It undergoes three rounds of DNA replication to generate a polyploid cell, which subsequently divides giving rise to eight haploid extracellular motile flagellated microgametes which ultimately fertilise the macrogamete [20].

The diploid zygote, formed by the fusion of the macro- and microgametes, develops over the course of approximately 18 hours into a motile ookinete which attaches and penetrates the midgut epithelium to reach the basal lamina where it develops extracellularly into an oocyst. There, during between 10 and 30 days the oocyst matures and repeatedly divides its nucleus multiplying its DNA content until thousands of haploid sporozoites are formed from a single mother cell [21]. The length of the oocyst maturation period varies considerably with each *Plasmodium* species and depends on the environment's temperature. The process of oocyst development is the only extracellular stage of the *P. falciparum* life cycle.

When sporozoites are fully formed, they actively egress from the oocyst and migrate through the mosquito's circulatory system eventually reaching and invading the salivary glands. However, only a subset (around 25%) of the released sporozoites will be able to invade the salivary glands, where parasite numbers range from 10^2 to 10^5 [22]. Finally, they are released from the salivary glands and accumulate in the salivary duct, becoming infective. During the subsequent mosquito blood-fed, a small proportion of the resident sporozoites (10-100) are deposited on the dermis and then migrate to a blood vessel, penetrate it and are carried in the bloodstream to the liver, where they infect hepatocytes starting the exoerythrocytic or liver-stage cycle of growth [23], [24].

2.2. Liver stage

Once the sporozoites reach the liver, they cross the endothelial barrier either by traversing the endothelial cells or their resident macrophages "Knupffer cells", or by squeezing through the gaps between them [25]. The interaction between the motile parasites and the hepatic cells is mediated by the contact between the circumsporozoite protein (CSP) from the surface of sporozoites and the heparan sulfate proteoglycans (HSPGs) specific to the hepatocytes [26]. This first interaction is a critical step for infection and, as such, is consequently targeted for the development of recombinant vaccines. During the hepatic stage of the parasites, the infection is asymptomatic, and it is known as the incubation period, with a variable duration depending on the *Plasmodium* species responsible for the infection.

During invasion, the plasma membrane of hepatocytes buds in to create a protective parasitophorous vacuole (PV) that will provide a suitable environment for the parasites to develop and start their first asexual replication, known as exoerythrocytic schizogony. Here, the parasite first develops into a liver stage trophozoite, which will undergo 13 to 14 rounds of DNA replication and nuclear division to produce a syncytial schizont containing thousands of nuclei. Eventually, tens of thousands of haploid daughter merozoites will be released into the bloodstream through the budding off of merosomes,

which are host cell membrane enclosed vesicles filled with the merozoites [27]. Finally, the merosome membrane breaks apart releasing the containing merozoites. These then attach to and invade red blood cells (RBCs) to start the intraerythrocytic cycle of infection, also known as the blood-stage cycle.

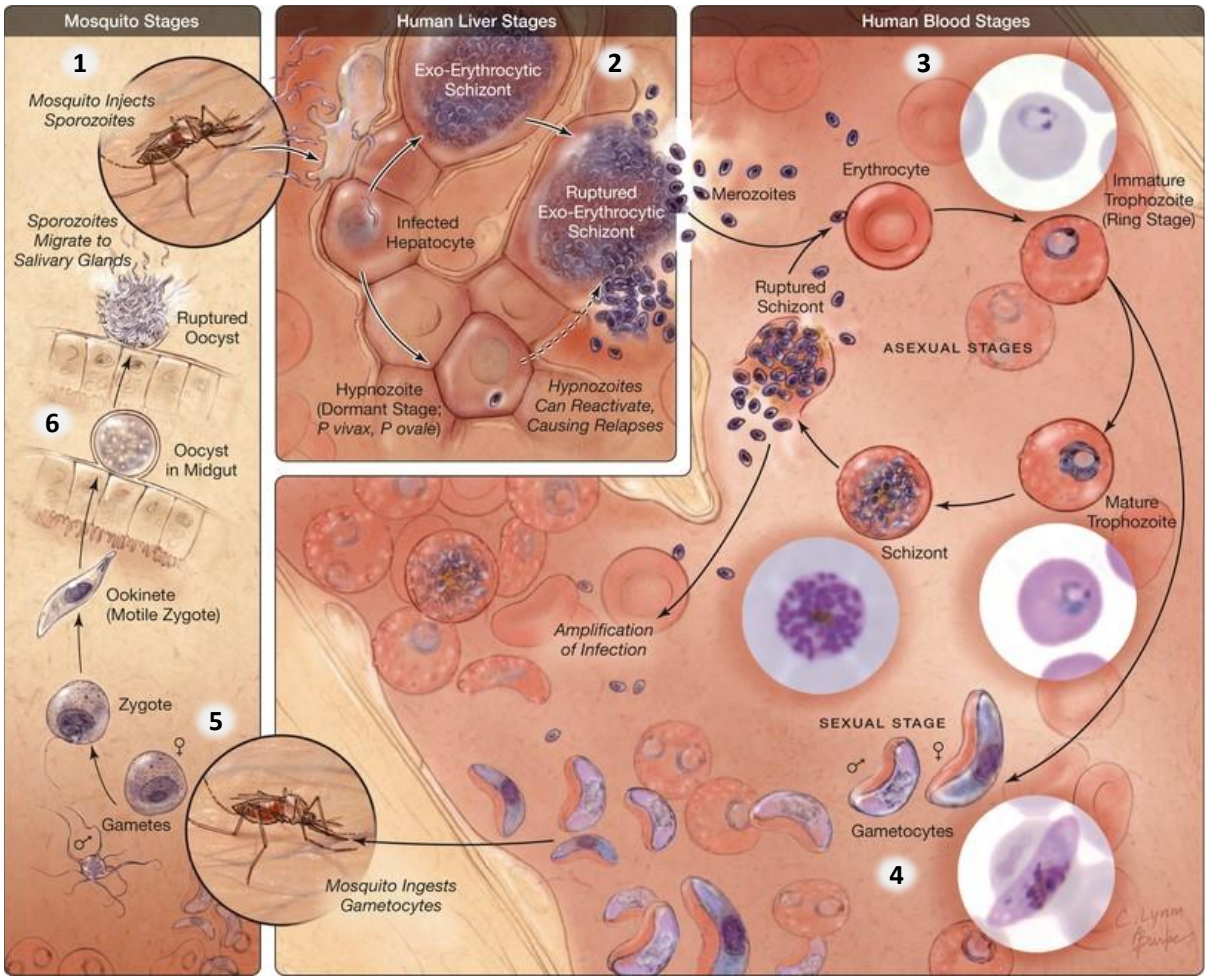


Figure 2. Plasmodium falciparum life cycle [28]***.*** *The main phases in the liver and in the red blood cells (asexual and sexual erythrocytic stages) of the human host, and in the mosquito gut and the salivary glands are depicted. With the first mosquito bite **(1)**, between 10 and 200 highly motile sporozoites are released along with salivary fluid inside the dermal tissues. They circulate within the blood to reach the liver cells and begin differentiating into small trophozoites and grow to become a multinucleated hepatic schizont in 5 days **(2).** Some Plasmodium species (P. vivax and P. ovale) can differentiate in a dormant form inside hepatocytes known as hypnozoites, which can later reactivate and cause relapses of infection. After 6 or 7 days of growth and development, single nucleated merozoites are released from the merosome and reach the blood circulation again, where they invade red blood cells to start the blood-stage cycle **(3).** During repeated rounds of a 48h asexual cycle, P. falciparum parasites invade and multiply inside erythrocytes, before rupturing their host cell releasing merozoites to invade new RBCs. Some merozoites can differentiate into the sexual stages known as gametocytes **(4),** which are taken up in the blood meal of the next Anopheles mosquito. In the mosquito gut, male and female gametes fertilise and produce a zygote **(5)** which then develops into a motile ookinete that crosses the midgut epithelium and starts the differentiation into an oocyst. Sporozoites develop inside the oocyst before invading the salivary glands **(6),** which become a reservoir to initiate the whole cycle all over.*

2.3. *Intraerythrocytic stage*

During the blood stage of infection parasites multiply and persist through alternating cycles of red blood cell (RBC) invasion and asexual replication within the RBC. This phase is entirely responsible for the clinical signs and pathogenicity of malaria. The different stages of the asexual intraerythrocytic development cycle (IDC) of *P. falciparum* are described below.

2.3.1. *Invasion*

The intraerythrocytic stage of infection starts with the invasion of the merozoite into the RBC (Fig. 3). A specialized structure at the apical end of the parasite known as the apical complex harbours specialized secretory organelles, named rhoptries, micronemes and dense granules, that contain proteins necessary for invasion [29]. Invasion starts with the attachment of the merozoite to the RBC's plasma membrane surface, in a process mediated by the interaction between merozoite GPI-anchored merozoite surface proteins (MSPs), with specific receptors present at the RBC's membrane, like heparin-like proteoglycans or the Band 3 protein [30]. The initial attachment to the RBC membrane can occur at any point on the surface of the merozoite, but a subsequent re-orientation mediated by the interaction of MSPs and RBC receptors is required to ensure that the apical end of the parasite makes contact with the RBC surface. From there, a number of adhesins, perforins and effectors stored in the apical organelles are secreted either onto the merozoite plasma membrane surface and bind to specific receptors on the erythrocyte or into the red blood cell, being directly involved in invasion [31]. For instance, the apical membrane antigen 1 (AMA-1) and the erythrocyte binding antigen 175 (EBA-175) are adhesin proteins released from the micronemes and translocated to the merozoite membrane. EBA175 binds to its RBC receptor, Glycophorin A [32], and causes the release of rhoptry proteins, such as the reticulocyte binding protein homologue 2b (*Pf*RH2b), inside the RBC [31]. Another RH secreted protein at the merozoite surface, *Pf*RH5, anchors the merozoite to the RBC by binding to the basigin protein receptor [33], and this anchoring allows AMA1 to initiate the tight junction formation by binding to the RON complex, which is formed by four rhoptry neck proteins (RON2, RON4, RON5, RON8) that are secreted from rhoptries and translocated to the RBC surface. These are in turn involved in sealing the engagement between the merozoite and the RBC in the invasion process [34] (Fig. 3).

Overall, these interactions between merozoite proteins and RBC's receptors promote an irreversible and tight adhesion of both cells, sealing the merozoite's plasma membrane with the one of the RBC forming a tight junction that acts as a platform for the parasites to enter the RBC, driven by an actin-myosin type motor [35]. The tight junction produces a circumferential ring around the parasite that

leads to the internalization of the merozoite into a parasitophorous vacuole (PV), allowing the merozoite to glide inside the RBC surrounded by the housing PV as it sheds its surface proteins. Finally, invasion is completed when the RBC membrane is resealed, releasing the PV containing the parasite into the host cell cytoplasm.

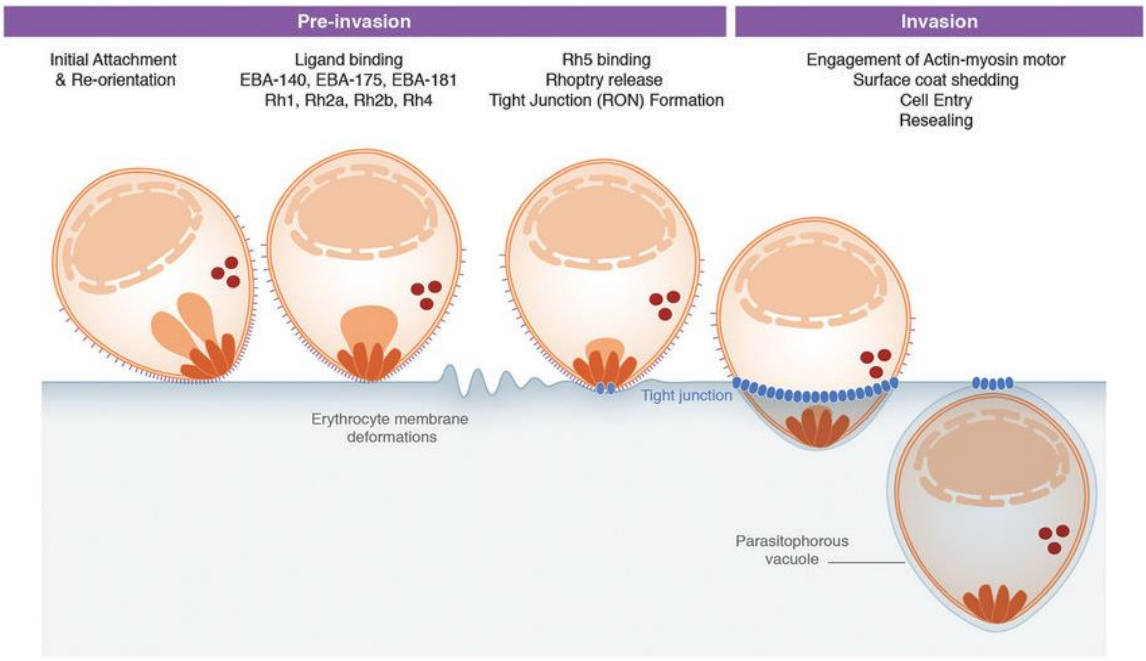


Figure 3. Invasion of a Plasmodium falciparum merozoite into an erythrocyte [36]. *The initial interaction and attachment of the merozoite to the RBC membrane (RBCM) can occur in any point of the merozoite's membrane. Later, it re-orients so that its apical pole is in close contact with the RBCM. At this point, proteins discharged from parasite's secretory organelles interact with RBC membrane proteins and allow for the establishment of the tight junction, through which the merozoite enters the host cell into the parasitophorous vacuole.*

2.3.2. Blood-stage schizogony

Following invasion, parasites proliferate through a process called blood-stage schizogony, in which they replicate their genome several times before cytokinesis. Then, the daughter cells are released during egress and subsequently invade other RBCs, starting the cycle all over again and leading to a build-up of parasitic load (Fig. 2). This repetitive process of erythrocyte infection is responsible for causing the typical symptoms and pathologies of malaria, as the repeated cycles of schizont rupture and merozoite release are reflected on the cyclical fever presentation during the infection.

The process of schizogony in *P. falciparum* can be summarised as follows:

1. The haploid invading merozoite (1n) begins its development within the red blood cell with a ring like appearance, a phase that will last from 0 to 22 hours post invasion (hpi). During this stage, the parasite feeds on small amounts of haemoglobin from the erythrocyte through the cytostome (a localized invagination of the PV membrane and the parasite's membrane), that

creates a budding of the membranes and forms double-membrane, haemoglobin-filled vesicles that are taken to the digestive vacuole. There, haemoglobin degradation produces amino acids that are taken up by the parasite [37]. As ring-stage parasites grow, they start to synthesize molecules that can be exported into the host cell surface, modifying it and leading to adherence to non-infected RBCs and to the linings of blood vessels [38].

2. After 23 hpi, the ring stage parasite enlarges into the form of a trophozoite (a period that will last from 23 to 34 hpi). In the trophozoite stage, the parasite will enter S phase and starts replicating its DNA at around 25-27 hpi. It is therefore the period of most active growth, feeding and RBC modification. At this point, the parasites continue feeding on haemoglobin, and the haem product of haemoglobin digestion crystallizes into particles of a non-toxic dark pigment, known as hemozoin, that accumulate within the digestive vacuole [37]. Continuous export of parasite proteins onto the RBC membrane leads to knob formation, presenting proteins such as the erythrocyte membrane protein 1 (*Pf*EMP1) or knob-associated histidine-rich proteins (KAHRP), which facilitates a strong adhesion to linings of blood vessels of viscera and placenta [39]. The expression of parasite proteins in the RBC membrane is facilitated by vesicular structures in the host cell termed 'Maurer's clefts', which are formed by the parasites to serve as platforms for the trafficking of proteins to the host cell surface [40].
3. At around 33-35 hpi, the parasites reach the schizont stage, which will last until the egress of the daughter merozoites around 44-48 hpi. There, after three to four consecutive rounds of asynchronous DNA replication and nuclear division (that started at the trophozoite stage) a multinucleated schizont is formed and a final rather synchronous round of replication will be followed by a massive cytokinesis [2]. Finally, 16 to 32 fully formed haploid merozoites within the large segmented schizont will be released to the bloodstream ready to infect other RBCs.

2.3.3. Egress

After the final round of nuclear division, a single cytokinetic event gives rise to the individual merozoites. Once daughter cell segmentation is complete, a Ca^{2+}-mediated signalling cascade triggers the sequential poration and rupture of the parasitophorous vacuole membrane (PVM) and RBC membrane (RBCM) in a short period of time to finally release the merozoites. The parasite protein phosphatase 1 (*Pf*PP1) plays an essential role in stimulating guanylate cyclase α (CGα) activity to synthesize cGMP, which in turn causes the activation of the GMP-dependent protein kinase G (*Pf*PKG) [41]. The *Pf*PKG activation causes a rapid mobilization of cytosolic Ca^{2+} that activates members of the CDPK (Ca^{2+} dependant protein kinase) family [42]. In addition to *Pf*PKG, *Pf*CDPK5 triggers the intracellular release and activation of several effector molecules that ultimately cause the rounding

up of the PVM [43]. Activation of *Pf*PKG and Ca^{2+} regulate the release of a parasite protease known as Subtilisin-like protease 1 (*Pf*SUB1) from the exonemes (apical secretory organelles) into the PV lumen [44]. *Pf*SUB1 has several substrates, such as the PV-resident cysteine protease serine-like repeat antigen-6 (SERA6), which upon processing by *Pf*SUB1, mediates the degradation of ß-spectrin [45], a protein of the RBC cytoskeleton, leading to the loss of RBCM structural scaffold [46]. This leads to the formation of pores that causes an increased permeabilization and osmotic swelling that ultimately results in the final RBCM rupture. Finally, the merozoites are ejected and dispersed into the bloodstream [47], [48]

2.3.4. Sexual commitment

During the intraerythrocytic cycle, some of the merozoites will differentiate into the sexual stages of the parasite: female (macrogametocytes) and male (microgametocytes), responsible for the transmission of infection when ingested by a female anopheline mosquito during a blood meal. Parasites within the human are obliged to maintain a balance between producing replicative asexual-stage parasites that will keep the infection going and developing enough non-proliferative gametocytes to guarantee transmission.

The commitment to gametocytogenesis is regulated in all *Plasmodium* species by a conserved transcription factor of the ApiAP2 family termed *Pf*AP2-G [49], [50]. In asexual stage parasites, the *Pf*AP2-G gene is silenced by epigenetic mechanisms that involve the presence of heterochromatin marked by the heterochromatin protein 1 (*Pf*HP1). Activation of the gene requires the displacement of *Pf*HP1 in a process that depends on the gametocyte development 1 protein (*Pf*Gdv1), resulting in sexual commitment [51]. This sexual progenitor later undergoes a secondary cell fate decision to become either a male or a female gametocyte [52]. The stimuli triggering sexual commitment is not fully understood and its rate varies between *Plasmodium* species and with environmental conditions (including temperature changes and the presence of an antimalarial treatment). Finally, sequestration of gametocytes in the bone marrow promotes maturation (through stages I to V) within the host RBCs [53], until both macro- and microgametocytes are ingested during the mosquito blood meal, starting the cycle all over.

3. Pathophysiology of malaria infection

The clinical symptoms of malaria are all associated with the blood stages of parasite development. In general, the symptoms of malaria might appear early or late in the course of infection and the disease can begin gradually or suddenly. The clinical picture of malaria is not uniform and it can be displayed in two ways: asymptomatic or uncomplicated; and severe [54].

3.1. Asymptomatic or uncomplicated malaria

In endemic locations, uncomplicated malaria is the most common form of the disease. People living in malaria endemic areas are repetitively exposed to *P. falciparum* infections throughout their lifetime, which enables them to develop immunity to the disease from childhood [55]. As a result, it is rare for adults to manifest clinical symptoms or die from a malaria infection.

The asymptomatic form of malaria may be achieved through a higher parasite clearance after a certain age or because the patients are able to carry higher parasite loads without showing any symptoms [56], [57]. In uncomplicated malaria typical symptoms are unspecific and resemble other febrile illnesses (fever, chills, sweats and headache), mediated by the human host's cytokine responses upon the rupture of infected erythrocytes [58]. Mild symptoms also include myalgia and minor gastrointestinal symptoms. In endemic areas and particularly in children, the presence of hepatosplenomegaly, thrombocytopenia and anaemia, regardless of the presence of fever, are also associated with malaria [59].

3.2. Severe malaria

In non-immune people, malaria caused by *P. falciparum* can progress very rapidly to severe malaria unless proper treatment is started. Severe malaria is caused by a rapid proliferation and multiplication of parasite load, iRBCs rupture and sequestration of iRBCs to endothelial cells of capillary vessels of vital organs [54]. Severe or complicated malaria causes severe anaemia, haemoglobinuria, and hypoglycaemia due to hemolysis, abnormal blood coagulation or spontaneous bleeding [60], metabolic acidosis (associated with respiratory distress), cardiovascular complications [61], renal failure [62] and cerebral malaria (CM). CM is caused generally by *P. falciparum* infections, and results in a coma caused by endothelial inflammation in the brain due to accumulation and sequestration of iRBCs, leukocytes, and platelets in the brain vasculature [63]. If left untreated, the progression of these complications can be rapid and fatal in most cases.

Malaria mortality in children under the age of five in Africa, where the heaviest burden of disease falls, is mostly attributed to CM and severe anaemia [64]. Anaemia is common mainly in African children and pregnant women. Organ failure, on the other hand, affects mainly adults in Asia and South America.

3.3. Determinant factors for severity of malaria infections

3.3.1. Host factors

Malaria exerts a considerable selective pressure on human populations in endemic areas. This is reflected in the accumulation of genetic traits that would normally be disadvantageous but confer reduced susceptibility to malaria infections.

The presence of polymorphisms in immune system genes or affecting RBCs have been associated with variation in susceptibility to severe malaria [55]. For example, mutations in several immunologically relevant genes like tumour necrosis factor α (TNF-α) or the interleukin IL-4 have been found with a significant frequency in cerebral malaria patients, suggesting a regulatory role of these factors in the pathogenesis of severe malaria [65], [66]. On the other hand, several conditions due to polymorphisms affecting RBCs have been shown to confer protection against the severe form of the disease, mainly because the patients display abnormal RBC shapes and membrane composition that challenge the entry and survival of the parasites within the host erythrocytes. For instance, several common phenotypes in malaria endemic regions are: the presence of a heterozygous mutation in the β-chain of the haemoglobin gene (referred to as haemoglobin S (HbS) and known as "sickle cell trait") [67] is associated with significant protection against severe malaria; α-thalassemia (loss of one of the duplicated α-globin genes) causes enhanced immune recognition of iRBCs and impaired parasite growth [68]; and the absence of the Duffy chemokine receptor confers resistance to *P. vivax* infections [69]. Finally, glucose-6-phosphate dehydrogenase (G6PD) deficiency [57], structural haemoglobin variants HbC and HbE [70], [71], ovalocytosis [72], polymorphisms in complement receptor 1 or the receptor CD-36 [73], [74], and the blood group O [75] are also common traits in malaria endemic regions that have been associated with a protective role against severe malaria.

3.3.2. Parasitic factors

Immune evasion is a key strategy for the continued survival of blood stage parasites within the human host. A range of biomolecular strategies help them evade the host's immune system starting from the asymptomatic pre-erythrocytic stages. As the sporozoites reach the liver and enter the hepatocytes,

they have developed mechanisms to suppress the phagocytic functions of the resident macrophages, called “Knupffer cells” [76], [77]. During maturation in the hepatocytes, they actively accumulate host intracellular Ca^{2+}, keeping the levels in the host cell low and inhibiting the translocation of phosphatidylserine residues onto the host cell membrane. This blocks the “eat me” signal which would prompt phagocytes to engulf these cells, permitting development and save release of the merosomes filled with merozoites into the bloodstream [78]. In addition, *P. vivax* and *P. ovale* parasites have the ability to differentiate into liver-stage hypnozoites that can remain dormant in hepatocytes without being eliminated by the host immune system for extended periods of time. Upon reactivation, these cause relapses of malaria infection [79].

During the asexual stage of development inside the RBCs, parasites display a range of evasion strategies to avoid the host immune defences ensuring its continued survival in the host and subsequent transmission [80]. These strategies include extensive modifications of the host RBCs surface and sequestration of iRBCs within the microvasculature. Parasites export proteins both into the erythrocyte’s cytoplasm and surface to escape host immune responses and to create a suitable environment for growth. When exported to the iRBC membrane, they form prominent knobs on the iRBC surface containing the knob-associated histidine-rich proteins (KAHRP) or the erythrocyte membrane protein 1 (*Pf*EMP1) [39]. The knobs contain parasite surface antigens during the advance stages of the parasite’s intraerythrocytic developmental cycle, being directly involved in rosetting and endothelial cytoadherence to prevent splenic clearance and phagocytosis [81]. *Plasmodium* parasites can sequester iRBCs in the microvasculature by binding to endothelial cells, proteins, or glycoproteins. This prevents the infected erythrocytes from passing through the spleen and being destroyed by resident macrophages. The sequestration of infected RBCs in the walls of the microvasculature leads to formation of aggregates that can ultimately disturb the blood flow and have fatal consequences for the host. For instance, excessive sequestration in the microvasculature of the brain is one of the pathophysiological manifestations of cerebral malaria [82], and sequestration of iRBCs in the placenta is associated with complications in infected pregnant women [83]. Another strategy that *Plasmodium* parasites use to escape the host immune response is known as rosetting, where iRBCs bind to many uninfected erythrocytes through surface receptors forming a rosette that protects the emerging merozoites from host antibodies and provides them with a favourable environment to rapidly invade the bound uninfected RBCs [80], [84].

4. The fight against malaria

Global collective efforts have led to a decrease in malaria cases and deaths as a result of more than a century of worldwide preventive measures and research aimed at improving protection, diagnosis and management of malaria.

4.1. Prevention

The WHO's recommended malaria prevention tools and strategies include vector control, by the application of indoor residual spraying and the use of insecticide-treated mosquito nets, and preventive chemotherapies. Measures taken to implement vector control have proved to be efficient in preventing infection and reducing disease transmission. However, they are increasingly threatened by mosquitoes' resistance to insecticides. Moreover, the WHO recommends the deployment of preventive chemotherapies, consisting in giving full treatments with antimalarials to vulnerable populations, *i.e.,* pregnant women and children under 5 years old during the seasonal periods of highest risk of malaria infection, regardless of whether they are infected or not [85], [86]. Also, chemoprophylaxis for non-immune travellers to high-risk areas has been proven effective for preventing potential fatalities.

4.2. Vaccine

Since 2021, the WHO recommends the use of the vaccine RTS,S/AS01 (Mosquirix^{TM}), developed by Glaxo Smith Kline (GSK), as a cost-effective strategy to eradicate malaria in the world's endemic regions in a one-size-fits-all policy. It is a pre-erythrocytic vaccine targeting the sporozoites to prevent invasion of the hepatocytes, consequently blocking transmission. It contains a peptide of the *P. falciparum* circumsporozoite protein (CSP) fused to the hepatitis B virus surface antigen which acts as a virus-like carrier particle along with an immunogenic liposome-based adjuvant. However, the vaccine is not effective against *P. vivax* (since it is based on the 3D7 *P. falciparum* strain sequence), and, depending on the parasite's strain, the effectiveness varies among population subgroups (being as low as 36% in children and about 26% in toddlers). These modest results, in addition to the reported decrease in the efficacy over time (antibody levels show a gradual decline at about 20 months [87]), indicate that the vaccine's effectiveness in preventing malaria is very limited [88]–[90]. For these reasons, although it is deployed as a complementary tool to other preventive measures, it is not considered an eradication means and more efforts are being implemented for the development of a more effective vaccine.

The R21/Matrix-M vaccine is showing more promising results [91]. Similarly to RTS,S, it targets the sporozoite stage of the parasite and contains a *P. falciparum* CSP peptide, but shorter, as well as a carrier protein derived from the hepatitis B virus. The Matrix-M adjuvant stimulates antigen-presenting cells entry at the injection site and enhances antigen presentation in local lymph nodes, hence improving immune response. This vaccine was tested in a Phase 2b clinical trial on children (5-17 months) in Nanoro, Burkina Faso, where it reached the WHO efficacy goal of 75% or greater over 24 months (showing 77% efficacy), and it was even greater when administered with a higher Matrix-M adjuvant dose, reaching 80% at 12 months following the booster vaccination [91], [92]. It has been approved for deployment in Ghana in children aged 5 to 36 months, who are at highest risk of death from malaria, and other countries are undergoing Phase III clinical trials to assess large-scale safety and efficacy.

4.3. Diagnosis

A prompt and accurate diagnosis of malaria infection is crucial for effective disease management. The main challenge regarding malaria diagnosis is that its signs and symptoms are non-specific and resemble those of other febrile illnesses. In areas with high malaria transmission, cases of suspected malaria are confirmed by diagnosing either by microscopy analysis of blood smears or by diagnosis with a rapid diagnostic test (RDT), in order to prescribe an appropriate treatment [93]. The standard examination method is still microscopic inspection, although it has low detection limits for low parasitaemia cases, and it requires a certain expertise in the hands of the person examining the sample. Currently, the recommendation of the WHO is to use RDTs if quality-assured microscopic tests cannot be ensured. RDTs rely on the detection of parasite antigens (mainly *Plasmodium* histidine-rich protein 2 or 3) and their sensitivity is limited, not allowing the quantification of parasitaemia [94]. Despite this, these commonly used tests help to better treat patients with febrile infections, which helps to reduce morbidity and death.

4.4. Treatments

Transmission-blocking treatments aim to prevent transmission to new uninfected individuals, by targeting the sexual stages of the parasite. Other treatments target the pre-erythrocytic liver stage aiming to prevent the establishment of malaria infection. Blood-stage antimalarial drugs aim to reduce the clinical symptomatology once the infection is established, and thus, reduce the burden on patients. Nevertheless, resistance exists to all available treatments, which poses a major threat to the control and elimination of malaria (Fig. 4a) [95]. Most reports of antimalarial drug resistance involve *P. falciparum* infections, but it also occurs in *P. vivax* [96]. Massive human and vector migration alter

local patterns of transmission and pose additional challenges for malaria control initiatives [97]. To prevent drug resistance, treatment includes combination therapy with at least 2 effective antimalarials that have distinct mechanisms of action. Depending on which stage the antimalarial compounds have an effect on, they are classified in four categories: blood schizonticides (which target the asexual forms of the parasites within the RBCs), tissue schizonticides (target the hepatic stage), hypnozoiticides (target the dormant forms in the liver), and gametocytocides (act on the sexual forms of the parasites within the blood). Depending on their pharmacological origin, the antimalarial compounds are also classified in different categories (Fig. 4b).

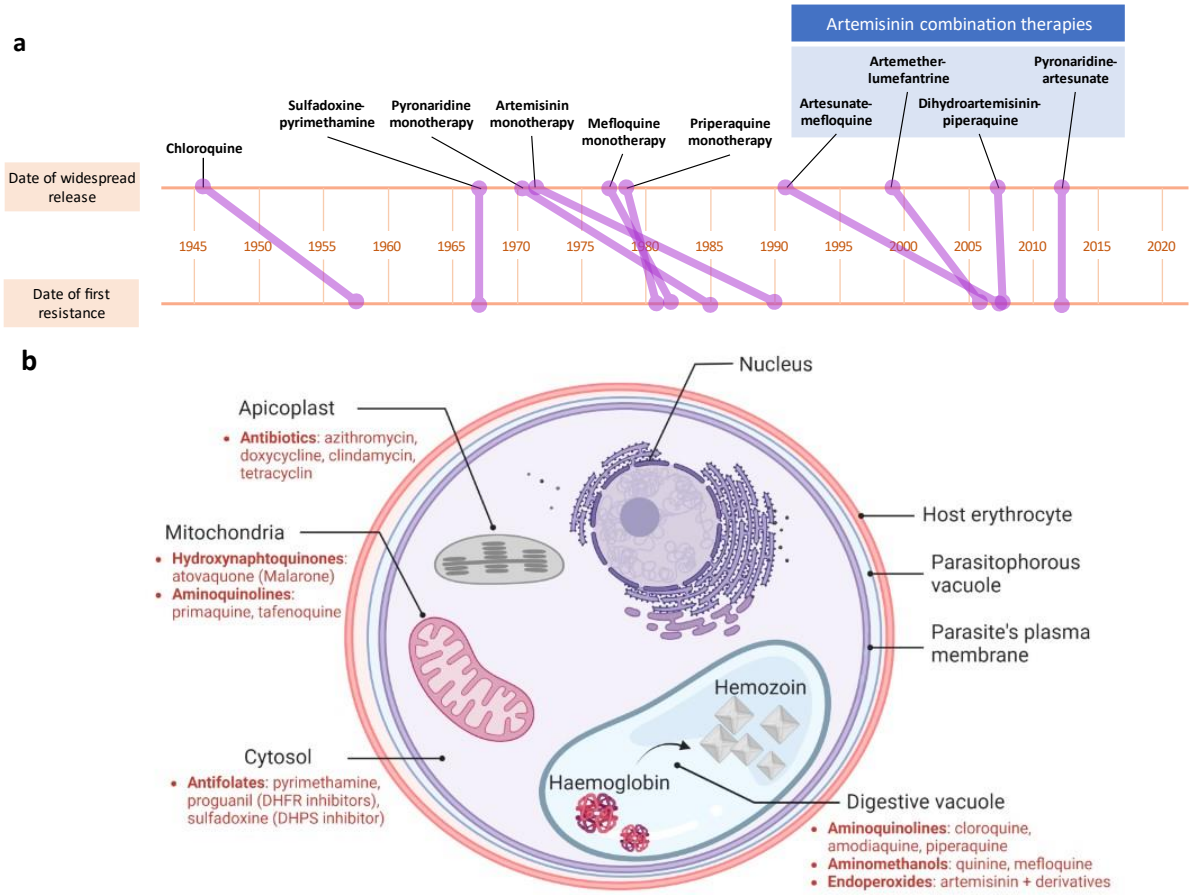


***Figure 4. Antimalarials**. a) Timeline of release of antimalarial drugs and the appearance of parasite resistance to them. Adapted from* [95]. *b) Overview of antimalarial drugs and their targets within a P. falciparum intraerythrocytic parasite. Adapted from* [98]. *Made with Biorender.com.*

4.4.1. Aminoquinoline drugs

Chloroquine, amodiaquine and piperaquine are 4-aminoquinolines indicated to treat infections of *P. vivax, P. malariae, P. ovale* and *P. falciparum*. These molecules are blood schizonticides and act by inhibiting the formation of hemozoin in the digestive vacuole by interfering with the detoxification process of haem (produced by the digestion of haemoglobin), leading to the accumulation of this toxic compound [99].

Primaquine is an 8-aminoquinoline, highly active against hypnozoites and gametocytes. Therefore, it is indicated for the prevention and treatment of relapses of malaria caused by *P. vivax* and *P. ovale,* and it is also used to prevent transmission of *P. falciparum* in areas of Artemisinin resistance. Its mechanism of action involves its oxidation in the liver, which in turn leads to the formation of hydroxylated primaquine metabolites and their oxidation, generating H_2O_2. The accumulated H_2O_2 displays anti-parasitic activity by damaging proteins essential for parasite survival and it can also generate reactive oxygen species (ROS), which induce oxidative stress and lead to parasite apoptosis [93]. It is the only medication in use that is effective against mature gametocytes as well as *P. vivax* infections. However, in individuals with Glucose-6-phosphate dehydrogenase (G6PD) deficiency (a genetic condition commonly spread in African populations that causes an impaired oxidative stress response), it is known to produce dose-dependent haemolysis [86], [100]. Tafenoquine is a metabolically stable and therefore slowly eliminating version of primaquine, preventing relapses in a single dose. Hence, it is a good candidate to accelerate the elimination of *P. vivax* [101].

4.4.2. Amino alcohol drugs

Quinine, an alkaloid derived from the bark of the cinchona tree, is a blood schizonticide in *P. falciparum* and also has gametocytocidal activity against *P. vivax* and *P. malariae*. Although their exact mechanism of action is unknown, quinine, mefloquine and lumefantrine also inhibit the detoxification of haem inside the digestive vacuole [102].

4.4.3. Antifolates

Pyrimethamine, proguanil and sulfadoxine are blood schizonticides mainly active against late stages of the parasite's asexual cycle. Pyrimethamine and proguanil inhibit the dihydrofolate reductase (DHFR), blocking the biosynthesis of purines and pyrimidines, which are essential for DNA synthesis, thus leading to the failure of DNA replication during schizogony. In contrast, sulfadoxine acts by inhibiting the activity of dihydropteroate synthetase (DHPS), which is required for the conversion of 4-aminobenzoic acid to folic acid; a vital compound for DNA synthesis, repair and methylation [93].

4.4.4. Hydroxynaphthoquinones

Atovaquone, an analogue of ubiquinone, is the only compound of this class currently on the market and shows antimalarial activity against all *Plasmodium* species. Its mechanism of action involves the inhibition of electron transfer in the mitochondria by binding to cytochrome bc1, leading to a loss of membrane potential, and indirectly blocking the synthesis of pyrimidines. This molecule is particularly

effective in the context of prophylaxis in *P. falciparum*, where it is combined with proguanil for its commercial use (Malarone^{TM}) [103].

4.4.5. Antibiotics

Clindamycin, azithromycin, doxycycline and tetracycline are several antibiotics with antimalarial activity. They share a common mode of action targeting the apicoplast in *Plasmodium*. By interfering with the 70S ribosome subunit, they inhibit the apicoplast translation machinery and peptide chain initiation [104]. The metabolic functions of the apicoplast are not immediately affected but this organelle is essential for parasite survival, hence, parasites die at the end of the second intraerythrocytic cycle after drug exposure.

4.4.6. Endoperoxides

Artemisinin and its derivatives artesunate, dihydroartemisinin and artemeter are the most widely used reference compounds in the treatment of malaria caused by *P. falciparum*. Artemisinin is a natural compound extracted from the *Artemisia annua* plant, which has been used as a herbal remedy in China for centuries [105]. Artemisinin is a very powerful blood schizonticide, and its derivatives act on the early intraerythrocytic stages (rings), allowing rapid clearance of parasites from the blood [106]. Although the exact mechanism of action is not fully known, dihydroartemisinin is the active metabolite common to artemisinin and its derivatives, and its antimalarial effect depends on binding haem following its incorporation into the parasite, hence inhibiting its detoxification. Additionally, artemisinin derivatives appear to disrupt a variety of organellar and cellular processes, including glycolysis, protein synthesis and degradation, cell cycle control, and haemoglobin endocytosis owing to the generation of reactive carbon radicals, which eventually results in cell death. Moreover, artemisinin derivatives appear to accumulate preferentially in infected red blood cells, concentrating the drug several hundred times compared to uninfected cells [93].

Currently, the first-line treatment for *Plasmodium* infected patients consists of oral Artemisinin based combination therapies (ACT) or chloroquine for uncomplicated malaria and intravenous artesunate for severe malaria. ACTs involve the use of a fast-eliminating artemisinin derivative with another slow-eliminating antimalarial. Artemisinin compounds have a fast-acting effect on asexual stages but do not hamper the sexual gametocytes, which can persist longer in blood.

Chapter 2. *Plasmodium* biology

1. Phylogeny and common structures of apicomplexans

Apicomplexan parasites are evolutionary distant from the model organisms contained in the group Opisthokonta, such as mouse, humans, *Xenopus* and yeast (Fig. 5a). *Plasmodium* species belong to the phylum Apicomplexa, which constitutes a phylum that comprises more than 5000 species, mostly with a parasitic lifestyle that causes diseases in humans or animals. Among these are the pathogens causing toxoplasmosis (*Toxoplasma gondii*), cryptosporidiosis (*Cryptosporidium parvum*) and malaria (*P. falciparum*). Apicomplexa belong to the superphylum Alveolata, along with Cilliates and Dinoflagellates, and they share a set of characteristic common features (Fig. 5b).

For instance, they share a common ultrastructure called the apical complex formed by the assembly of specific secretory organelles at the apical tip of the cell (rhoptries, micronemes, and dense granules), which facilitate the invasion of the parasite into host cells. Furthermore, they display a group of alveoli forming a set of sacs beneath their plasma membrane that form a stabilizing protein structure called the Inner Membrane Complex (IMC), which maintains the shape and structure of apicomplexan parasites as well as allowing them to glide and cross biological barriers to actively penetrate host cells. In addition, the majority of apicomplexan parasites carry an apicoplast, a plastid organelle that arose through secondary endosymbiosis, when the ancestor of all apicomplexan parasites engulfed a eukaryotic algae and retained the algal plastid [107]. Although over the course of evolution it has lost its photosynthetic ability, it is required for parasite survival and it contributes to the metabolism of isoprenoids (critical for diverse cellular process such as prenylation of proteins involved in vesicular trafficking), fatty acids (precursors for membrane lipid synthesis), iron-sulphur clusters (co-factors for essential biological functions), and haem biosynthesis [108].

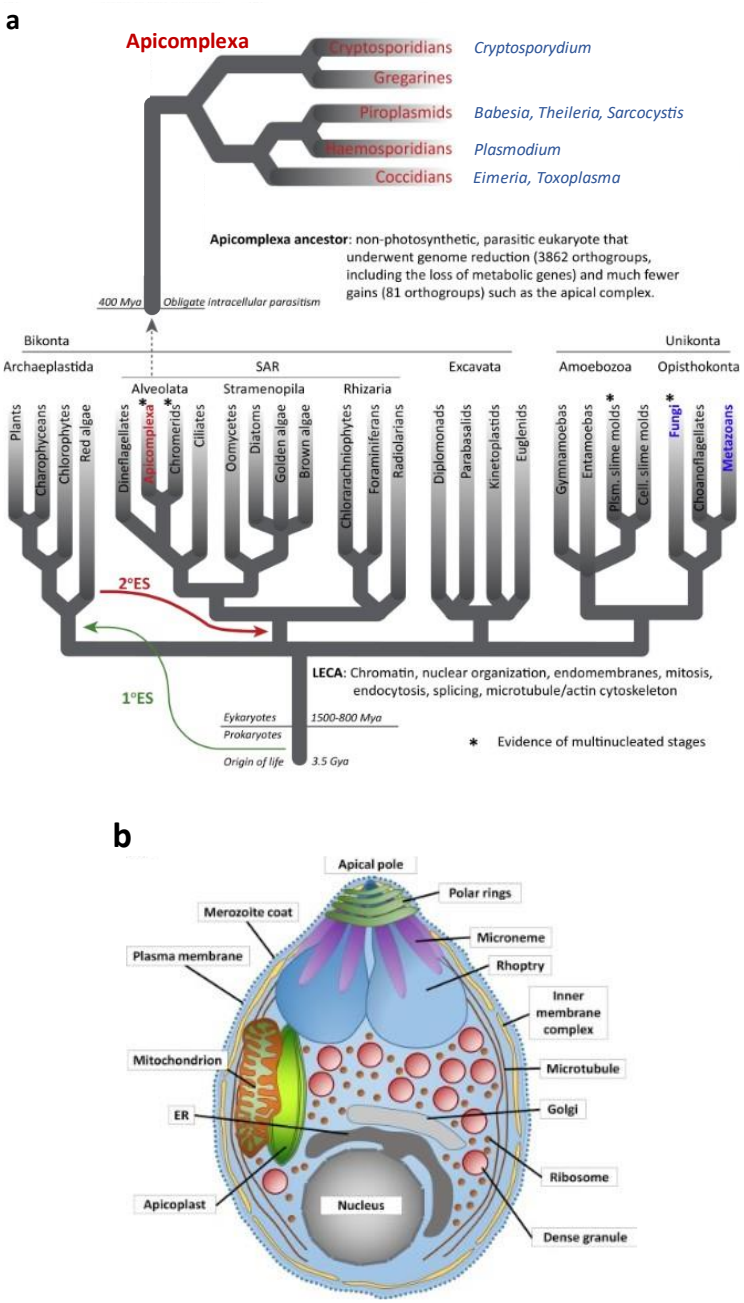


Figure 5. Phylum Apicomplexa and basic structures*. a) Phylogenetic tree showing the major eukaryotic groups, with a focus on apicomplexan organisms and examples of species. Green and red arrow point to endosymbiosis events (ES) that led to acquirement of plastids. Adapted from* [109] *b) Basic structures of a Plasmodium falciparum merozoite. ER: Endoplasmic reticulum. Adapted from* [110].

2. *Plasmodium* species infecting humans

As previously mentioned, there are five human-infecting *Plasmodium* species: *P. falciparum*, *P. vivax*, *P. ovale, P. malariae* and *P. knowlesi.* The main differences between these species are detailed in table 1 and include the existence of a dormant form of the parasite in *P. vivax* and *P. ovale.* They are able to differentiate into liver-stage hypnozoites which remain inactive inside the hepatocytes for weeks,

months or even years and eventually reactivate, leading to a restart of the blood infection with potential for transmission and a relapse of clinical malaria [79], [96]. Another variation between species is the differential preferences they have towards the distinct erythrocyte maturation stages: *P. vivax* and *P. ovale* invade immature reticulocytes whereas *P. malariae* prefers senescent old erythrocytes, *P. falciparum* invades erythrocytes of all ages (although preferably reticulocytes [111]), and *P. knowlesi* shows no preference. The length of the asymptomatic liver-stage incubation period (interval between infection and the onset of symptoms) also shows variability between species, ranging from 9 days in *P. falciparum* infections, to 30 days in *P. malariae* cases and even years in cases of *P. vivax* and *P. ovale* hypnozoites. Finally, the intraerythrocytic stage of parasite's development has a variable duration depending on the *Plasmodium* species. For *P. falciparum* and *P. vivax*, this process lasts between 44-48h, whereas *P. knowlesi* parasites present a cycle duration of 24h, *P. ovale* of 50h and *P. malariae* of 72h. Furthermore, the number of merozoites produced per schizont also varies from ~15 in *P. knowlesi* to as many as ~32 in *P. falciparum*, and the exact reason of the merozoite number limitation is not known [112], [113].

Morphologically, the different species are distinguishable, and this is the main diagnostic form through microscopic examination of Giemsa-stained blood smears [114], [115]. *P. falciparum* smears are characterized by the presence of mostly rings and young trophozoites (the early forms of their asexual cycle). The most recognizable signs of *P. vivax* are larger infected erythrocytes and the presence of grains called 'Schüffner's dots' in the RBC's cytoplasm. *P. ovale* also exhibits Schüffner's dots and tends to cause the elongation of RBCs, making it challenging to distinguish from *P. vivax*. However, *P. ovale* is slightly smaller than *P. vivax* and contains smaller merozoites per schizont. The smallest of all, *P. malariae,* produces around 8 - 10 merozoites that are frequently grouped in a rosette pattern within the iRBCs.

Table 1. Characteristics of the different human infecting Plasmodium species [112], [116].

Characteristics	*P. falciparum*	*P. knowlesi*	*P.vivax*	*P.ovale*	*P. malariae*
Erythrocyte preference	Reticulocytes	All	Reticulocytes	Reticulocytes	Mature erythrocytes
Erythrocytic cycle	44-48h	24h	44-48h	50h	72h
Parasitaemia *in vivo*	Around 1%, can be high	Around 1%, can be high	>2%	>2%	>1%
Incubation period	7-9 days	8-12 days	7-12 days, up to years (hypnozoites)	12-20 days, up to years (hypnozoites)	16-60 days
Relapses	No	No	Yes	Yes	Not fully cured
Pathology	Highly pathogenic, most deadly, mainly in Africa	Potentially lethal, South-East Asia	Not life-threatening, most prevalent outside Africa	15 M cases annually, mainly in Africa	Scattered worldwide
Morphology	Mostly rings, multiple infected RBCs	16 merozoites produced per schizont	Schüffner's dots, enlarged infected RBCs	Schüffner's dots, elongated RBCs, smaller and fewer merozoites per schizont	Band form in trophozoites, merozoites in rosette within RBCs
Genome	23.3 Mb, 14 chr, 80% AT content	24.4 Mb, 14 chr, 62% AT content	29.1 Mb, 14 chr (isochore), 59% AT content	33.5 Mb, 14 chr, 70% AT content	33.6 Mb, 14 chr, 75% AT content

3. *Plasmodium falciparum* genome

In general, *Plasmodium* species possess compact haploid genomes of 23 – 33 Mb, which encode 4,600–5,000 core protein-coding genes spread among 14 linear chromosomes, as well as a variable number of multigene families, mostly subtelomeric [117]. *Plasmodium* species contain two additional genomes: a linear mitochondrial genome that is one of the smallest known (only 6 kb) [118] and a ~34 kb circular plastid genome housed in the apicoplast [119].

The first full *Plasmodium* genome sequence published was *P. falciparum* in 2002 [117], followed by the analysis of *P. vivax* and *P. knowlesi* genomes in 2003 and 2008 [120], [121], respectively. More recently, the genomes of *P. malariae* and *P. ovale* were sequenced [122], completing the genomes of the five human infecting *Plasmodium* species.

3.1. *P. falciparum genome organization*

The nuclear genome of *P. falciparum* is comprised of 14 chromosomes (ranging from 640 kb for chr. 1 to 3.3 Mb in chr. 14), has a total size of 22.8 Mb, and encodes around 5300 protein-coding genes [117]. One of the most distinct characteristics of the *Plasmodium falciparum* genome is the high content of adenine (A) and thymine (T) (80.6%) [117]. Such AT richness is not present in other *Plasmodium* species like *P. vivax* or *P. knowlesi*, despite the high synteny between them [123]. To compare it with other well-studied eukaryotes, like *S. cerevisiae* [124] and *S. pombe* [125], the *P. falciparum* genome contains roughly a similar number of genes, but it is about twice as large. In addition, the coding regions of *P. falciparum* genes are, on average, longer than those of *S. pombe* (2.3 kb vs. 1.4 kb), which could be explained by the fact that many *P. falciparum* proteins that often display long repeats of the same amino acid (most frequently asparagine [126]). Moreover, around 60% of the encoded *P. falciparum* proteins have little or no similarity to proteins in other organisms and are annotated as "hypothetical" or with "unknown function" [127]. The proportion of genes encoding for these proteins is higher in *P. falciparum* than in other model organisms, reflecting the greater evolutionary distance between Apicomplexa and other sequenced eukaryotes. Finally, the *P. falciparum* genome holds 43 transfer RNAs (tRNA) and a few genes of ribosomal RNA units (rRNA) that are classified in two categories according to the expression stage of parasite life cycle: type S rRNAs, mainly expressed in the mosquito, and type A, preferentially expressed in the human host [128].

Proteomic characterization of the life cycle stages of the parasites (sporozoites, merozoites, trophozoites and gametocytes) revealed an extremely diverse proteome, with only 152 proteins (6%) shared between the four phases, whereas approximately 50% of the sporozoite proteins are unique to this stage, and the proportion of unique proteins for trophozoite, merozoite and gametocytes ranges between 20 % and 30% [112].

3.2. *Promoters, UTRs, TSSs*

Plasmodium parasites show a tightly regulated gene expression program across their intraerythrocytic life cycle. Transcriptomic studies have reported that the expression of the ~5300 genes varies considerably between the different intra-erythrocytic stages in a continuous cascade of gene expression, where most genes show a single peak of maximum expression [129]–[131]

Untranslated regions (UTRs) are known to play crucial roles in the post transcriptional regulation of gene expression. A study of the transcriptome of *Plasmodium* parasites through amplification-free RNA-seq [130] revealed that the parasite's genome comprises particularly large 5' UTRs, with an

average size of 600 bp and reaching up to 5.4 kb, unlike in *S. cerevisiae* where UTR size ranges between 100-200 nucleotides [132]. In another transcriptomic study of the landscape of *P. falciparum* transcription initiation through 5' cap sequencing [133], the authors show that Transcription Start Sites (TSS) appear to be clustered across fairly wide regions in the genome, indicating a broad promoter architecture similar to other eukaryotic organisms. The majority of the genes are preceded by clusters of more than one transcription initiation sites, allowing them to use multiple promoter regions as well as multiple TSSs [133]. On a genome-wide scale, TSSs of actively transcribed genes are surrounded by a region depleted of nucleosomes, facilitating the recruitment of the transcription machinery, and enabling the formation of pre-initiation complexes. Accordingly, nucleosomes show a dynamic depletion in the 5'UTR regions of the most highly expressed genes throughout the IDC [134].

3.3. Euchromatin vs. heterochromatin

The chromatin of the nuclear *P. falciparum* genome exists in two distinct functional states: a relaxed euchromatin state linked to active transcription and a condensed heterochromatin state associated with gene silencing. The majority of the parasite's genome is in a non-compact, transcriptionally permissive euchromatic state. In contrast, the regions where heterochromatin can be found, defined by an increased nucleosomal occupancy, are restricted to centromeres, subtelomeric regions, and a few chromosome-internal heterochromatic islands [135]. Inside the nucleus, these heterochromatin DNA regions are clustered towards the nuclear periphery [136].

Subtelomeric regions comprise most of the compact heterochromatin. The chromosome ends are made up of telomeric tandem repeats (GGGTTT/CA) followed by an array of DNA elements at the subtelomeric regions [137]. These telomere-associated sequences (TAS) are species-specific and consist of a coding and a non-coding region. The non-coding region is composed of a set of six blocks of conserved tandem repeats known as telomere-associated repetitive elements (TAREs) located between the telomere and the coding regions, with a variable length between 20-40 kb [138]. Adjacent to the non-coding TAREs are the subtelomeric coding regions, which are structurally highly conserved and are known to play a role in parasite virulence, as they contain antigenic variation multigene families like *var/rifin/stevor* which encode variant surface antigens (VSA) (Fig. 6) [139]. The majority of *stevor* genes are localized at the chromosome ends; in contrast to *var* and *rifin* genes that can also be found in internal regions of the chromosomes.

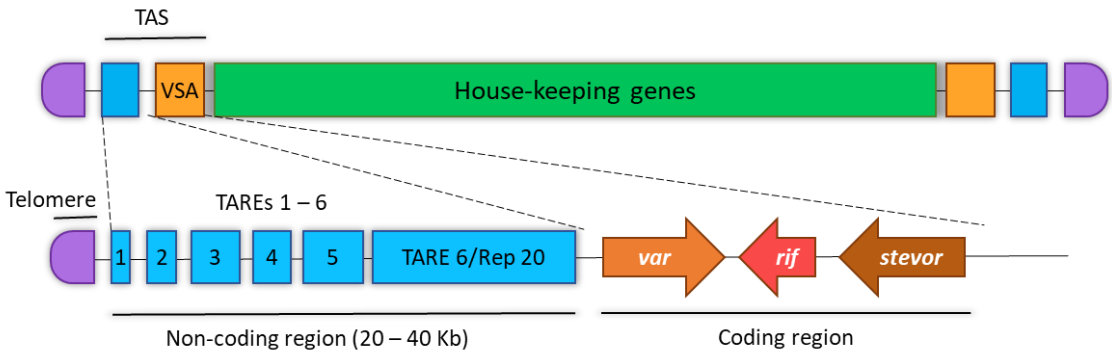


Figure 6. Structure of P. falciparum subtelomeric regions*. TAS: Telomere associated sequences. TARE: Telomere associated repetitive element. VSA: variant surface antigens. Adapted from* [137].

The centromeres of *P. falciparum* have been mapped through the Chromatin Immunoprecipitation of the centromeric histone variant *Pf*CENH3 and sequencing of its associated DNA [140]. With a ranging size around 4–4.5 kb, they are extremely AT-rich (98%) and contain multiple repeats of varying size and copy number. The chromatin status of centromeres and pericentromeric regions is mostly heterochromatin, although they do not display a significant enrichment in heterochromatin marks [140].

3.4. Post-translational modifications as means to regulate transcription.

The chromatin is organised in nucleosomes, and each nucleosome consists of an octamer of core histones wrapped with a 147 bp DNA fragment in 1.75 super helical turns around [141] (Fig. 7). The genome of *P. falciparum* encodes a canonical form of each core histones (two each of H2A, H2B, H3 and H4) and four histone variants (H2A.Z, H3.3, centromere specific H3 (CenH3), and H2Bv). One characteristic of *Plasmodium falciparum* parasites is the absence of the H1 linker histone [142], [143]. An extensive variety of covalent modifications in the flexible N-terminal tails of the histones play an important role in transcription regulation, and they display a dynamic enrichment that varies throughout the intraerythrocytic development cycle (IDC), suggesting a cycle-associated regulation [139], [144].

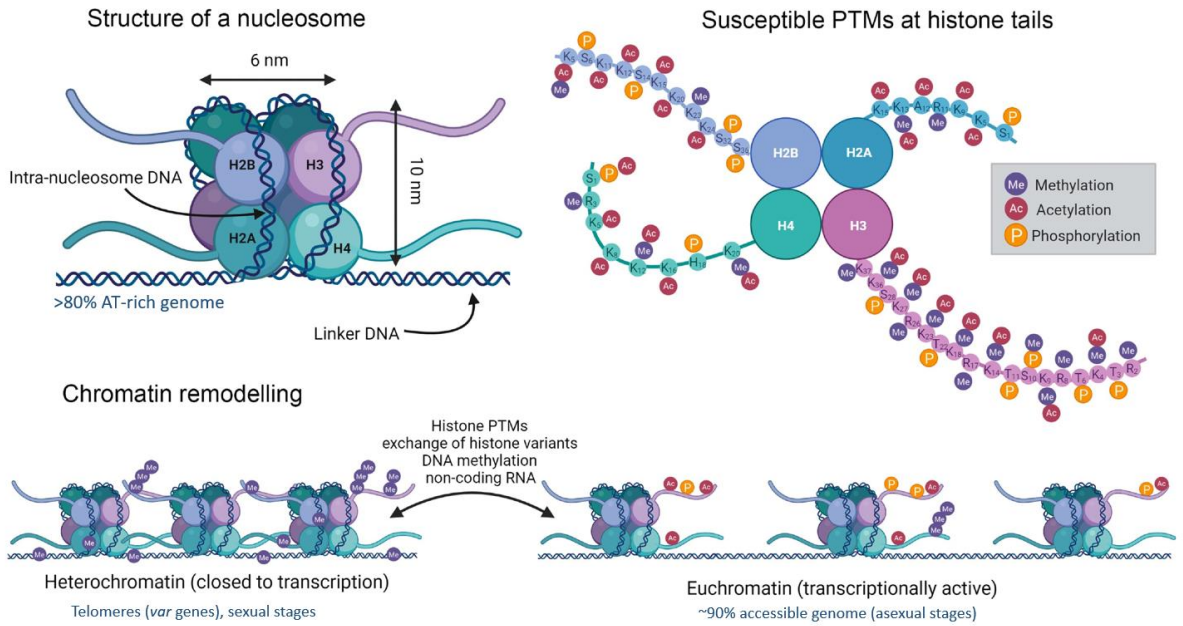


Figure 7. Nucleosome structure and post-translational modifications at histone tails*. The DNA is wrapped around the nucleosome, that comprises intranucleosome DNA and the core eukaryotic histone proteins, including H2A, H2B, H3, and H4 (the linker Histone 1 is absent). Histone tails undergo PTMs at different amino acids and play a role in chromatin remodelling and DNA accessibility to transcription. In asexual parasites, the majority of chromatin exists in an euchromatic state, while gametocyte chromatin is enriched in H3K9me3-marked heterochromatin. Adapted from* [143], [145].

Euchromatin regions, promoters and enhancer regions are marked by the presence of post-translational modifications associated with active transcription. For instance, H3K4me3 and H3K9ac have been shown to be enriched in the promoters of active genes during the schizont stage of the asexual cycle, whereas they show no discrimination between active and inactive genes in rings [144], [146]. These dynamically marked euchromatic regions are also associated with an enrichment of the histone variant H2A.Z, suggesting that the H3K4me3 and H3K9ac marks are preferentially localized on nucleosomes with the H2A.Z variant [147]. In human cells, euchromatin markers like H3K27ac and H3K18ac are enriched at promoters and H3K4me1, H3K27ac and H3K18ac in enhancers [148], [149]. A genome-wide study of the correlation between *P. falciparum* gene expression profiles and histone modifications associated with their cognate promoter and enhancers showed that there is a dynamic enrichment of the acetylation at transcription start sites (TSS) of genes correlating with their gene expression [150]. On the other hand, levels of H3K4me1 remain constant across the IDC, showing an enrichment in the transcribed sequences of genes and a depletion before their TSS and after their transcription termination site (TTS) [144].

The regions of heterochromatin maintain clusters of genes in a silent state and are marked by the presence of several specific post-translational modifications including histone deacetylation, H3K9me3 and the heterochromatin protein *Pf*HP1 that binds to H3K9me3 [151], [152]. H3K9me3/*Pf*HP1 show a constant localization across the IDC, clustered on subtelomeric and some

chromosome internal regions, and they are specifically associated with clonally variant gene families, like *var, rifin* and *stevor*.

3.4.1. The var gene family

The variant surface antigens (VSA) are encoded by a number of diverse multigene families, such as *var, rifin, surfin* and *stevor* genes. The *var* gene family is the best characterized VSA member, it consists of 60 genes that code for the *P. falciparum* Erythrocyte Membrane Protein 1 (*Pf*EMP1) and have a mutually exclusive expression (only one gene is expressed at a time). The ability to switch the expression between these genes allows the parasite to evade the host immune system and to ensure adherence to the microvasculature, increasing the severity of disease [153].

The case of *var* genes is a clear example of epigenetic control of transcription, although the molecular mechanisms controlling mutually exclusive expression are not fully understood. Most of the members of the family remain repressed while a single *var* gene at a time is able to escape silencing and is abundantly transcribed [154]. An epigenetic modification mechanism allows the parasites to switch between the exclusively transcribed *var* genes, allowing them to alter the presentation of surface antigens and evade the host immune system [153]. The TSS region of the transcribed *var* gene shows an enrichment in the euchromatic marks H3K9ac and H3K4me3, and the stably silent *var* genes are enriched in H3K9me3 and *Pf*HP1 in their 5' upstream and coding regions [152]. This suggests a competition between acetylation and trimethylation in H3K9 at the promoter regions of *var* genes to influence transcription activity.

3.5. G-quadruplexes

G-quadruplexes (G4s) are non-canonical secondary structures formed in G-rich DNA or RNA sequences. They are formed by four tracts of at least three guanines located in close proximity in the same strand, forming three or more G-quartets stacked on top of each other that are stabilized by monovalent cations and held together by Hoogsteen bonds between the guanines (Fig. 8). These structures can be formed intramolecularly or intermolecularly and, depending on the combinations of strand direction and loop composition, they can adopt a wide range of topologies. Additionally, their GC-rich signature motif ($G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$) allows *in silico* predictions of their genomic location and distribution [155], [156].

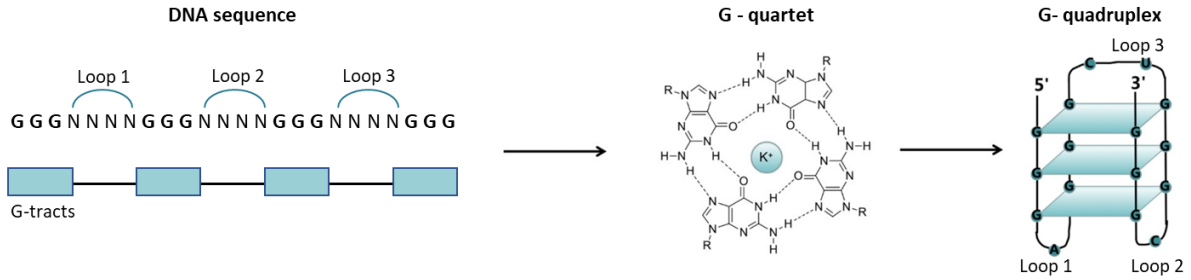


***Figure 8. G-quadruplexes**. DNA sequence and structure. Adapted from* [157], [158].

In humans, G4s show a non-random distribution, mostly accumulating at telomeric repeats and enriched in regulatory regions such as transcription start sites (TSS), promoters, replication origins (the sites of the genome where DNA replication starts) and nucleosome-depleted regions. Multiple roles have been attributed to these non-canonical secondary structures, including regulation of replication [159], gene transcription [160] and telomere maintenance [155], [161], [162].

Regarding DNA replication, G-quadruplexes have been suggested to partake in origin specification through the exclusion of nucleosomes, which could subsequently favour the binding of replication components such as the Origin Recognition Complex (ORC). In fact, genome wide analyses of active origins of replication in mouse, *Drosophila* and human cells have shown that more than 60% of replication origins contain GC-rich regions with potential to form G4s, known as Origin G-rich Repeated Elements (OGRE) [161], [163]–[165], and their presence has been linked in mammalian cells to their cognate origin efficiency [7].

In *P. falciparum*, G4 forming sequences (G4FS) have been mapped throughout the genome using two different algorithms [166], [167]. The QGRS Mapper [168] searches for canonical ($G_3N_mG_3N_mG_3N_mG_3$) G4-forming sequences in the genome, whereas the G4Hunter algorithm [169] identifies potential G4 forming sequences considering their G-richness and G-skewness, instead of a specific consensus sequence, and assigns a G-quadruplex propensity score to each site. In *P. falciparum,* an enrichment of potential G4 sites was found at nucleosome-depleted regions and towards chromosome ends [167], consistent with the increased G content of telomeric repeats, that makes them prone to form G4s [170].

Chapter 3. DNA Replication

The ability of the cells to replicate their DNA accurately and efficiently is essential for their survival and adaptation in changing environments. Understanding the mechanisms driving DNA replication is thus a key area of research in microbiology and genetics. DNA replication involves a complex interplay of effectors and regulatory factors that coordinate the unwinding, synthesis, and fidelity of the newly synthetized DNA molecule.

1. DNA replication in prokaryotes

Prokaryotic cells display generally a small genome organized in a single or multiple circular chromosomes. DNA replication begins at a specific site on the single chromosome termed the origin of replication, or *oriC,* a highly AT-rich region [171]. At this spot, specialized proteins and helicases initiate the unwinding of the two strands of the double helix, and replication proceeds in both directions along the chromosome. The prokaryotic DNA polymerases synthesize the new complementary strands of DNA using the existing ones as template. As the new strands are synthesized, they coil up to form two complete identical copies of the genome [172], [173].

2. DNA replication in model eukaryotic organisms

In eukaryotic cells, DNA replication begins at specific sites on the DNA known as the origins of replication. During the G1 phase of the cell cycle, these are recognized by the Origin Recognition Complex (ORC), a hexameric protein complex that binds to them and afterwards acts as a landing platform for the sequential binding of other proteins, including the Cell division cycle 6 protein (Cdc6) and the Cdc10-dependent transcript 1 (Cdt1), that forms a complex with the mini-chromosome maintenance (MCM2-7) hexameric DNA helicases and together constitute the pre-replicative complex (pre-RC) [173]–[175]. The assembly of this complex is termed origin licensing. Only a subset of these licensed origins will be activated in S phase while the rest stay dormant. Their activation involves the interaction of the MCM2-7 hexamers with other proteins that promote the assembly and activation of additional replication factors, like Cdc45, which eventually lead to the loading of the PCNA sliding clamp and the DNA polymerase machinery. This process can be divided in 4 distinctive steps which are explained in more detail below.

2.1. Step 1: Recognition of origins of replication

The origins of replication are defined as the specific sites in the chromosomes where double-stranded DNA unwinds to generate single-stranded DNA templates for genome duplication. One of the best characterised model organisms, the budding yeast *Saccharomyces cerevisiae*, displays a set of sequence-specific motifs at the origins of replication known as autonomously replicating sequence (ARS). This sequence, if expressed solely in a plasmid, confer it with the ability to replicate as a mini chromosome like the rest of the genome, once per division cycle. Two key elements form the ARS: the 11 bp consensus sequence (ACS) [5'-(A/T)TTTA(T/C)(A/G)TTT (A/T)-3'] (known as the A domain), that is essential for replication initiation as it is recognized by the eukaryotic replication initiator Origin Recognition Complex (ORC) [176], and additional sequences of ~100-200 bp 3' to the consensus T-rich strand (3'-flanking sequence), known as the B domain, that are also required for origin function [6]. Despite this specific sequence defining all the 12,000 ACSs present in the *S. cerevisiae* genome, only a subset of them (~400) are used, indicating that the presence of the ACS is a necessary but not sufficient element for origin specification [177].

In multicellular organisms, because of the larger size of their genomes, thousands of replication origins must be activated in every cell cycle (up to 30,000–50,000 in human or mouse cells) [161], [178]–[180]. Furthermore, replication origins in metazoans do not display consensus sequences and exhibit a high grade of heterogeneity. The sole element that has been repeatedly found in metazoan origins is a G-rich repeat element (OGRE) that, although not displaying sequence specificity, exhibits a high tendency to form G-quadruplexes. In *Drosophila* cells, mouse and human cells this G-rich element is present in more than 60% of the origins of replication, generally in several copies per origin; and initiation of replication is fired at specific positions downstream of this OGRE element [161], [163]–[165], [181].

Proteins involved in origin recognition are relatively conserved but the main origin binding factor, the origin recognition complex (ORC), varies in its sequence specificity. The ORC is a hexameric protein complex composed of five subunits (ORC1 to 5) forming a crescent moon shape, plus the ORC6 subunit bound in the periphery of the complex (Fig. 9) [182]. This complex binds to the replication origins in an ATP-dependent manner [183]. The ORC1-5 subunits are members of a class of ATPases known as the AAA+ family (ATPases Associated with various cellular Activities), and they contain a C-terminal DNA-binding domain, and an N-terminal ATPase domain [184]. In the case of the ORC, first, one molecule of ATP binds to ORC1 and another ATP to ORC5, and this interaction stabilizes the bound state to the origin DNA, leading to a tight attachment of the ORC complex to the DNA strand with the ORC subunits forming a spiral around the DNA (Fig. 9) [185]. Once the ORC is bound to origin DNA, the

ATPase activity is inhibited, the ATP remains stably bound to ORC1/5 and the process continues to promote the recruitment of the Cdc6 subunit [186], [187].

2.2. Step 2: Pre-replicative complex assembly

Following the stable formation of the ORC complex around the origins of replication, the Cdc6 (cell division cycle 6 protein) specifically recognizes the ATP-bound state of ORC1 and becomes transiently associated with ORC through a mechanism controlled by further ATP binding of Cdc6. Finally, the Cdc6 protein will close the ring-like structure of the ORC around the origin, topologically trapping the double stranded DNA (Fig. 9) [182], [183], [188].

Once the ORC-Cdc6 complex is stably bound to origin DNA during early G1 phase, the complex formed by Cdt1 and the MCM2-7 helicase will be loaded (Fig. 9). Cdt1 (Cdc10-dependent transcript 1 protein) is essential for the loading of the eukaryotic replicative DNA helicase, formed by two copies of the mini-chromosome maintenance (MCM) protein complex positioned head-to-head. Each MCM2–7 complex contains one copy of the six essential MCM2, MCM3, MCM4, MCM5, MCM6, and MCM7 proteins. All the MCM proteins belong to the AAA+ family and carry the well-conserved domains for ATP binding and hydrolysis, that provide them with the energy needed for the helicase motor to unwind the DNA. The mechanism by which the MCM2-7 complex is loaded is facilitated via an interaction with ORC6 [189].

ORC6 is located distal to the central channel of the ORC complex containing the DNA; it is different from the other ORC subunits as it is the only subunit that is not related to the AAA+ family of ATPases and is not involved in the recognition of the origin DNA. ORC6 binds Cdt1 directly, promoting the loading of the MCM2-7 hexamer onto DNA through sequential ATP hydrolysis by Cdc6 [186], [187]. ATP hydrolysis by Cdc6 triggers the closure of the MCM2-7 ring around DNA [186]. The first MCM2-7 complex encircles the DNA in a ring-shape structure oriented with the C-terminus end of the hexamer binding to ORC-Cdc6 and the double stranded DNA inserted into the channel, forming the intermediate OCCM complex (ORC-Cdc6-Cdt1-MCM) [190](Fig. 9). Shortly after, ATP hydrolysis by the ORC and MCM2-7 triggers the expelling of Cdc6 and then Cdt1 to form the OM complex (ORC-MCM) [186], [188].

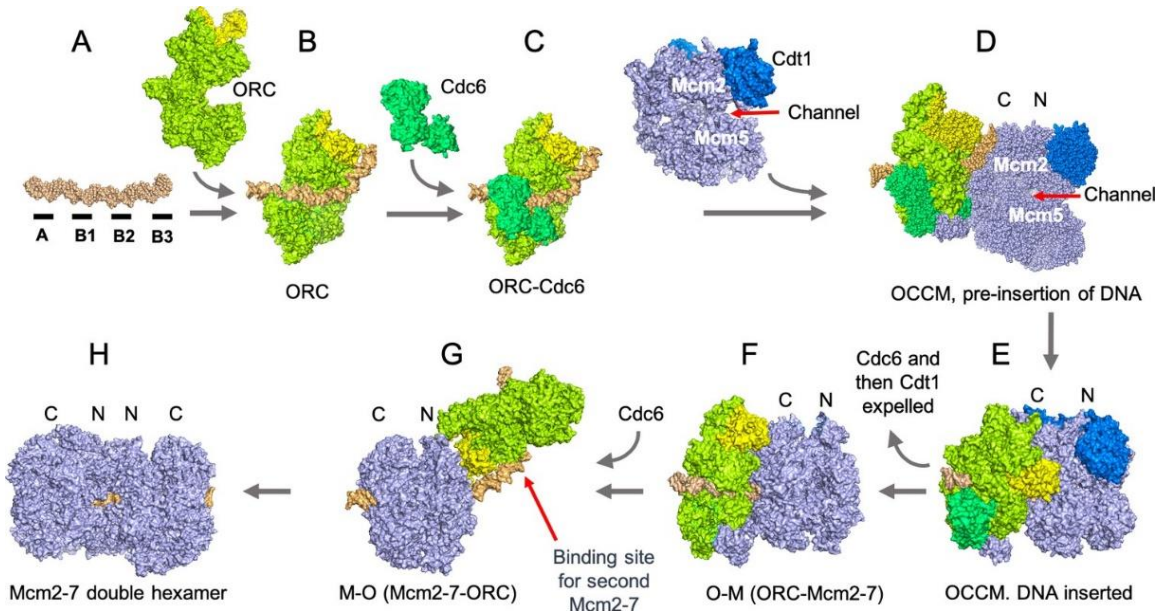


Figure 9. The assembly of the pre-replicative complex in S. cerevisiae [190]***.*** *(A) The genetic sequence ARS (formed by the A-B1-B2-B3 elements) acts as the starting point for replication in S. cerevisiae (DNA segment in orange). (B) The ORC (lime green, ORC6 yellow) binds to it. (C) ORC recruits Cdc6 (green). (D) ORC-Cdc6 recruits the Cdt1-MCM2-7 complex (Cdt1 in blue and MCM2-7 in violet). The C-terminal end of the MCM2-7 complex binds ORC-Cdc6. (E) The double stranded DNA is inserted into the channel of the MCM complex, and the hexamer is partially closed, creating the OCCM intermediate complex. (F) ATP hydrolysis by the MCM2-7 expels Cdc6 and then Cdt1, creating the OM complex. (G) ORC flips to the other side of MCM2-7, creating the MO complex. As a result, ORC6 is now orientated towards MCM2-7-Nt side. ORC recruits a second Cdc6, creating a binding site for a second Cdt1-MCM2-7 complex that is loaded in an opposite orientation to the first MCM2-7. (H) The pre-replicative complex is established, with the MCM2-7 double hexamer and the ORC still bound to the DNA.*

The bidirectional nature of DNA replication requires the loading of two MCM2-7 helicases around DNA in a head-to-head conformation. This is actually accomplished by a single ORC molecule recruiting the two MCM complexes sequentially [190]–[192]. Following the release of the Cdt1, a new interaction site on MCM is created, which will be the landing pad of ORC6, that bends over and forms a new interaction with the opposite face of the MCM2-7 [191] (Fig. 10). The ORC is then released from the initial DNA binding site and forms an open ring around DNA in the opposite orientation, forming the MO complex (MCM-ORC). ORC then recruits a second Cdc6, which offers a new binding site for a second Cdt1-MCM2-7 complex that will be loaded in the opposite orientation to the first MCM2-7 [189] (Fig. 10). Consequently, head-to-head interactions between the two helicases are established. Finally, another round of ATP hydrolysis triggers the release of the second Cdt1 and Cdc6 and leads to formation of the stable double helicase hexamer, completing the pre-Replicative Complex assembly (pre-RC) [192]. This process occurs along the G1 phase of the cell cycle, and it is known as “origin licensing”. Licensed origins contain the pre-RC bound to DNA and are ready to be activated in S phase, triggering the unwinding of the double stranded DNA.

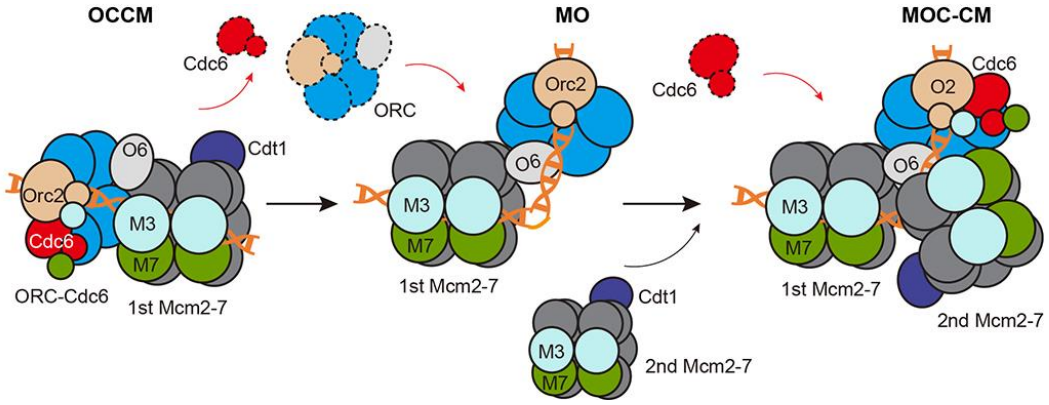


Figure 10. Molecular mechanism of loading the second Cdt1-bound Mcm2–7 hexamer onto origin DNA [192]***.*** *After the formation of the OCCM complex, ORC and Cdc6 dissociate from the C-terminal side of the loaded MCM2–7. ORC binds to the origin DNA and to the N-terminal side of the first MCM2–7, forming the MO complex. Cdc6 next binds to MO to recruit the second Cdt1- bound MCM2-7, leading to the eventual assembly of the MCM2–7 double hexamer.*

The formation of the pre-RC is a tightly controlled process, and it is restricted to G1 phase to prevent any origin of replication from triggering DNA replication more than once in each cell cycle. This regulation varies among organisms, and it is mediated by the inhibition of loading of the double helicase hexamer outside of G1 phase [193]. In human cells, the expression of Cdt1 is tightly regulated throughout the cell cycle, being the protein levels high in G1 phase and low in S phase. In addition, Cdt1 is inhibited by direct binding of Geminin [194], a substrate that mediates the proteolysis of Cdt1 through ubiquitination and accumulates during the S, G2 and M phases, being destabilized during G1 phase. This tight regulation ensures thereby that the MCM complex is recruited only in G1 phase, and that replication occurs only once during the cell cycle [195], [196].

2.3. Step 3: DNA replication activation and fork progression

After the pre-RC is assembled and the origins are licensed, they remain inactive until some, in an apparent stochastic manner [180], [197], will trigger the activation of DNA replication during S phase. What determines which origins will be activated and which remain dormant is still not well understood. Depending on when during S phase they are activated, origins can be classified as early-, mid-, or late-replicating. Depending on their use, they fall into three classes: constitutive, dormant or inactive, and flexible [198]. Constitutive origins are those that are always located in the same position and are always activated in any cell cycle; they are the minority of the classes. Flexible origins are licensed origins that will potentially be activated stochastically or under specific conditions that affect S phase, such as DNA damage or changes in growth conditions. Finally, inactive or dormant origins are licensed origins that are usually never used in normal conditions, but DNA replication can be triggered from them in specific cell programmes or under stress conditions.

The proportion of licensed origins that remain dormant varies substantially depending on the organism. For example, in fission yeast, slightly less than 50% of the origins are considered constitutive while the rest are mostly inactive and are replicated passively by forks that elongate from actively replicating origins [199]. In contrast, in higher eukaryotes, like humans, only between 5-20% of the licensed origins are activated in an seemingly stochastic way and they are localized in efficient and constitutive replication initiation "zones" rather than in specific origin "sites" [197], [200].

The activation of the MCM helicase at the at the G1-S phase transition relies on the activity of two critical cell-cycle dependent protein kinases: the S-phase cyclin-dependent kinases (CDK) and the Cdc7-Dbf4 (DDK), who both bind directly and phosphorylate proteins from the pre-RC like the MCMs [201]–[204]. Cdc7-Dbf4 activation is required for the loading of additional essential factors to the pre-RC like the Cdc45, which interacts with MCM2 and Sld3 [205]. Then, the phosphorylation of the latter along with Sld2 by CDK creates a binding site for Dpb11. This in turn serves as an anchor for the Replication Protein A (RPA) and the GINS complex (go-ichi-ni-san, composed of the proteins Sld5, Psf1, Psf2 and Psf3) to establish the two CMG complexes (Cdc45, MCM2-7, GINS) that encircle the double-stranded origin DNA and together form the pre-Initiation Complex (pre-IC) [201]. Finally, MCM10 binding is required as it binds to DNA and to the N-terminal regions of MCM2 and 6, disrupting the tight interaction between the two MCM2-7 hexamers [206]. This, along with the ATP hydrolysis by CMG, provides sufficient force to untwist open several turns of the dsDNA, leading to the extrusion of the lagging strand from the central channel and enabling the CMGs to transition into encircling only the leading strand ssDNA [201] (Fig. 11). Following the unwinding process of DNA by CMG-MCM10, the two CMG helicases translocate and pass each other with their N-terminal ring ahead and the C-terminal ring pushing from behind through ATP hydrolysis as a motor in 3' → 5' direction, migrating away from each other, establishing the bidirectional replication forks [207], [208].

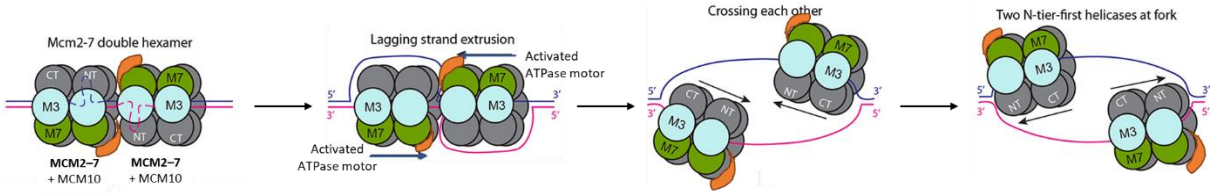


Figure 11. Bypass of the two helicases to start replication*. The lagging strand is extruded from the central channel of each helicase to be able to pass each other because the eukaryotic CMG helicase translocate with the N-terminal region first on the leading strand DNA. Cdc45 and GINS are omitted but they form the CGM complex with MCM2-7. MCM2–7 hexamers are shown in grey, light blue and green, and MCM10 is shown in orange. Adapted from* [192].

After origin unwinding, the Replication Protein A (RPA) associates with the ssDNA to stabilize it and it is required for loading of the DNA polymerase α machinery. In eukaryotic DNA replication, DNA polymerases (Pol) cannot initiate new chains of nucleic acids without the presence of an RNA primer. Therefore, a DNA-dependent RNA polymerase is required. The DNA polymerase α-primase complex

(Pol α/primase) binds to ssDNA and synthesizes a short (~12 nucleotides) RNA primer that later extends with ~25 nucleotides of DNA to form a hybrid RNA/DNA primer. The Pol α/primase acts once on the leading strand and repeatedly primes the lagging strand (Fig. 12) [209]. This, in turn, allows for the recruitment of the replication factor C (RFC), which loads the proliferating cell nuclear antigen (PCNA), that forms a ring around the DNA in an ATP-depentant manner [210]. PCNA recruits DNA polymerases (Pol δ and Pol ε) and tethers them to the template DNA progressing as a sliding clamp through the DNA molecule [211]. After the priming, the leading strand is replicated by Pol ε and the lagging strand is elongated with Pol δ. Pol ε binds directly to the CMG helicase and forms the leading strand replisome synthesizing continuously the complementary strand of DNA [212], [213]. Pol δ extends the lagging-strand primers discontinuously forming the Okazaki fragments until it reaches the downstream RNA primer (Fig. 12). Then, it is recycled to the RNA primer of the next Okazaki fragment and starts synthesizing again [214], [215]. PCNA also helps organize Okazaki fragment maturation through promoting the activity of the flap-endonuclease FEN1, which removes the RNA primers and creates ligatable nicks that will be resolved by the DNA ligase afterwards [216], [217]. Similarly to FEN1, the RNAse H protein also degrades the RNA moiety in RNA-DNA hybrids, thus, participating in the removal of the RNA primers of Okazaki fragments [218].

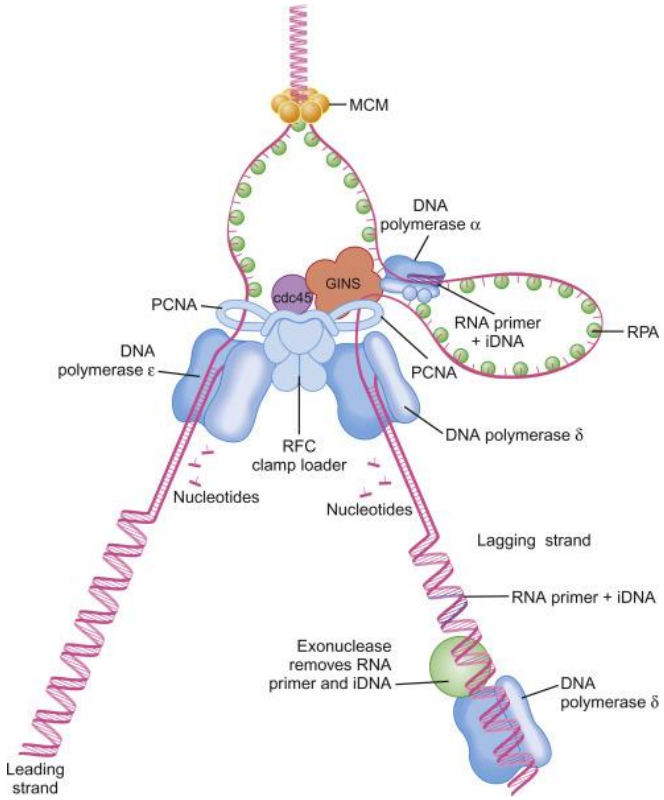


Figure 12. Eukaryotic replisome [219]. *The eukaryotic replication fork is generated when the GMC complex helicase unwinds the two strands of DNA, and the RPA protein coats ssDNA to prevent reannealing. Next, the DNA polymerase α complex makes an RNA primer followed by initiator DNA (iDNA). Then, the RFC clamp-loader assembles the sliding clamp (PCNA) around the DNA and recruits DNA polymerases. DNA polymerase ε is loaded onto the leading strand, and DNA polymerase δ onto the lagging strand. On the lagging strand, sections of RNA primer followed by iDNA are removed and then replaced with new DNA by DNA polymerase δ.*

During the progression of replication forks, as helicases unwind the parental duplex DNA, this leads to overwinding of the unreplicated DNA and to the formation of positive supercoiling ahead of the moving fork [220]. If unresolved, supercoils accumulate and generate topological stress, hampering replication progression. There are two ways to dissipate supercoils: by topoisomerases, that create a transient nick on one of the DNA strands (or both) and pass the complementary one (or the full helix) through it [221]; or by the generation of pre-catenanes, that are formed when the entire fork rotates clockwise counteracting the overwinding of unreplicated DNA [222].

The rate at which individual replication forks synthesize DNA and the total number of origins that are activated in each cell cycle determines the overall duration of the cell cycle. Generally, eukaryotic cells show a relatively constant average fork speed ranging from 1-2 kb per minute, varying depending on the species, cell type and S-phase progression stage. Human cell replication is coordinated between a variable fork progression, with an average fork speed of 1.46 kb/min, and a variable origin spacing, as activation of additional origins compensate any possible decrease in replication fork speed due to fork stalling or other factors [223]. In the budding yeast *Saccharomyces cerevisiae*, replication forks advance at a relatively steady pace of 2 kb/min, and they are significantly slowed down at known replication fork barriers like tRNA genes (their high transcription rate can cause transcription-replication conflicts [224], [225]) or regions like centromeres and telomeres [5].

There are several known natural impediments to DNA replication in eukaryotic and prokaryotic genomes that can lead to inhibition of replication and genomic instability. These obstacles include unusual DNA secondary structures, slow replication zones, DNA binding proteins and clashes with transcriptional machinery in highly transcribed genes. One example of a DNA secondary structure that may pose a challenge to replication in eukaryotes is the G-quadruplexes. These are formed in G-rich regions from ssDNA. The action of unwinding the double stranded DNA by the CMG helicase leaves portions of the lagging strand template become single stranded, providing a chance for the formation of ssDNA secondary structures like hairpin configurations or G-quadruplexes, although they can also occur in the leading strand. These structures could interfere with the progression of the DNA polymerase, ultimately leading to the stalling of the whole replication fork [226], [227]. Consequently, replication fork speed is known to slow down at specific regions of the chromosomes displaying an overrepresentation of G-rich repeats prone to form G-quadruplexes structures like centromeres and telomeres. In addition, telomeres pose challenges to replication fork progress because subtelomeric sequences are very repetitive and are very enriched in heterochromatin (and heterochromatin binding proteins), which can obstruct the passage of the replisome [228]. The centromeric DNA mediates the attachment of the spindle microtubules and promote the kinetochore formation, a structure that may

also present a significant obstacle to replisome progression in addition to the presence of the specific centromeric histones and heterochromatic histone marks [229]. Apart from chromosomal structures, there are genomic locations that also constitute replication fork barriers (RFB): the multicopy tandem repeats of ribosomal DNA (rDNA) separated by non-transcribed spacers that host specific replication blocks [230]; and tRNA genes in the case of yeast, as these are an extremely highly transcribed set of genes and the presence of the RNA polymerase machinery represents an obstacle that may hamper replication fork progression [224], [225].

Conflicts between the transcription and replication machinery are a significant source of genomic instability. Large protein complexes perform both processes by moving processively along the genome; and, sometimes, both machineries may move in opposite directions, resulting in head-on collisions. Sometimes, even if they move in the same direction, they do so at different velocities, so co-directional collisions can also occur [227]. In actively transcribed loci, upon encountering the transcription machinery, replication forks pause and may eventually collapse, leading to the activation of DNA damage response pathways and genomic instability.

As replication forks progress through the genome and the replisome machinery advances, the nucleosomes ahead of the DNA replication fork are disassembled and removed from the DNA to facilitate duplex unwinding. Then, a mix of newly synthesized and recycled histones is assembled into nucleosomes on the daughter DNA strands in a process facilitated by the replisome components themselves and histone chaperones [231]. Finally, the epigenetic marks present on the pre-existing histones are restored in the new ones by the lysine methyl-transferases (KMT) and demethylases (KMD) [232].

2.4. Step 4: DNA replication termination

Replication forks advance in opposite directions from the origin of replication and progress through the DNA molecule until two converging replication forks coming from neighbour origins meet. This process is known as DNA replication termination. It involves the disassembly of the replisome and the synthesis completion and resolution of the daughter DNA molecules.

During replication fork convergence the length of unreplicated DNA between them becomes shorter, leaving less room to form supercoils and less chances for the topoisomerases to act to resolve them. As the converging replisomes approach one another, they depend on the formation of pre-catenanes to manage the topological stress and this could imply a gradual slowdown in replication fork progression, or require additional factors [233]. When converging replication forks encounter, CMG

helicases associated with their cognate leading strand bypass each other (without leading to fork stalling), progressing through the unwound ssDNA until they reach the dsDNA of the downstream Okazaki fragment (Fig. 13). The replisome passes over this ssDNA-dsDNA junction by translocating along the double strand without further unwinding to allow the processing of the final Okazaki fragment by Pol δ and FEN1 [217]. When the last fragment of parental DNA is replicated, a final catenane is formed and will later be resolved by topoisomerases after the disassembly of the replisome [234].

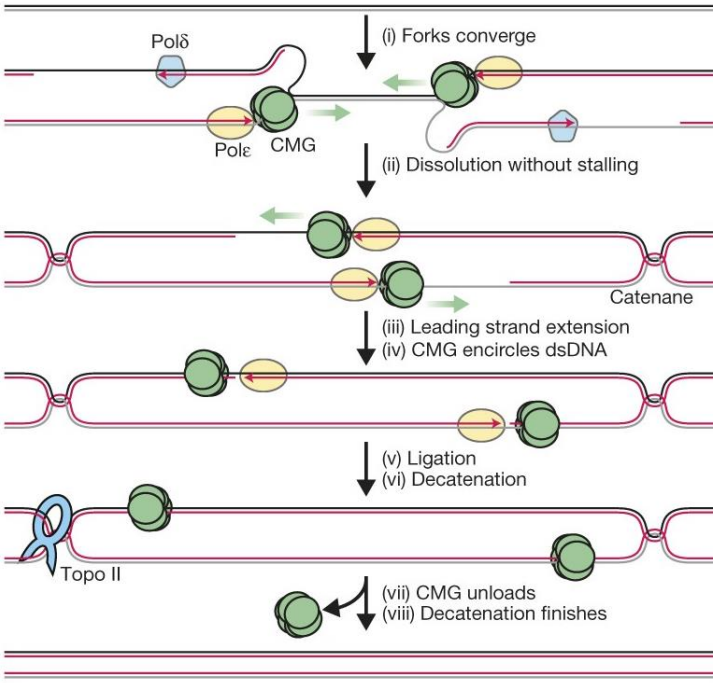


Figure 13. Replication termination mechanism in human cells. *Leading strand CGM complexes pass each other unhindered and remain associated with DNA, and they are disassembled only after the leading strand of one fork is ligated to the lagging strand of the opposing fork. Finally, decatenation is accomplished by topoisomerase II.* [235].

Finally, replisomes are actively dissociated from the DNA in a process termed disassembly, that involves the polyubiquitination of the MCM7 subunit and the removal of the CMG complex by an ATPase [236]. The reminder replisome components are then passively disassembled from the DNA as an indirect consequence of CMG unloading.

The termination sites are not sequence specific and their distribution is generally defined as the midpoints between the two adjacent replication origins, with their relative firing timing being key to define the termination areas. Interestingly, termination regions (TER) have been described in bacteria and budding yeast and they contain fork pausing elements that can influence and hamper fork progression and merging [234].

3. DNA replication in Apicomplexa

As obligate intracellular parasites, Apicomplexa rely on highly flexible and adaptable replication strategies to multiply within the hosts and generate progeny. The cell division programs change between species, involving different mechanisms of DNA replication, nuclear division and cytokinesis that produce in the end between two and several thousand progeny cells [237]. Apicomplexan parasites are divergent eukaryotes, as they branched very early during evolution. Therefore, conventional eukaryotic DNA replication principles might not be appropriate to describe them.

3.1. Mitosis in Plasmodium falciparum

During their intricate life cycle, they undergo multiple rounds of extreme population expansion within short periods of time in both their vertebrate hosts and anopheline mosquito vectors. There are four critical steps during their life cycle where parasites must rapidly multiply to generate abundant progeny. Two of them take place within the human host (liver stage development and intraerythrocytic reproduction), and two of them within the mosquito vector (male gamete formation and sporozoite formation).

3.1.1. Hepatic stage

Following an infectious mosquito bite, a small number of the inoculated parasites reaches the host liver for further development. There, each sporozoite inside a hepatocyte will first replicate into a trophozoite and continue undergoing 13 to 14 rounds of successive DNA replication and karyokinesis until an hepatic schizont is formed containing tens of thousands of nuclei. Then, from the surface of the infected hepatocyte, a vesicle termed merosome containing up to 90,000 haploid exoerythrocytic merozoites assembles and is eventually released into the bloodstream where the merozoites will invade new RBCs to continue with the blood-stage schizogony.

During liver stage schizogony, *Plasmodium* parasites display several mechanisms such as nutrient scavenging from the host hepatocytes and stage-specific protein expression patterns that allow them to massively proliferate and produce high numbers of daughter cells [130], [238], [239].

3.1.2. Blood stage schizogony

Once inside the RBCs, the merozoites begin the intra-erythrocytic developmental cycle (IDC), where they begin as a ring stage, progress into a trophozoite and then undergo alternating rounds of asynchronous DNA replication and nuclear division to form a schizont with a variable number of nuclei

(16 to 32) (Fig. 14) [2]. Finally, a single mass cytokinesis step gives rise to up to 32 haploid daughter cells that will be released after the rupture of the host erythrocyte. The ability of the parasites to replicate their genome and multiply inside RBCs is liked to their pathogenicity, as all the clinical symptoms from malaria are associated with the blood stage of the parasite's life cycle. In fact, one of the hallmarks of malaria is the cyclic manifestation of chills and fevers, which is a consequence of the cyclic rounds of synchronous schizont maturation, rupture of infected cells and, hence, massive release of parasites into the bloodstream [240].

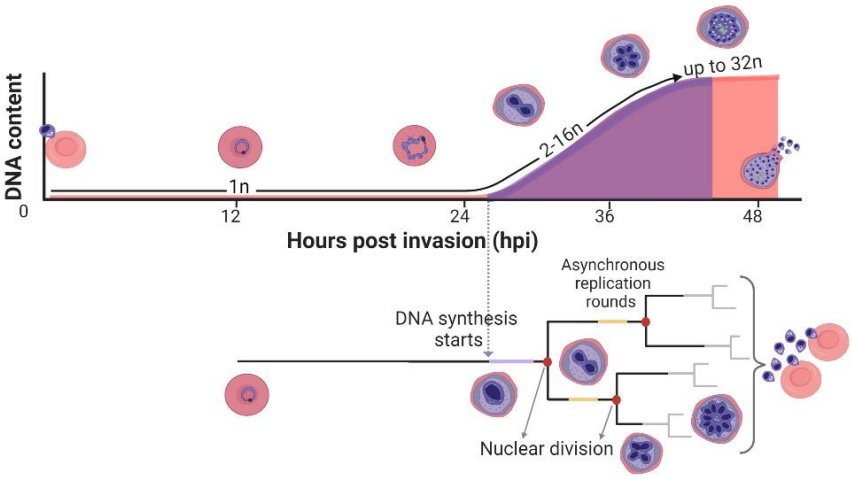


Figure 14. Progression of DNA content variation through the P. falciparum cell division cycle. *A haploid (1n) merozoite inside the red blood cell (RBC) develops as the "ring" form from 0h to about 24 h post invasion. As the parasite transitions from rings to trophozoites, it undergoes a first round of DNA replication and nuclear division. Then, multiple asynchronous DNA replication rounds followed by nuclear division, but without cytokinesis, produce the multinucleated schizont containing 16 to 32 nuclei. A final mass cytokinetic event releases the haploid daughter cells (merozoites) to the bloodstream. Done with Biorender.com*

A study characterizing DNA replication dynamics during *P. falciparum* IDC [2] has shown that the first round of DNA replication and nuclear division is slower (length of the S-phase is longer) and as schizogony progresses the duration of the S-phases is shorter. In fact, longer first division cycles have been shown to lead to a delayed end of schizont formation, suggesting that the number of nuclei formed are what drives the progression through schizogony, regardless of the time needed to reach a certain number of daughter nuclei [2]. Finally, the asynchronicity of DNA replication of the different nuclei as schizogony proceeds suggests the presence of a limiting factor that has to be shared between the simultaneously dividing nuclei.

In contrast, other studies involving DNA combing experiments or immunofluorescence microscopy [113], [241] have concluded that the S-phase duration is lower in the initial DNA replication rounds and it increases throughout schizogony, potentially due to shared factors between nuclei that become increasingly limiting in advanced stages of schizogony, when several nuclei are replicating simultaneously. These findings enhance the need for the investigation of DNA replication dynamics over the course of *P. falciparum* schizogony.

3.1.3. Replication in male gametes

To ensure transmission, *P. falciparum* parasites within RBCs differentiate into male micro- and female macrogametocytes which are ingested by mosquitoes during a blood meal. Female macrogametocytes go through several morphological changes, but not replication, to form the macrogamete. In the case of the male microgametocytes, once in the mosquito gut, they undergo three rounds of DNA replication and mitosis in under 15 minutes to form eight haploid male microgametes. This extremely fast mechanism of cell multiplication is highly regulated under a specific cascade of kinases (like the Ca^{2+}-dependent protein kinase 4, CDPK4) that phosphorylate proteins related to different cell cycle phases almost simultaneously [242]–[244].

3.1.4. Sporozoite formation

Within the mosquito midgut, a subset of male and female gametes will fuse and create diploid zygotes. Zygotes then develop into motile ookinetes that migrate to the midgut epithelium and differentiate into oocysts. Oocysts then undergo 10 to 11 rounds of DNA replication to generate thousands of haploid sporozoites in an extracellular process that in *P. falciparum* lasts around 10-15 days known as sporogony [20], [245], [246].

3.2. DNA replication machinery in P. falciparum

The core eukaryotic DNA synthesis associated factors are conserved in *P. falciparum* parasites, including all subunits of the MCM helicase complex [247], DNA polymerases α, δ and ε [248], [249], the replication protein A (RPA) and the PCNA1 sliding clamp [250], [251]. Interestingly, a *P. falciparum* specific additional copy of PCNA (*Pf*PCNA2) is also present, although it is not involved in DNA replication and participates mainly in the DNA damage response pathway [250]. From the origin recognition complex (ORC), homologues of ORC1, ORC2 and ORC5 subunits have been found [252]–[255]; as well as a putative ORC3 domain-containing protein (PF3D7_1029900) and a putative ORC4 (PF3D7_1334100) [244]. Besides, there is a *Plasmodium* CDK homolog, the Cdc2-related protein kinase-1 (Crk-1) [256], [257], a putative Cdt1-like protein (PF3D7_1343300), and several *P. falciparum* proteins have been predicted to be putative equivalents of subunits of the GINS complex based on sequence homology and structural similarity [258]. *P. falciparum* replication machinery lacks homologues of essential eukaryotic factors like Cdc6 (although *Pf*ORC1 C-terminal domain shows homology with Cdc6 and displays ATPase activity [254]), Geminin, Cdc45 and DDK-kinases, which suggests either the presence of additional parasite specific replication proteins or a less complex mechanism of DNA replication initiation than model eukaryotes.

Most of the members of the replisome that have been found in *P. falciparum* have been reported to be essential in both an *in vitro* and *in vivo* essentiality screen [259], [260], enhancing the importance of DNA replication in the establishment of infection. These essential factors are thus potential candidate targets for drug development aiming to block proliferation of the parasites in their blood stage.

The unusual cell division cycles in *Plasmodium falciparum* must be accurately controlled in a temporal and spatial manner, which calls for both global and local specific regulators that may also be unusual. The repertoire of regulatory factors are involved in multiple mechanisms of parasite development and include CDKs (PK5 and PK6 are involved in initiation of DNA replication during the IDC [261], and Mrk1 is required for cytokinesis [262]); CDK-related kinases (CRKs, Crk1 and Crk3 have roles in transcriptional regulation [256], whereas Crk4 has an essential role in schizogony [4]) and cyclins (Cyc1 is involved in schizont segmentation, Cyc3 in oocyst formation and Cyc4 in Crk-5 activation [263]).

In model eukaryotic organisms, there are certain established checkpoints that regulate cell cycle progression in addition to the cyclin-CDK regulatory pathways. They ensure that each cell is replicated only once and act as quality control during the different phases detecting: appropriate cell development (G1/S boundary checkpoint), effective DNA replication or presence of DNA damage (S/G2 boundary checkpoint), and correct chromosomal attachment to the spindle (G2/M boundary checkpoint) [264]. If the conditions for successful cell division are not met, progress will be halted, and the cell cycle is interrupted through the activation of specific cell cycle inhibitors. Through these mechanisms the cells prevent the spread of damaged or imperfectly replicated daughter genomes.

The existence of such checkpoints in *Plasmodium falciparum* is still not clear, as homologues of checkpoint proteins such as p53 [265], [266], Rb (retinoblastoma tumour suppressor family) [267], ATM (Ataxia Telangiectasia Mutated) and ATR (Ataxia Telangiectasia and Rad3-Related) [268] have not yet been identified. However, *P. falciparum* parasites enter a state of dormancy when exposed to nutrient starvation or drug pressure, that is reversible [269], [270], suggesting the presence of potential sort of checkpoints that induce cell cycle blockage under unfavourable conditions. In fact, in the related apicomplexan parasite *T. gondii* tachyzoites, progression through G1 phase is regulated by the Cdk-related kinase *Tg*Crk2, whereas the licensing of DNA replication in S phase is controlled by *Tg*Crk5, and the spindle microtubules assembly at the metaphase to anaphase transition in M phase is regulated by *Tg*Crk6 [271]. In *P. falciparum*, *Pf*Crk4 has been suggested to be the master regulator that controls progression through the cell cycle during schizogony [4] by phosphoregulating proteins involved in firing of origins of replication at each of the consecutive rounds of genome duplication. However, robust evidence for cell cycle checkpoints in *P. falciparum* IDC remains elusive and further

investigations need to be done to elucidate the regulation of the parasite's cell cycle as well as how it might be affected by host extrinsic factors.

3.3. DNA replication as druggable target

DNA replication is a process transversal to all cells and organisms as it is essential for cellular proliferation. Failure to accurately duplicate their genome leads to generation of imperfect progeny cells that might not be viable or display malignant phenotypes. As DNA replication stress is a major driver of genome stability in cancer cells, proteins involved in response to deleterious replication are a common target for inhibitors used in cancer therapy. Replication stress-inducing anticancer therapies include topoisomerase inhibitors, alkylating agents, platinum compounds, and poly ADP-ribose polymerase (PARP) inhibitors [272], [273].

In addition, DNA replication is a targeted mechanism for the development of antibiotics to inhibit bacterial proliferation (Fig. 15). The most common clinically used antimicrobials are fluoroquinolones, that inhibit DNA replication by selectively inhibiting topoisomerases and/or gyrases, leading to an increased supercoiling of the DNA helix and inability to decatenate it [274]. There are other DNA replication inhibitors with potential to be used as antimicrobials: DNA ligase inhibitors (adenosine analogues), bacterial DNA polymerase inhibitors (guanine inhibitors), inhibitors of the sliding clamp loader (griselimycins) and inhibitors of the stabilizer SSB protein (Fig. 15). The increasing resistance of bacteria to commonly used antibiotics enhances the need for developing novel antimicrobial drugs, and the DNA replication pathway provides a promising range of essential proteins that may serve as potential targets for new antibiotics.

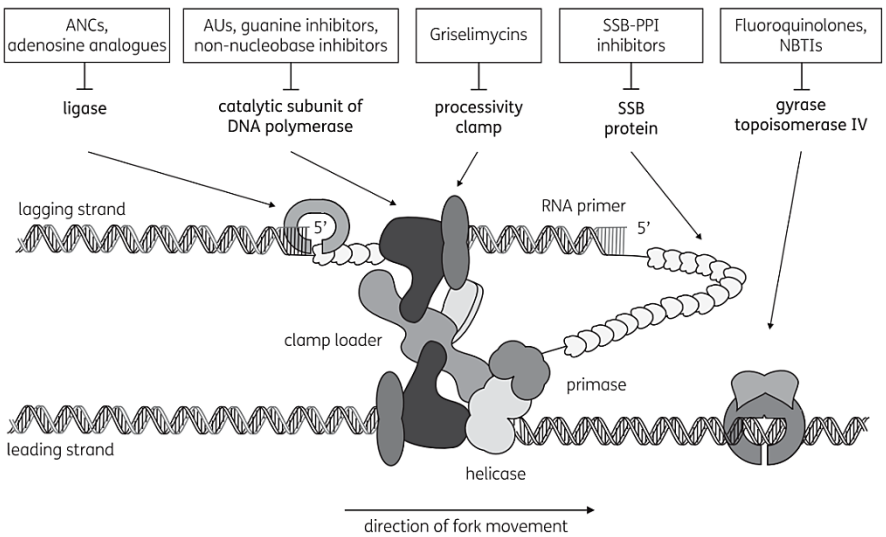


Figure 15. Drugs targeting microbial DNA replication [274]. *The core of the replisome machinery and the other proteins targeted by antimicrobial compounds are shown. Important classes of drugs inhibiting specific proteins are boxed. PPI, protein–protein interaction.*

3.4. DNA replication in P. falciparum: a potential drug target

The presence of unique replication pathways or proteins not found in the human host presents promising targets for developing novel antimalarial compounds. In fact, clinically used antibiotics targeting bacterial replication have been shown to inhibit DNA replication of *Plasmodium* organelles with prokaryote origin, like the apicoplast. For example, by targeting the *P. falciparum* gyrase from the apicoplast, fluoroquinolone antibiotics inhibit its ATPase activity, hindering parasite growth *in vitro* [104]. As the proteins involved in DNA replication or repair of the apicoplast have no orthologs in the human host cells, they pose a perfect target for new antimalarials development. Similarly, proteins unique to *P. falciparum* with no homology to their human counterparts offer a whole range of potentially druggable targets. Further identification and characterization of the *P. falciparum* replication components will improve our understanding of the mechanisms driving parasite proliferation and potentially provide us with new candidates for novel drug development, which is becoming increasingly urgent as new strains appear resistant to current antimalarials.

As an essential process for the development of the parasite and establishment of infection, DNA replication has historically been targeted for antimalarial drug development. For instance, anti-folate drugs like pyrimethamine target DNA replication by blocking the production of reduced folate cofactors, which are essential for nucleotide production. However, resistance to these inhibitors caused by mutations in the targeted enzymes is spread worldwide [275]. One example of an antimalarial drug that indirectly induces DNA damage is the frontline antimalarial Artemisinin through the production of free radicals, but has been affected by resistance [95]. Drugs targeting key cell cycle regulators and DNA repair mechanisms in *P. falciparum* have the potential to be highly effective. For example, the *Pf*PK5 inhibitor flavopiridol leads to a decrease in DNA replication in *P. falciparum*. Besides, several CDK inhibitors like Roscovitine, olomoucine or the DHFR inhibitor WR99210 have shown potential inhibitory effects *in vitro* on cell cycle progression of parasites that were treated with artemisinin and entered dormancy, likely preventing reactivation following artemisinin treatment [276].

4. Methods to map origins of replication throughout the genome

To understand the regulation of DNA replication in malaria parasites, one crucial step is to identify the set of DNA replication origins in the genome that will drive the multiple S phases in each intra-erythrocytic cycle. Their distribution throughout the chromosomes and variable firing efficiency will determine the duration of each replication round and, thus, the overall duration of the cell cycle. A

variety of experimental approaches have been used to characterise the nature and positioning of replication origins in other eukaryotes such as yeast [10], [199], *Drosophila* [163], [181] or mammalian genomes [161], [197]. Most of the methods developed to map initiation of replication are population-based methods, where the whole cell lysate is subjected to the same experimental protocol and, thus, the replication origins detected are not necessarily found in all the cells. Still, they provide reproducible and robust spatial and temporal replication origin patterns.

4.1. Isolation of replisome members or intermediates

Replication origin mapping can be achieved directly by isolating key replication components or replication intermediates. One example of this is the mapping of the binding sites of pre-RC components, such as the ORC or MCM complex, in a technique known as ChIP-seq (Chromatin Immunoprecipitation followed by sequencing) [277], [278]. By pooling down individual members of these complexes along with the DNA that they are bound to (after crosslinking the proteins and DNA), we can isolate those DNA fragments and sequence them to map the location of the binding sites of these proteins. In the case of ChIP-seq of ORC proteins, what is obtained is a cartography of the licensed initiation sites, which may or may not be later activated.

An additional replication intermediate structure that can be isolated for further mapping is the whole replicative bubble, with the Bubble-seq method [178]. Here, the circular replication bubble containing the two diverging forks is trapped in agarose gels and isolated for sequencing.

4.2. Mapping of newly synthesized DNA

Active DNA synthesis in proliferating cells can be mapped through isolating nascent strands from active replication forks or mapping the incorporation of nucleoside analogues in newly synthesized DNA. SNS-seq (Short Nascent Strands sequencing) [161] relies on the isolation and purification of the RNA-primed nascent strands of DNA during initiation of DNA synthesis in S-phase. This is achieved through a purification step with λ-exonuclease, which digests all DNA that does not contain the RNA primer in their 5' end (because only the newly synthesized DNA strands will contain this RNA cap). The short DNA strands are then sequenced and mapped, allowing for the generation of a genome-wide map of ORIs that were activated during replication.

In addition, active origins can be studied by mapping the incorporation of nucleoside analogues (like BrdU or EdU) into the DNA as replication progresses. FORK-seq/NanoForkspeed and DNAscent [5], [10], [279] use thymidine analogues like BrdU (5-Bromo-2-deoxyuridine), IdU (5-iodo-2-deoxyuridine) or EdU (5-ethynyl-2-deoxyuridine), which are added to the parasites in pulse-chase experiments and

later detected by Oxford nanopore sequencing technology (ONT). ONT relies on the passage of a ssDNA molecule through nanopores embedded in an electro-resistant membrane. Each nanopore is connected to a sensor chip, which measures the electric current that flows through it. When each nucleotide of DNA passes through the nanopore, the current is altered to produce a characteristic disruption that is then decoded using basecalling algorithms to determine the DNA sequence in real time [280]. Then, specific algorithms can be used for mapping the incorporation of the thymidine analogues, allowing the detection of active replication forks and replication initiation sites [5], [279]. The advantage of this long read sequencing technology is that it confers single molecule resolution, because when several replication forks and initiation sites are mapped from a single sequencing read, this means that these events took place in the same DNA molecule, *i.e.,* in the same nucleus.

Finally, origin efficiency and the replication timing program can be studied with Repli-seq [281], a technique in which actively replicating cells are labelled with BrdU and the BrdU-immunoprecipitated DNA is measured and sequenced. If the labelling is done at different stages of S-phase, early replicating or late replicating origins can be mapped, and assessment of differences in origin efficiency can be done. This method can even be used in a single cell approach (scRepli-seq [282]) to obtain a map of replication timing at the single molecule level.

4.3. Mapping replication fork progression

The actively replicating forks can be analysed to measure DNA synthesis directionality, replication fork speed as well as to identify the sites of initiation of replication. DNA combing or fibre spreading allow the analysis of DNA replication at the level of individual single molecules that are stretched along glass slides [283]. Thanks to the usage of clickable nucleoside analogues like EdU, CldU or IdU, their incorporation into DNA can be observed by immunofluorescence. These methods allow the direct observation of newly synthesized DNA along individual molecules and measure replication fork directionality and speed. Although the genomic coordinates of initiation events cannot be mapped, fibre stretching methods allow to discriminate between elongating forks and initiation or termination events.

Objectives of this thesis

Plasmodium falciparum parasites rely on effective and accurate DNA replication to ensure growth and multiplication inside the human host. The mechanisms and factors involved in genome replication are well described both for prokaryotic and eukaryotic cells. However, many of these elements have not been found in *Plasmodium* and the DNA replication strategy is therefore not completely understood.

The main objective of my PhD project is to shed light on key aspects of the initiation of DNA replication in *Plasmodium falciparum,* namely the origins of replication and the composition of the replicative complex. To achieve this goal, my thesis project focuses on the following aims:

1. Study of the genetic landscape of origins of replication:
    a. Mapping of potential sites of initiation of DNA replication
    b. Mapping of active origins of replication during schizogony
    c. Exploration of genomic determinants of origin specification
2. Determination of the composition of the replicative complex:
    a. Establish the nascent DNA and replisome interactome
    b. Investigate the role of non-canonical components of the replicative complex

Materials & Methods

1. Parasitology

1.1. Plasmodium parasites in vitro culture

P. falciparum 3D7 strain and transgenic lines were cultured in human erythrocytes (type O or A) in a 5% haematocrit suspension in RPMI 1640 medium (Gibco Life Technologies, 52400 RPMI 1640, 25 mM HEPES) supplemented with 50 mg/L hypoxanthine (C.C.Pro GmbH), 5% AB+ human serum and 0.5% Albumax and 10 µg/mL gentamicin (Sigma). Cultures were kept at 37ºC under a controlled trigaz atmosphere containing 5% O_2, 5% CO_2 and 90% N_2.

Parasite development was monitored on Giemsa stained thin-blood smears.

1.1.1. Blood washing

Blood was acquired from the French blood bank (Etablissement Français du sang, 21PLER2018-0057). Before using it for parasite culture, the blood was washed twice with RPMI medium supplemented with gentamicin (termed washing media). Washes were performed by diluting the blood 1:3 in washing media followed by a centrifugation step (5 min, 1811 g). Then, the buffy coat was removed by aspiration and the blood was washed a second time. Finally, a 50% blood suspension was prepared in washing media and used to supplement parasite cultures at a final concentration of 5% (*i.e.* 5% haematocrit).

1.1.2. Thawing

Glycerol frozen parasites were warmed for 30 seconds in a 37ºC water bath. Then, 0.2V of pre-warmed 12% NaCl was added slowly and incubated for 5 min at room temperature (RT). Next, 10V of pre-warmed 1.6% NaCl was added dropwise and then centrifuged for 5 minutes at 340 g. Parasites were then washed in pre-warmed washing media once and resuspended in culture media. Blood was added to reach the desired haematocrit [284].

1.1.3. Freezing

In a culture of asexual blood-stage parasites, only the ring stages can be frozen by current methods. A culture of young parasites (< 10 hpi) between 3-8% parasitemia was centrifuged for 5 minutes at 340 g and then pre-warmed freezing solution (6.2 M Glycerol, 0.14 M sodium lactate, 5 mM KCl in PBS, final pH= 7.2) was added dropwise in three rounds: 0.4V, 1.2V and 2.4V, each followed by a 5 min incubation at RT. Aliquots of 1 mL of the above mixture were stored in cryovials at -80ºC.

1.1.4. Synchronization

1.1.4.1. Percoll gradient

This procedure uses a density gradient to selectively isolate RBCs infected with late-stage parasites which are less dense than uninfected RBCs or RBCs infected with younger parasites. Pelleted RBCs infected with late-stage parasites (segmented schizont stages) were isolated on a density gradient of 63% percoll. Up to 1 mL of material was used per 4 mL of percoll solution. The gradient was spun at 1360 g for 11 min (acceleration = 4, brake = 2) and the layer of schizonts present at the interface between media/percoll was collected and washed with washing media. The isolated RBCs containing the late schizonts were then transferred into a new culture plate with fresh media and blood where parasites were allowed to reinvade new RBCs.

1.1.4.2. Sorbitol treatment

The sorbitol treatment allows the selection of young stage parasites as it induces osmotic lysis of RBCs infected with later stages due to presence of permeation pathways in the membranes of these cells [285]. A culture of young ring stage parasites was centrifuged, and the pellets were mixed vigorously with 9V of warm 5% D-Sorbitol prepared in water. This suspension was incubated for 5 min at 37ºC and then centrifuged for 5 minutes at 340 g. Cells were washed once with washing media and resuspended into the new culture media.

To obtain a semi-synchronous culture I combined both approaches. Specifically, after obtaining a suspension of mature parasites using the percoll method, parasites were allowed to reinvade in fresh RBCs for a period ranging from 1h to 3h, depending on the experiment, and then a sorbitol treatment would selectively lyse the RBCs containing schizonts that had not reinvaded in that time window. This allowed studies on age-matched parasites cultures.

1.2. Generation of transgenic lines

1.2.1. Genomic DNA extraction

Before extraction, erythrocytes were lysed in a solution of 0.15 % saponin (Sigma) in PBS for 5 min on ice, and then parasites were harvested by centrifugation (3250 g at 4ºC, 5 min). Then, genomic DNA used for PCR amplification for cloning or genotyping purposes was extracted using the NucleoSpin Blood kit (Macherey Nagel).

1.2.2. Generation of targeting vectors

The genomic DNA used in all cloning reactions was extracted from the wild type 3D7 *P. falciparum* strain. PCR reactions intended for molecular cloning used the KAPA HiFi polymerase (Roche), following the manufacturer's recommendations but using a modified cycling programme: 95ºC 3 min / 25-30x 98ºC 20 secs, 55ºC 30 secs, 72ºC 1 min per kb / 5 min 72ºC.

1.2.2.1. CRISPR/Cas9 approach

Several plasmids were constructed and used during my project following the same strategy. To introduce a triple haemagglutinin (HA_3) tag and/or the biotin ligase sequence (BirA*) at the C terminus of the protein of interest, two donor sequences were cloned. They targeted the region immediately upstream (homology region 1; HR1), and downstream of the stop codon (homology region 2; HR2). Both HRs were cloned sequentially into the plasmid pLN-HAx3 by In-Fusion HD cloning (Clontech) following digestion with AvrII and AflII/NaeI (HR1) or AflII and BamHI (HR2). At each cloning stage bacterial colonies were screened using the GoTaq Green DNA Polymerase master mix (Promega) following the manufacturer's instructions. The cloned regions were verified in the colonies screed by Sanger sequencing (Eurofins).

Guide RNAs were ordered as oligonucleotides, annealed with a descending temperature gradient, phosphorylated with T4 PNK for 30 min at 37ºC and cloned into the pDC2-Cas9-hDHFR-yFCU plasmid [286] after BbsI digestion using T4 DNA ligase (ligation reaction performed overnight at 16ºC). All restriction enzymes, T4 DNA ligase and T4 PNK were purchased from NEB. The sgRNAs targeting the 3' end of the genes were selected and the sgRNA sequences were cloned individually into the pDC2-Cas9-hDHFR-yFCU plasmid [286], which contains the drug resistance marker (human Dihydrofolate Reductase (hDHFR), conferring resistance to WR99210), along with the Cas9 endonuclease sequence and the U6 promoter. This and the plasmid containing the HRs of the target gene were transfected simultaneously to obtain a scarless insertion of the tag via a double crossover homologous recombination event.

The three *Pf*ORC1,2,5::HAx3 parasite lines were generated by my supervisor before my arrival.

To generate the Ty containing vectors, the Ty tag was cloned from the pLN_Ty plasmid and the HA_3 tag was excised from the pLN_HA_C vector using restriction enzymes (AflII and AvrII) that also excised the HR1 of the *Pf*ORC1 or *Pf*ORC5 gene. The HR1 containing the shield mutations from the pLN_HA_C vector was re-cloned and inserted in a 2-way infusion reaction of both PCR products (the HR1 and the

Ty tag) into the AflII/AvrII digested pLN_HA_C plasmid, enabling the assembly of the pLN_Ty vector containing both original HRs with the Ty tag between them.

A list of all primers used for cloning and guide RNA sequences used can be found in Appendix 1.

1.2.2.2. Episomal expression

TK^+: A plasmid containing the viral Thymidine kinase was provided by Catherine Merrick [287].

*Pf*PCNA1: to clone the complete CDS region preceded by the full length 5'UTR for episomal expression, both sequences were included in HR1 and cloned in the plasmid pln_HA_C by In-Fusion HD cloning (Clontech) following the same strategy as described above.

1.2.3. Transfection of parasites

To generate transgenic parasites, a pelleted suspension of ~5% ring-stage-infected RBCs (100 µL) was transfected with a total of 60 µg of circular plasmids as previously described [288]. Electroporation was performed using a Bio-Rad Gene Pulser with settings 310 V, 950 µF and 200 Ω in 0.2 cm cuvettes. Time constants ranged between 9 – 15 milliseconds. After electroporation, 100 µL of fresh 50% RBCs were added to the suspension of parasites. These were grown under agitation (50 rpm) and the tagged lines containing the pDC2-Cas9-hDHFR-yFCU plasmid [286] were selected 8-12 h after with 2.5 nM of WR99210 (from Jacobus Pharmaceuticals; New Jersey, USA) for 12 days and subsequently grown without drug selection. The propagation of transgenic parasites was monitored by Giemsa-stained thin-blood smears. Clones were isolated from these mutants by limiting dilution in the absence of drug pressure. The line expressing the thymidine kinase episomally (TK^+) was continuously grown under 2.5 nM of WR99210 drug pressure. The line expressing the tagged *Pf*PCNA1 episomally was grown continuously with 2.5 µg/mL of blasticidin S.

1.2.4. Genotyping of transgenic parasite strains

For each transfection, both mixed population and clones were screened for the correct gene edition and absence of WT locus by PCR. For each mutant line, three genotyping PCR reactions were performed on the respective genomic DNA:

1. **Integration 5':** primers annealed on the region upstream the cloned HR1 and on the HA_3 tag.
2. **Integration 3':** primers annealed on the region downstream the cloned HR2 and on the HA_3 tag.

3. **WT PCR:** one primer annealing on the stop codon paired with another one upstream the cloned HR1. It would only work if the integration had not been successful, and the WT locus was still intact.

PCR reaction mixtures were prepared using Gotaq Green master mix (Promega) according to the manufacturer's guidelines with annealing temperatures ranging from 48 to 58ºC and elongation time varied according to the size of the amplicon with 1 min per kb.

The details of all the primers used to genotype are listed in Appendix 1.

1.2.5. Dilution cloning

When genotyping PCR results were positive, the transgenic population was subjected to limiting dilution cloning in the absence of drug pressure to obtain the individual clones with the desired endogenous locus modification.

Limiting dilution cloning was performed in 96-well plates at 2% hematocrit. Starting from a late-stage parasite culture with 0.5-1% parasitemia, the culture was diluted and distributed so that each well contained 200 µL of medium with an average of 0.3 parasites. Media was changed on days 8, 10 and 12 and presence of parasites was probed on Giemsa stained thin-blood smears every two days from day 12. Positive wells were screened for the correct genotype and three clones were frozen.

1.2.6. Protein expression detection

1.2.6.1. Western blot

A suspension of schizont-infected RBCs containing a minimum of 10^8 parasites was harvested and RBCs were lysed with 10V of 0.15% saponin. After a PBS wash, nuclei were isolated by sequential washes with different buffers. First, cells were lysed on ice with 1 mL of cell lysis buffer (CLB: 20 mM Hepes pH 7.9, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.64% NP40, 1 mM DTT and protease inhibitor cocktail (PIC, cOmplete™ EDTA free, ref. 4693132001 from Roche)) for 5 minutes. Next, the nuclear fractions (*i.e.,* pellets) were washed three times in CLB and then lysed in 25 µL of Low Salt buffer (LSB: 20 mM Hepes pH 7.9, 0.1 M KCl, 1 mM EDTA, 1 mM EGTA, 0.64% NP40, 1 mM DTT and PIC) for 20 minutes at 4ºC. Finally, nuclear lysates were resuspended in 1x SDS-laemmli buffer and sonicated to reduce viscosity. DTT was added to a final concentration of 100 mM and samples were boiled for 5 minutes before being loaded on a 10% SDS–PAGE gel. Gels were transferred onto PVDF membranes. Membranes were then blocked with 3% BSA in PBS 1X, for 1h and then incubated overnight with the primary antibody monoclonal rat anti-HA (Roche, 3F10) diluted in 3% BSA at 1:2000. This was followed

by three washes with PBS-Tween 0.1% (10 min), and then incubated for 1 hour with the secondary antibody anti-rat HRP (ab6845) diluted in 3% BSA at 1:5000 and finally washed three times with PBS-Tween 0.1%. Blots were incubated with the enhanced chemiluminescence (ECL) solution Clarity Max Western ECL Substrate (BioRad) and imaged on a Biorad ChemiDoc Imager device. Equal amounts of proteins were loaded in each well. To confirm this, polyclonal rabbit primary antibodies anti-aldolase (*Pf*FBPA, fructose-bisphosphate aldolase) (ab207494) at 1:2000 or anti-*Pf*Histone 3 (ab1791) at 1:5000 were used as loading control. In those cases, the secondary antibody anti-rabbit HRP (ab97051) was used at 1:10000 followed by ECL detection.

1.2.6.2. Immunofluorescence

For parasite lines containing triple HA tagged proteins, thin blood smears were fixed with 4% PFA for 15 minutes and quenched with 0.1 mM Glycine for 10 minutes. After, they were permeabilised with 0.01 % Triton x-100 and washed 3 times with 1x PBS. Then samples were blocked in 1.5 % BSA in PBS for 1 hour at room temperature. Samples were incubated with the primary antibody monoclonal rat anti-HA (Roche, 3F10) diluted in 0.15% BSA at 1:2000, for one hour at room temperature, and then washed three times in PBS 1x. Next, samples were incubated with the secondary antibody anti-rat Alexa 488 (Invitrogen, A11006) at 1:5000, for 1h at room temperature and then washed three times in PBS 1x. Before mounting using 20 µL Prolong Diamond Antifade (Molecular Probes), the DNA was stained by incubation with 2 µg/mL 4',6-diamidino-2-phenylindole (DAPI) for 5 min and washed twice in 1x PBS.

Slides were imaged using a Confocal ZEISS 880 FastAi microscope with a 63x immersion oil objective. Z-stacks were taken (0.16 µm inter-slice distance), initially processed with the default Airyscan Processing method and images in 2-D were constructed from the maximum intensity projection on ImageJ [289]. Colocalization analysis for each cell was done using the JACOP plugin from ImageJ [290].

1.3. Replication timing monitoring

1.3.1. FACS

To determine the timing of DNA replication of our *P. falciparum* 3D7 strain, the DNA content of parasites was measured by flow cytometry. Briefly, following synchronization with a 2-hour invasion window as detailed above, parasites were harvested hourly from 24 hours post-invasion (hpi) to 36 hpi and stained with VybrantTM DyeCycleTM Green (VG; 5 µM) (V35004, ThermoFisher Scientific) to stain the DNA of the parasites and MitoTrackerTM Deep Red FM (MT; 1 µM) (M22426, ThermoFisher Scientific) to label viable mitochondrion according to their membrane potential. Fluorescence in each

timepoint was measured in triplicate samples on the BD FACSAriaTM III (Becton Dickinson), using the red laser, 670/14 nM filter for MT and the green laser, 530/30 nM filter for VG. Data analysis was done using FlowJoTM v.10.8.1 software. A ring-stage sample and an asynchronous sample were used to gate parasites having one nucleus (gate 1N, *i.e.,* non-replicating parasites) from those having two or more nuclei (gate ≥2N). These gates were applied to all timepoints. This experiment was done prior to my arrival.

1.3.2. Immunofluorescence of EdU incorporation into DNA

Synchronised TK^+ parasites at ~4% parasitaemia were incubated with 100 µM 5-ethynyl-2-deoxyuridine (EdU, Invitrogen, ref. number A10044) for 45 min in standard culture conditions at either 25-, 29- or 35 hours post invasion. After, 200 µL of parasitized RBCs were spun down at 900 g and then washed once with 1% BSA in 1x PBS. Pellets were then fixed with paraformaldehyde (1% final concentration) for 15 min at room temperature and then washed three times with 1% BSA in 1x PBS. Next, cells were permeabilised in 0.01% Triton x-100 and 0.015% saponin in 1 mL PBS 1x for 15 minutes at RT. After, cells were washed 3 times in 1 mL of 1% BSA. EdU signal was detected with click chemistry (Click-iT® EdU flow cytometry assay kit from Molecular Probes, catalogue number C10419, containing all click reaction reagents). Parasite pellets were incubated in click reaction buffer containing 2 mM $CuSO_4$, 1x Buffer additive provided with the kit (which contained Sodium ascorbate), and Alexa Fluor 488 dye azide (CLK-1275-1) for 30 minutes at RT, protected from light, and washed 3 times with 1% BSA in 1x PBS. Pellets were washed twice in 1x PBS, incubated with 2 µg/mL 4',6-diamidino-2-phenylindole (DAPI) for 5 min, washed twice in 1x PBS, mounted using 20 µL Prolong Diamond Antifade (Molecular Probes) and set overnight at room temperature.

Slides were imaged using a Confocal ZEISS 880 FastAi microscope with a 63x immersion oil objective. Z-stacks were taken (0.16 µm inter-slice distance), initially processed with the default Airyscan Processing method and images in 2-D were constructed from the maximum intensity projection on ImageJ. Colocalization analysis for each cell was done using the JACOP plugin from ImageJ. 3-D reconstructions were done with the Imaris software (RRID:SCR_007370) [291].

2. Next Generation Sequencing

2.1. Chromatin Immunoprecipitation followed by sequencing (ChIP-seq)

2.1.1. Protocol

Chromatin Immunoprecipitation followed by sequencing of *Pf*ORC1::HA_3, *Pf*ORC2::HA_3 and *Pf*ORC5::HA_3 parasites was prepared using a previously described [292] protocol with slight changes. For each timepoint (25 hpi ±1h, 32 hpi ±1h, 35 hpi ±1h) two replicates were prepared with a total of 5 x 10^9 infected RBCs per sample. For each sample, the parasite culture was cross-linked with methanol-free paraformaldehyde (final concentration 1%) at 37ºC for 10 min and then quenched with 0.125 mM glycine for 5 min at 37ºC. Erythrocytes were washed once with 1x PBS, lysed with 0.15% saponin and washed again with 1x PBS. Parasites were then incubated on ice with 4 mL of Lysis buffer (10 mM Hepes pH 8.0, 10 mM KCl, 0.1 mM EDTA pH 8.0, and PIC) for 5 min, and then nuclei were disrupted with 150 strokes on a pre-chilled douncer homogenizer. Subsequently, these were pelleted by centrifugation for 10 min at 13500 g, 4ºC and resuspended in 3 mL of SDS lysis buffer (1% SDS, 10 mM EDTA pH 8.0, 50 mM Tris–HCl pH 8, and PIC). Next, the chromatin was sheared into fragments of 200-600 bp by sonication (Bioruptor; Diagenode) for 30 minutes (30 cycles, 30 seconds ON, 30 seconds OFF) and debris was removed by centrifugation for 10 minutes at 13500 g, 4ºC. Ten percent of the sample was kept as “input” and frozen until DNA purification.

For each immunoprecipitation, 6 mL of chromatin were pre-cleared using 2 µg of rat IgG antibody for 2h at 4ºC, with gentle rotation, followed by incubation with 20 µL of Magna ChIP Protein A/G Magnetic Beads (Sigma-Aldrich) for 2h at 4ºC, with gentle rotation. The beads were harvested using a magnetic rack and processed as the “IgG control”. After, the pre-cleared chromatin was diluted tenfold in ChIP dilution buffer (16.7 mM Tris–HCl pH 8.0, 0.01% SDS, 1.0% Triton X-100, 1.2 mM EDTA pH 8.0, 150 mM NaCl and protease inhibitor cocktail) and immunoprecipitated overnight at 4ºC, with 20 µL of anti-HA antibody (Abcam 9110) per reaction followed by a 2h incubation at 4ºC with 20 µL of Magna ChIP Protein A/G Magnetic Beads, with gentle rotation.

The IP and IgG fractions were washed at 4ºC once with 1 mL of Low Salt Wash Buffer (20 mM Tris–HCl pH 8.0, 2 mM EDTA pH 8.0, 150 mM NaCl, 1% Triton X-100, 0.1% SDS), once with 1 mL High Salt Wash Buffer (20 mM Tris–HCl pH 8.0, 2 mM EDTA pH 8.0, 500 mM NaCl, 1% Triton X-100, 0.1% SDS), once 1 mL with LiCl Wash Buffer (10 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5 % NP-40, 0.5 % sodium deoxycholate), and once at room temperature with 1 mL TE buffer (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0). Immune complexes were eluted in 300 µL of elution buffer (10 mM Tris pH 8.0, 1 mM EDTA

pH 8.0, 1 % SDS) and incubated at 65ºC for 30 min in an agitating thermal block (900 rpm 1 min, 30 sec still).

Eluted IP DNA, IgG and input DNA were reverse crosslinked by incubating at 65ºC for 15 hours. 300 µL of TE buffer was then added to each sample and RNAse A was added to a final concentration of 0.2 mg/mL, mixed by inverting and incubated for 2 hours at 37ºC. Proteinase K was added to a final concentration of 0.2 µg/mL and samples were incubated at 55ºC for 2h. DNA was precipitated using the phenol-chloroform method. Briefly, 600 µL of phenol:chloroform:isoamyl alcohol (P:C:IA – 25:24:1) were added and mixed thoroughly and the aqueous phase was separated through centrifugation at 13500 g, for 5 min at 4ºC. The aqueous layer was then transferred into a tube containing 600 µL of chloroform:isoamylalcohol (24:1), and spun at 13500 g, for 5 min at 4ºC. The aqueous layer was transferred to a new tube containing and 800 µL of cold isopropanol. DNAs were pelleted by centrifugation at 20,000 g for 20 min at 4ºC and washed twice with cold 80% EtOH, before being air-dried and resuspended in 10 µL of 10 mM Tris-HCl, pH 8.0.

2.1.2. Library preparation, sequencing and peak calling

DNA sequencing libraries were prepared using the MicroPlex Library Preparation Kit v3 (Diagenode) following the manufacturer's instructions and sequenced on a Hiseq2500 or NovaSeq SP flowcell, for a minimum of 5 million reads per sample.

Quality control of fastQ files was performed using the FastQC software (version 0.11.9) [293]. Sequencing reads were aligned and mapped to the *P. falciparum* 3D7 genome (v.57 from PlasmoDB [127]) using Burrows–Wheeler Alignment BWA MEM (version 0.6) [294] with default settings to generate SAM files. These were further processed using Samtools (version 1.17) [295] to generate their corresponding BAM files. Peaks for the IP and IgG samples were called using MACS2 (version 2) [296] with a q-value cut-off of 0.05 and using the input as control sample. The BEDtools suite was used to compute the intersection between any datasets throughout [297], [298]. The narrow peaks datasets for the two biological replicates of each ORC ChIP were intersected and only common peaks were kept. Unspecific peak fragments (*i.e.*, also present in the IgG sample) were removed if the overlap was >50 bp with the IgG control peaks. A final filtering step to discard peaks smaller than 50 bp produced the final datasets that were used for further analysis and to generate the figures in this study.

For visual representation of the colocalization between two datasets, deepTools v.2 [299] was used. Matrix files were generated from bigwig files of *Pf*ORC1 and *Pf*ORC2 using *ComputeMatrix*, and then

were used for plotting the individual reads enriched with *plotHeatmap* from BEDtools (colocalization of both replicates per ORC protein and colocalization of the final datasets of ORC1 and ORC2) [297], [298].

The Integrative Genomics Viewer [300] was used to visualize the genomic localization of the different datasets of this study. Circos plot was made with the online tool https://venyao.shinyapps.io/shinyCircos/ using the datasets of *P. falciparum* chromosome sizes from PlasmoDB (v. 57) [127], *Pf*ORC$_{1\text{-}2}$, SNS-seq ORIs, NFS ORIs and the coordinates of G4FS from *P. falciparum* [167] as input.

2.2. Sequencing of Short Nascent Strands (SNS-seq)

2.2.1. Protocol

Short Nascent Strands were purified before my arrival in the lab as previously described [7]. The cell number was increased to 10^8 to ensure enough starting material. Synchronous parasites were harvested at 29hpi (+/-1h). In each experiment three samples were prepared:

A. Negative control – high molecular weight DNA devoid of SNS,
B. SNS isolated in RNAse-free conditions and digested with λ-exonuclease
C. λ-exo background control: SNS isolated as in condition ii) but subjected to a digestion with a cocktail of RNAses prior to λ-exo nuclease. This last sample was treated bioinformatically as the IgG control in the ChIP-seq experiment.

2.2.2. Library preparation, sequencing and peak calling

The Illumina TruSeq ChIP Sample Prep Set A was used for preparation of sequencing libraries, which were sequenced either on a HiSeq 2000 or NovaSeq SP flowcell, for a minimum of 5 million reads per library. Mapping, peak calling and the various correlation and statistics analyses were performed following the same procedure as with ChIP-seq.

2.3. NanoForkSpeed (NFS)

2.3.1. Protocol

Each sample started from a culture with at least $5x10^8$ parasites at 29 or 35 hpi. We incubated the cultures with a short pulse of 2 min with 100 µM of 5-bromo-2-deoxyuridine (BrdU, ref. B23151 from Invitrogen) and followed by a “chase” of 1 mM of thymidine (Sigma, TI895-1G) for 45 minutes. After,

RBCs were lysed in 0.15% saponin and pellets were washed once in PBS. The DNA was extracted with the Monarch® HMW DNA Extraction Kit, following the manufacturer's instructions.

2.3.2. Library preparation, sequencing and initiation sites detection

A minimum of 3 µg of DNA were purified using the Short Reads Eliminator kit (Circulomics) and libraries were prepared according to the "Native barcoding genomic DNA (with EXP-NBD104, EXP-NBD114 and SQK-LSK109)" protocol provided by Oxford Nanopore. Each library was sequenced separately on a PromethION sequencer using R9.4.1 pore version. Basecalling, read mapping, fork directionality and speed analyses were performed with NFS as previously described [5].

Individual replication fork speed was computed at each genomic feature by intersecting the genomic coordinates of the Start (X0) and end (X1) of each fork with the coordinates of the mentioned features. Centromere coordinates were obtained from the result of ChIP-seq of the CenH3 protein [140]; telomeres were considered as the first and last 3 kb of each chromosome, tRNA and rRNA coordinates were extracted from the latest release of PlasmoDB (v. 57) [127], heterochromatin was defined as H3K9me3 enriched regions [151] and euchromatin as H3K9me3 depleted regions. The top and bottom 25% expressed gene list was obtained from the timepoint t24h from [130]. The genes whose expression at 24 hpi was at the top or bottom quartile were selected and their genomic coordinates were used to intersect with the replication forks and obtain the individual fork velocities. All statistical tests were performed using the programming language R [301]. Datasets were assessed for normality by Shapiro–Wilk testing with α-level=0.05. Statistical analysis of the differences between fork speeds and the difference between inter-origin distances was performed with a two-sample Wilcoxon rank sum test.

For the single molecule speed computation, I calculated the distance between two initiation sites mapped from the same read, and only the speed of the upstream incoming replication fork was considered.

2.4. Data analysis

Plots were created with R programming language using ggplot2 and tidyr packages [301]–[303].

For genome-wide representation of the origins, the coordinates of the *Pf*ORC$_{1\text{-}2}$ ChIP-seq peaks, SNS-seq peaks or NFS initiation sites and their clusters were converted into Genomic Ranges in R and represented with the *covplot* function of ChIPseeker [304], [305].

To check if the amount of overlap between 2 sets of intervals was more than what would be expected by chance given their coverage and the size of the genome, Fisher's exact test from BEDtools was used (null hypothesis of independence) [297], [298]. Statistical significance was set to $p \leq 0.05$. To calculate the overlap between the *Pf*ORC$_{1\text{-}2}$ binding sites, or SNS-seq origins, and several *P. falciparum* post-translational modifications [144], [147], [306], the Jaccard similarity index was computed with BEDtools [297], [298]. This statistic represents a measure of the ratio of the amount of intersecting base pairs between two sets over the number of base pairs in the union of the two sets. Consequently, the final statistic ranges from 0.0 to 1.0, with 0.0 meaning no overlap and 1.0 total overlap. The results of the pairwise comparisons with Jaccard were used to generate the heatmaps on R.

Inter origin distance (IOD) was computed with 'bedtools closest' using a single file as input with the 'ignore overlaps (-io)' and 'ignore upstream (-iu)' specifications to ensure that distances were computed between consecutive peaks. Empirical cumulative distribution functions (ECDF) of the distance between datasets were computed with the result of the 'bedtools closest' calculations between the datasets and represented using the *ECDF* function from Hmisc (version 4.7-0) on R [301]. As a control, the genomic coordinates of each dataset were first randomly redistributed by shuffling them while keeping constant the number of sites per chromosome with the *shuffleGRgen* function from the RepNano repository (RepNanoFunction_GB.r) [10], and then they were subjected to the same analysis of computation of distances as the experimental sites.

The FASTA sequences of the peaks were obtained with the bedtools *getfasta* function and used as input for the online MEME motif analysis [307], [308].

2.4.1. Cluster analysis

To investigate whether the *Pf*ORC$_{1\text{-}2}$ sites or active ORIs were located in clusters, first the whole genome was divided into 10 kb windows, with a sliding window of 5 kb, with the *makewindows* function from BEDtools. Then, the coverage of the *Pf*ORC$_{1\text{-}2}$ sites, SNS-seq or NFS origins was calculated over the 10 kb windows. A cluster was then considered if a given 10 kb window contained 3 or more *Pf*ORC$_{1\text{-}2}$ sites, or 5 or more SNS-seq or NFS origins.

2.4.2. G-quadruplex forming sequences

The FASTA sequences of the peaks were used for the analysis of GC content on the emboss infoseq tool (https://www.bioinformatics.nl/cgi-bin/emboss/help/infoseq). They were also used as input for the online G4hunter algorithm analysis [169] (window size: 25, threshold: 1.2).

The average enrichment profile of G4FS around ORC sites was determined and visualized after assigning a constant score to all G4 sites [167], converting the datataset to bigwig (with *bedGraphToBigWig*) and using *ComputeMatrix* and the *plotProfile* function from deeptools2.

The distance between *Pf*ORC$_{1\text{-}2}$ sites or SNS-seq origins and the closest G4FS were computed with the standard mode of 'bedtools closest' and ECDFs were computed and visualized as described above.

2.4.3. Comparison with gene coordinates

Meta-gene plots were computed with the *plotPeakProf2* (ChIPseeker). Each of the *P. falciparum* core gene coordinates (regions of the genome described as core by Miles *et al*. [309], and devoid of H3K9me3 [151]) or *var* gene coordinates (retrieved from PlasmoDB v.57) were scaled from 0 to 100% and converted into a segment of equal length. The start and end of each CDS was labelled by TSS (Transcription Start Site) and TTS (Transcription Termination Site). Additionally, a region of the same dimensions as the gene size (or 3000 bp, in the case of *var* genes) was added both upstream and downstream. Then, the enrichment of *Pf*ORC$_{1\text{-}2}$ or active origins was computed for the internal and flanking sequences and compared to a randomized control generated as mentioned above.

For associations of features with local gene expression levels (percentile) I used the timepoint t24h from Chappell *et al.* [130]. I analysed the expression level of the genes that contained a *Pf*ORC$_{1\text{-}2}$ origin, a SNS-seq origin or at least 80% of an NFS origin (considered as 3 kb around the midpoint between the diverging forks). Datasets were assessed for normality on R by Shapiro–Wilk testing with α-level = 0.05. Statistical analysis of the differences between gene expression percentile was performed with a two-sample Wilcoxon rank sum test also on R.

3. Proteomics

3.1. Protein immunoprecipitation using anti-HA beads

Immunoprecipitation of *Pf*ORC2::HA_3 parasites was performed from a parasite culture of at least 10^8 parasites tightly synchronised with Percoll-sorbitol (with 2-hour invasion window) to obtain 35 hpi schizont stage parasites. Parasites were harvested (1000 g, 5 min centrifugation) and RBC membrane was lysed with 0.15% saponin in PBS at 4ºC and spun for 5 min at 3000 g. After a PBS wash, cells were lysed on ice with 1 mL of cell lysis buffer (CLB: 20 mM Hepes pH 7.9, 10 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.65% NP40, 1 mM DTT and PIC) for 5 minutes. Next, nuclei were isolated by washing the nuclear fractions (*i.e.,* pellet) three times in CLB and then lysed in 25 µL of Low Salt buffer (LSB: 20mM Hepes pH 7.9, 0.1 M KCl, 1 mM EDTA, 1 mM EGTA, 0.64% NP40, 1 mM DTT and PIC) for 20 min at 4ºC. The nuclear insoluble fraction debris was collected by centrifugation (30 min, 12000 rpm, 4ºC) and resuspended in 1x SDS-laemmli buffer. The lysate was incubated with 40 µL of anti-HA magnetic beads (Thermofisher, ref. 88836) overnight in a rotor at 4ºC. Immune complexes were collected on a magnet and a fraction of the flowthrough was saved and resuspended in SDS-laemmli buffer. The beads were washed 5 times in lysis buffer and then resuspended in 30 µL of 1x SDS-laemmli buffer. Prior to loading on SDS-PAGE gel, the eluate was boiled at 95ºC for 5 min and either blotted as described on section 1.2.6.1 or the SDS-PAGE gel was subjected to Silver Staining for the detection of the whole protein content.

3.2. Silver Staining

Silver Staining of the SDS-PAGE gel containing the immunoprecipitated protein complexes was done with the Pierce Silver Stain kit from Thermofisher scientific (ref. 24612), following manufacturer's instructions. Briefly, the polyacrylamide gel was washed twice in ultrapure water (5 min), fixed with two washes with 30% ethanol: 10% acetic acid solution (15 min each wash), and washed twice (5 min) with 10% ethanol and twice more (5 min) in ultrapure water. Then, the gel was incubated for 1 minute with Sensitizer Working Solution (25 µL Sensitizer with 12.5 mL water) and washed twice (1 min) in water. The gel was stained for 30 minutes with the Stain Working Solution (250 µL Enhancer with 12.5 mL Stain), washed twice (20 seconds) with ultrapure water and developed for 7-10 minutes with the Developer Working Solution (250 µL Enhancer 12.5 mL Developer) until bands appeared, when reaction was stopped by drowning the gel in 5% acetic acid for at least 10 minutes.

3.3. iPOND

Isolation of proteins on nascent DNA strands was performed as described previously [310] with several changes. All samples started from a synchronised culture of the TK^+ parasite line at the beginning of the replicative stage (29 hpi) with at least $1.5x10^8$ parasites. Four replicates of each sample were harvested in each of the experiments, except for the negative control in the first experiment where only duplicates were available.

- **EdU sample**. 2-minute EdU pulse: aimed at capturing replisome proteins
- **Thy chase sample**. 2-minute EdU pulse + 45-minute Thymidine chase: aimed at capturing bystander chromatin associated proteins not related to the replisome
- **No Click Control sample**. Negative control: 2-minute EdU pulse, followed by click reaction without the azide component, to permit detection of unspecific binding to the beads.

Parasite cultures were pulsed (2 min) with 100 µM 5-ethynyl-2-deoxyuridine (EdU, Invitrogen, ref. number A10044) and the "chase" samples were then incubated for 45 min with 1 mM Thymidine (Sigma, ref. number TI895-1G). After, DNA-protein complexes were crosslinked with methanol-free paraformaldehyde (final concentration 1%) at 37ºC for 15 min and quenched with 0.125 mM glycine for 5 min at 37ºC. Erythrocytes were washed once with 1x PBS, lysed with 0.15% saponin and washed again with 1x PBS. Pellets were stored at -80ºC.

On the day before mass spectrometry analysis, the parasite pellets were resuspended in lysis buffer at a concentration of 1×10^7 cells/mL (Jena biosciences, CLK-1065) supplemented with protease inhibitor cocktail (PIC). The lysate was incubated on ice for 30 min and then spun down for 2 min, 900 g at 4ºC. Supernatant was discarded, and pellets were washed once with 0.5% BSA in PBS 1x and once in PBS 1x (both containing PIC). Click reactions were prepared in PBS 1x containing either 0.5 mM biotin azide (Jena biosciences, CLK-1167-5.1) or agarose resin containing sodium azide (Thermofisher, ref. 29201) in the first iPOND experiment – omitted in both cases in the No Click Control, 10 mM sodium ascorbate (Jena biosciences, CLK-MI005-1G), 1 mM $CuSO_4$ (Jena biosciences, CLK-MI004-50.1), and the copper stabilizer TBHAT 5 mM (Jena biosciences, CLK-1010-25). Click reactions were rotated for 2 hours at RT. Then, samples were spun for 4 mins at 2000 g, 4ºC and the supernatant was transferred to a separate tube (Flowthrough). Pellets were washed once with 0.5% BSA in PBS 1x and once in PBS 1x (both containing PIC). Samples were then resuspended in lysis buffer (1% SDS in 50 mM Tris-HCl pH 8) containing PIC at a concentration of 2.5×10^8 cells per 1500 µL. Next, the chromatin was sheared into fragments of 200-600 bp by sonication (Bioruptor; Diagenode) for either 5 minutes (in the first experiment) or 20 min (2 x 10 cycles, 20 seconds ON, 40 seconds OFF,

with a 5 min waiting period between each 10 cycles; in the second experiment) and debris was removed by centrifugation for 10 min at 13500 g, 4ºC. 10% of this sample was kept as "input". In the first iPOND experiment, the agarose resin containing the pulldown proteins was directly resuspended in SDS-laemmli buffer (containing DTT) and proteins were eluted at the mass spectrometry facility (FPP) for downstream processing. In the second experiment, the remaining lysate after sonication was incubated overnight with an equal volume suspension containing 20 µL of pre-washed streptavidin magnetic beads per 1 × 10^8 cells (Thermofisher, ref. 88816). Streptavidin-biotin complexes were harvested on a magnet and washed three times with cold lysis buffer. Finally, proteins were eluted in SDS-laemmli buffer (containing DTT). Crosslinking was reversed by boiling the samples at 95ºC for 25 min. Next, samples were either loaded on an SDS-PAGE gel or submitted to the mass spectrometry facility (FPP) for downstream processing.

3.3.1. Mass Spectrometry

Protein digestion was performed at FPP (Plateforme de Protéomique Fonctionnelle) on S-Trap™ micro columns (Protifi, USA) following the manufacturer's instructions. In brief, protein extracts were treated in 5% SDS/50 mM triethylammonium bicarbonate (TEAB), reduced with 20 mM dithiothreitol (DTT) and incubated for 10 min at 95ºC. Samples were cooled to room temperature and alkylated with 40 mM iodoacetamide (IAA) for 30 min at room temperature in the dark. Samples were then acidified with phosphoric acid at final concentration of 1.2% and diluted 6 times in S-Trap binding buffer (90% methanol/100 mM TEAB). The resulting protein suspension was transferred to the S-Trap filter via centrifugation at 4000 g for 1 min. Trapped proteins were washed five times with 150 µL S-Trap binding buffer. One µg of trypsin (Trypsin GoLd, Promega) in 50 mM TEAB was added to the filter surface and incubated for 2 hours at 47ºC. Tryptic peptides were eluted sequentially with 40 µL of 50 mM TEAB, 50 mM TEAB / 0.2% aqueous formic acid, and then with 0.2% aqueous formic acid/ 50% of acetonitrinile (ACN) via centrifugation at 4000 g. Eluted peptides were vacuum-dried and stored before mass spectrometry analysis.

Peptides were resuspended in 100 µL of Loading buffer and 1µL (4 µL for NCC_R4) was injected on nanoflow HPLC (RSLC U3000, Thermo Fisher Scientific) coupled to a mass spectrometer with nanospray source (Qexactive HF-X, Thermo Fisher Scientific). Peptides were separated on a capillary column (reverse phase C18, NanoViper, Thermo Scientific) using a 2-40 % buffer B gradient in 123 min (A= 0,1 % formic acid; B= 0,1 % formic acid, 80 % acetonitrile) at flow rate of 300 nL/min. Spectra were registered via Xcalibur 4.1 software (Thermo Scientific). Spectral data of RAW file was analysed with MaxQuant software v2.0.3.0 and Perseus v1.6.15.0. The database for *Plasmodium falciparum* specific peptides and contaminants used was RefProteome_PlasmodiumFalciparum_UP000001450_2022-

05.fasta; with Carbamidomethyl (C) as fixed modification and variable modifications Oxidation (M) and Acetyl (N-term). Data validation was realised with protein and peptide FDR of 0,01.

These procedures were performed by Khadija El Koulali and overseen by Mathilde Decourcelle at the FPP (Plateforme de Protéomique Fonctionnelle) in Montpellier.

3.3.2. Data analysis

Differential enrichment analysis of proteomic data was done with the raw output of MaxQuant software using the DEP package on R [301], [311]. Briefly, contaminant proteins and decoy database hits (annotated as "Reverse") were removed. Proteins were kept for analysis if they were identified in all replicates of at least one condition. The data was background corrected and normalized by variance stabilizing transformation (*vsn*), and missing values were imputed using random draws from a Gaussian distribution centered around a minimal value (*MinProb*, q=0.01). Differential enrichment analysis was then done with the *test_diff* function, which uses Limma linear modelling, and all plots were generated on R using ggplot2 and tidyr packages [302], [303].

Protein-protein interaction analysis was done using STRING (http://string-db.org) [312] (minimum required interaction score = 0.4), with the list of candidate proteins significantly enriched in the EdU sample vs Thy (adjusted p-value < 0.1) as input (n=19, table 5). Finally, Gene Ontology analysis was done using the same list as input to the Gene Ontology analysis build-in tool from PlasmoDB (Biological Process) and the output was plotted with ggplot2 on R.

Results

Chapter 1. Genetic landscape of the origins of replication

The canonical eukaryotic cell cycle has four distinct phases (Fig. 16a): starting with a preparatory G1 phase (gap phase), during which the mother cell undergoes a number of metabolic changes to be ready for division. Then, during the synthesis (S) phase, DNA replication is activated, and the genome is duplicated. Next, a second gap (G2) phase allows for the final maturation before cell division. Finally, the cell divides during Mitosis (M phase) generating two daughter cells [313].

Plasmodium parasites, in contrast, during intra-erythrocytic development multiply using a strategy termed schizogony which deviates greatly from the typical binary fission approach. In schizogony, following a long G1-like phase during which the cell grows, parasites undergo four to five cycles of DNA replication and nuclear division (S/ND), thus creating a temporary syncytium. Subsequently, a general cytokinetic event leads to the formation of up to 32 daughter cells [2], [113](Fig. 16b). During the G1 phase the members of the pre-replicative complex are recruited to the replication initiation sites (origins) and act as a landing platform for the DNA polymerase machinery until replication is triggered in S phase. We first aimed to describe the timing of the initial DNA replication round in order to define the time window during late G1 at which the pre-replicative complex assembled at the origins of replication.

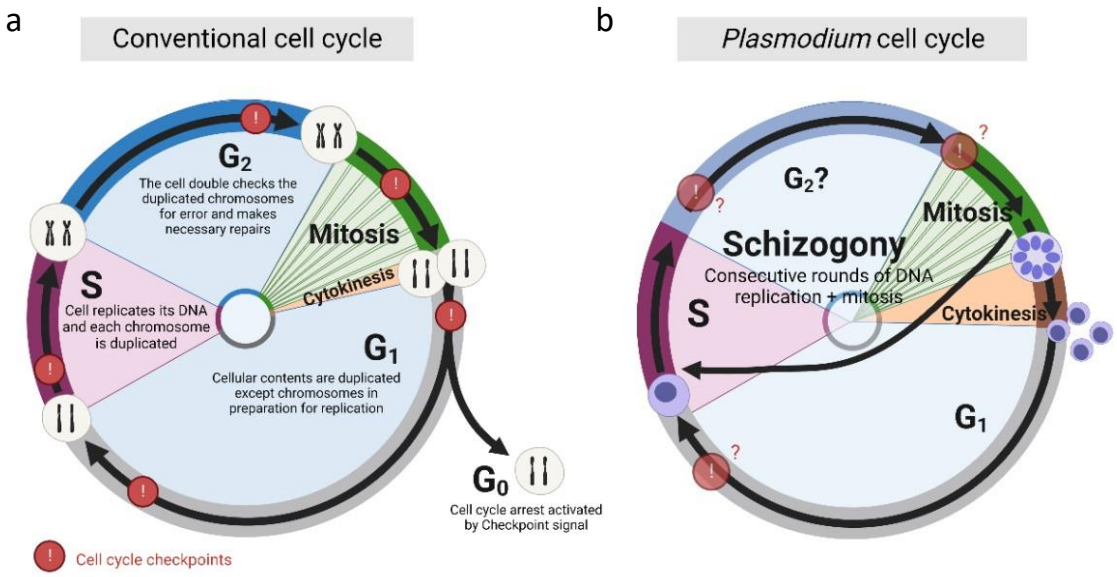


Figure 16.Conventional eukaryotic cell cycle vs Plasmodium falciparum cell cycle*. The canonical eukaryotic cell cycle (a) starts with the G1 phase, where the cell undergoes a period where all the necessary machinery for DNA replication is synthesized. Then, the cell replicates its DNA in S-phase. Following DNA replication, the cell enters in another gap phase termed G2 where all the machinery necessary for cell division is synthesized and necessary repairs are done in the DNA. The following M-phase is composed of two discreet steps: mitosis, which constitutes the pairing and separation of the duplicated chromosomes, and cytokinesis which is the physical process whereby the cell splits into two daughter cells. In some situations, the cell will exit the cycle and enter a quiescent G0 state. P. falciparum cell cycle (b) displays a long G1 phase and consecutive rounds of DNA replication and mitosis, with an apparent G2 phase, followed by a single cytokinesis. Made with Biorender.com*

1. Timing of replication initiation

Currently, there are not good synchronisation chemicals that can efficiently arrest the *Plasmodium* cell cycle at G1 phase immediately before the onset of S-phase [314]. However, by exploiting the fact that *Plasmodium* parasites exit and re-invade new RBCs every 44-48h and combining the use of density gradients and chemicals that selectively promote the lysis of certain parasite stages, it is possible to obtain a semi-synchronous population of parasites.

To map the origins of replication of malaria parasites we first determined the time during the intra-erythrocytic cycle when DNA synthesis is initiated. We allowed a RBC reinvasion window of 2h, after which the remaining late-stage parasites that had not yet reinvaded were selectively lysed using the sorbitol method (see Materials & Methods section) [285]. Next, we used flow cytometry to measure the average DNA content of the population of parasites over a 12h period and used the timing of increase in DNA content as a proxy for onset of the first S-phase following RBC invasion. This set of experiments was performed by another PhD student prior to my arrival.

As observed in Fig. 17, the percentage of parasites before 26 hpi that had undergone genome duplication was extremely low, ranging between 2.5-4.2% in the three analysed replicates. Between the timepoint 26-28 hpi and 30-32 hpi, about 33% of the population entered S-phase. The majority of parasites were thus in a state akin to a G1 state up to roughly 26 hpi. As a result, we considered the onset of DNA replication to take place at approximately 29 hpi±1h (*i.e.,* the midpoint between the timepoints 26-28 hpi and 30-32 hpi).

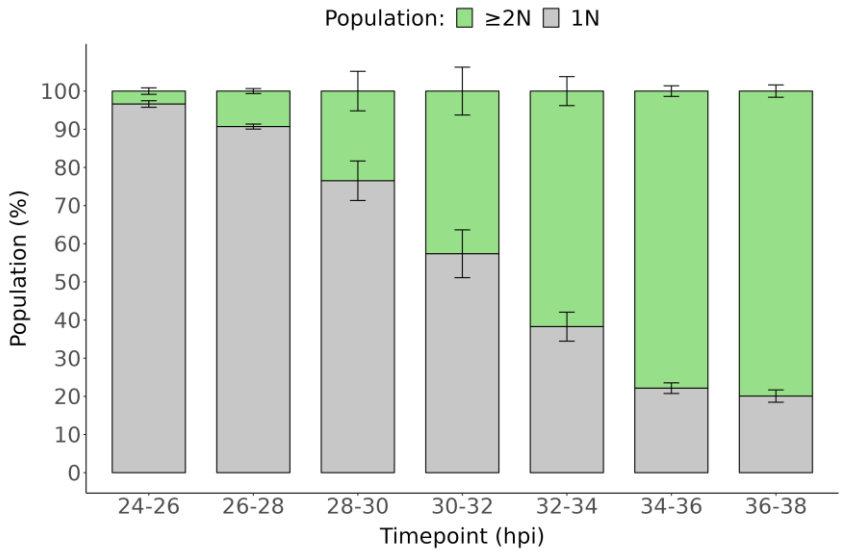


Figure 17. Plasmodium falciparum replication timing measured with flow cytometry. *Total DNA content (N) of P. falciparum 3D7 parasites was measured by flow cytometry over schizogony. Invasion window 2h.*

To further confirm the established timing, I performed several immunofluorescence assays (IFAs) using the nucleoside analogue EdU (5-ethynyl-2-deoxyuridine), since its incorporation into the genome should only occur upon DNA synthesis. However, since *P. falciparum* parasites rely on *de novo* synthesis of pyrimidines and cannot salvage exogenous nucleotides due to the absence of a key enzyme (Thymidine Kinase) [287], I used a *P. falciparum* transgenic line carrying a viral thymidine kinase (TK^+), previously generated in the lab, which is able to incorporate nucleoside analogues such as BrdU (5-Bromo-2-deoxyuridine) and EdU (5-ethynyl-2-deoxyuridine) into DNA during replication. To detect the incorporation of EdU by IFA, I used Click chemistry (Copper(I)-catalyzed azide-alkyne cycloaddition, CuAAC), which consists on the formation of a covalent bond between an alkyne (EdU in the DNA) and an azide modified dye (in my case, a fluorophore azide – Alexa fluor 488 green azide), in a reaction catalysed by Copper (Fig. 18) [315]. After the click reaction, the fluorescence signal can be detected by microscopy.

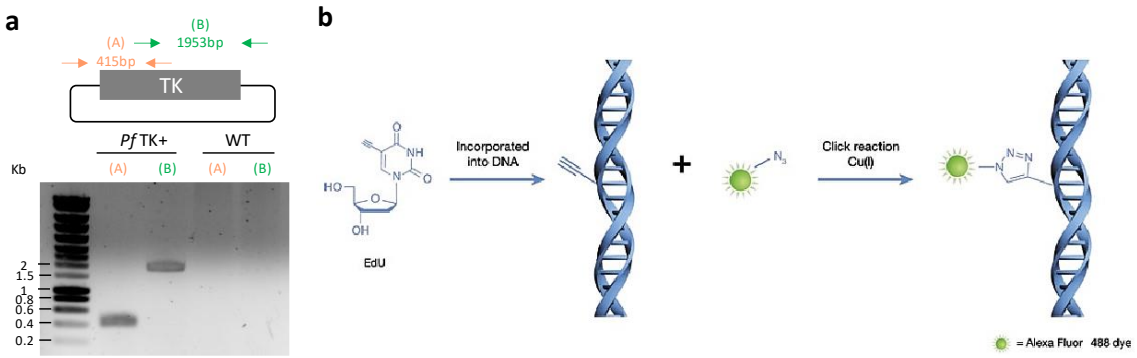


Figure 18. The Thymidine Kinase allows for incorporation of labelled nucleotides into DNA. *a) Genotyping PCR of the TK gene. Expected sizes are indicated above the plasmid map. The same PCR reactions, termed A and B, were performed in the TK^+ parasite line and in the WT Pf3D7 line. b) Click reaction between the incorporated EdU in the DNA and the fluorophore azide* [315]*. The ethynyl group of the EdU (alkyne) incorporated in the double-stranded DNA forms a triazole bond with the Alexa Fluor 488 azide.*

As this approach had not been previously used in malaria parasites, I first confirmed the incorporation of EdU in the genome by incubating 35 hpi schizont-stage parasites with 100 µM of EdU or 1% DMSO (negative control). Since human cells are TK^+, I used Human Foreskin Fibroblast (HFF) cells as a positive control. To pinpoint the minimal duration of the EdU pulse required to detect enough fluorescence signal, I tested different incubation times ranging from 15 min to 45 min and I observed that 15 minutes were not sufficient to observe EdU incorporation into DNA. Although 30 min yielded some fluorescence signal, only after a 45 min incubation did I satisfactorily detect EdU signal both in the parasite and HFF samples (Fig. 19). I then confirmed the significant colocalization of the EdU and DAPI signals in the 45 min-incubation experiments using the JACOP colocalization plugin from ImageJ and plotted the profile of the two signals.

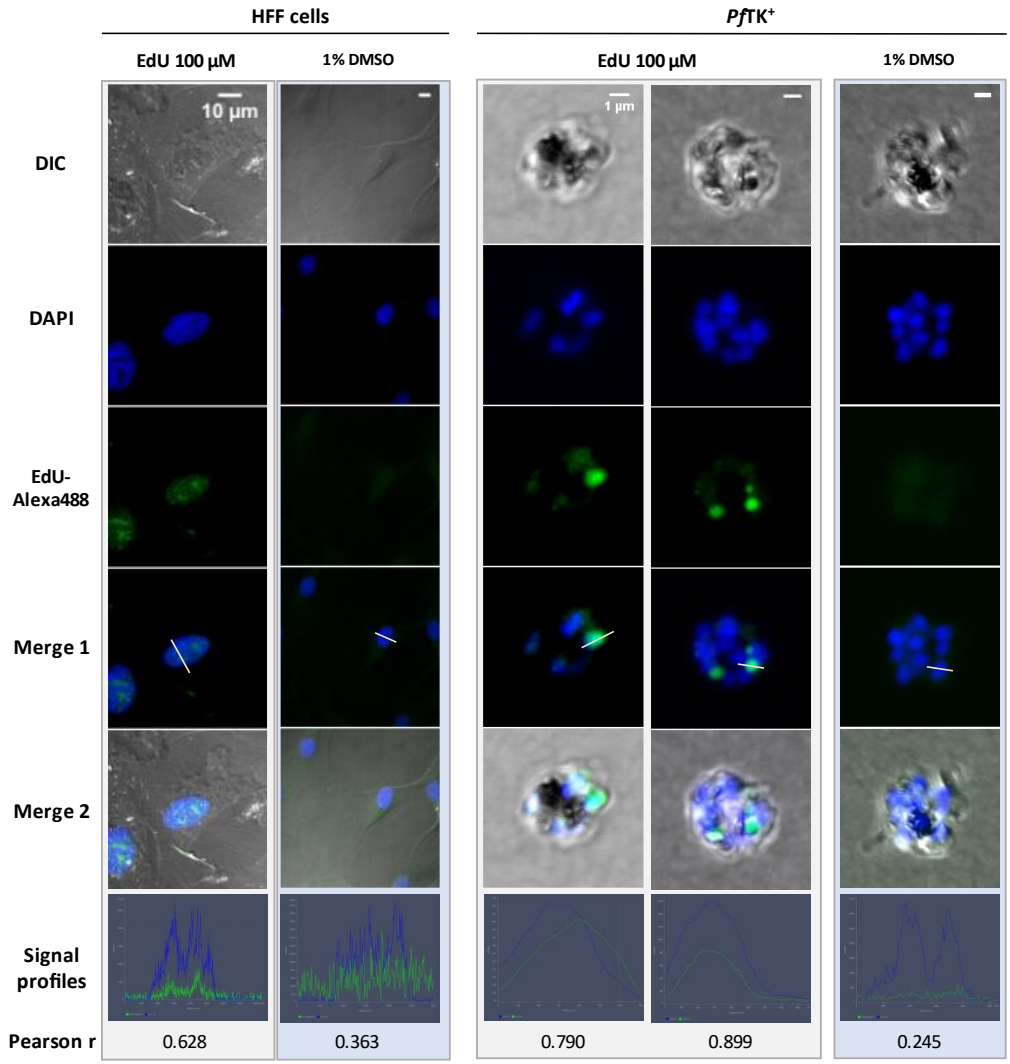


Figure 19. Visualization of active replication during schizogony after a 45-minute incubation 100 µM of EdU*. Replicated DNA is stained with Alexa Fluor 488 (green) and total DNA is stained with DAPI (blue). HFF cells were used as TK+ control and DMSO as negative control. The light pink dashed bar in the Merge 1 images represents the line drawn to plot the profiles of the signals shown below. Scale bar represents 1 µm for P. falciparum cells and 10 µm for HFF cells. Pearson's coefficient of correlation between the green and blue signal is shown below the profile plots.*

I further explored the nature of the EdU signal in the parasite's nuclei by performing 3D reconstructions of several IFA Z-stack images after a 45-minute EdU incubation period. For example, in Fig. 20 we can observe a parasite with 2 nuclei that have incorporated EdU into their DNA, meaning that they were actively replicating, and DNA synthesis occurred in both nuclei, during the incubation with EdU. Accordingly, in the 3D reconstruction generated with Imaris, the green signal from the EdU-fluorophore is located inside the volume generated from the DAPI signal (Pearson's correlation coefficient = 0.826), which is also seen in the profiles of both signals plotted with the Zen software, that display a colocalized increase in the intensity of the signals.

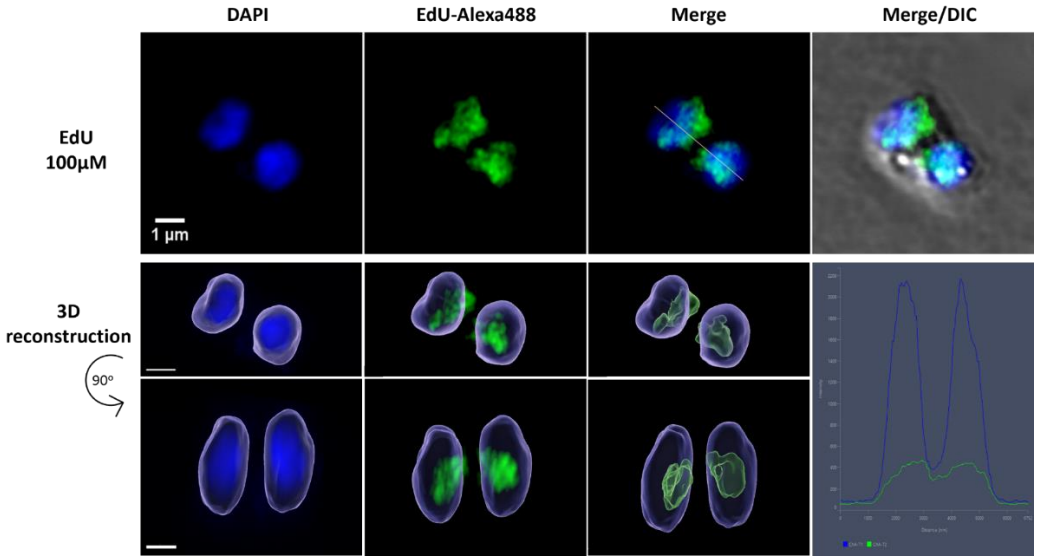


Figure 20. Visualization of active replication during schizogony after a 45-minute incubation 100 µM of EdU and 3D reconstruction*. Replicated DNA is stained with Alexa Fluor 488 (green) and total DNA is stained with DAPI (blue). The light pink dashed bar in the Merge 1 image represents the line drawn to plot the profiles of the signals shown below. Scale bar represents 1 µm.*

Knowing that a 45-min incubation with EdU provided reliable detection of DNA synthesis, I next performed a time-course experiment to better define the timing to apply the origin mapping techniques. I examined three different timepoints: 25, 29 and 35 hpi (Fig. 21). In agreement with the cytometry data, DNA replication is almost negligible at 25 hpi (Fig. 21a), with only 1.6% of all the analysed cells showing positive EdU signal (Fig. 21b, c). At 29 hpi, at least 18.9% of the population had actively replicated DNA in the preceding EdU incubation time (Fig. 21b), and most of the EdU+ cells displayed EdU labelling in a single nucleus (Fig. 21a, c). Finally, midway through progression of schizogony (at 35 hpi) we observed a sharp increase in the percentage of cells that had replicated their DNA, with 96.6% of the parasites showing positive EdU signal, and an average of 5 labelled nuclei per schizont (Fig. 21a-c) (Appendix 2).

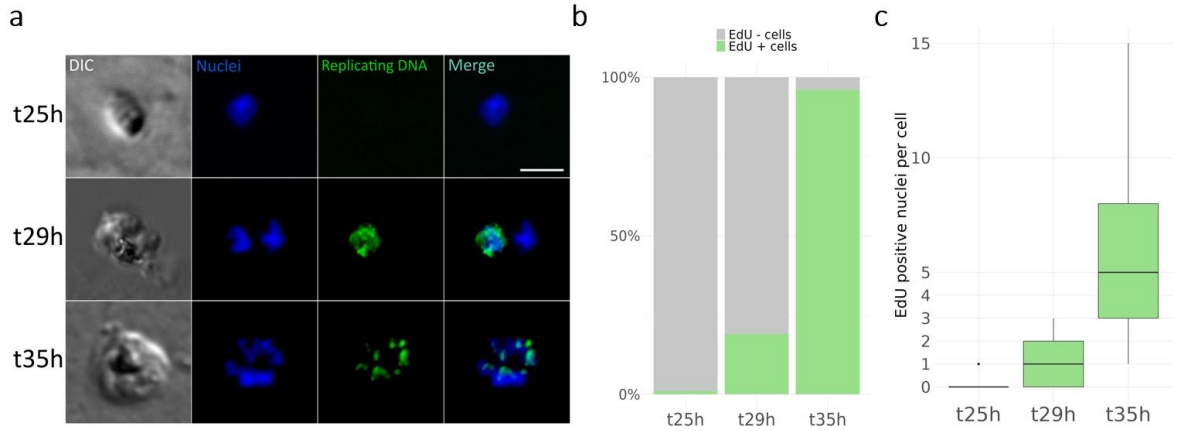


Figure 21. Immunofluorescence to measure DNA replication timing*. a) Visualization of active replication during schizogony after a 45 min incubation with 100 µM of EdU followed by click-chemistry labelling with a fluorophore at 25, 29 and 35 hpi. Total DNA is stained with DAPI and replicating DNA represents signal of EdU-Alexa fluor 488. Scale bar represents 2 µm. b) Percentage of parasites in which DNA replication took place (i.e., EdU positive) at 25, 29 and 35 hpi. At least one hundred parasites were counted for each timepoint (Appendix 2). c) Number of nuclei showing positive EdU signal per parasite. More than 40 parasites were counted for each timepoint (Appendix 2).*

Altogether, these results allowed us to set defined windows to map the origins of DNA replication of *P. falciparum*: 25 hpi (± 1h) as the window to study *Pf*ORC occupancy throughout the genome prior to initiation of DNA synthesis; and 29 hpi (± 1h) as the window to study replication initiation events during the initial S-phases of the intraerythrocytic cycle.

2. Licensed origins mapping through ChIP-seq of ORC1, and ORC2 at 25 hpi.

2.1. Introduction

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a widely used technique to map origins of replication [277], [278], [316], [317]. By pulling down proteins that form the pre-replicative complex we can map their binding sites throughout the genome (Fig. 22). Subunits of the ORC and MCM complexes are common targets of such experiments since they bind tightly to DNA and recruit other members of the replisome, directing the localization of the initiation of replication. ChIP-seq of *Pf*ORC proteins at the onset of replication allowed us to obtain a set of all the potential origins of replication throughout the genome, which helped us deepen our understanding of DNA replication initiation in *Plasmodium falciparum*.

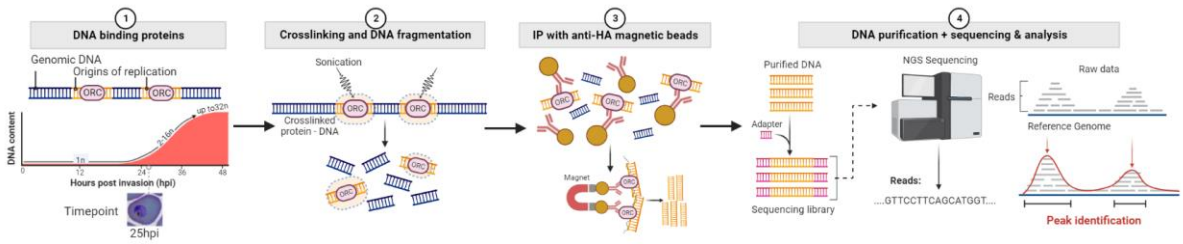


Figure 22. Schematic representation of the ChIP-seq protocol*. The workflow starts with a culture of parasites before the onset of replication (25 hpi). Proteins are crosslinked to DNA and subjected to sonication to shear DNA fragments. PfORC::HA proteins are immunoprecipitated along with the chromatin they are bound to with magnetic beads coupled to anti-HA antibodies. Chromatin is then purified, and libraries are prepared for sequencing. The bioinformatic pipeline includes mapping the sequencing reads to the reference genome and identifying the peaks (i.e., the binding sites of PfORCs). Created with BioRender.com*

The genome of *Plasmodium falciparum* encodes for homologues of the subunits ORC1, 2 and ORC5 of this complex, as well as for a putative ORC4 (PF3D7_1334100) and a putative ORC3 domain-containing protein (PF3D7_1029900). The expression level and localization throughout the asexual IDC of the subunits *Pf*ORC1, *Pf*ORC2 and *Pf*ORC5 have been previously characterized and they were found to be diffused in early stage parasites but forming intense distinct foci exclusively within the nucleus (perinuclear foci in the case of *Pf*ORC2) as the parasites mature into trophozoites and schizonts, when DNA replication takes place [252], [254], [255]. Additionally, *Pf*ORC5 was found to colocalise with *Pf*ORC1, suggesting their interaction to form the ORC complex like in other organisms. *Pf*PCNA1, a

protein commonly used in immunofluorescence experiments as a proxy for active replication visualisation, was also found to colocalise with *Pf*ORC5 during the replicative phases of the parasites cycle (from 26 hpi until 35 hpi) [252]. In addition, yeast complementation assays have validated the conserved role for *Pf*ORC1 and *Pf*ORC2 in DNA replication [255], [318]. All these results suggested that the role of *Pf*ORC proteins is likely to be conserved in *Plasmodium* parasites and ratified the choice of *Pf*ORC1, *Pf*ORC2 and *Pf*ORC5 proteins as targets to investigate initiation of DNA replication.

2.2. Results

2.2.1. Expression and localization of ORC subunits in Plasmodium falciparum

2.2.1.1. Comparison between *Pf*ORC1,2, 5 and their human/yeast homologues

A bioinformatic analysis using BLAST alignment to compare the human and yeast ORC protein sequences with their *P. falciparum* putative homologues (Fig. 23) showed a 38% and 32% sequence identity of *Pf*ORC1 with the *S. cerevisiae* and human ORC1 proteins, respectively (Fig. 23a); 22% and 27% of identity in the case of *Pf*ORC2 (Fig. 23b), and 20% and 27% for *Pf*ORC5 (Fig. 23c). Interestingly, *Pf*ORC1 was also found to have a AAA+ motif with ATPase activity [184] and high similarity to human Cdc6 (58% similarity and 37% sequence identity) and *Sc*Cdc6 (46% similarity and 27% sequence identity), suggesting that the *Pf*ORC1 subunit could act as an equivalent of the Cdc6 protein since the *P. falciparum* genome lacks a true homologue of this protein.

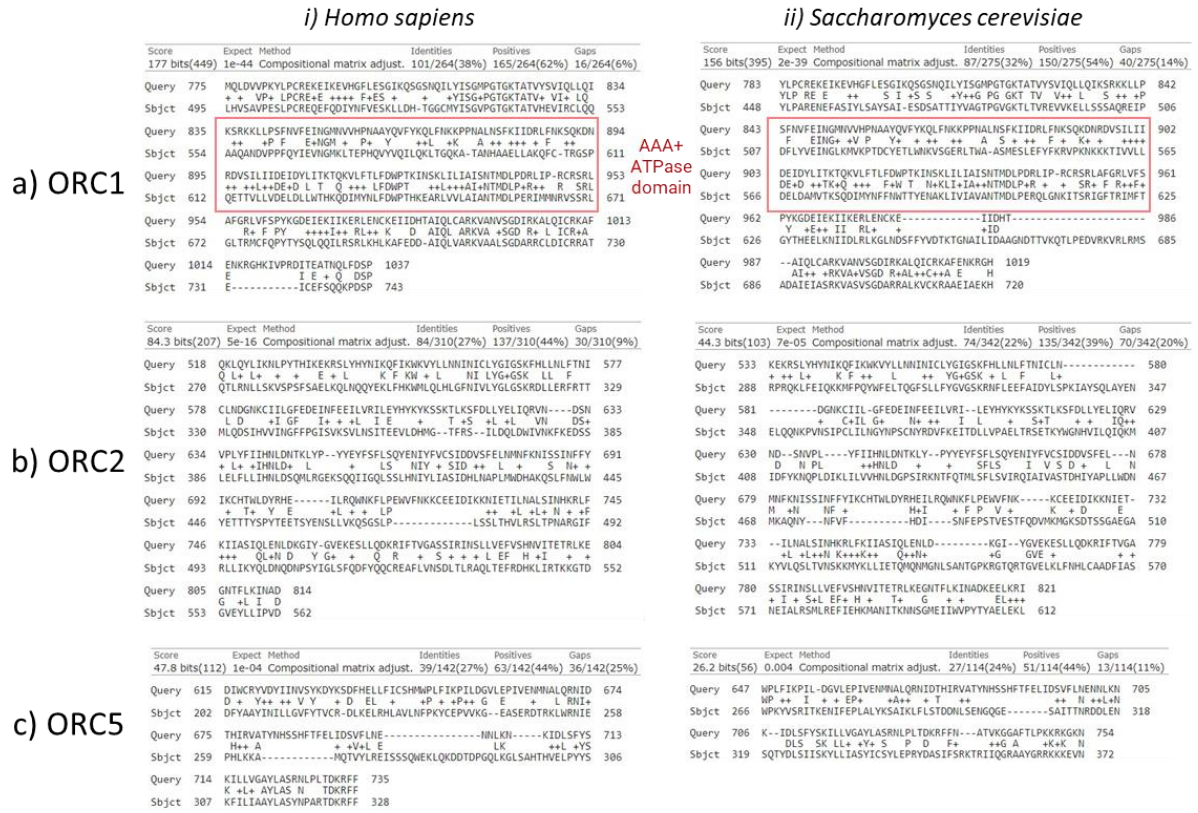


Figure 23. BLAST analysis of the protein sequence of the putative Plasmodium falciparum ORC proteins and their yeast and human homologues. *a) PfORC1 aligned to Homo sapiens ORC1 (i); or Saccharomyces cerevisiae ORC1 (ii). The AAA+ ATPase domain is indicated with a red rectangle. The result of alignment between HsORC1 and PfORC1 showed 38% of identity and 62 % similarity. When PfORC1 was aligned to ScORC1, the values obtained were 32% identity and 54% similarity. PfORC1 length is 1189 amino acids residues with a predictive size of 138.76 kDa. HsORC1 is 861 aa long and has a predicted molecular weight of 97.37 kDa. ScORC1 is 914 aa long and its predicted molecular weight is 104.4 kDa. b) The result of alignment between HsORC2 and PfORC2 showed 27% of identity and 44% similarity (i). When PfORC2 was aligned to ScORC2, the values obtained were 22% identity and 39% similarity (ii). PfORC2 length is 825 amino acids residues with a predictive size of 98 kDa. HsORC2 is 577 aa long and has a predicted molecular weight of 66 kDa. ScORC1 is 620 aa long and its predicted molecular weight is 61.25 kDa. c) The alignment between HsORC5 and PfORC5 resulted in 27% of identity and 44% similarity (i). In contrast, the alignment between PfORC5 and ScORC5 showed 24% identity and 44% similarity (ii). PfORC5 length is 899 amino acids residues with a predictive size of 104 kDa. HsORC5 is 435 aa long and has a predicted molecular weight of 50.29 kDa. ScORC1 is 479 aa long and its predicted molecular weight is 55.3 kDa.*

2.2.1.2. Generation of triple HA-tagged parasite lines

With the aim to perform the chromatin immunoprecipitation experiment followed by sequencing (ChIP-seq) to map the binding sites of the ORC complex in *Plasmodium falciparum*, the lab had generated three parasite lines, each containing the *Pf*ORC1, *Pf*ORC2 or *Pf*ORC5 proteins tagged with a triple Hemagglutinin (HA_3) epitope tag in the C-terminal region using CRISPR-Cas9 technology [319].

To do so, a two plasmid strategy was used: the pDC2-Cas9-hDHFR-yFCU [286] vector carried the Cas9 endonuclease along with the guide RNA and the WR99210 resistance cassette (hDHFR) while the pLN-HA_3 vector carried the homology regions (HR) necessary for the homologous recombination event, as well as the HA_3 tag. The principle of CRISPR/Cas9 technology relies on the Cas9-mediated double

strand break, which in *P. falciparum* is repaired by homologous recombination in a double recombination event between the homology regions of the plasmid and the endogenous wild type (WT) locus, leading to the integration of the tag [319], [320] (Fig. 24a), since the non-homologous end joining mechanism is absent in *P. falciparum* [321]. Transgenic parasites were selected with the drug WR99210, clones were isolated from these mutants by limiting dilution in the absence of drug pressure and their genotype was confirmed by PCR (Fig. 24b).

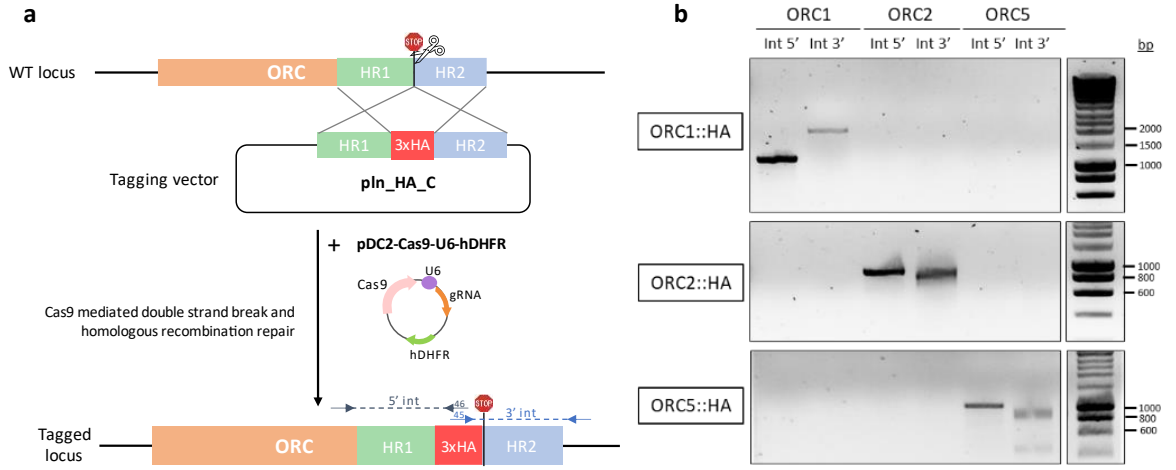


Figure 24. Generation of triple HA tagged parasite lines of ORC1, ORC2, ORC5*. a) Using the CRISPR/Cas9 system, the triple HA tag is inserted in the C-terminal end of the gene of interest. The pLN_HA_C plasmid contains the HA_3 tag along with the flanking 5' and 3' homology regions. The Cas9 mediated edition results in the generation of endogenously HA_3 tagged transgenic lines. b) Genotyping PCR analysis of genomic DNA of the transgenic parasite lines confirming the successful modification of the endogenous loci. Primers 45 and 46 are located within the HA_3 tag in opposite directions and pairing with opposite primers outside the HRs. Expected sizes are for PfORC1 1076 bp (5' int) and 1815 bp (3' int); for PfORC2 864 bp (5' int) and 793 bp (3'int); and for PfORC5 1044 bp (5'int) and 947 bp (3'int).*

2.2.1.3. Expression of ORC1, ORC2 and ORC5 in *P. falciparum*

After confirming the integration of the triple HA tag by genotyping PCRs, I next confirmed the expression of the *Pf*ORC::HA_3 proteins using western blot with anti-HA antibodies in the middle of schizogony (35 hpi) (Fig. 25). The observed size of *Pf*ORC1 is slightly smaller than what was expected whereas *Pf*ORC2 and *Pf*ORC5 show a bigger size, suggesting that the latter proteins undergo some type of post-translational processing. In particular, in the case of *Pf*ORC2, the more intense band that can be observed above the main thin one is probably due to the ubiquitinated status of *Pf*ORC2, since the presence of ubiquitin would increase the size 8 kDa, matching the difference between both bands. This was confirmed by the detection of ubiquitin in a *Pf*ORC2 sample previously done in the lab with anti-ubiquitin specific antibodies.

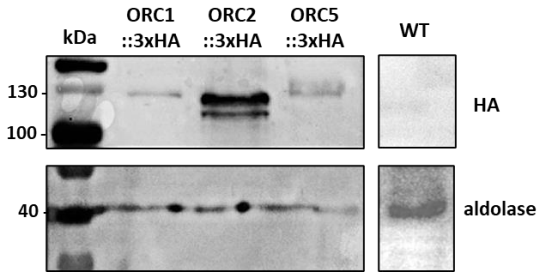


Figure 25. Expression of the PfORC::HA$_3$ proteins. *Detection of the triple HA tag in the three transgenic parasite lines and wild type Pf3D7 as negative control. The predicted sizes of the tagged orcs were 142.5kDa for PfORC1::HA, 102 kDa for PfORC2::HA and 108 kDa for PfORC5::HA. Pf Aldolase (FBPA, fructose-bisphosphate aldolase) was used as a loading control, with a predicted size of 38 kDa.*

I next investigated the subcellular localization of *Pf*ORC1 and *Pf*ORC2 by immunofluorescence microscopy at two timepoints: 25 hpi (G1 phase), and 35 hpi (when multiple nuclei per cell are undergoing DNA replication) (Fig. 26). A clear HA signal for *Pf*ORC1 and *Pf*ORC2 was detected mostly inside the nucleus, with some additional perinuclear accumulation, both at the beginning and the middle of schizogony (Fig. 26a). In addition, I measured the colocalization between the HA signal and the DNA in the nuclei stained with DAPI and confirmed the correlation between them, with a slightly higher correlation coefficient at t35h than at t25h (Fig. 26b).

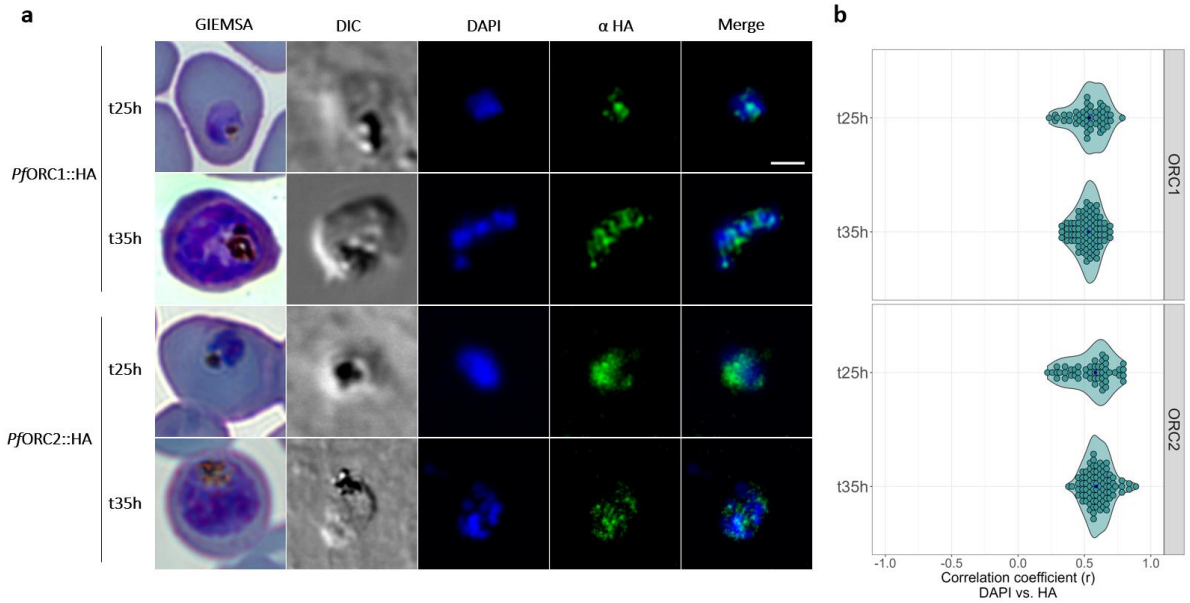


Figure 26. Immunofluorescence assay of PfORC1::HA$_3$ and PfORC2::HA$_3$. *a) Subcellular localization of the tagged proteins at 25 and 35 hpi. Parasites pictures from Giemsa smears are shown at the left. Total DNA is stained with DAPI in blue and the PfORC::HA$_3$ proteins in green. Scale bar is 2 µm. b) Correlation coefficient measured between the HA$_3$ signal from the PfORC proteins and the DAPI signal from the nuclei. Median correlation coefficients are 0.538 and 0.541 for PfORC1 at t25h and t35h, respectively; and 0.583 and 0.591 for PfORC2 at the respective timepoints.*

These observations are highly concordant with the subcellular localization of their human homologues, being found in distinct foci within the nucleus and also showing a certain amount of perinuclear localization [252], [254], [255]. Having confirmed the localization of the *P. falciparum* ORC proteins within the nuclei of the parasites, I then performed the ChIP-seq experiment with the

parasites that endogenously express HA-tagged *Pf*ORC1, *Pf*ORC2 and *Pf*ORC5 proteins at different stages of schizogony: t25h, t32h and t35h. These series of timepoints represent key points of the replicative cycle, as we have previously demonstrated: t25h before the first replication round, t32h before the second replication round and 35h when there are many nuclei replicating inside the same cell. I performed the experiment with two biological replicates for each protein at each timepoint following the protocol described in the Materials & Methods section. In the following section, I will focus on the description of the results obtained with the first timepoint.

2.2.2. ChIP-seq t25h: Quality Control

Following sequencing, a minimum of 10 million reads per sample were obtained. I performed quality control of the raw fastQ files and mapped the sequencing reads to the genome, after which I called the peaks for the IP and IgG samples using the input as control. The IgG sample consists on the incubation with a non-specific control immunoglobulin G (IgG) instead of the protein-specific antibody, allowing for the detection of unspecific signal and gives an indication of the assay background [322]. Peaks were called with a quality threshold of q ≥ 0.05. Finally, only the common peaks between the two replicates of each sample were kept and the unspecific peaks were removed (overlapping fragments were subtracted if ≥50 bp overlapped with the IgG control). With a final filtering step to discard peaks smaller than 50 bp I produced the final datasets that were used for further analysis. The specific numbers of reads and peaks obtained are shown in the table below.

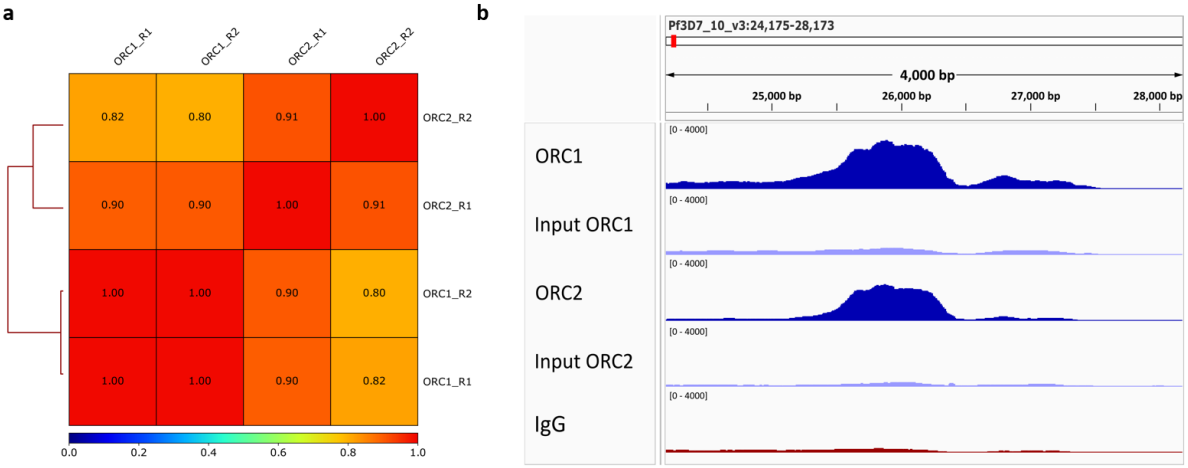
Table 2. Quality control of ChIP-seq t25h experiment.

	***Pf*ORC 1**		***Pf*ORC 2**		***Pf*ORC 5**	
	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2
Total sequences	60735162	20273562	18209482	21918172	14164295	12015206
Reads passed QC	60423748	20167714	18136874	21810938	14164295	12015206
Mapped reads	51073099	17684126	13993179	15461465	7013187	8698779
Genome coverage (≥1x)	93.7%	88.9%	87.5%	88.8%	79.2%	82.1%
Number of peaks called (q ≥ 0.05)	5867	3870	4994	4853	313	224
Pearson correlation r	1		0.91		0.99	
Overlapping peaks	3560		4215		171	
Peaks after - IgG	3801		3924		78	
Peaks ≥ 50bp	**2693 peaks**		**2748 peaks**		**53 peaks**	

As it can be observed in table 2, unfortunately, the data obtained after ChIP-seq of *Pf*ORC5 did not yield results with enough quality, since the number of peaks called was very low and a big proportion of them was removed as they overlapped with the IgG control. Therefore, I decided to not consider it for further analysis. From this moment on, the analysis will focus on the results obtained with the peaks from the *Pf*ORC1 (2693) and *Pf*ORC2 (2748) experiments (Appendix 3).

2.2.3. Binding sites of PfORC1 and PfORC2 at t25h

The biological replicates of *Pf*ORC1 and *Pf*ORC2 were highly correlated (Fig. 27a), and there was a strong association between the final sets of binding sites of *Pf*ORC1 versus *Pf*ORC2 (containing only the common peaks between both replicates), with a Fisher's test p-value < $1.6e^{-22}$. In fact, 70% of the peaks of *Pf*ORC1 overlapped with the *Pf*ORC2 dataset and they displayed highly concordant binding profiles (Fig. 27b-e), suggesting that these proteins bind to the DNA in a non-random manner, which was expected as these proteins are meant to bind to the origin DNA in a complex.



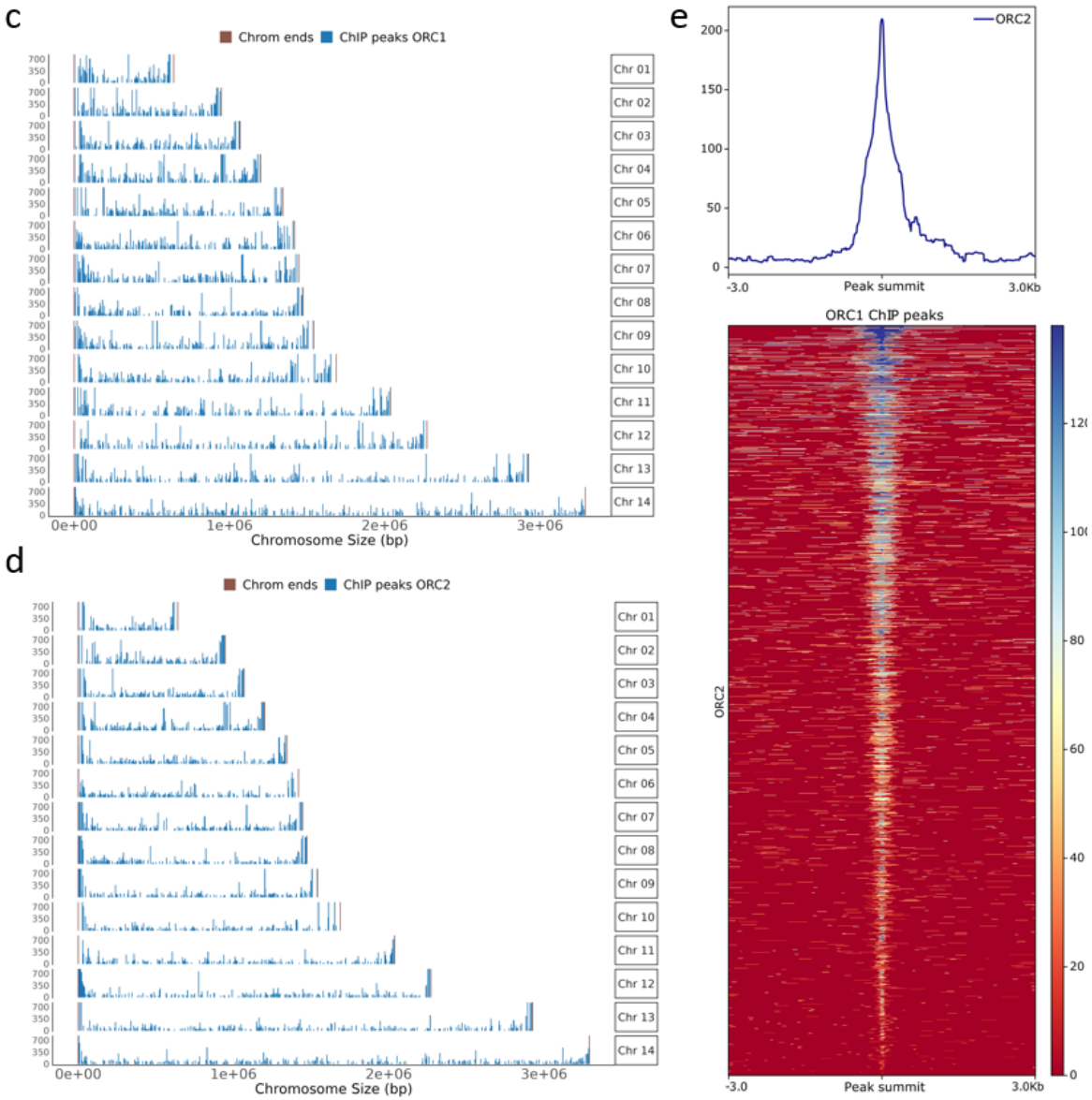


Figure 27. PfORC1 and PfORC2 show highly concordant binding sites. *a) Heatmap showing Pearson's correlation coefficient between each ChIP-seq replicate of PfORC1 and PfORC2. b) IGV snapshot showing an example of overlapped PfORC1 and PfORC2. The input signal used to call the peaks and the IgG unspecific control signal are also shown. c) Genome wide distribution of PfORC1 and d) PfORC2 sites. e) Colocalization between PfORC1 and PfORC2. Heatmap of enrichment of PfORC1 over PfORC2 with 3 kb upstream and downstream flanking regions. The colour bar indicates the range of intensities based on ChIP enrichment, from red to blues for lower to high enrichment values. On the y axis, each row shows the PfORC2 ChIP peaks centered in their summit and they are sorted in descending order based on the mean colocalization of the datasets.*

The intersection of the peaks of *Pf*ORC1 with *Pf*ORC2 generated the final dataset used for downstream analyses comprising 1861 putative binding sites of the ORC complex throughout the genome, forming the licensed origins of replication (Appendix 3) (Fig. 28a). The set of *Pf*ORC$_{1\text{-}2}$ binding sites displayed a median peak length of 442 bp, which is slightly shorter than the median peak sizes of the human and yeast ORCs [317], [323], [324]. Also, the median *Pf*ORC$_{1\text{-}2}$ peak size was around half the size of *P. falciparum* histone post-translational modifications (PTMs), which are characterized by broad peaks; and longer than typical DNA sequence-dependent transcription factor binding sites like those of the ApiAP2 transcription factors (Fig. 28b).

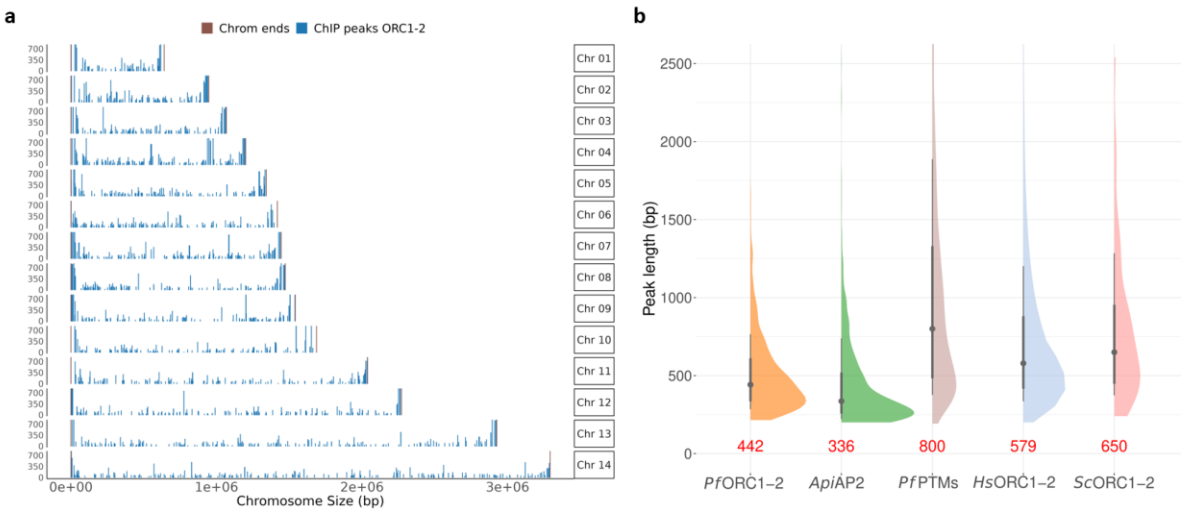


***Figure 28. Distribution of the PfORC$_{1-2}$ binding sites**. a) Genome wide distribution of the peaks common to PfORC1 and PfORC2. b) Distribution of peak sizes of the PfORC$_{1-2}$ sites and comparison to the lengths of binding sites of either transcription factors (PfAP2-G* [325] *and PfAP2-I* [326]*)* [327]*, post-translational modifications (PTMs) (including H3K9me3, H3K4me1 and H2A.Z)* [150] *and binding sites of ORC 1 and 2 in other species (Homo Sapiens, human; and Saccharomyces cerevisiae, yeast)* [176], [317], [323], [324]*. The grey dot points the median speed (indicated in red below) and the thick and thin grey lines indicate the 50 and 75% intervals, respectively.*

Moreover, the *Pf*ORC$_{1-2}$ peaks were distributed throughout the genome, with a median distance between each other of 5378 bp for the *Pf*ORC1 peaks, 5437 bp for *Pf*ORC2 and 7246 bp for the intersected *Pf*ORC$_{1-2}$ dataset (Fig. 29a). The final dataset maintained for downstream analysis contained only the peaks common to both *Pf*ORCs (Fig. 28a). This distance between *Pf*ORC binding sites was as long as 106 655 bp, although only 10% of the *Pf*ORC$_{1-2}$ sites (3.6% for *Pf*ORC1 and 3.3% for *Pf*ORC2) were separated by more than 30 kb.

Since 25.6% of the *Pf*ORC$_{1-2}$ sites were separated by less than 3 kb (*Pf*ORC1: 32.7%; *Pf*ORC2: 32.5%), I investigated if the *Pf*ORC$_{1-2}$ complex could be found in clusters. Indeed, I detected 261 clusters with 3 or more *Pf*ORC$_{1-2}$ binding sites within 10 kb (*Pf*ORC1: 557 clusters; *Pf*ORC2: 577 clusters). These clusters were also found distributed throughout the genome and included a total of 635 *Pf*ORC$_{1-2}$ sites, which comprises 34.1% of the whole dataset (Fig. 29b).

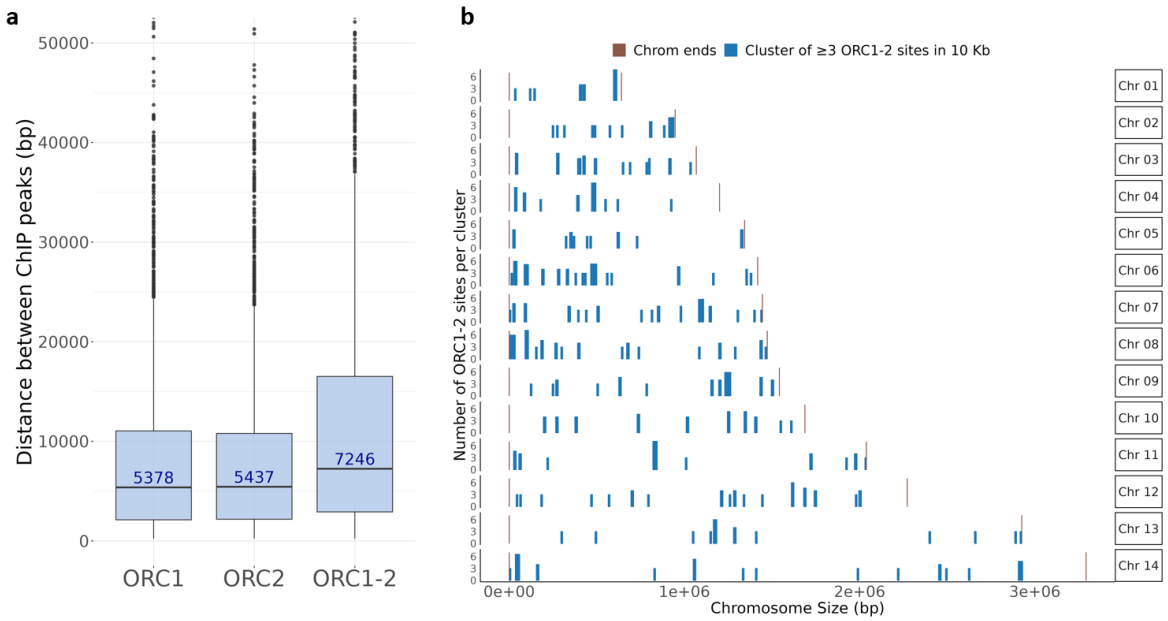


Figure 29. PfORC$_{1-2}$ binding sites are found closer than expected by chance. *a) Inter-origin distance (bp) measured for PfORC1 and PfORC2 and the PfORC$_{1-2}$ intersected dataset. Median distance is shown. b) Distribution of PfORC$_{1-2}$ clusters throughout the chromosomes. A cluster is defined as a 10kb region of the genome containing 3 or more PfORC$_{1-2}$ sites.*

2.2.3.1. Analysis of association with G4FSs

Since the origins of replication in the budding yeast *Saccharomyces cerevisiae* are sequence dependent, being marked by autonomously replicating sequences (ARS), I wanted to explore whether there was any sequence specificity found in *P. falciparum* ORC binding sites, as it had been previously suggested [328]. After scanning the *Pf*ORC$_{1-2}$ dataset in the MEME suite motif finder, no consensus sequence was found, which is consistent with what has been reported in metazoans [163], [164]. Nevertheless, I found that, despite *P. falciparum* harbouring one of the most skewed base pair compositions with >80% of the genome being AT-rich, the average GC content of the *Pf*ORC$_{1-2}$ peaks was 27%, which is strikingly higher than the whole genome GC content (19.3%) (Fig. 30a). This could be explained by the fact that the *Pf*ORCs, like in other metazoans, might be preferentially recruited to DNA led by certain DNA characteristics such as GC-rich DNA secondary structures like G-quadruplexes rather than the DNA sequence itself. As a result, I verified if the *Pf*ORC$_{1-2}$ dataset was enriched in potential G-quadruplex forming sequences (G4FS). To do so, I used the G4Hunter online tool [169], [329], which predicts the propensity to form G-quadruplexes (G4) and assigns a score to a given nucleic acid sequence on the basis of G-richness and G-skewness. As per calculated by the G4Hunter algorithm, the 23.3 Mb of the whole *P. falciparum* genome, with a GC content of 19.3%, harbours 6033 genomic sites with propensity to form G4s (0.3/1000 bp). Comparatively, the *Pf*ORC$_{1-2}$ binding sites had 564 potential G4FS over 0.65 Mb (0.9/1000 bp), which corresponded to a 3-fold enrichment (hypergeometric p-value= $2.47e^{-132}$). In fact, when I compared the *Pf*ORC$_{1-2}$ binding profile with the

genome-wide distribution of G4FS in *Plasmodium falciparum* (G4H1.2) [167], I obtained a significantly positive association value (odds ratio= 3.04, p-value= $2.1e^{-28}$) and found that 342 of the annotated G4FS either overlapped, or were within 500 bp of a *Pf*ORC$_{1\text{-}2}$ site (253 peaks, which represents 13.6% of the total *Pf*ORC$_{1\text{-}2}$ binding sites) (Fig. 30b-d). Additionally, all high-score *Pf*ORC$_{1\text{-}2}$ peaks either overlapped or were in close proximity of a G4FS (Fig. 30e, f), suggesting that G-quadruplexes probably play an important role in the recruitment of the ORC complex in *P. falciparum* parasites.

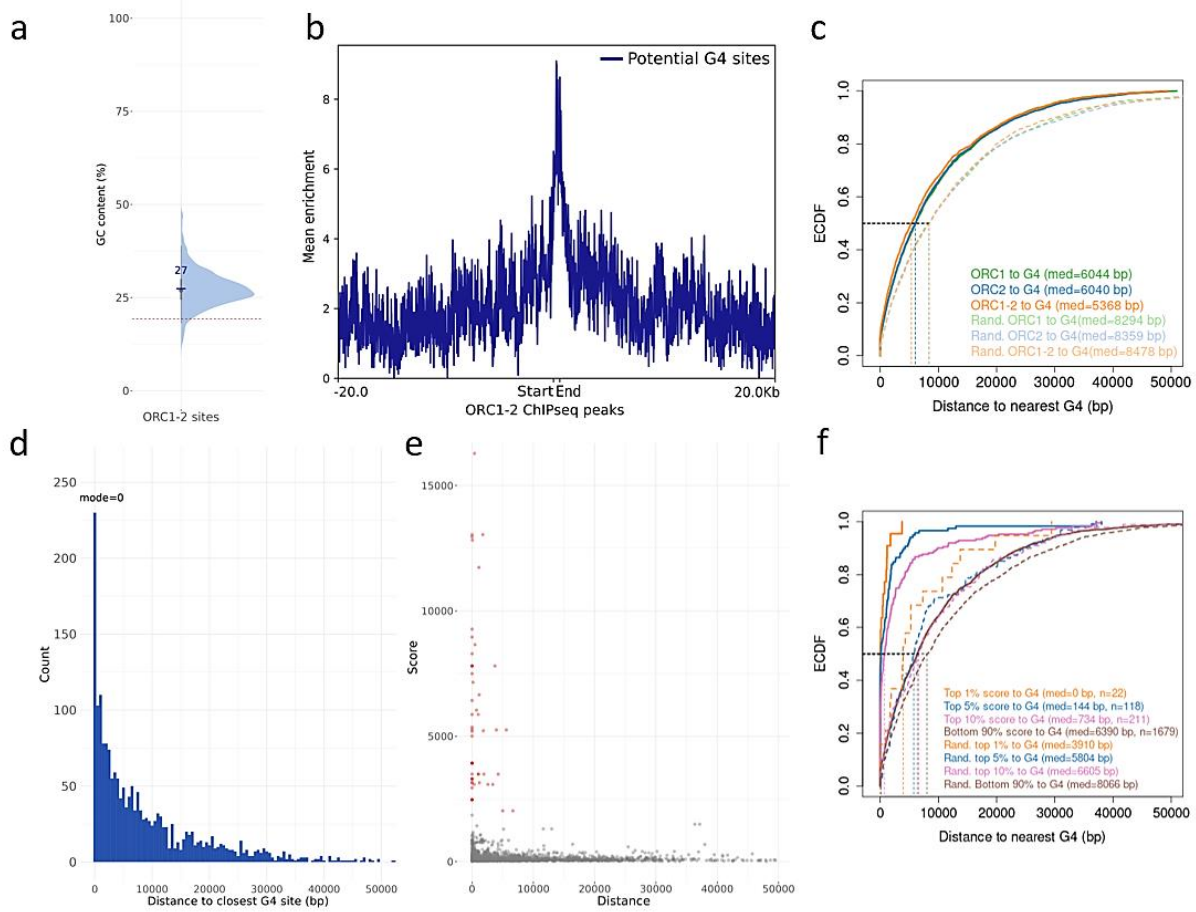


Figure 30. PfORC$_{1\text{-}2}$ binding sites are associated with G-quadruplexes*. a) Percentage of GC content within the PfORC$_{1\text{-}2}$ binding sites. The horizontal bar depicts the GC content of the whole P. falciparum genome (19.3%), and the average GC content of PfORC$_{1\text{-}2}$ sites is indicated above; thick and thin vertical lines indicate the 50% and 95% intervals of the data. b) Enrichment of G-quadruplex forming sites (G4FS) within and around the PfORC$_{1\text{-}2}$ sites. Peak lengths were scaled and are defined by "start" and "end" labels. c) Empirical Cumulative Distribution Function (ECDF) of the distances between ChIP peaks (PfORC1 in green, PfORC2 in blue, and PfORC$_{1\text{-}2}$ in orange) to the closest G4FS. Randomized datasets are represented as dashed lines. Median distances are indicated. d) Histogram of the distances between a given PfORC$_{1\text{-}2}$ site and the closest G4FS. e) Distribution of the enrichment score of PfORC$_{1\text{-}2}$ peaks in relation to its distance to the nearest G4FS. Scores 1 to 2000 are represented in grey and 2000 onwards are shown in red (max score = 16 800). f) Empirical Cumulative Distribution Function (ECDF) of the distances between the ChIP peaks of different MACS2 enrichment scores (top 1 %, top 5%, top 10 %, or the bottom 90 %) to the closest G4FS. Randomized datasets are represented as dashed lines. Median distances and number of peaks are indicated.*

2.2.3.2. *Pf*ORC$_{1\text{-}2}$ binding sites and transcription

The distribution and selection of replication origin sites in other eukaryotic systems has been shown to be associated with the presence of epigenetic features such as nucleosome positioning, histone

post-translational modifications (PTMs) and CpG islands [330]. For instance, ORC binding sites in humans are located in regions of open chromatin, marked by specific active chromatin PTMs (H3K27ac, H3K9ac, H3K4me1, and H3K4me3) and colocalize with transcription start sites (TSS) [317], [323]. In contrast, in *S. cerevisiae,* ORC binding sites are generally located in intergenic regions and are excluded from TSSs [176].

To address the contribution of different histone PTMs to *Pf*ORC recruitment I calculated the Jaccard statistic between the distribution of *Pf*ORC$_{1\text{-}2}$ binding sites and several available ChIP-seq datasets of *P. falciparum* histone PTMs captured at a cognate timepoint. The Jaccard statistic measures similarities by calculating the amount of overlap between datasets: 1 meaning full overlap, and 0 no overlap at all. The datasets I used for the comparisons included H3K9me3 [151], H3K9ac [147], H3K4me3 [147], H3K27ac [150], H3K18ac [150], H3K4me1 [150] and the histone variant H2A.Z [150]. As it can be observed in Fig. 31, the Jaccard statistic results between the *Pf*ORC$_{1\text{-}2}$ and *Pf*PTMs were low, ranging from 0.005 to 0.063, suggesting that PTMs are not likely to be directly involved in *Pf*ORC recruitment.

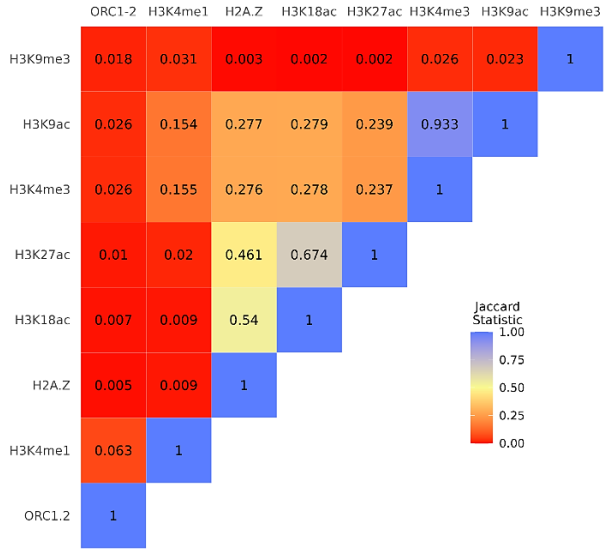


Figure 31. Association between PfORC$_{1\text{-}2}$ and PfPTMs*. Heatmap showing the Jaccard statistic of overlap between different P. falciparum histone modifications and the PfORC$_{1\text{-}2}$ peaks. Values range from 0 (red, no overlap) and 1 (blue, complete overlap).*

To investigate whether active transcription played a role in the recruitment of the ORC complex, I studied the distribution of the *Pf*ORC$_{1\text{-}2}$ sites relatively to the coordinates of core genes (euchromatin regions that show no presence of H3K9me3) [130]. I found that both the TSS (transcription start sites) and TTS (transcription termination sites) of those genes were depleted in *Pf*ORC$_{1\text{-}2}$ binding (Table 3, TSS: p-value= $6.4e^{-54}$ (anticorrelation), odds ratio= 0.4; TTS: p-value= $8.7e^{-25}$ (anticorrelation), odds ratio= 0.43) (Fig. 32a). Instead, *Pf*ORC$_{1\text{-}2}$ sites accumulated within the body of core genes (Fig. 32a and Table 3; p-value= $4.3e^{-45}$, odds ratio = 2.055). Additionally, the association between *Pf*ORC$_{1\text{-}2}$ sites and

the body of active genes was stronger when the expression percentile cut-off was increased to include only the 25 % most expressed genes, which were three times more likely to contain a *Pf*ORC$_{1\text{-}2}$ binding site that what would be expected by chance (p-value = 1.82e^{-61}, odds ratio= 2.9). Accordingly, the median percentile of expression of the 889 genes enriched in *Pf*ORC$_{1\text{-}2}$ binding sites was 71, in contrast to the percentile of a random subset of core genes, which was 53 (Fig. 32b).

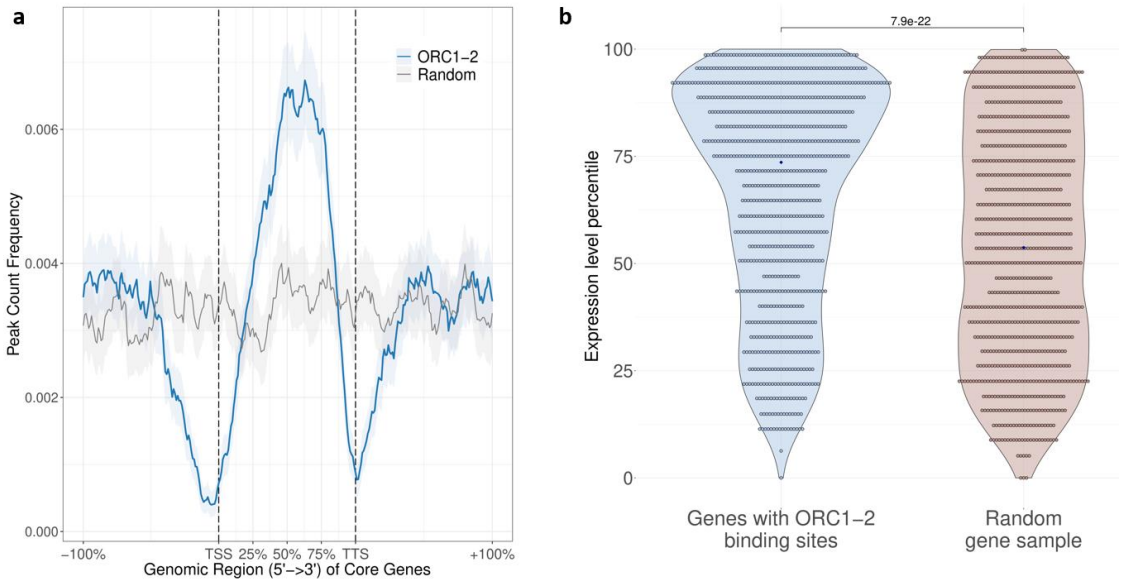


Figure 32. Enrichment of PfORC$_{1\text{-}2}$ within P. falciparum core genes. *a) Enrichment of PfORC$_{1\text{-}2}$ sites (blue) over scaled coordinates of P. falciparum core genes (H3K9me3 depleted genome). A randomized (control) dataset is shown in grey. TSS corresponds to the Transcription Start Site and TTS to the Transcription Termination Site of each gene. Regions of equal length to each gene are included upstream and downstream the TSS and TTS, respectively. b) Violin plots depicting the expression percentile of the genes displaying an enrichment of PfORC$_{1\text{-}2}$ binding sites (n= 889 genes) and a randomized dataset of equal number in brown. The result of a two-sided Wilcoxon test is shown.*

Moreover, ORC proteins are known to be associated with the heterochromatin protein HP1 in other organisms like *Drosophila* and humans [331]–[335], and even the N-terminal region of *Pf*ORC1 has been shown to bind to telomeric DNA and regulate *var* gene silencing in *Plasmodium falciparum* [253], [336]. For these reasons, I analysed *Pf*ORC$_{1\text{-}2}$ binding positioning relatively to the coordinates of several variant antigen gene families found in heterochromatin: *var* genes and *rifin/stevor* genes. I found a strong association between *Pf*ORC$_{1\text{-}2}$ sites and *var* genes, which was true for both *Pf*ORC subunits (Table 3). Regarding the distribution of the *Pf*ORC$_{1\text{-}2}$ sites relatively to the coordinates of these genes, there was a strong enrichment of *Pf*ORC$_{1\text{-}2}$ sites preceding the TSS of *var* genes, specifically at around -1.5 kb (Fig. 33a, Table 3). Whilst there was also an enrichment within *var* gene bodies, this was localised to three sharp and distinct regions including one at the very end of the coding sequence, spanning the TTS. Interestingly, this pattern was seen for both *Pf*ORC1 and *Pf*ORC2 and it was not just a feature of *Pf*ORC1 [253]. In total, 51 out 61 var genes contained *Pf*ORC$_{1\text{-}2}$ binding sites. In addition, since both *var* genes [337] and the *Pf*ORC$_{1\text{-}2}$ binding sites are associated with G4FS I next looked at the overlap between the three of them. The enrichment of *Pf*ORC$_{1\text{-}2}$ binding sites in *var* genes was

adjacent to the G4FS enrichment on the *var* gene promoters, suggesting that the *Pf*ORCs may be recruited by the G4 structures and bind to DNA directly beside them (Fig. 33a). To ensure that this enrichment in promoters was not a common feature of the heterochromatin environment, I looked at the *Pf*ORC$_{1\text{-}2}$ distribution on the other H3K9me3 genes which include other virulence factors such as *rifin* and *stevor* genes. In contrast, there was a depletion of *Pf*ORC$_{1\text{-}2}$ binding sites in TSS and TTS, akin to the *Pf*ORC$_{1\text{-}2}$ binding profile in core genes (Table 3, Fig. 33b).

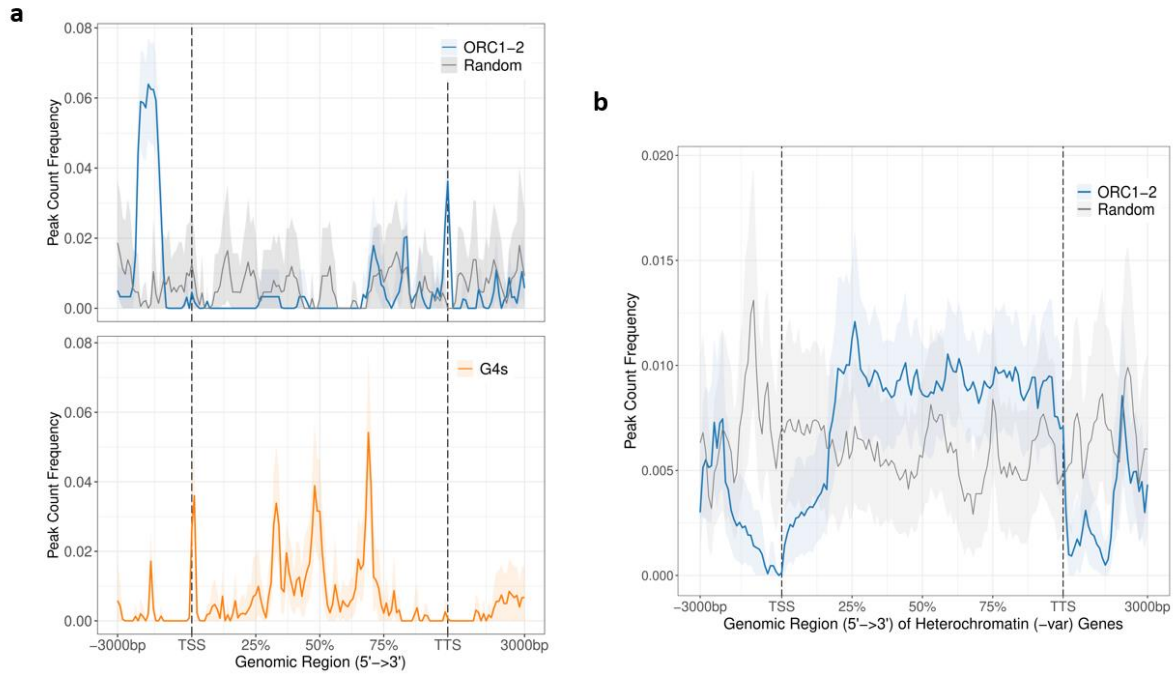


Figure 33. Enrichment of PfORC$_{1\text{-}2}$ in heterochromatin. *a) Enrichment of PfORC$_{1\text{-}2}$ sites (blue) and G4FS (orange) over scaled coordinates of the 60 P. falciparum var genes. A randomized (control) dataset is shown in grey. TSS corresponds to the transcription Start Site and TTS to the transcription termination site of each gene. Regions of 3000 bp are included upstream and downstream the TSS and TTS, respectively. b) Enrichment at Heterochromatin genes excluding the 60 var genes.*

Table 3. Correlation between PfORC$_{1\text{-}2}$ binding sites and core, var, and heterochromatin (except var) genes. *Fisher's test results are shown.*

		Core active genes t25h		*var* genes		H3K9me3 (- *var*)	
		Gene body	TSS 3kb	Gene body	TSS 3kb	Gene body	TSS 3kb
***Pf*ORC$_{1\text{-}2}$**	P-value	$4.3e^{-45}$	$6.4e^{-54}$ (antic.)	0.0614	$5.12e^{-06}$	$4.05e^{-16}$	0.835
	Odds ratio	2.055	0.4	Inf*	3.266	2.613	0.973
***Pf*ORC1**	P-value	$3.42e^{-63}$	$3.6e^{-76}$ (antic.)	0.118	0.00275	$7.88e^{-19}$	0.998
	Odds ratio	2.177	0.411	Inf*	2.115	2.545	0.997
***Pf*ORC2**	P-value	$1.02e^{-96}$	$4.93e^{-84}$ (antic.)	0.628	0.000265	$5.91e^{-16}$	0.006 (antic.)
	Odds ratio	2.623	0.393	Inf*	2.543	2.361	0.764

*inf: all var genes overlapped with the *Pf*ORC$_{1\text{-}2}$ dataset. Green color indicates positive correlation and red no correlation or anticorrelation.

2.2.4. Additional timepoints throughout schizogony

Furthermore, I performed the ChIP-seq experiment at two additional timepoints during schizogony: just before the second replication round (t32h) and at the middle of schizogony, when there are many nuclei replicating simultaneously (t35h). Unfortunately, the results obtained were of low quality (Table 4), since after removing the unspecific IgG signal, a very low number of peaks was retained. As a result, the data could not be used for subsequent analyses.

Table 4. ChIP-seq of PfORC1, PfORC2, and PfORC5 at t32h and t35h.

Dataset			Peaks (q ≥ 0.05)	Overlapping peaks	Pearson correlation r	Peaks after - IgG	**Peaks ≥ 50bp**
***Pf*ORC 1**	T32h	Rep 1	470	322	0.99	230	**155**
		Rep 2	628				
	T35h	Rep 1	1666	933	0.98	482	**207**
		Rep 2	1507				
***Pf*ORC 2**	T32h	Rep 1	367	221	0.73	142	**92**
		Rep 2	360				
	T35h	Rep 1	251	183	0.79	90	**54**
		Rep 2	587				
***Pf*ORC 5**	T32h	Rep 1	489	366	0.91	208	**157**
		Rep 2	627				
	T35h	Rep 1	595	266	0.99	145	**104**
		Rep 2	445				

However, before my arrival to the lab, a pilot ChIP-seq experiment of the *Pf*ORC2::HA_3 line was performed at 35 hpi, so the results of this experiment were included in the analysis. After the intersection of the two replicates, the subtraction of the unspecific IgG peaks and the filtering of the small (< 50 bp) peaks, I obtained a dataset of 4501 *Pf*ORC2 t35h peaks. They were 439 bp long on average, similarly to the t25h *Pf*$ORC_{1\text{-}2}$ peaks, and displayed a median IOD of 1827 bp, much lower than the t25h *Pf*$ORC_{1\text{-}2}$ dataset, which can be expected since the number of binding sites detected in the former is more than two times higher (Fig. 34a). Again, these peaks did not show a random distribution and were grouped in clusters. I found 1417 clusters of 3 or more *Pf*ORC2 peaks at t35h (Fig. 34b). Up to 14 peaks were found in a single cluster, with a median of 4 peaks per cluster. In total, 3690 *Pf*ORC2 sites were contained inside a cluster, making up for 82% of the total number of peaks.

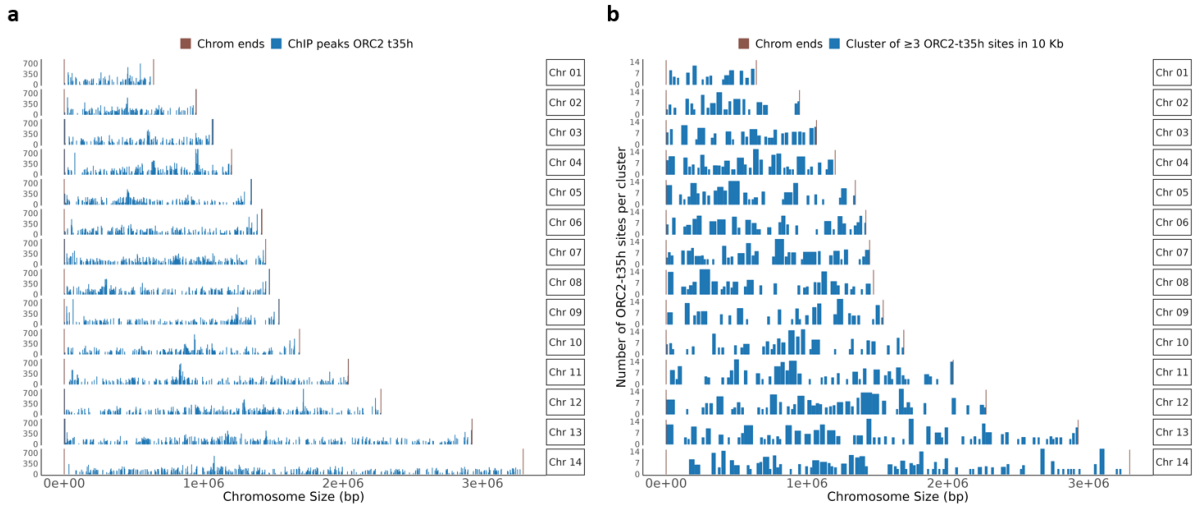


Figure 34. Distribution of binding sites of PfORC2 at t35h. *a) Individual ORC2 t35h peaks across the chromosomes. b) Distribution of clusters of ≥3 peaks within 10 kb. Height of bars represents the number of peaks found in each cluster.*

However, the binding sites of *Pf*ORC2 at t35h show a significantly different distribution than those at t25h (p-value=2.36e^{-27} (anticorrelation), odds ratio = 0.427), with only a small proportion of peaks that overlap between the two timepoints (109 overlapping peaks) and a median distance of 2706 bp between each t35h peak to the closest t25h peak. Moreover, in contrast to the t25h sites, the results of the analysis of the GC content of the *Pf*ORC2-t35h binding sites show that these are not especially GC rich but display a GC content similar to that of the whole genome (Fig. 35a). Still, they are located significantly closer to G4FS than expected if the distribution was random (p-value= 1.52e^{-16}, odds ratio = 2.142), with a median distance of 6681 bp between each *Pf*ORC2 site and the closest G4FS. Similarly to the t25h dataset, the peaks overlapping or located in closest proximity to these secondary structures were the peaks that display the highest scores (Fig. 35b, c). These results are consistent with what we obtained with the t25h timepoint, suggesting that although the binding sites of the ORC complex may change over the course of schizogony, G-quadruplexes play an important role in the recruitment of the ORC complex in *P. falciparum* parasites both at the beginning and at a more advance stage of schizogony.

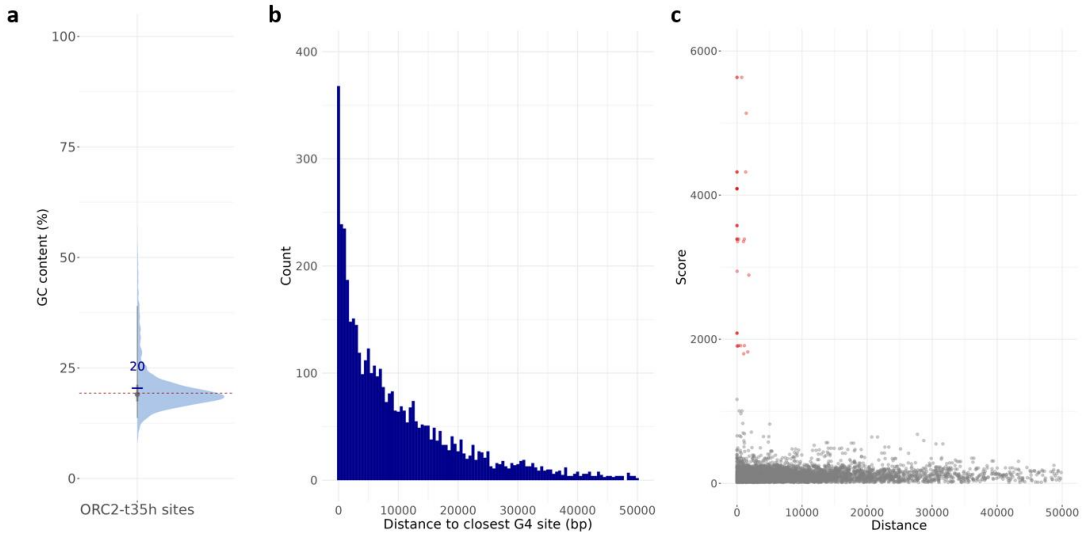


Figure 35. PfORC2 t35h binding sites are not GC rich but show an association with G4FS*. a) Percentage of GC content within the PfORC2 binding sites. The horizontal bar depicts the GC content of the whole P. falciparum genome (19.3%) and the median content of the PfORC2 t35h sites is depicted in blue. b) Histogram of the distances between a given PfORC2 site and the closest G4FS. c) Enrichment scores of PfORC2 peaks in relation to its distance to the nearest G4FS. Scores 1 to 1500 are represented in grey and 1500 onwards are shown in red (max score = 5630).*

In addition, I also found a depletion of *Pf*ORC2 t35h peaks both at TSS and TTS of core genes, and a very strong enrichment within the gene bodies (p-value = $2.15e^{-102}$, odds ratio = 2.376), akin to what I observed with the t25h dataset (Fig. 36a). At this advanced stage of schizogony, higher expression levels correlate, although not as much as at the beginning of schizogony, with *Pf*ORC2 binding, since the median percentile of the 1294 genes enriched in *Pf*ORC2 was 57 (vs. 52 in the random gene sample) (Fig. 36b). These results are consistent with what we observed with the t25h dataset; despite the binding sites of both timepoints being different, high transcriptional activity probably leads to a more accessible chromatin that promotes the replisome machinery binding to DNA.

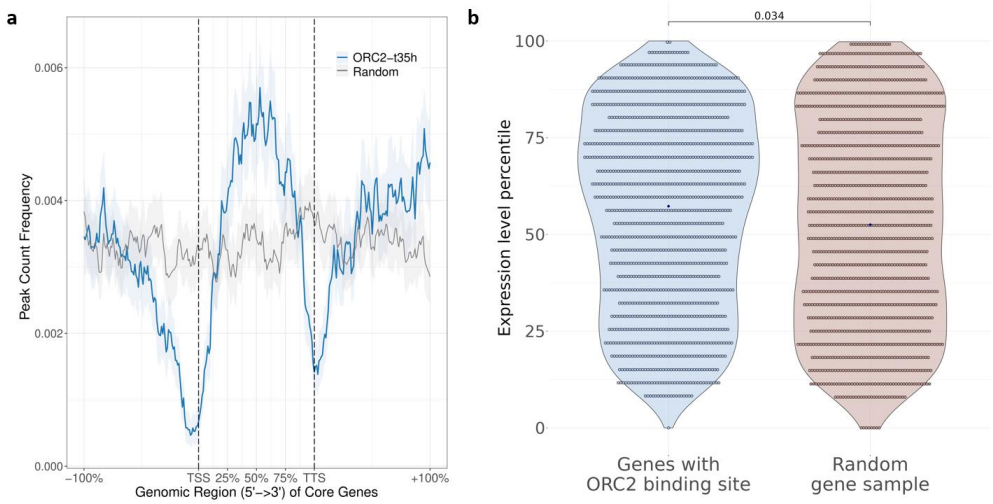


Figure 36. Enrichment of PfORC2 t35h binding sites within P. falciparum core genes. *a) Enrichment of PfORC2 sites (blue) over scaled coordinates of P. falciparum core genes. A randomized (control) dataset is shown in grey. TSS corresponds to the transcription Start Site and TTS to the transcription termination site of each gene. Regions of equal length to each gene are included upstream and downstream the TSS and TTS, respectively. b) Violin plots depicting the expression percentile of the genes displaying an enrichment of PfORC2 binding sites (n= 1294 genes) and a randomized dataset of equal number in brown. The result of a two-sided Wilcoxon test is shown.*

Finally, I investigated whether *Pf*ORC2 binding sites at t35h show any association with clonally variant genes in heterochromatin such as the *var* gene family. Indeed, there was a strong association between *Pf*ORC2 sites and heterochromatin regions characterised by the presence of H3K9me3 (p-value= $8.58e^{-27}$, odds ratio= 2.620). In fact, this correlation was marked by the *var* genes, as all 61 *var* genes overlapped with a *Pf*ORC2 t35h site. In turn, if the *var* genes were excluded from the heterochromatin regions, this association was almost completely lost (p-value = 0.0422, odds ratio = 1.242). This strong correlation between *Pf*ORC2 and *var* genes confirms the t25h results and suggests an potential role of this subunit on *var* gene silencing, both at the beginning and middle of schizogony [253], [336].

3. Active origins mapping

Since not all the licensed origins are activated in each replication round, to obtain a genome-wide cartography of the active origins of replication, I have mapped the sites of active DNA synthesis using two different strategies: sequencing short DNA nascent strands (SNS-seq) at the onset of the first S-phase of *P. falciparum* schizogony (t29h); and mapping the incorporation of the thymidine analogue BrdU into replicating DNA at two different timepoints during schizogony (t29h and t35h). By combining the data from these strategies, I have been able to assemble a robust set of putative active origins of replication.

3.1. SNS-seq

3.1.1. Introduction

Short Nascent Strand sequencing (SNS-seq) [161] consists on the isolation and purification of the newly synthesised RNA-primed strands of DNA. Small DNA fragments are isolated on a sucrose gradient gel and subsequently digested with λ-exonuclease which will digest the DNA fragments that do not carry the 5' RNA cap. These nascent DNA fragments are then sequenced and mapped (Fig. 37). In this technique, an additional negative control sample is sequenced to exclude any signal due to a possible bias in the λ- exonuclease treatment, consisting on a control DNA sample treated with RNAse A prior to λ-exonuclease treatment [165]. This results on the removal of the RNA 5' cap of nascent strands, which will then be digested with the λ-exonuclease, obtaining a negative control sample devoid of signal that represents the potential background noise.

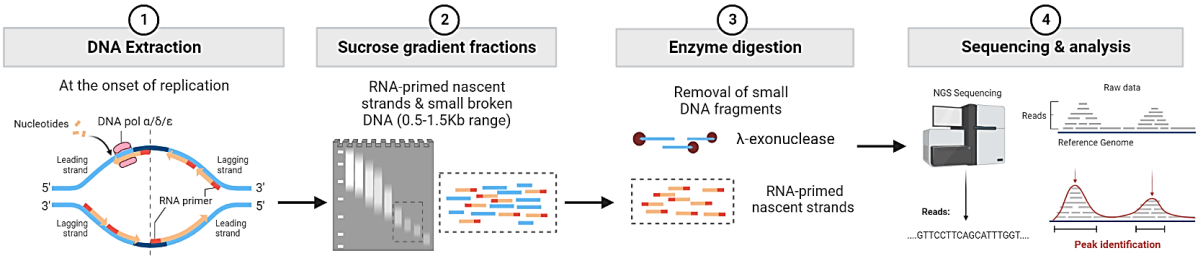


Figure 37. Scheme depicting the SNS-seq protocol overview. *DNA is extracted from a culture of replicating parasites at the onset of schizogony (29 hpi). RNA-primed short nascent strands are isolated from sucrose gradient gels and small broken DNA fragments are removed by digestion with λ-exonuclease, allowing to retain and purify only the RNA-primed nascent strands, which are then sequenced and mapped to the reference genome. Done with Biorender.com*

3.1.2. Results

Short nascent DNA strands were purified from samples harvested at 29 hpi before my arrival in the lab, following the protocol specified in the Materials & Methods section. I performed the analysis and called the peaks with the same process as with ChIP-seq (using the negative control as the input), obtaining 4826 peaks. Then, I removed the unspecific signal (using the DNA strands that were digested with RNAse prior to λ-exonuclease treatment treated like the IgG control from ChIP-seq) and filtered the short peaks. Finally, I identified 4796 initiation sites (Appendix 4) distributed throughout all chromosomes (Fig. 38a), and although it roughly increased, the number of origins mapped per chromosome did not mirror chromosome length (Fig. 38b).

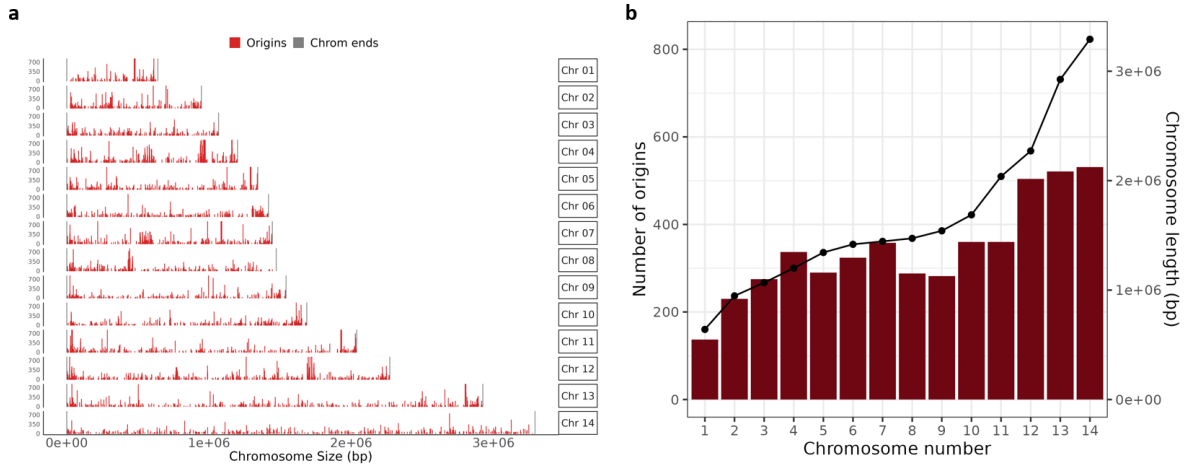


Figure 38. Distribution of origin sites identified with SNS-seq. *a) Genome wide distribution of SNS-seq origins throughout the chromosomes. b) Number of origins in each chromosome is represented with the red bars. Individual chromosome lengths are indicated by the black points.*

The inter-origin distances (IOD) were non-random and heterogeneous, with a median length of 1922 bp. In fact, 1976 origins were separated by less than 1 kb (41% of all origins) (Fig. 39a). I therefore investigated if there were clusters of origins, similar to what was observed for *Pf*ORC$_{1\text{-}2}$. Indeed, I found

523 origin clusters with at least five origins within 10 kb, which were distributed throughout the whole genome (Fig. 39b).

Figure 39. Clusters of SNS-seq origins. *a) Inter-origin distances between SNS-seq origins. Frequency of the various distances was calculated in 1 kb intervals. b) Genome wide distribution of the clusters of SNS-seq origins. The different heights of the bars represent the number of SNS-seq origins found in each cluster.*

3.1.2.1. Comparison with *Pf*ORC$_{1\text{-}2}$ binding sites.

ORC binding sites determine where the replisome is to be recruited and, thus, where DNA replication can be triggered from. Therefore, I analysed the relationship between the distribution of *Pf*ORC$_{1\text{-}2}$ and the SNS-seq origins. The two datasets were strongly associated (p-value: 1.23e^{-148}, odds ratio= 4.2), with 35.2% of the *Pf*ORC$_{1\text{-}2}$ peaks overlapping with an SNS-seq peak (Fig. 40a, b). Additionally, since MCM helicase binding and replication machinery spreading after loading into DNA has been reported in other systems that could lead to an expansion or shift of the replication initiation area [338], [339], I re-calculated the overlap between the two datasets within a window of 2 kb. Interestingly, 72.2% of the *Pf*ORC$_{1\text{-}2}$ sites were within 2 kb of an active origin (Fig. 40c). In addition, active origin clusters highly correlated with the *Pf*ORC$_{1\text{-}2}$ clusters (p-value= 7.74e^{-36}, odds ratio= 6.62) overlapping with nearly half

of them (44.4%, Fig. 40d). In fact, the totality of *Pf*ORC$_{1\text{-}2}$ clusters contained 900 active origins (versus 484 in a randomized clusters dataset), with an average of 3.8 active origins per *Pf*ORC$_{1\text{-}2}$ cluster (Fig. 40e). This strongly suggests that clustered licensing is more likely to lead to initiation events.

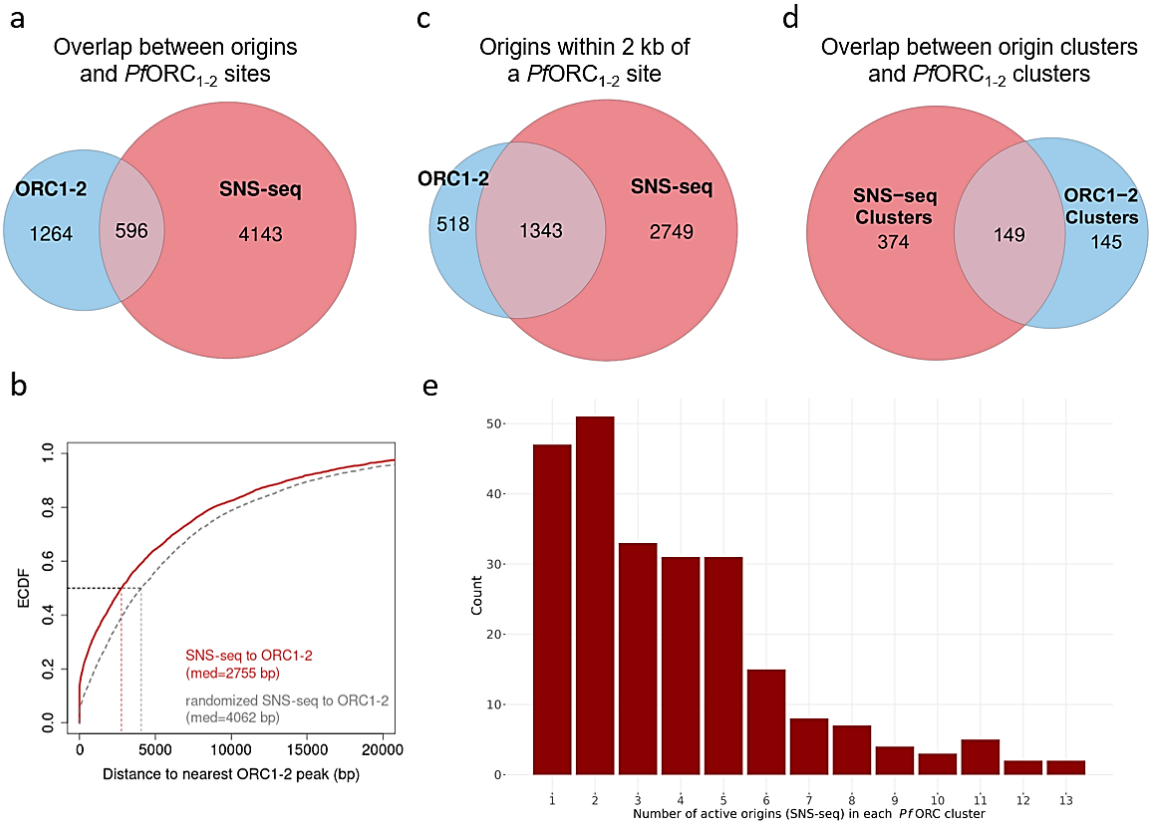


Figure 40. Comparison between active SNS-seq origins and PfORC$_{1\text{-}2}$ sites. *a) Overlap between the active origins (SNS-seq) and the PfORC$_{1\text{-}2}$ binding sites. b) Empirical Cumulative Distribution Function (ECDF) of the distances between SNS-seq origins and the closest PfORC$_{1\text{-}2}$ site. Median distances are indicated below. c) Number of SNS-seq origins within 2 kb of a PfORC$_{1\text{-}2}$ binding sites. d) Overlap between SNS-seq clusters (5 or more origins in a 10 kb window) and PfORC$_{1\text{-}2}$ clusters (3 or more PfORC$_{1\text{-}2}$ sites in a 10 kb window). e) Number of SNS-seq origins found in a given PfORC$_{1\text{-}2}$ cluster, y axis represents number of clusters containing that amount of SNS-seq origins.*

3.1.2.2. Analysis of association with G4s

I next investigated whether the active origins mapped with SNS-seq displayed any sequence specificity. Similarly to the set of *Pf*ORC$_{1\text{-}2}$ binding sites, I did not find any significant consensus sequence in the active origins. In addition, the SNS-seq ORIs displayed an increase in GC content relatively to the rest of the genome (32% vs 19.3%, Fig. 41a). They also harboured an increased proportion of potential G4FS as per calculated by G4Hunter: 1925 potential G4FS found within the ORIs, corresponding to a 5-fold enrichment in comparison to the whole *P. falciparum* genome (1.5 G4FS/1000 bp vs 0.3/1000 bp in the whole genome, hypergeometric p-value=1e^{-208}). Accordingly, I detected a strong association with G4FS (p-value= 5.24e^{-344}, odds ratio= 10.4, Fig. 41b), which was

even stronger than what was seen for the *Pf*ORC$_{1\text{-}2}$ binding sites. A total of 984 origins (~21%) had at least one G4FS within 500 bp (Fig. 41c).

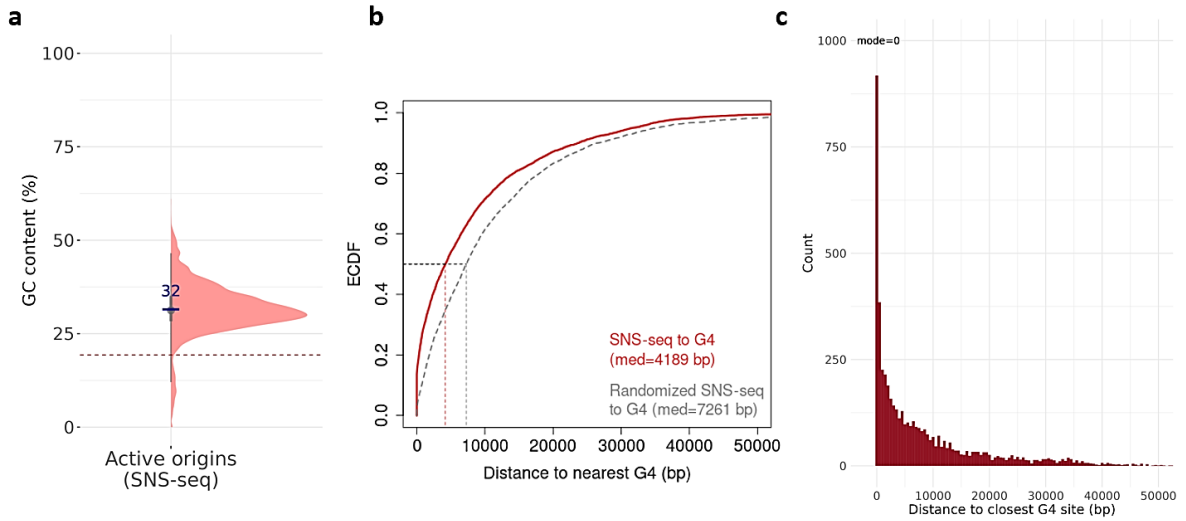


***Figure 41. Increased GC content in SNS-seq origins and proximity to G4FS**. a) Percentage of GC content within the SNS-seq origins. The horizontal bar depicts the GC content of the whole P. falciparum genome (19.3%) and average GC content of SNS-seq ORIs is indicated above; thick and thin vertical lines indicate the 50% and 95% intervals of the data. b) ECDF of the distances between a given SNS-seq origins and the closest G4FS. c) Histogram of the distances between a given SNS-seq origin and the closest G4FS site.*

3.1.2.3. SNS-seq origins and transcription

Similarly to what was observed for *Pf*ORC$_{1\text{-}2}$, there was no significant overlap between active origins of replication and *P. falciparum* histone PTMs (Jaccard index range: 0.002 to 0.14, Fig. 42).

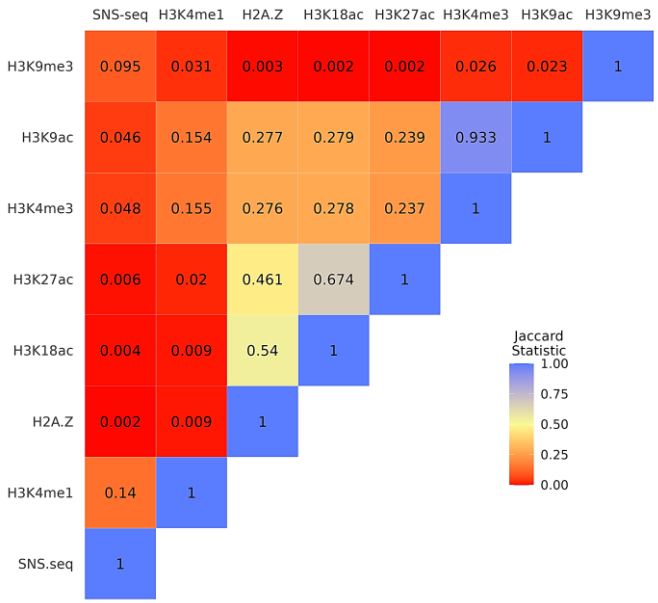


***Figure 42. Association of SNS-seq origins and PfPTMs**. Heatmap showing the Jaccard statistic of overlap between different P. falciparum histone modifications and the active origins detected by SNS-seq. Values range from 0 (red, no overlap) and 1 (blue, complete overlap).*

Additionally, I investigated the enrichment of active ORIs in the bodies of core genes. Active origins were even more depleted than *Pf*ORC$_{1\text{-}2}$ sites in TSS and TTS of core genes (TSS: p-value=3.4e^{-115} (anticorrelation), odds ratio= 0.23; TTS: p-value= 8.13e^{-182} (anticorrelation), odds ratio= 0.21), whereas the gene bodies were enriched in SNS-seq ORIs (Fig. 43a). This enrichment was seen regardless of gene expression levels, although it was stronger in actively transcribed genes (all genes: p-value=5.42e^{-56}, odds ratio=1.882; transcribed genes at 24h [130]: p-value=1.05e^{-71}, odds ratio=2.127). In fact, the expression percentile of the 1544 genes enriched in SNS-seq ORIs was significantly higher than the percentile of a randomly selected gene set (71 vs 55, Fig. 43b). Also, the gene bodies of the 25% most expressed genes of *P. falciparum* were strongly associated with the ORI sites (p-value < 10^{-4}, Chi-squared test) and the majority of genes of the top expression quartile contained, on average, two SNS-seq origins. In turn, nearly 300 of these highly expressed genes were found in the SNS-seq clusters (p-value < 1e^{-300}) and no enrichment was seen for genes of average or low expression levels. This indicates that while the association between origin licensing and gene bodies was significant irrespective of gene expression levels (although stronger with the top quartile genes), origin firing is enhanced by strong expression activity.

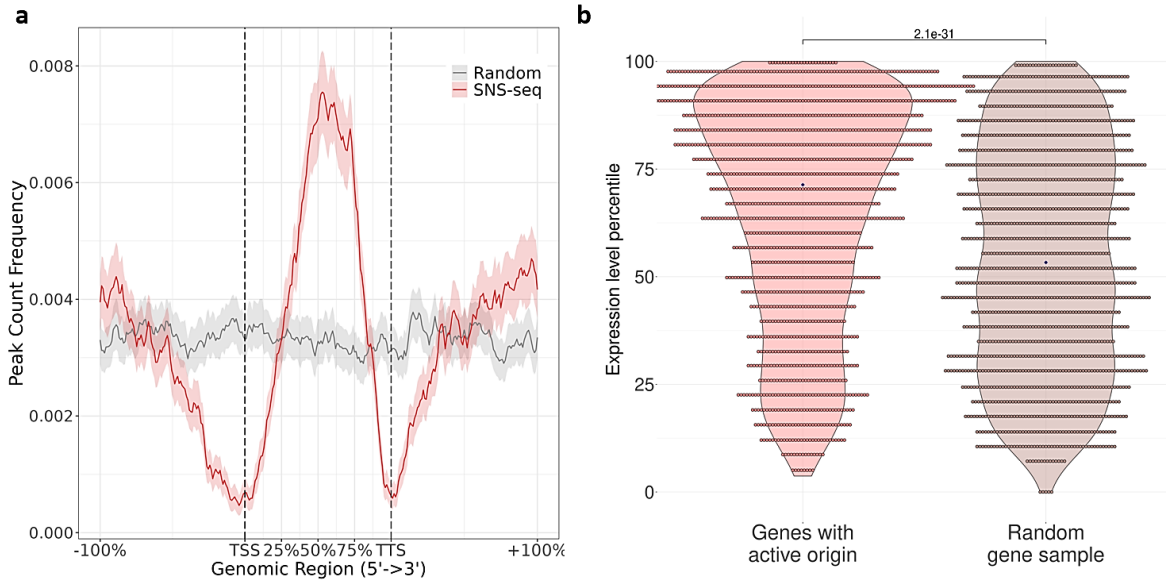


Figure 43. Enrichment of SNS-seq ORIs within P. falciparum core genes. *a) Enrichment of SNS-seq origins (red) over scaled coordinates of P. falciparum genes. A randomized (control) dataset is shown in grey. TSS corresponds to the transcription Start Site and TTS to the transcription termination site of each gene. Regions of equal length to each gene are included upstream and downstream the TSS and TTS, respectively. b) Violin plots depicting the expression percentile of the genes displaying an enrichment of SNS-seq origins (n= 1544 genes) and a randomized dataset of equal number in brown. The result of a two-sided Wilcoxon test is shown.*

3.2. NanoForkSpeed

3.2.1. Introduction

Nanopore sequencing (Oxford nanopore technologies, ONT) is a long-read sequencing technology that can be used as another tool to map origins of replication in the genome. This is achieved by exploiting the fact that thymidine analogues, such as BrdU, incorporated in the genome during DNA synthesis, display a different profile, comparatively to thymidine, when analysed by ONT. As a single-stranded DNA molecule passes through the nanopore embedded in a membrane subjected to an electrical current, the disruption in the electric signal caused by each passing nucleotide is detected and can be translated into bases and mapped to the genome. The thymidine analogue BrdU causes a shift in the signal which, while similar to that of the thymine, it is different enough to allow its identification. One tool that allows the differential identification of the four standard DNA nucleotides plus BrdU is the NanoForkSpeed algorithm [5]. With NanoForkSpeed (NFS), the thymidine analogue BrdU (5-Bromo-2-deoxyuridine) is added to the parasites in pulse-chase experiments and detected by nanopore sequencing (Fig. 44) [10]. BrdU is added to replicating parasites during a short incubation period (a "pulse" of 2 minutes), where it will be incorporated into DNA at the place of Thymine only in forks that were actively replicating during the pulse. This is followed by a "chase"; a longer incubation with thymidine, which will progressively substitute the BrdU as DNA synthesis progresses (Fig. 44). This allows the mapping of the incorporation of BrdU in the genome and, consequently, the detection of replication forks that were active during the BrdU pulse [5], [10].

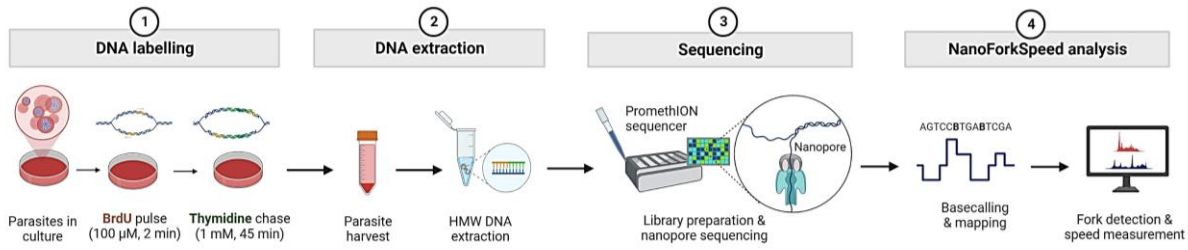


***Figure 44. Scheme depicting the NFS protocol overview**. Replicating parasites are incubated with a short pulse of BrdU followed by a long Thymidine chase. Then, High Molecular Weight (HMW) DNA is extracted, libraries are prepared and sequenced with ONT. Basecalling, mapping and fork detection are done with NanoForkSpeed* [5].

By using NFS, we can map where this nucleoside analogue has been incorporated in the DNA during replication and infer the sites of initiation of replication in the genome by identifying the midpoint between two diverging replication forks. A replication fork is detected by NFS where there is a sharp increase in the BrdU signal downstream of a segment of zero BrdU content (*i.e.*, before the initiation of the BrdU pulse), followed by a longer and milder downward slope of progressively lower BrdU signal intensity (representing the lower incorporation of BrdU in DNA during the thymidine chase, since it is

being diluted by the thymidine) (Fig. 45). The orientation of these asymmetric BrdU level signatures by the detection of the different segments (no BrdU signal, ascending slope, and descending slope) allowed to infer replication fork directionality and this information was then used to map the initiation and termination events. An initiation event is defined as the region between two divergent forks (in the same read) and a termination event as the region between two converging forks (in the same read).

Another advantage of ONT is that it allows sequencing of extremely long reads (up to 100 kb) which may cover very large genomic regions within a single read, thereby conferring single molecule resolution as several replication forks and initiation sites can be mapped from a single read, which comes from a single nucleus, meaning they were present in the same cell.

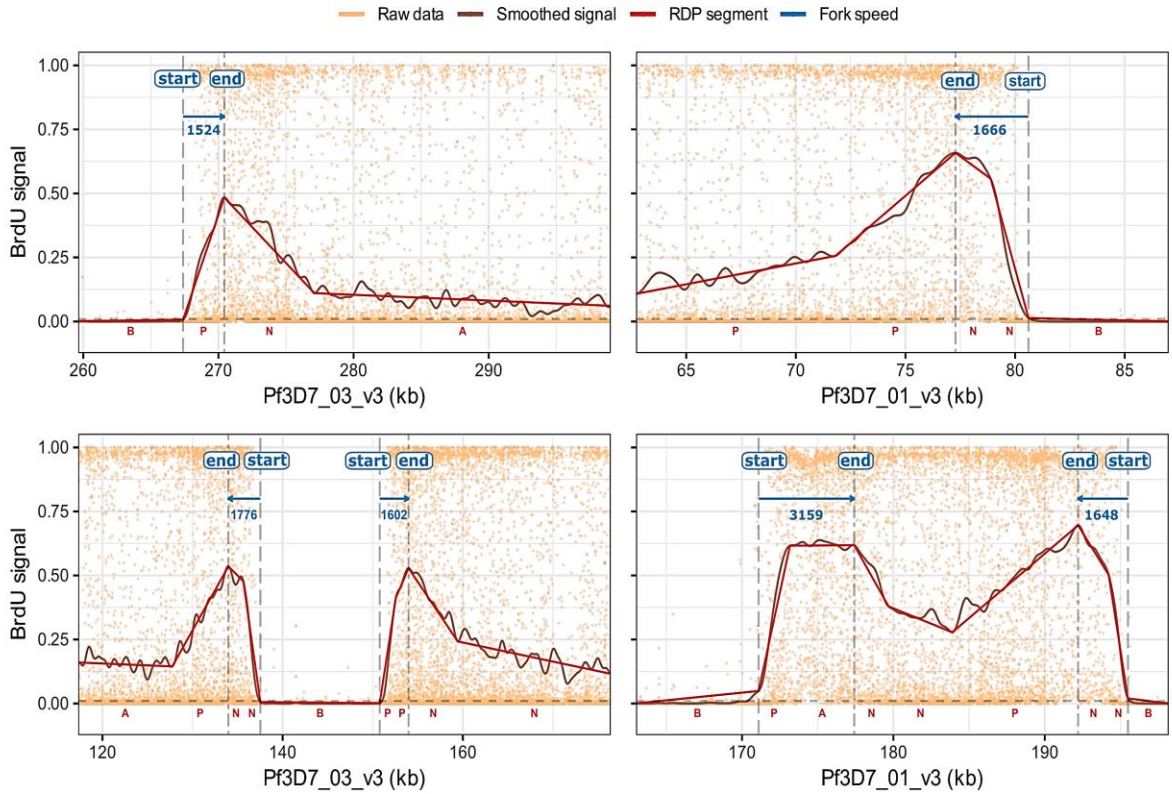


***Figure 45*. Detection of replication forks and initiation and termination events by NanoForkSpeed.** *BrdU content profiles of nanopore sequencing reads of genomic DNA from pulse-labelled P. falciparum TK⁺ parasites processed by NFS. Panels show typical replication signals, namely rightward, leftward, diverging, and converging forks. Orange dots, raw data (each dot represents the probability of BrdU at each Thymine position); brown curve, smoothed signal; red lines, segments resulting from the piecewise linear simplification method using the Ramer-Douglas-Peucker algorithm (RDP) to detect and orient BrdU tracks* [5] *(B, flat segments with background BrdU level; A, flat segments with a BrdU level above background; P, segments with a positive slope; N, segments with a negative slope); start, estimated position of the start of BrdU incorporation; end, estimated position of the start of the thymidine chase; blue arrow, fork direction, with fork velocity (bp/min) indicated below.*

3.2.1. Results

I performed the NFS experiment using the thymidine kinase-expressing (TK^+) *P. falciparum* line that incorporates thymidine analogues into DNA at two different stages of schizogony: t29h, the onset of the first S-phase; and t35h, in the middle of schizogony. Parasites were incubated with a short pulse (2 min) of BrdU, followed by a long thymidine chase (45 minutes; 10-fold excess), and then I extracted high molecular weight DNA to prepare ONT sequencing libraries (Fig. 44).

Since the diverging or converging forks need to be detected in the same read, only very long reads that harboured multiple forks allowed me to map initiation or termination events. In fact, for the timepoints t29h and t35h, I was able to map ~7.6M and ~13.6M reads, of which 47 421 and 43 801 contained at least one replication fork. From these, 3 722 (0.5% of all t29h reads) and 3 232 (0.24% of all t35h reads) contained at least one initiation event, respectively. The median size of a read containing at least an initiation event was 37 816 bp (t29h) and 37 548 bp (t35h) (Fig. 46).

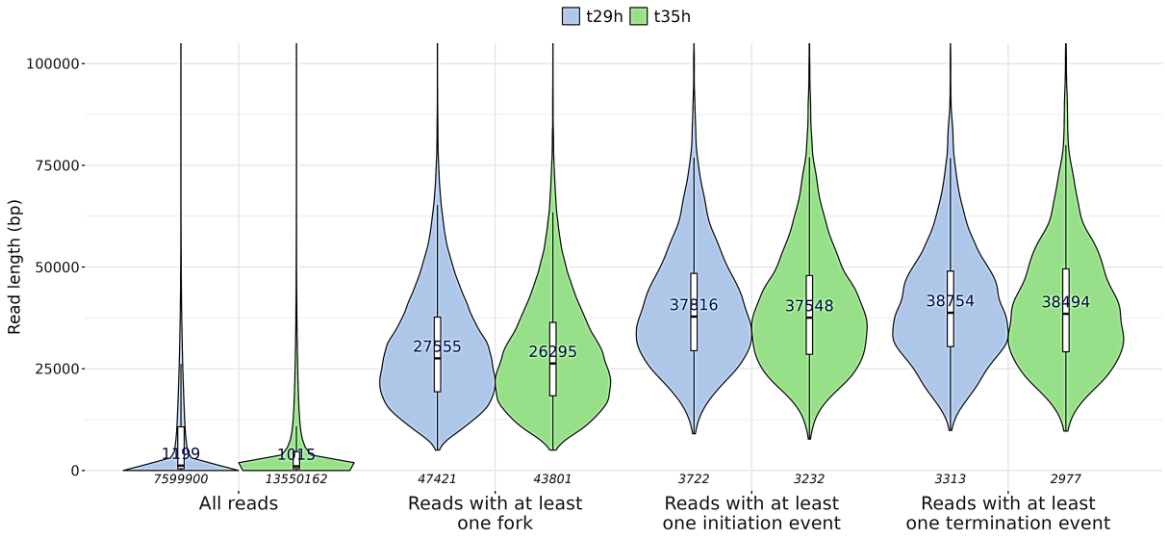


Figure 46. Length (bp) of the sequenced reads at t29h and t35h. *The number of reads sequenced in each group is indicated in black below the violins and the median read length in blue over the boxplot.*

A total of 54 590 (t29h) and 50 117 (t35h) forks were mapped, which corresponded to 2.34 and 2.15 forks per kilobase, and showed a coverage of at least 2x in 95% and 93% of the genome for each timepoint. These allowed me to map 3 786 (Fig. 47a) and 3 286 (Fig. 47b) active origins of replication at t29h and t35h, respectively (Appendix 5).

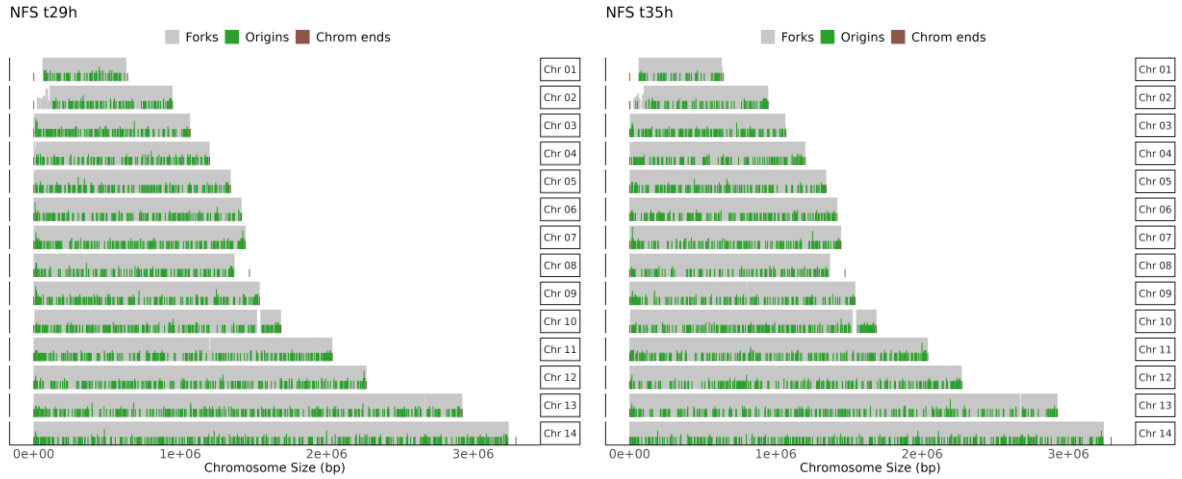


Figure 47. Distribution of replication forks (grey) and initiation sites (green) detected with NFS at t29h and t35h. *Different heights indicate enrichment of forks or initiation events detected in certain genomic regions.*

The distribution of the origins of replication mapped with NFS (defined as the midpoint between two diverging forks ±1.5 kb) throughout the chromosomes was non-random, with a median IOD of 4270 bp for t29h and 4788 bp for t35h. In addition, NFS origins group in clusters, consistent to what I found for SNS-seq ORIs and *Pf*ORC$_{1\text{-}2}$ binding sites. I detected 489 clusters of 5 or more NFS origins within 10 kb at t29h (Fig. 48).

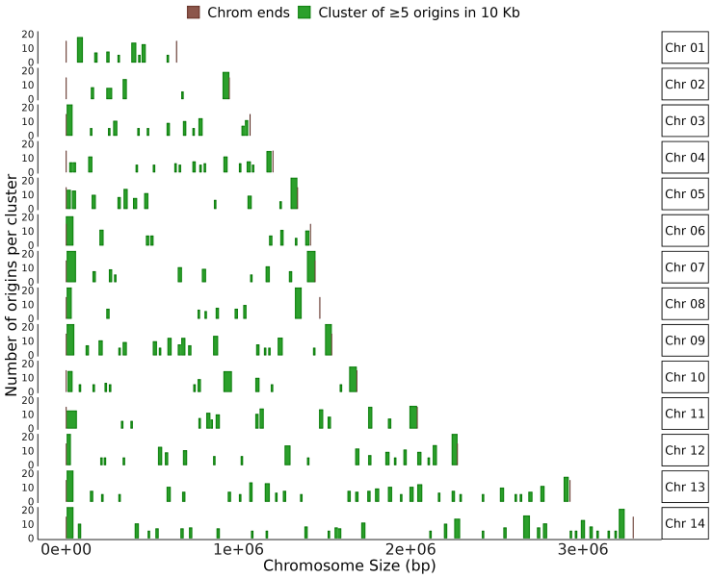


Figure 48. Clusters of initiation events detected with NFS t29h. *Genome wide distribution of NFS origin clusters. A cluster is defined as a 10 kb region containing 5 or more ORIs. The different heights of the bars represent the number of NFS origins found in each cluster.*

3.2.1.1. NFS origins and transcription

I next investigated the enrichment of NFS initiation sites within the body of core genes. Similarly to *Pf*ORC$_{1\text{-}2}$ and SNS-seq ORIs, the NFS origins displayed a depletion in the TSSs and an enrichment, although milder, in gene bodies (Fig. 49a). Accordingly, the percentile of expression of the 550 genes

that contained an NFS initiation site was higher than the percentile of a set of genes randomly selected (57 vs 50) (Fig. 49b).

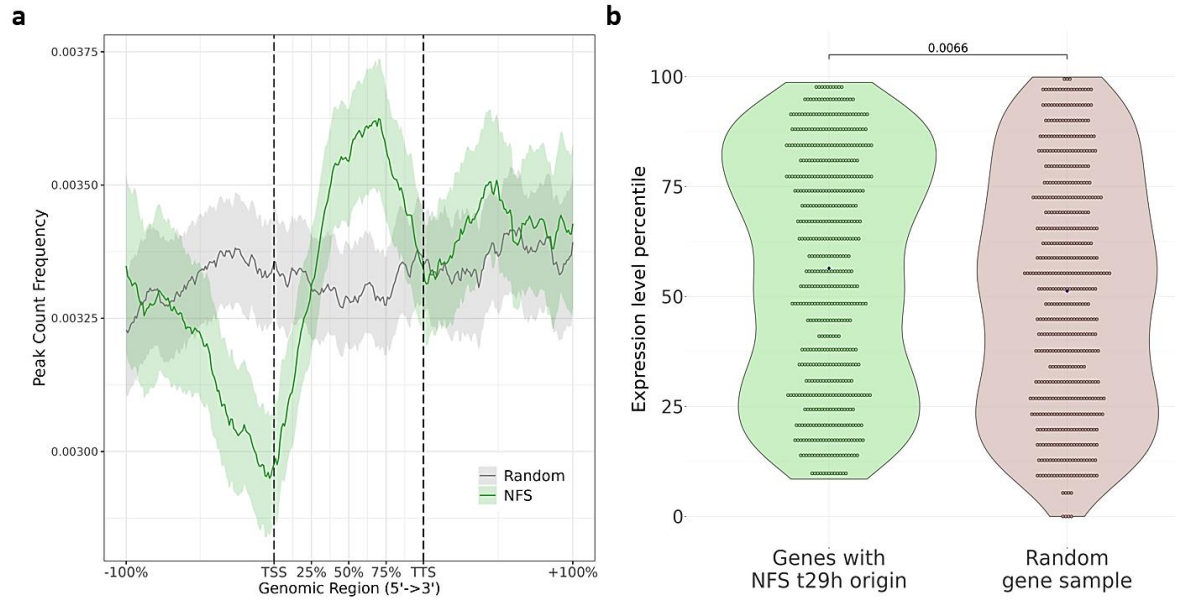


Figure 49. Enrichment of NFS t29h in core genes*. a) Enrichment of NFS origins (green) over scaled coordinates of P. falciparum core genes. A randomized (control) dataset is shown in grey. TSS corresponds to the transcription Start Site and TTS to the transcription termination site of each gene. Regions of equal length to each gene are included upstream and downstream the TSS and TTS, respectively. b) Violin plots depicting the expression percentile of the genes displaying an enrichment of* NFS t29h origins *(n= 550 genes) and a randomized dataset of equal number in brown. The result of a two-sided Wilcoxon test is shown.*

3.2.1.2. Comparison with licensed and active origins

Sometimes the limited overlap between different origin mapping methods presents a challenge for accurate identification of replication initiation sites [161], [340]. Having used three methods, I next compared the overlap between the different datasets.

Interestingly, about two-thirds of the *Pf*ORC$_{1\text{-}2}$ and SNS-seq sites overlapped with the initiation regions of NFS t29h (Fig. 50 a,b). This was further confirmed by fisher test: p-value NFS t29h vs. *Pf*ORC$_{1\text{-}2}$= 1e^{-2}; p-value NFS t29h vs. SNS-seq= 1.5e^{-4}. To further explore the relationship between these different methods, I also compared the t29h NFS inter-origin distances per chromosome with those identified using SNS-seq and found an overall agreement between the IODs of these two methods, suggesting that NanoForkSpeed is a reliable method for active origin mapping (Fig. 50c).

In addition, NFS clusters were associated and overlapped with 30% of both the *Pf*ORC$_{1\text{-}2}$ clusters and SNS-seq origins clusters (Fig. 50 d, e) (NFS t29h vs. *Pf*ORC$_{1\text{-}2}$ clusters: p-value=3.1e^{-4}, NFS t29h vs. SNS-seq clusters: p-value= 4.12e^{-12}). Furthermore, 234 of the 261 *Pf*ORC$_{1\text{-}2}$ clusters (90%) contained between one and 18 NFS origins, with a mean of 3.88 origins per cluster.

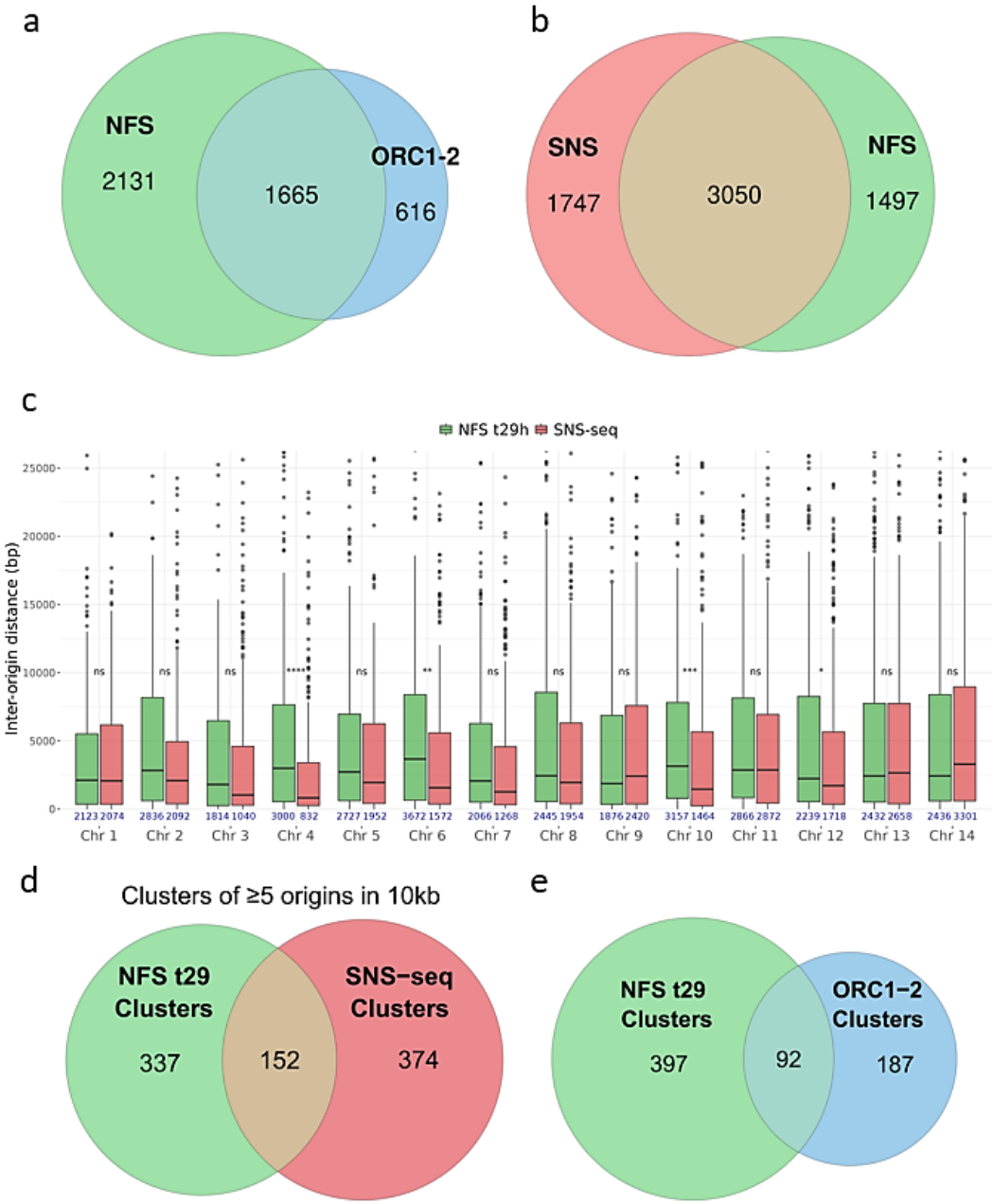


Figure 50. Comparison of NFS t29h ORIs with PfORC$_{1-2}$ sites and SNS-seq ORIs. *a) Overlap between NFS active origins PfORC$_{1-2}$ binding sites. b) Overlap between NFS and SNS-seq origins. c) Distribution of inter-origin distances measured by SNS-seq or NFS at 29h for each chromosome. Median distances are shown in blue as well as the results of two-sided Wilcoxon tests (ns: p > 0.05, *: p <= 0.05, **: p <= 0.01, ***: p <= 0.001, ****: p <= 0.0001). d) Overlap between NFS and SNS clusters. e) Overlap between NFS and PfORC$_{1-2}$ clusters.*

Taken together, although a correlation with G4FS was not apparent in the NFS dataset there was a high level of concordance between the datasets obtained in this study. By using multiple approaches and comparing their results, I have gained a comprehensive picture of the genetic landscape of replication origins in *Plasmodium falciparum*, which will be useful for future studies in this field.

3.2.1.3. Replication fork speed

Another advantage of the NanoForkSpeed algorithm [5] is that it allows to measure the velocity at which replication forks progress through the genome. The pace at which replicating forks advance is calculated as the length of the labelled DNA region divided by the duration of the BrdU labelling pulse (2 minutes). Accordingly, the NFS algorithm detects the start and end coordinates of the BrdU pulse as the points of initiation and end of the positive steep slope of BrdU signal (Fig. 45) and calculates the length of the DNA sequence produced during the BrdU pulse [5]. By compiling the speeds of all forks for each timepoint, I calculated the median genome-wide fork speed to be of 1542 bp/min for t29h and 1496 bp/min for t35h. These results were in agreement with the DNA combing data from *P. falciparum,* where replication speed was reported to be 1.4 kb/min for 24-30 hpi parasites and 1.2 kb/min for 32-38 hpi parasites [341]. Even though there was a >30-fold difference between the fastest and slowest forks, more than half moved at ±35% of the median speed (54.7% of t29h forks and 52.9% of t35h forks) and highly extreme speeds were exceptional and accounted for only 1.5% of the forks.

I next explored if replication fork speed changed depending on whether the fork travelled on leading or lagging strand of the replication bubble. Interestingly, lagging strand forks exhibited a slightly slower speed compared to leading strand forks at t35h (p-value= $4.9e^{-6}$). However, at t29h, seemingly similar speeds were seen for both leading and lagging strand forks (Fig. 51). These results suggest that replication fork speed may be influenced by strand directionality, particularly at later stages of schizogony. Further studies are needed to fully understand the underlying mechanisms governing these observations.

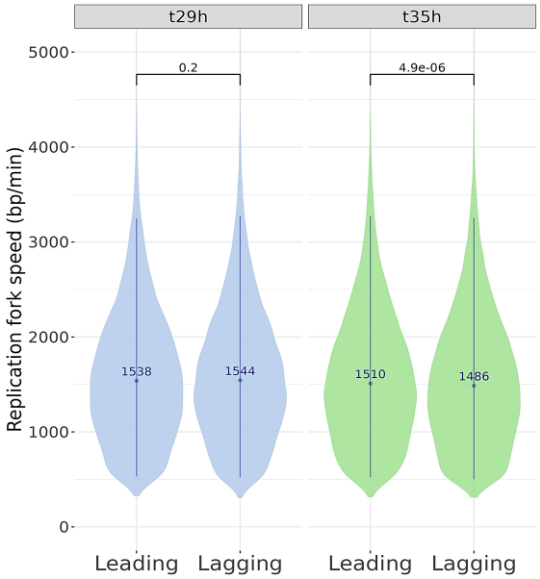
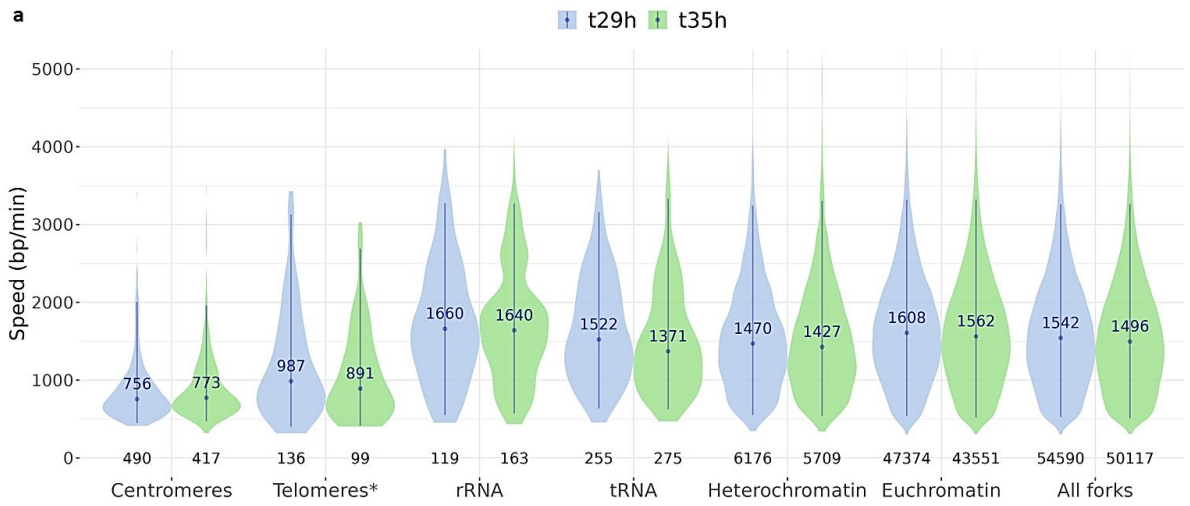


Figure 51. Replication fork speed on leading and lagging strands at t29h and t35h. *Median fork speeds are indicated above the blue point and the thin blue lines indicate the 95% data interval. Two-sided Wilcoxon comparisons between the median speed at the two timepoints are indicated above.*

To investigate whether local chromatin environment had any impact on fork speed, I examined specific regions of the genome, including centromeres, telomeres, rRNA genes, tRNA genes, heterochromatin (regions marked by the presence of H3K9me3) and euchromatin (H3K9me3 negative regions) (Fig. 52a). In centromeres there was a drastic reduction in fork speed at both timepoints, with a median speed of 756 bp/min and 773 bp/min, respectively. This decrease in speed was not due to a decrease in read coverage (Fig. 52b), indicating that it is likely due to specific chromatin conditions present at centromeres that might hamper replication fork progression. Additionally, forks travelling over telomeres (considered as the first and last 3 kb of each chromosome) displayed similar speeds of 987 bp/min and 891 bp/min, respectively, suggesting that these regions may also have different chromatin features that difficult fork progression [342]. Interestingly, I found that rRNAs showed the highest median fork speed of all regions analysed, with speeds of 1660 bp/min and 1640 bp/min at t29h and t35h, respectively. This finding suggests that the chromatin structure of rRNA genes may be permissive for rapid fork movement. In contrast to rRNAs, tRNAs displayed a median fork speed slightly below the genome average, with speeds of 1522 bp/min and 1371 bp/min. This observation suggests that the chromatin structure of tRNAs may have a moderate but lower impact on replication fork progression. In addition, there was a significant difference in fork speed between H3K9me3-rich (heterochromatin) and H3K9me3-depleted (euchromatin) regions, with heterochromatin displaying significantly slower fork speeds than average in both timepoints (p-value heterochromatin vs. euchromatin: t29h = $5.7e^{-36}$, t35h = $1.4e^{-29}$). This finding is consistent with previous studies showing that heterochromatin is generally more compact and less accessible to the DNA replication machinery than euchromatin [5], [342], [343].



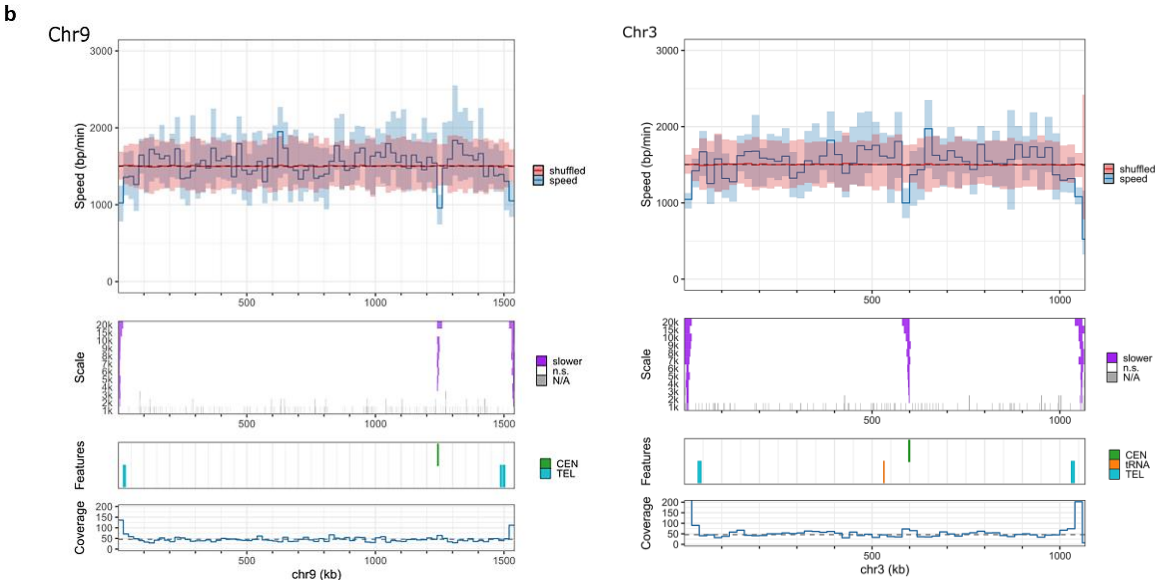


Figure 52. Replication fork speed measured by NFS*. a) Local fork speeds detected at t29h and t35h, over specific genome features or gene families. Number of forks detected is noted in black below the violins; median fork speeds are indicated above the blue point and the thin blue lines indicate the 95% data interval. b) Example of NFS replication fork progression map for chromosome 9 and 3. Pannels from top to bottom: (i) median of detected fork speeds (blue line) with 98% confidence interval of the median (light blue), and median of reshuffled speeds (red line) with 98% confidence interval of the median (light red) computed in 20kb windows (dotted line represents the median fork speed of the whole genome); (ii) Results of Mann–Whitney-Wilcoxon tests with Holm correction (one-sided) performed along the chromosome to compare the speed distribution in windows of a given width (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 and 20 kb) to the speed distribution on the whole genome (purple, regions of lower fork speed; white, n.s., not significant; statistical significance was set to p < 0.01; grey, N/A, not applicable, regions with no fork); (iii) position of selected genomic features (CEN, centromere in green; TEL, telomeres in blue); (iv) Coverage of individual replication fork velocities (dotted line, median coverage of the genome). Although fork speed decreases towards the ends of the chromosome and over the centromere, this is not due to a depletion in coverage reads. Figure done by Laurent Lacroix.*

To explore deeper the relationship between replication and transcription, I then investigated the progression of the fork with respect to directionality of transcription. The results showed that replication speed at both timepoints is not significantly different depending on whether transcription is happening in the same strand or in the opposite (Fig. 55a). Also, fork progression was found to be significantly slower at t35h in genes having a direction of transcription opposite to the direction of replicating forks (head-on) than in genes where replication and transcription were co-directional; a difference that could be due to conflicts between the replication and transcription machineries but that was not seen at t29h (Fig. 55b). Finally, gene expression level was also a factor influencing fork speed, with the genes displaying lower expression levels being associated with higher speeds both at t29h and t35h (Fig. 55c).

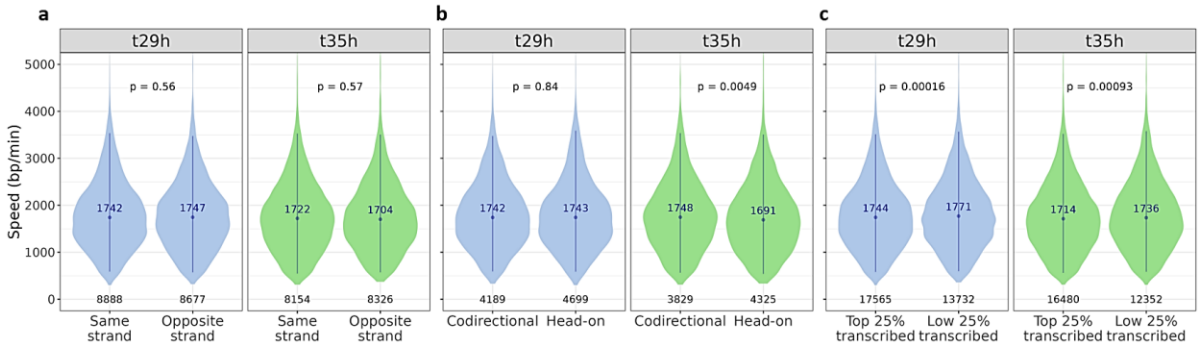


***Figure 53. NFS replication fork speed measurement in relation with transcription**. Number of forks are indicated in black below the violins and median fork speeds are indicated above the blue point. The thin blue lines indicate the 95% data interval. Two-sided Wilcoxon comparisons within each timepoint between the speed at the two conditions are indicated above: a) same strand transcription – replication vs. opposite; b) Co-directional replication-transcription vs. Head-on; c) Top 25% transcribed genes vs. Lowest 25% transcribed genes.*

3.2.1.4. NFS measurement at single molecule level

Since it had previously been reported that the first S-phase of *P. falciparum* schizogony runs slower than the posterior ones [2], but other studies have suggested differently [341]. Using our NFS fork speed data I set out to investigate whether a faster first S-phase was due to a higher number of origins or rather due to a faster pace of DNA synthesis by replication forks. Since the number of initiation events and replication fork density detected for both timepoints was quite similar (Fig. 54), these data suggest that the previously detected difference in duration of the S-phases across schizogony is not likely due to a difference in the number of active origins.

To further investigate this, I examined the inter-origin distances (IODs) at t29h and t35h, since a big reduction in the IOD could still facilitate faster completion of replication rounds. However, calculating IODs from bulk mapping methods, such as SNS-seq, can be misleading as they represent population averages. To overcome this limitation, I examined only the IODs between initiation events captured in the same read, which could have only come from individual nuclei. There were 64 and 52 reads containing more than one origin in the t29h and t35h NFS datasets, respectively. While reads containing one initiation event were on average 37 kb long, the ones containing 2 or more initiation events were on average 56 kb and 57 kb for t29h and t35h, respectively. As expected, the single cell IODs were considerably longer (22 823 bp at t29h and 21 943 bp at t35h) than the IODs previously calculated for the total population (4270 bp and 4788 bp, Fig. 54a). In addition, the single molecule data showed that an increase in IOD correlated with an increase on the incoming replication fork speed (Fig. 54b), in both timepoints (p-value = $9.2e^{-6}$). An ANCOVA test of variance comparing the slopes of the linear regression lines generated for both timepoints revealed no significant difference between the slopes (slopes for t29h: 0.02786, t35h: 0.03136), and a 2-way ANOVA test revealed that there was no significant effect of the timepoint in the variation of the speed vs. distance (p-value=0.6995).

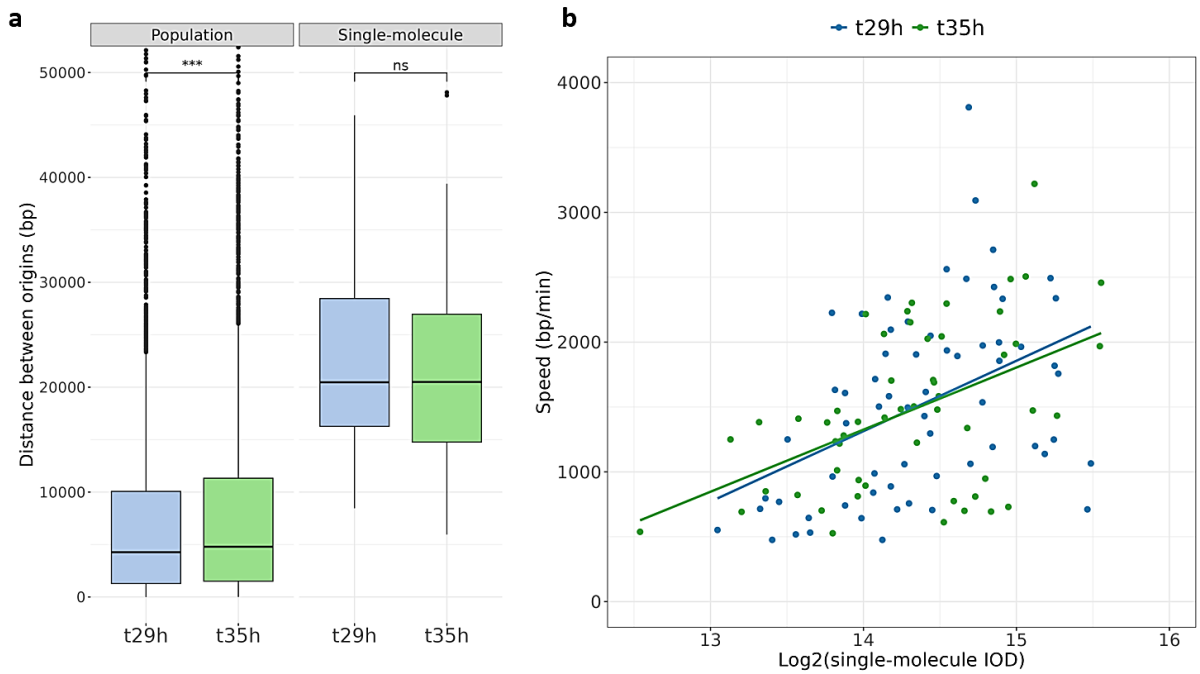


Figure 54. Single-cell level measurement of IOD and speed with NFS. *a) Inter-origin distance calculated for the total population data (p-value = 4,7x10^{-4}) or from sequencing reads containing multiple initiation events which could have only come from one nucleus (p-value = 0.64). b) Relationship between single cell IOD and the respective incoming upstream leading strand fork speed. These two variables were found to be strongly associated (Two-way ANOVA, p-value = 9.2x10^{-6}), and the factor timepoint was found not to significantly influence them (Two-way ANOVA, p-value = 0.7).*

These results indicate that the replication program of *P. falciparum* balance having more spaced origins of replication (reflected in higher IODs) with a higher replication fork speed in those areas of the genome, and vice versa, to complete each round of DNA in a timely manner, both at the beginning or middle of schizogony.

Klaus *et al.* [2] measured S-phase as the period of time during which their nuclear cycle sensor (*Pf*PCNA1::GFP) was visible in a given nucleus and their results showed that the first S-phase takes ~50 min whereas the following rounds ~35 min. Overall, our results indicate that asexual *Plasmodium falciparum* parasites can complete each round of DNA replication in 15 minutes, both at the beginning or middle of schizogony. The explanation of this difference might be that there could be an increased proportion of fork stalling in the first S-phase or that each nucleus could require an additional preparatory phase following recruitment of *Pf*PCNA1 into the nucleus before actually triggering replication activation. This preparation phase could then be longer in the first replicative cycle, and active DNA synthesis might be restricted to a short period within each S-phase.

4. Conclusion and perspectives

Altogether, this study provides a comprehensive investigation of the origins of replication and their genetic landscape in the human malaria parasite *Plasmodium falciparum*. By leveraging three different origin mapping methods, I have obtained highly concordant results (Fig. 55) that validate their efficacy to identify origins of replication throughout the genome.

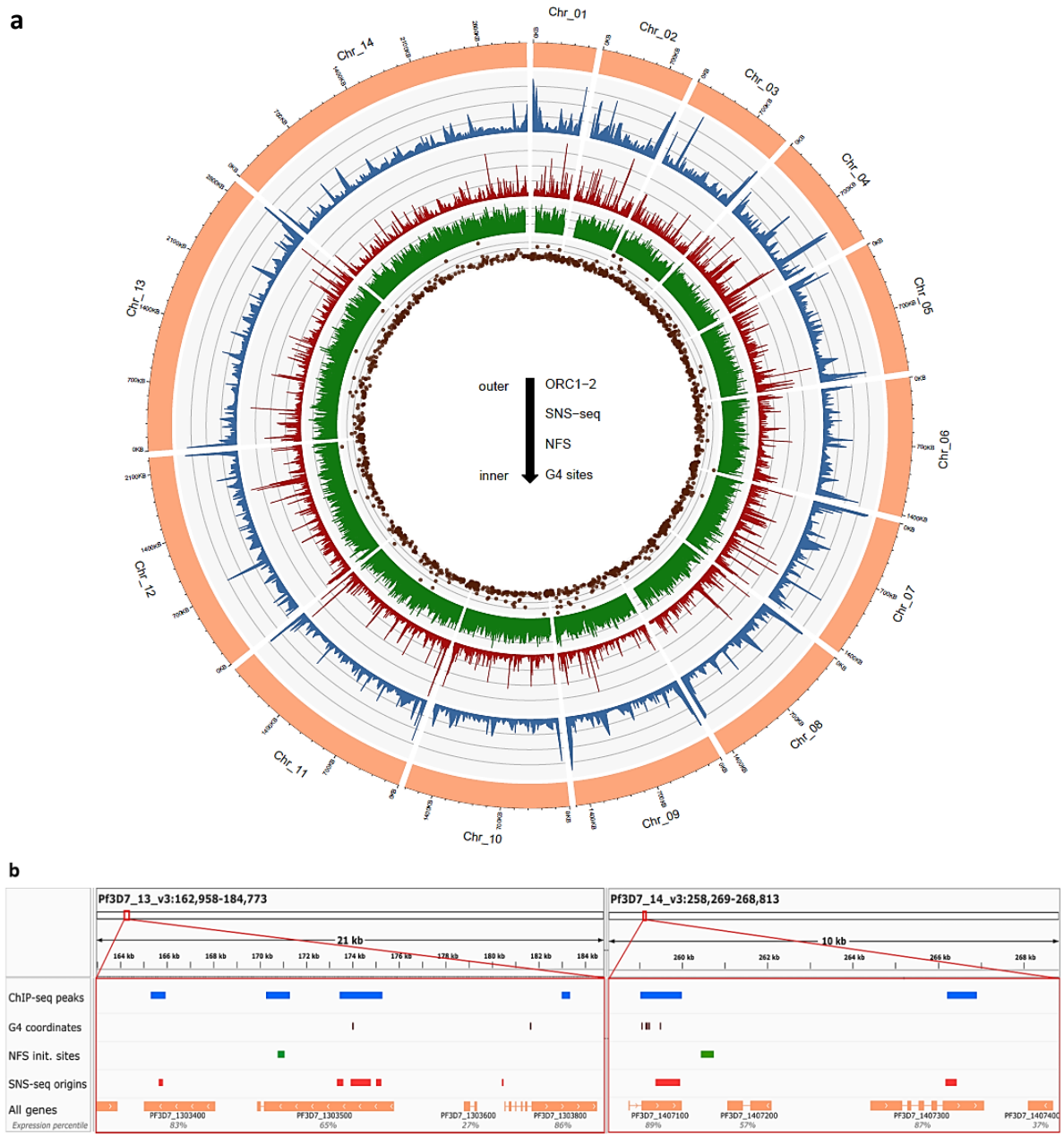


Figure 55. The three origin mapping methods yielded highly concordant results. *a) Circos plot depicting, from outer layer to inner: The length of the 14 P. falciparum chromosomes (orange), PfORC$_{1-2}$ ChIP-seq peaks (blue), SNS-seq origins (red), NFS t29h origins (green), G4FS coordinates from* [167] *(brown). b) IGV snapshot of two chromosome regions depicting the different datasets analysed. A section of chromosomes 13 and 14 are shown along with the different datasets and features. From top to bottom: PfORC$_{1-2}$ ChIP-seq peaks (light blue), G4FS coordinates from* [167] *(brown), NFS t29h origins (green), SNS-seq origins (red), and the position of encoded genes (orange) along with their expression percentile from* [130].

Most of the ORIs identified (~75%) by the three different methods showed a non-random distribution and appear clustered together in 523 "initiation zones" (Fig. 56). These clusters appear more frequently in some specific chromosome regions, especially closer to telomeres and centromeres. In these regions we have also detected a decrease of replication fork velocity. Regarding our observations at the single molecule level, the decrease of replication fork speed in these regions, probably due to the highly packed state of chromatin that hampers replication fork progression or the presence of chromatin-bound proteins, could be compensated by an increase in the number of origins of replication that would allow timely completion of the replication program (Fig. 54b: lower speed is associated with a shorter IOD, *i.e.,* more origins of replication in a given region).

Figure 56. Clusters of origins in 10 kb windows throughout the 14 chromosomes*. A cluster of origins is considered as a 10 kb window of the genome containing 3 or more Pf*ORC$_{1\text{-}2}$ *sites, or 5 or more SNS-seq or t29h NFS origins. Chromosome ends and centromere coordinates are indicated in brown, Pf*ORC$_{1\text{-}2}$ *clusters in blue, NFS t29h clusters in green and SNS-seq clusters in red.*

In addition, the set of origins mapped in this study displayed a strong association with G-quadruplex forming sequences (G4FS), akin to what has been demonstrated in mammalian cells [7]. This observation is consistent between the *Pf*ORC$_{1\text{-}2}$ sites and the SNS-seq active origins. It is still unclear whether G-quadruplexes contribute directly to the recruitment of ORC proteins or simply offer areas of nucleosome exclusion that ultimately facilitate binding of the *Pf*ORC and DNA polymerase machinery. In fact, *Pf*ORC$_{1\text{-}2}$ binding sites tend to appear in regions of relatively high GC content, such as the internal regions of genes, as compared to the rest of the AT-rich genome. It would be interesting to explore whether it is the presence of G4FS in these GC-rich regions what attracts *Pf*ORC binding, or if it is just this more balanced nucleotide content what promotes binding of the complex. A possible perspective to study contribution of G-quadruplexes to *Pf*ORC binding would be to employ G4

stabilizing or destabilizing agents to study the impact on *Pf*ORC binding efficiency. The use of a G4 stabilizing compound, pyridostatin, has already proved to cause large scale perturbations in transcription in *P. falciparum* [167], enhancing the potential of this G4-targeting approach to study pleiotropic downstream effects.

Finally, the correlation found between origin specification and active origins with strong transcription, which is true for the three methods employed in the study, suggests that these divergent organisms might have developed coordinated strategies to avoid conflicts between the replication and transcription machineries, although the mechanism by which they bypass these obstacles remains unknown. Despite active origin specification being associated with active transcription, there was a significant depletion of *Pf*ORC$_{1\text{-}2}$ sites and active origins in TSSs and TTSs. This is in contrast to what has been observed in human origins, where active TSS are necessary and sufficient for ORC binding [317], [323]. Perhaps, in *P. falciparum* transcribed genes, the presence of the RNA polymerase machinery leads to a more relaxed chromatin state that facilitates binding of the ORC and replication machinery. For this reason, I propose a model of replication origin specification (Fig. 57) where origins of replication are preferentially found in the genome in clusters or initiation zones, and within gene bodies of actively transcribed genes.

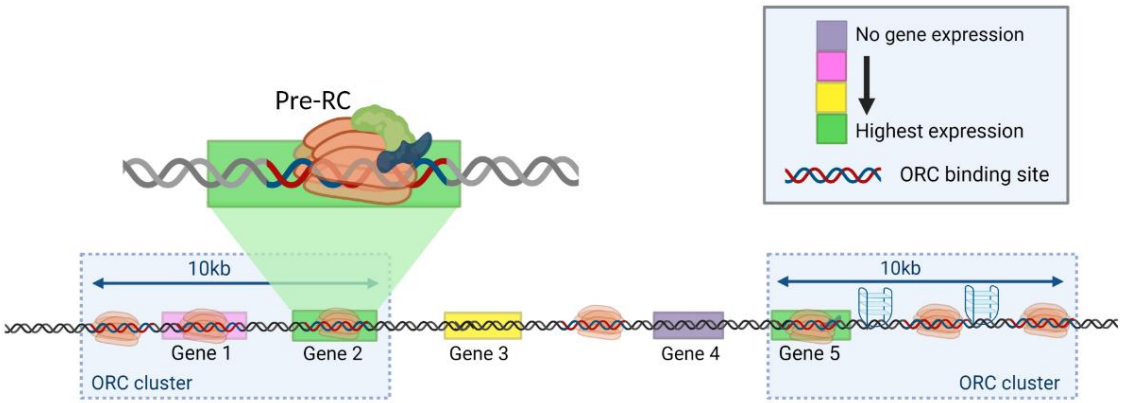


Figure 57. Model of replication origins in Plasmodium falciparum.

It is also worth pointing out the distinct pattern of *Pf*ORC$_{1\text{-}2}$ enrichment found in promoters of *var* genes. ORCs have already been shown to have a role in transcriptional silencing in different organisms. For instance, in *S. cerevisiae,* ORC binding is required for silencing of the *HM* mating type loci [344], [345] and in other model eukaryotes like *Drosophila, Xenopus* and mammals, both ORC1 and ORC2 interact directly with the heterochromatin protein HP1 [331], [334]. Even in *P. falciparum*, *Pf*ORC1 has been shown to be implicated in regulation of *var* gene expression through an association with the telomere-associated histone deacetylase *Pf*Sir2 [253], [336]. Here, I have shown that this role in *var* gene silencing is probably extended to *Pf*ORC2, although the exact mechanism of the regulation of mono-allelic expression in *P. falciparum* parasites is still not fully understood.

Finally, one recently published study of the origins of replication in *P. falciparum* [3] investigated *Pf*ORC1 binding and reported a much higher number of binding sites at 24 hpi (8843), which also displayed a preference for sites with high GC content (28.7%) and enrichment in subtelomeric regions, consistent with our findings. However, it is likely that the absence of replicates and, especially, the absence of an IgG control in their dataset may have led to an overestimation of binding sites. Indeed, my IgG dataset of unspecific peaks overlapped with a high number of their *Pf*ORC1 t24h dataset [3]. Although the number of *Pf*ORC1 binding sites described is higher, the low signal to noise ratio made it hard to perform a robust comparison with my *Pf*ORC1 dataset. Furthermore, this study reported a much slower average fork speed than what I obtained with the NFS measurements, ranging between 0.5-0.6 kb/min at 30 and 36 hpi, respectively. Although Garcia-Totañes *et al.* also use nanopore sequencing to map the incorporation of pulse-chase added nucleotide analogues, the algorithm used for replication initiation sites is different [279] and the length of the labelling pulse is 15 minutes, which is what I calculated as the time needed for the parasites to replicate their whole genome. It is also notable the low number of forks detected in comparison to our NFS datasets (1749 forks at t30 vs. 54590 forks from the t29h NFS dataset; and 25898 forks at t36 vs. 50117 forks from the t35h NFS dataset) and the low degree of overlap with our set of active ORIs (only 2% of the NFS t29h ORIs overlap with the t30 dataset, and 30% of the NFS t35h ORIs overlap with the t36 results).

Overall, I am confident that the three different origin mapping methods employed in this study have allowed me to make a comprehensive and thorough analysis of the potential and active sites of replication initiation in *Plasmodium falciparum*. By investigating their genetic landscape, I have made significant observations that contribute to the field of study of DNA replication in this pathogen. I found that the ORIs of *P. falciparum* display some features common to mammalian replication origins such as the lack of sequence specificity, GC-richness, and the association with G4FS. However, I also discovered some unusual features in *Plasmodium* origins, such as the striking enrichment of ORIs within gene bodies of strongly transcribed genes, and depletion from TSS/TTS. Such unconventional origin usage sheds light into the evolution amongst eukaryotes of origin specification and activation, which might be due to the extreme AT-richness of the genome, driven by their parasitic lifestyle through evolution [123], [130], [239]. Overall, this represents an important step towards understanding the replicative process that allows these parasites to expand their population by several orders of magnitude in a matter of days.

These results have led to the preparation of a research article, which has been submitted to the journal Nucleic Acids Research and is currently under review.

Chapter 2. Composition of the replicative complex

1. Introduction

The reliable and efficient replication of the parasite's genome lies on the coordinated action of multiple proteins that form the replisome machinery. Unravelling the composition and dynamics of this complex is crucial for gaining insights into the mechanisms underlying DNA replication in *P. falciparum* and identifying potential targets for therapeutic intervention. Currently, more than 60% of the genes of *P. falciparum* show weak to no homology in other eukaryotes [117] and lack functional annotation, being considered “hypothetical” or with “unknown function” [127].

In model eukaryotic systems the pre-replicative protein complexes assemble at replication origins and are composed of a set of highly conserved proteins. Although some of the core eukaryotic DNA synthesis associated factors are conserved, such as the MCM helicase complex [247], homologues of the ORC1, ORC2 and ORC5 subunits [252]–[255], DNA polymerases α, δ and ε [248], [249], the replication protein A (RPA) and PCNA1 [250], [251]; *Plasmodium* parasites lack many of the known factors. This suggests that regulation of DNA synthesis in this divergent organism is either less complex than in other systems, or probably orchestrated, at least partly, by unidentified parasite-specific factors. Putative homologs of other key elements have been found through sequence homology analysis, but their role in DNA replication has not yet been validated. These include a putative ORC3 domain-containing protein (PF3D7_1029900) a putative ORC4 (PF3D7_1334100) [244], a putative Cdt1-like protein (PF3D7_1343300), and several putative equivalents of subunits of the GINS complex [258].

Most of the members of the replisome that have been found in *P. falciparum* have been reported to be essential, in both an *in vitro* and *in vivo* essentiality screen [259], [260], suggesting that antimalarials targeting them would have the potential to be highly effective.

In this chapter, I explored different techniques to investigate the composition of the replicative complex in *Plasmodium falciparum*. For that I used two different strategies: first, I used proteins as bait molecules in immunoprecipitation experiments and aimed to perform proximity labelling experiments to selectively capture the interacting proteins within the replisome complex. Secondly, I explored an approach based on the isolation of proteins on nascent DNA (iPOND), which relies on pulling down the proteome crosslinked to EdU (5-ethynyl-2-deoxyuridine)-labelled DNA, from active replication forks.

2. Proteins as baits

One approach to investigate the elements that constitute the replicative complex involves using validated elements of the complex as "baits" in co-immunoprecipitation experiments.

2.1. Immunoprecipitation

Protein immunoprecipitation (IP) followed by mass spectrometry is a powerful technique used to selectively isolate and identify proteins that interact with a protein of interest. This approach enables the comprehensive characterization of protein-protein interactions, and has successfully been used in malaria parasites [244], [346]–[348]. The process relies on specifically targeting a protein of interest using antibodies that recognize and bind to it with high affinity and specificity (Fig. 58). Subsequently, the proteins that co-purify with the protein of interest provide valuable clues about potential interactors within the complex.

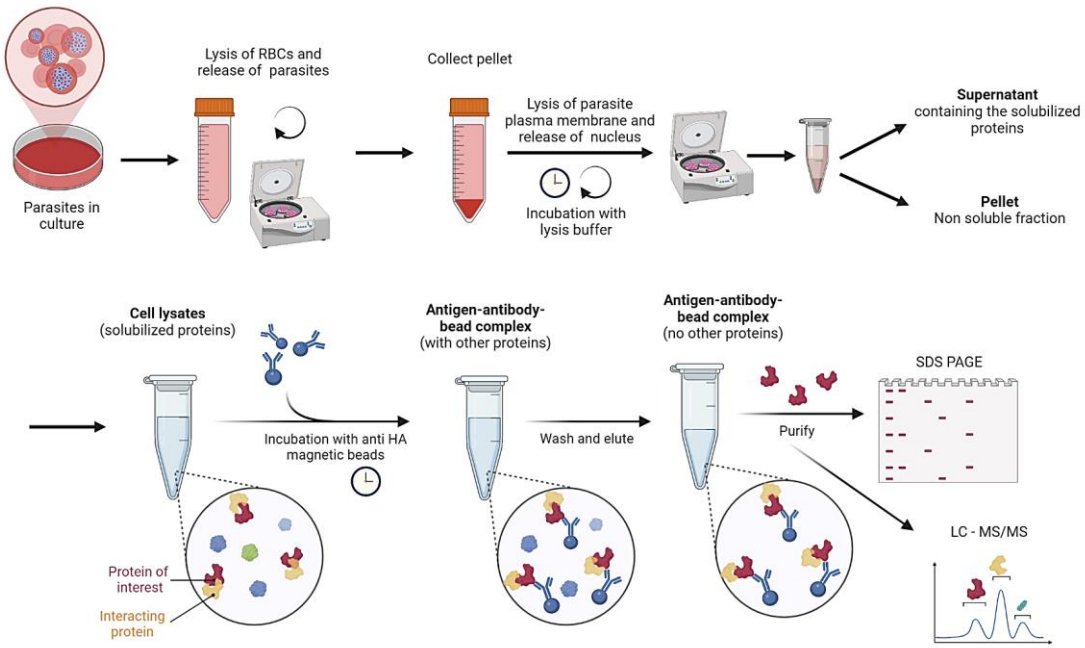


***Figure 58. Protein Immunoprecipitation (IP) protocol overview**. The workflow starts by the lysis of the RBC membrane and release of the parasites by centrifugation. Sequential lysis of parasite's plasma membrane and nuclear membrane release the solubilized proteins. Incubation of the cell lysate with anti-HA magnetic beads allows the isolation of the protein of interest along with potential interactors, that can be later analysed by MS or immunoblot. Done with Biorender.com*

Taking advantage of the *Pf*ORC2::HA expressing parasite line previously generated in the lab, I used *Pf*ORC2 as a bait in co-Immunoprecipitation (IP) assays to investigate the composition of the pre-replicative complex. Schizont stage parasites were harvested, and protein complexes were incubated with magnetic beads coated with anti-HA antibodies after lysis. As shown in Fig. 59a, although the signal in the input lane (positive control) was absent, we can observe a clear signal of the

immunoprecipitated *Pf*ORC2, with a double band at ~ 120 kDa, suggesting that the protein might be subjected to post translational modifications such as ubiquitination (size of ubiquitin is ~8 kDa). This was previously seen in the western blot result of the *Pf*ORC::HA proteins (Fig. 25). Additionally, the bands corresponding to the heavy and light chains of the anti-HA antibodies are also visible. In contrast, all signal was absent in the wild type lanes, confirming the specificity of the anti-HA antibodies. However, since the *Pf*ORC2::HA protein was not visible in a silver-stained gel (Fig. 59b), we did not proceed to mass spectrometry analysis, as we did not reach the criteria set by the platform. After several attempts to obtain a sufficient amount of protein to analyse by mass spectrometry, unfortunately the experiment never yielded the desired quantity. Therefore, I decided to approach the identification of members of the replication machinery from a different perspective: using proximity labelling.

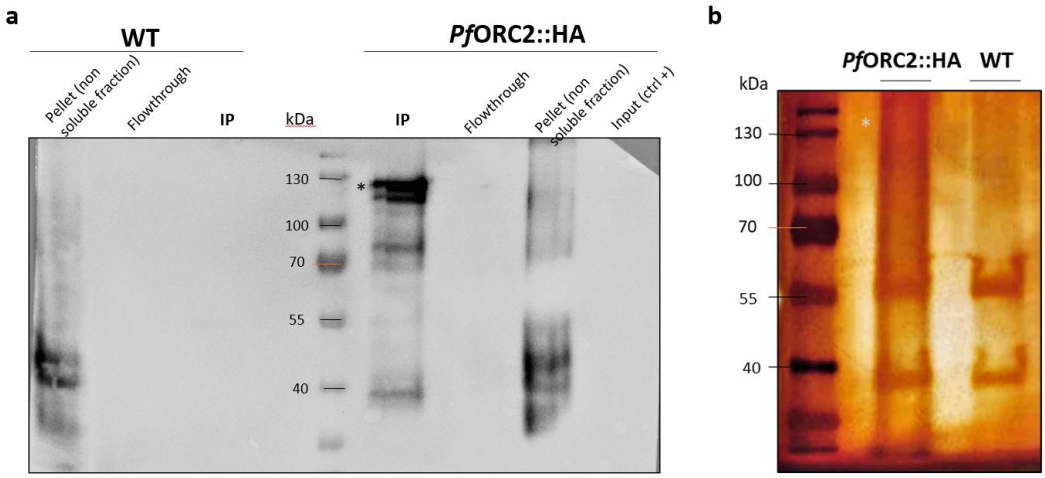


Figure 59. Result of the Immunoprecipitation attempt of PfORC2::HA*. a) Immunoblot of wild type P. falciparum 3D7 and PfORC2::HA parasites using anti-HA antibodies. Equivalent amounts of proteins were loaded per well. b) Silver Staining of the whole protein content of an SDS-PAGE gel containing equal amounts of proteins of PfORC2::HA parasites and wild type P. falciparum 3D7.*

2.2. Proximity labelling

Since the immunoprecipitation attempts did not yield the desired results, I decided to approach the exploration of the members of the replisome differently and aimed to perform proximity labelling (Turbo-ID [349]) for the identification of the ORC2 interactome.

The Turbo-ID technique has emerged as an improved version of the Bio-ID approach [349], [350] and it relies on the use of proximity-dependent biotinylation (Bio-ID) technology to study protein-protein interactions and identify interacting partners within a cellular context. A protein of interest (POI) is fused to a promiscuous biotin ligase enzyme from *Escherichia coli* that can biotinylate proteins in its

proximity [349], [351]. Upon expression of the POI-BirA fusion protein and addition of biotin to the cell culture medium, the biotin ligase biotinylates interacting proteins in close proximity to the POI (Fig. 60). The biotinylated proteins can then be captured using streptavidin coated magnetic beads, followed by elution and subsequent identification by MS. The Turbo-ID technique uses a mutated version of the Bio-ID BirA* (the R118G mutation improves its biotinylating activity, marked by the asterisk in BirA*), which exhibits enhanced catalytic activity, leading to faster and more efficient biotinylation of the interacting proteins than its predecessor [352]. This improved enzymatic activity allows for shorter labelling times that can be as short as 10 minutes, in comparison to up to 16 hours that were needed in Bio-ID, allowing for the detection of weaker or transient protein-protein interactions that would have been missed by the previous approach [353].

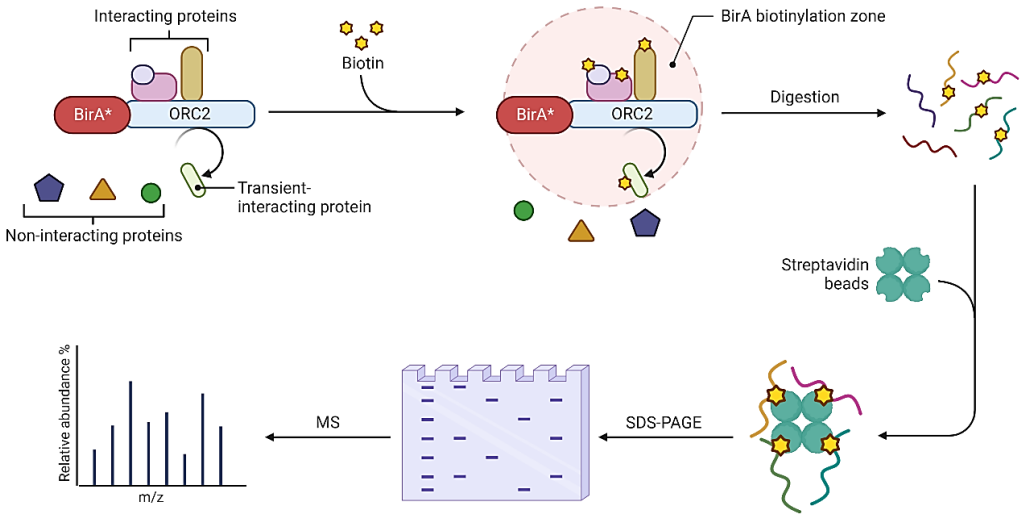


Figure 60. Scheme of the protocol of Turbo-ID of PfORC2. *The biotin ligase attached to the protein of interest (PfORC2) biotinylates interacting proteins when biotin is added to culture. Then, sequential lysis of RBC membrane, parasite's plasma membrane and nuclear membrane release the solubilized proteins. Incubation of the protein lysate with streptavidin magnetic beads allows for the pulldown of the biotinylated proteins, that can later be analysed on SDS-PAGE gels or MS. Done with Biorender.com*

In order to perform TurboID of *Pf*ORC2, I first attempted to generate a parasite line expressing *Pf*ORC2 fused with the TurboID biotin ligase BirA* in addition to a triple hemagglutinin (HA_3) tag at the C-terminal end, using CRISPR/Cas9 (Fig. 61a). I transfected wild type *P. falciparum* 3D7 parasites with these two plasmids in 6 different occasions, using the same guide RNA used to generate *Pf*ORC2::HA parasites, as well as a second one, but never obtained transgenic parasites.

As an alternative approach to CRISPR/Cas9, I decided to utilize the Selection Linked Integration (SLI) method for the generation of transgenic *P. falciparum* parasites [354]. In the SLI approach, a plasmid carrying a promoterless targeting region (an homology region covering the end of the gene) along with the desired tag is coupled with an additional selectable drug marker, separated by a skip peptide (T2A) in addition to the hDHFR drug resistance marker [354] (Fig. 61b). Initially, parasites carrying the pSLI

vector as an episome are selected with WR99210. Then, by applying a second round of drug pressure with the additional integration-linked drug (Neomycin), only parasites that will have integrated the whole plasmid via a single recombination event will survive. This additional selectable drug resistance cassette (Neomycin) will only be expressed after a single crossover integration event into the target locus and, thanks to the skipping peptide T2A, it will not be attached to the POI itself [355]. Unfortunately, after generating the tagging vector and transfecting wild type parasites several times I did not obtain parasites that carried the tag.

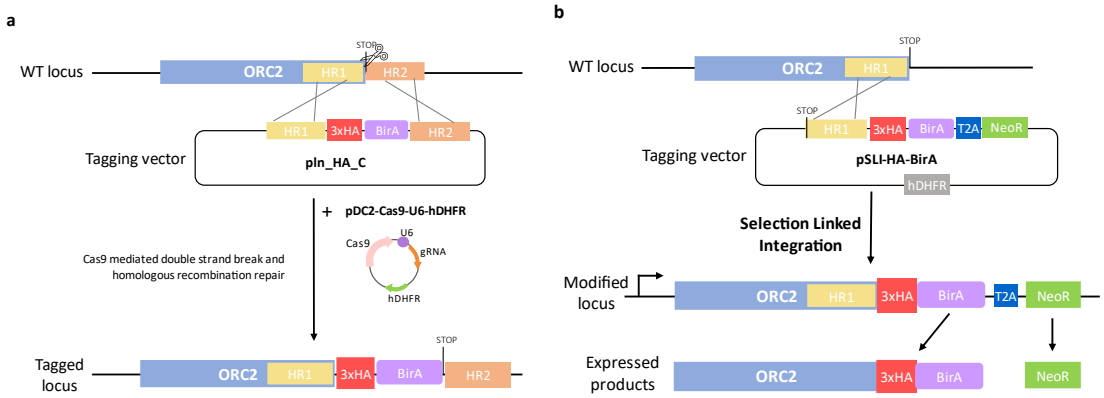


Figure 61. Strategy to generate the PfORC2::HA$_3$-BirA* line. *a) Vector construction for CRISPR/Cas9 strategy. b) Vector construction for the Selection linked integration strategy.*

3. DNA as bait: iPOND

The isolation of proteins on nascent DNA (iPOND) is an alternative approach for selectively purifying the proteins that play a role in DNA replication by using the newly replicated DNA as the bait, instead of proteins [310], [356], [357]. This provides a method to examine protein recruitment and interactions at active replication forks. It relies on the use of the thymidine analog 5-ethynyl-2-deoxyuridine (EdU), which is incorporated into newly synthesized DNA. To perform the iPOND experiments, I took advantage of the previously mentioned Thymidine Kinase expressing parasite line that allows the incorporation into DNA of EdU or BrdU [287]. The alkyne group in EdU enables the biotin-azide to be tethered to this nascent DNA fragment in a highly efficient cycloaddition reaction [315]. Then, after shearing the DNA-protein complexes by sonication, they can be isolated using streptavidin-based affinity purification and analysed using common protein detection techniques, such as mass spectrometry (MS) [358] (Fig. 62).

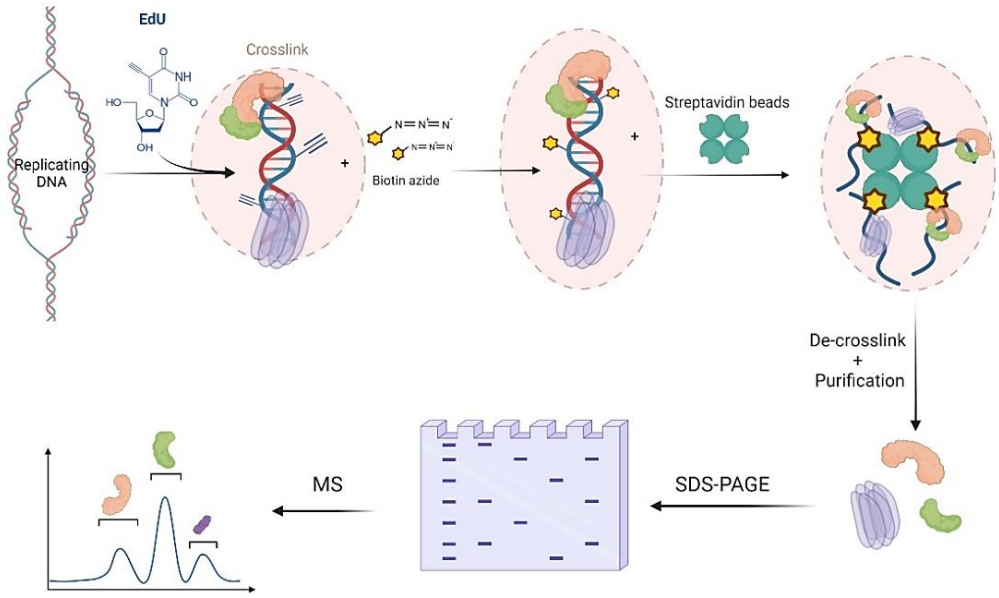


Figure 62. iPOND protocol overview. *Replicating parasites are incubated with a short pulse of EdU followed by crosslinking of the DNA with the proteins attached to it. Incubation with biotin, which binds to EdU within the DNA, followed by the streptavidin-based purification allows for the isolation of the proteins that were bound to nascent DNA, which are the putative members of the replisome complex. Done with Biorender.com*

To obtain a robust identification of proteins involved in replication, I performed the iPOND experiment using three different types of samples (Fig. 63). In the first, the "EdU" experiment (Fig. 63a), crosslinking of nascent DNA-protein complexes was performed immediately after the EdU pulse. This allowed for the isolation of proteins present on the replication fork*, i.e.,* the replisome components. The other two types of samples were two different negative controls to ensure the specificity of identification of only replication fork-associated proteins. The first consisted on performing the same treatment as the standard replication fork sample (EdU pulse followed by crosslinking), while omitting the azide component in the Click reaction (Fig. 63b) - "no click reaction control" (NCC). In the second negative control, I performed a pulse-chase type of experiment. Following the EdU pulse, cells were incubated with Thymidine (*i.e.,* chase) for 45 minutes (Fig. 63c). As DNA replication proceeded after the pulse, this allowed the replication complex to move away from the EdU-containing DNA. This sample thus aimed at capturing chromatin-bound proteins which are not directly involved in DNA synthesis. In all cases, the procedure started by labelling replicating parasites (trophozoite stage, 29 hpi) with EdU for a short period of time (2 min).

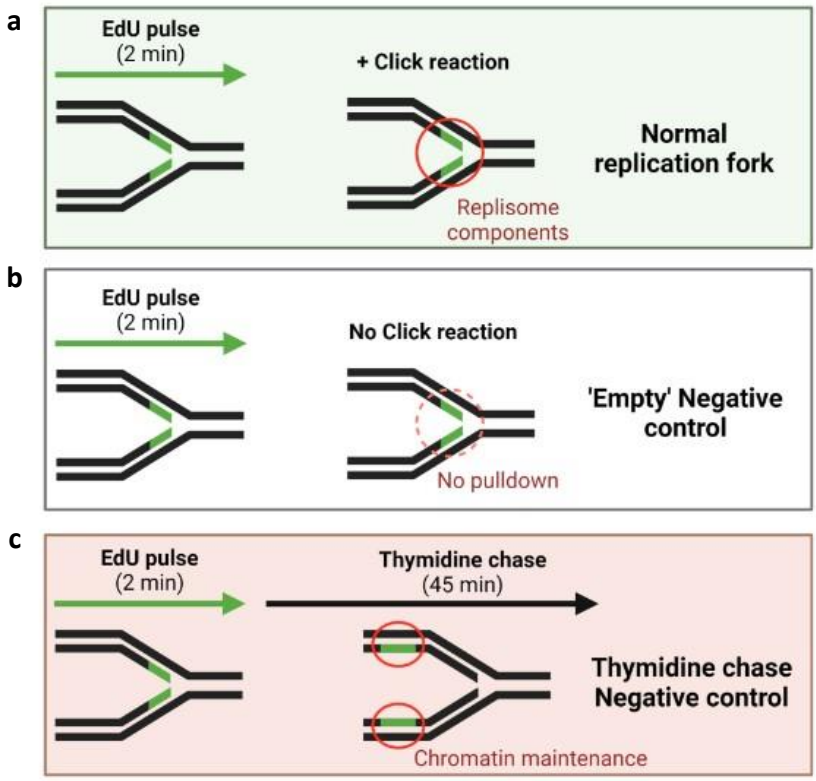


***Figure 63. Diagram of the three types of samples from the iPOND experiment**. The black lines represent unlabelled DNA and green segments represent DNA labelled with EdU. a) Experiment to detect proteins that localize at elongating replication forks. Parasites are incubated for two minutes with EdU prior to fixation. Click reaction allows to pull down EdU-labelled nascent DNA along with replisome components. b) Cells labelled for two minutes with EdU prior to fixation are processed without the click reaction agent (biotin-azide). No proteins should be obtained after pulldown. c) Cells labelled for two minutes with EdU were then treated with Thymidine for 45 minutes prior to iPOND. Proteins pulled down here should be involved in chromatin maintenance, whereas true replication proteins will no longer be enriched in this chase control. Made with Biorender.com. Adapted from* [356].

3.1. Results

Before proceeding to mass spectrometry analysis, I did a pilot experiment in which I omitted the thymidine chase sample to analyse our ability to capture proteins from EdU labelled chromatin. This was analysed with three methods: i) detection of the whole protein content on Ponceau staining of the Click pulldown vs. the flowthrough (Fig. 64a); ii) detection of the purified protein samples (Click, and No Click Control) on an SDS-PAGE gel followed by Silver Staining of the whole protein content (Fig. 64b); and iii) immunoblotting analysis of the immunoprecipitated fraction using a nuclear and a cytoplasmic marker to investigate the specificity of the method (Fig. 64c).

A substantial enrichment of proteins in the pulldown fraction was detected compared to the corresponding flowthrough fraction (Fig. 64a), demonstrating the effectiveness of the EdU-based labelling strategy in capturing chromatin associated proteins. The enrichment observed in the Click fraction, compared to the “No Click Control” fraction suggests that the click reaction successfully

targeted and captured proteins on nascent DNA (Fig. 64b). I next validated the specificity of this method in the capture of nuclear proteins using two different markers: Histone H3 (nuclear marker) and fructose-bisphosphate aldolase (FBPA; cytoplasmic marker). Only the H3 marker was detected in all the samples, indicating that the protocol specifically enriches nuclear proteins that interact with or are in close proximity to nascent DNA. (Fig. 64c).

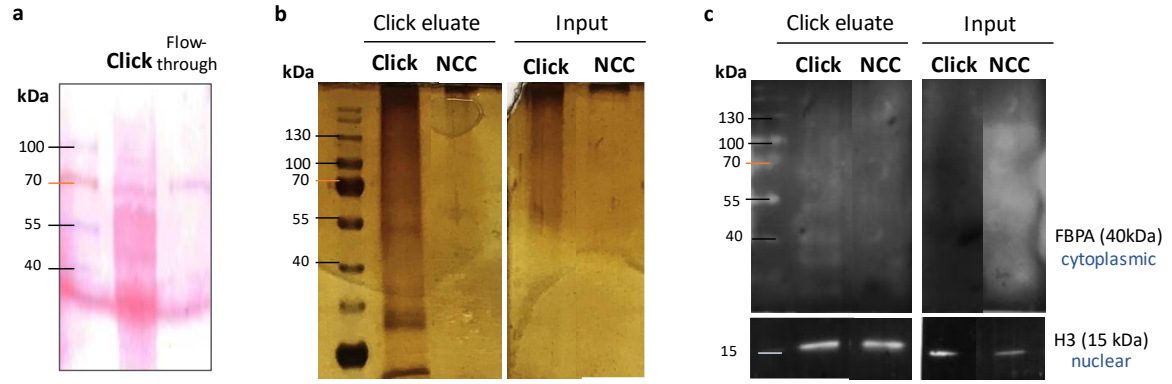


Figure 64. Validation of the iPOND method to purify proteins attached to newly synthesized DNA. *a) Ponceau staining of the whole protein content in an immunoblot membrane of pull-down proteins of the click reaction vs the flowthrough. b) Silver staining result of an SDS-PAGE gel comparing the whole protein content obtained after protein elution versus the input sample both in the Click and the No Click control (NCC, which omitted the azide from the click reaction). Equal amount of proteins were loaded in each well. c) Immunoblot of TK⁺ parasites using anti-H3 and anti-FBPA antibodies. Equal amount of proteins were loaded per well.*

Overall, the successful detection of proteins purified using the click chemistry-based pulldown approach provided a solid foundation for proceeding to downstream analysis with mass spectrometry to identify and characterize the proteins associated with nascent DNA. For this, replicating TK⁺ parasites (29 hpi) were subjected to a short pulse of 100 µM EdU (2 minutes) followed by direct crosslinking with paraformaldehyde or by a 45-minute incubation with thymidine 1 mM in the case of the Thy chase samples.

In the first iPOND experiment, I prepared four replicates for the EdU and thymidine chase samples, and two No Click Control (NCC) samples. In an attempt to simplify the protocol, I first performed the click reaction using an agarose resin as the azide, which allowed direct recovery of the interacting proteins by centrifugation of the "clicked" complexes. A total of 2529 proteins were identified, but only proteins that were present in all replicates of at least one sample were kept for differential enrichment analysis (2130 proteins). Although in this experiment the results of the replicates of each sample type were consistent and similar amongst each other (Fig. 65a), the amount of proteins detected in the NCC sample was too high, even higher than in the other pulldown samples, when it should have been negligible (Fig. 65b).

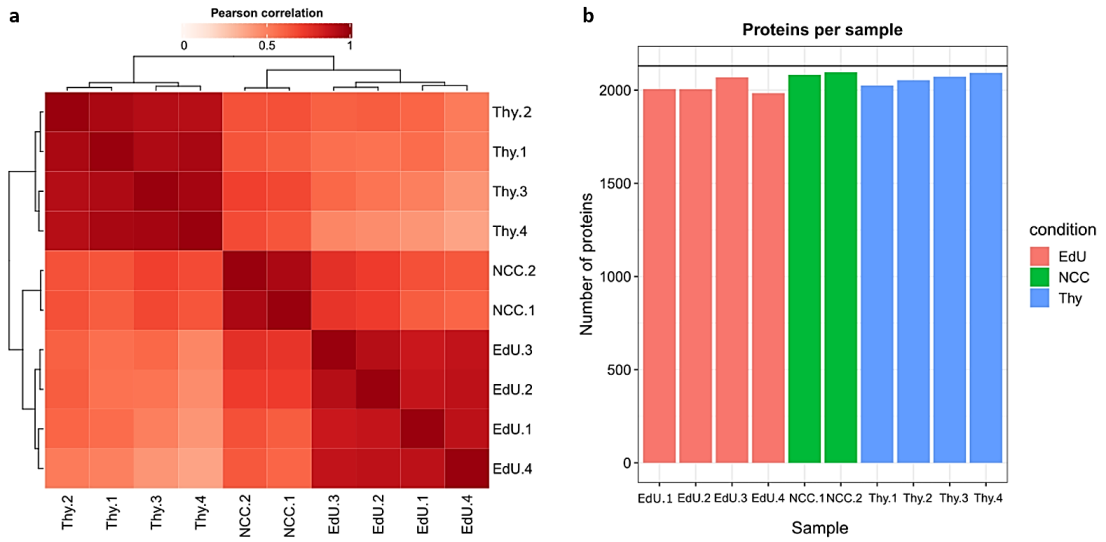


Figure 65. Quality control of the Mass Spectrometry results of the first IPOND experiment. *a) Pearson correlation between the samples. White represents no correlation (Pearson's coefficient = 0) and dark red complete correlation (Pearson's coefficient = 1). b) Number of proteins identified in each MS sample. Each sample type is represented with a different colour and replicates are also indicated.*

In addition, to validate the set of proteins obtained in each sample type, I selected two proteins known to be involved in either DNA replication or DNA repair to use as controls of the EdU and the thymidine chase samples, respectively. These two proteins were *Pf*ORC5, which is a part of the origin recognition complex [252] and, thus, a key element of the replisome machinery that should be highly enriched in the EdU sample; and *Pf*Rad51, a protein with a role in DNA repair involved in homologous recombination [359], which should display a significant enrichment in the thymidine chase sample. In the case of *Pf*ORC5, we can observe a slight enrichment in the EdU sample as compared to the Thy (difference in the median intensity EdU vs. Thy = 0.0716) (Fig. 66a). However, in the case of *Pf*Rad51, although there is an enrichment in the thymidine chase sample as compared to the EdU sample, the highest enrichment was observed in the NCC (difference in the median intensity EdU vs. NCC = - 0.312, Thy vs. NCC= - 0.138), which should have been devoid of most of the protein signal (Fig. 66b), suggesting that the detected signal could be entirely due to unspecific binding.

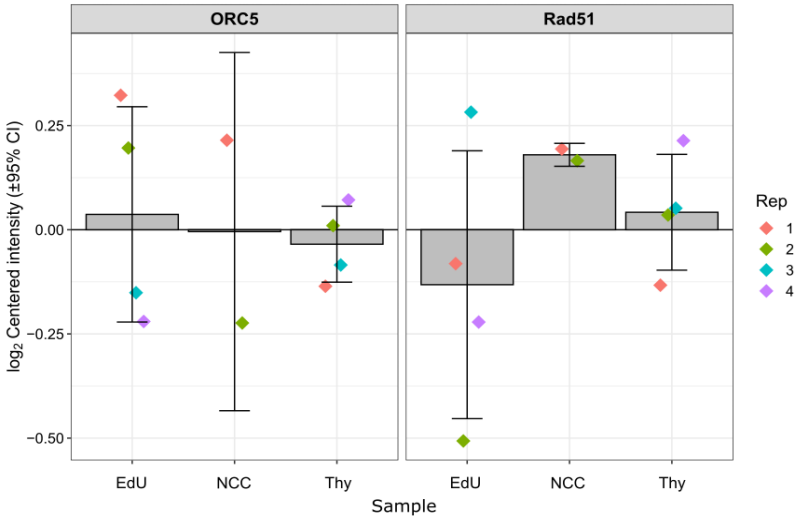


Figure 66. Differential enrichment of PfORC5 and PfRad51 in the different iPOND-MS samples. *Y axis represents the log2 of the median intensity of the signal detected in each sample with 95% confidence interval. Replicates are indicated with points of different colours.*

The high number of unspecific proteins detected in the NCC sample made it impossible to detect proteins specifically and significantly (p-value < 0.05) enriched in the EdU sample as compared to the NCC or Thy sample. I therefore hypothesised that the agarose slurry was perhaps too porous and allowed excessive unspecific binding which generated high levels of noise in the dataset. As a result, I next changed the strategy and used a biotin-azide for the click reaction with the EdU-labelled DNA and then performed the IP using streptavidin magnetic beads. In addition, I increased the sonication time to 20 min. The latter was intended as an additional measure to prevent background noise originating from unspecific (passenger) proteins bound to long fragments of DNA containing non-EdU labelled regions.

The analysis of the second experiment revealed a lower number of identified proteins (1071 proteins). Only proteins that were present in all replicates of at least one sample were kept for differential enrichment analysis, leaving a total amount of 811 proteins (Fig. 67). In this experiment, the number of proteins identified in the No Click Control sample (NCC), although lower than in the EdU and Thy pulldown samples, was still very high (all NCC replicates identified more than 600 proteins), indicating that this protocol requires further optimisation. In a future experiment I would recommend perhaps an even longer sonication step (30 min) in order to obtain a lower number of proteins identified in the NCC.

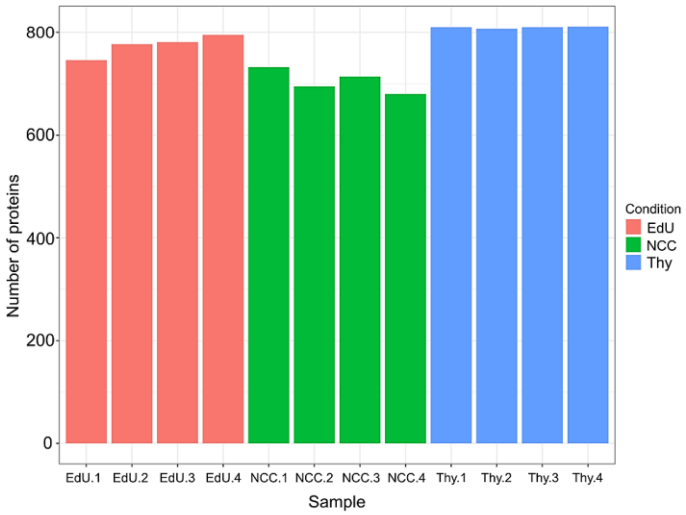


Figure 67. Number of proteins detected per MS sample in the second iPOND experiment*. Each sample type is represented with a different colour and replicates are also indicated.*

Still, I was able to confirm that this experimental approach is promising by verifying the enrichment of proteins that are known to be involved in active replication (in the EdU sample), such as *Pf*PCNA1 (difference in the median intensity EdU vs. Thy = 0.215); and proteins associated with more mature chromatin (in the Thy chase sample), such an RNA binding protein (Q8IJZ3) (difference in the median intensity EdU vs. Thy = - 0.509) (Fig. 68).

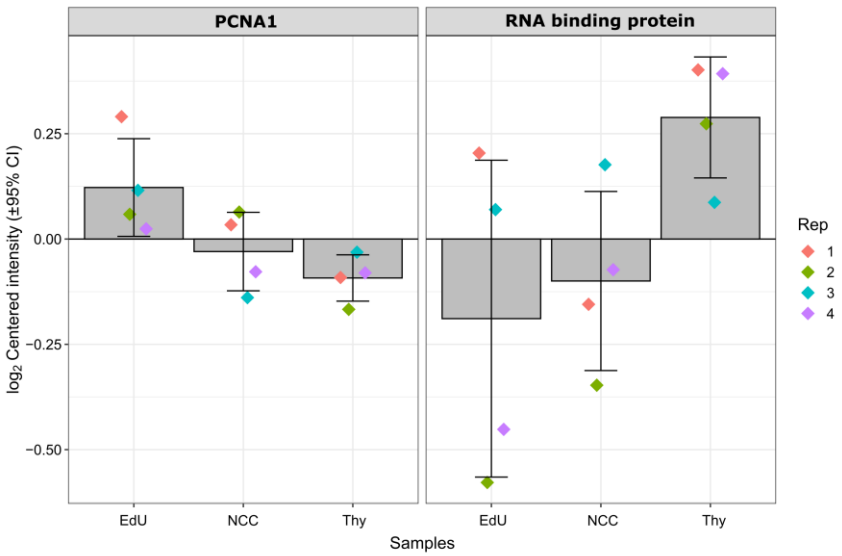


Figure 68. Differential enrichment of specific proteins in the EdU, NCC and Thymidine chase sample*. PCNA1, known to be a part of the replisome machinery; and an RNA binding protein (Q8IJZ3).*

Several known replisome proteins such as members of the ORC or MCM complex and DNA polymerases expected to be enriched in the EdU fraction were not detected as enriched. For instance, in the case of MCM2, MCM3, MCM6, DNA polymerase δ, Replication protein A, and replication factor C, although identified by MS, not enough peptides were identified, or the fold enrichment values and statistical reproducibility did not meet the stringent filtering criteria.

By establishing an adjusted p-value threshold of 0.05, 11 proteins (out of the 811) were significantly enriched on the EdU sample as compared to the Thy (Fig. 69). Unfortunately, all the fold change values between the two conditions ranged between 1,1 and 1,5 (Table 5).

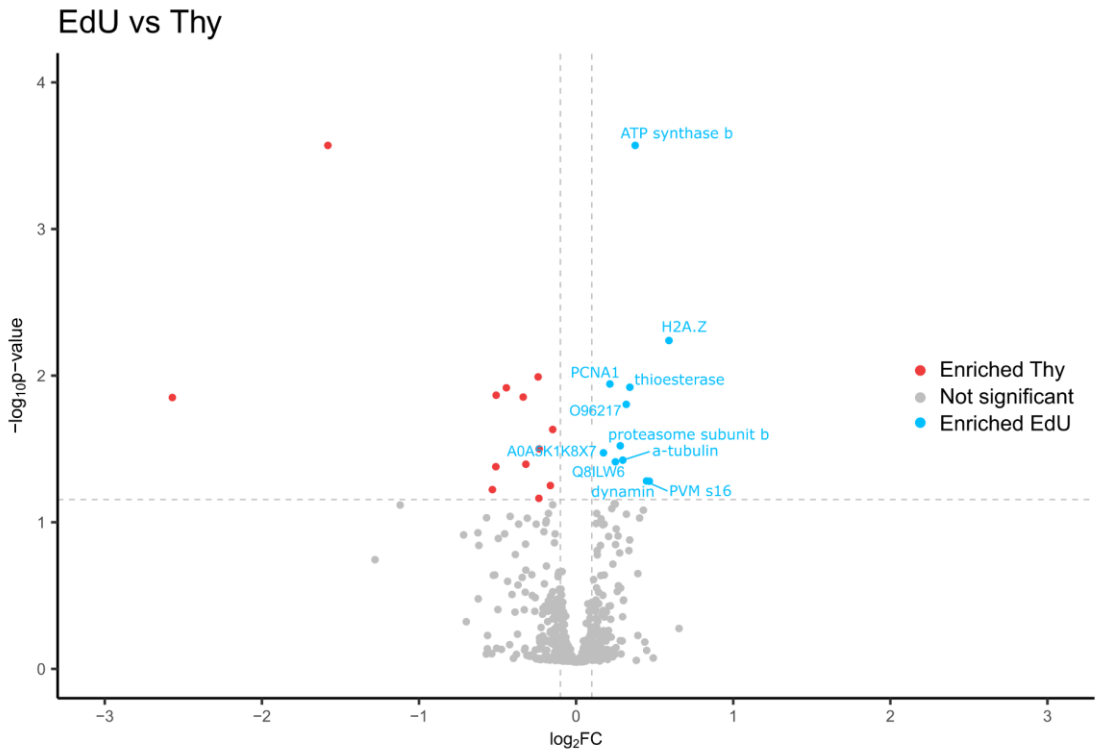


Figure 69. Volcano plot of the proteins differentially enriched in the EdU vs. Thy samples. *The log2 fold enrichment of the EdU sample proteins relative to the Thy chase negative control is plotted against the adjusted p-value of both samples. Proteins enriched in EdU over Thy with a p-value over the threshold of 0.05 are indicated in blue; the complete gene names are indicated in table 4 below. Proteins enriched in the Thy sample over EdU with p-value < 0.05 are indicated in red.*

Among the set of proteins showing the highest enrichment in the EdU fraction compared to the Thy sample there was *Pf*PCNA1 and the histone variant H2A.Z. In addition, this analysis revealed the enrichment on the EdU sample of additional proteins which were not readily linked to replication, such as a conserved protein of unknown function (O96217), but could be interesting candidates for future characterization (Table 5).

To further investigate the potential role in replication of the set of proteins enriched in the EdU sample, I performed a STRING analysis for known and predicted protein-protein interactions (PPI) [312]. For this analysis, I applied a less stringent p-value cut-off and included proteins that, although enriched in the EdU sample over Thy, displayed a less significant p-value ($p < 0.1$), making a total list of 19 proteins (Table 5).

Table 5. Proteins enriched in the EdU fraction over Thy sample. *The proteins shaded in light blue are the ones with a significant adjusted p-value (< 0.05) identified as EdU enriched in the volcano plot (Fig. 69).*

Protein ID	Gene	Adjusted p-value	Ratio EdU/Thy
Q8I0V2	ATP synthase subunit beta, mitochondrial	0.000269	1,298
O97320	Histone H2A.Z	0.00576	1,506
P61074	PCNA1	0.0114	1,161
O96259	thioesterase, putative	0.012	1,268
O96217	conserved protein, unknown function	0.0157	1,247
Q7K6A9	proteasome subunit beta type-4	0.0301	1,215
A0A5K1K8X7	nuclear cap-binding protein subunit 1, putative	0.0336	1,128
Q8IFP3	alpha tubulin 2	0.0377	1,229
Q8ILW6	Glycylpeptide N-tetradecanoyltransferase	0.0387	1,189
Q8I5M3	dynamin-like protein, putative	0.0423	1,364
Q6ZMA7	Parasitophorous vacuole membrane (PVM) protein S16	0.0425	1,381
A0A5K1K8W0	Polyubiquitin binding protein, putative	0.0745	1,183
O97241	ubiquitin conjugating enzyme	0.0752	1,188
Q8IEU2	gamete antigen 27/25	0.0807	1,170
Q8IM71	choline kinase (CK)	0.0827	1,345
Q8ILI6	Acidic leucine-rich nuclear phosphoprotein 32 (ANP32)	0.087	1,095
Q8IB14	High mobility group protein B2 (HMGB2)	0.0881	1,248
C6KT25	malate dehydrogenase	0.0936	1,323
Q8IHY0	Protein phosphatase PPM2	0.0948	1,117

The resulting protein-protein interaction (PPI) network revealed abundant functional interactions between most of the members, with an average node degree of 2.42 and a PPI enrichment p-value of 1.13e-05 (Fig. 70), indicating that this set of proteins have more interactions among themselves than what would be expected for a random set of proteins of the same size. Such an enrichment allowed me to conclude that the proteins are at least partially biologically connected.

The *Pf*PCNA1 and the histone variant H2A.Z were in the centre of the proposed PPI network displaying connections with many other proteins, including some indirectly related to DNA replication like the high mobility group protein B2 (HMGB2), suggested to be involved in chromatin remodelling for transcriptional regulation after replication [360], and an acidic leucine-rich nuclear phosphoprotein 32-related protein (ANP32) that has not yet been characterized in *P. falciparum* but shows predicted

interactions on an individualized STRING analysis with a nucleosome assembly protein (Q8I2W3), a histone acetyltransferase (Q8III2), the DNA methyltransferase 1-associated protein 1 (C6KTC1) and an ATP-dependent helicase (C0H4W3).

In addition, other proteins have been identified in nascent DNA, like a putative ubiquitin-conjugating enzyme E2 (O97241), and a putative polyubiquitin binding protein (A0A5K1K8W0), which could be involved in the ubiquitination of certain members of the replisome to regulate their activity. In fact, in human cells, PCNA activity is regulated through poly- and de-ubiquitination by the ELG1 protein [361].

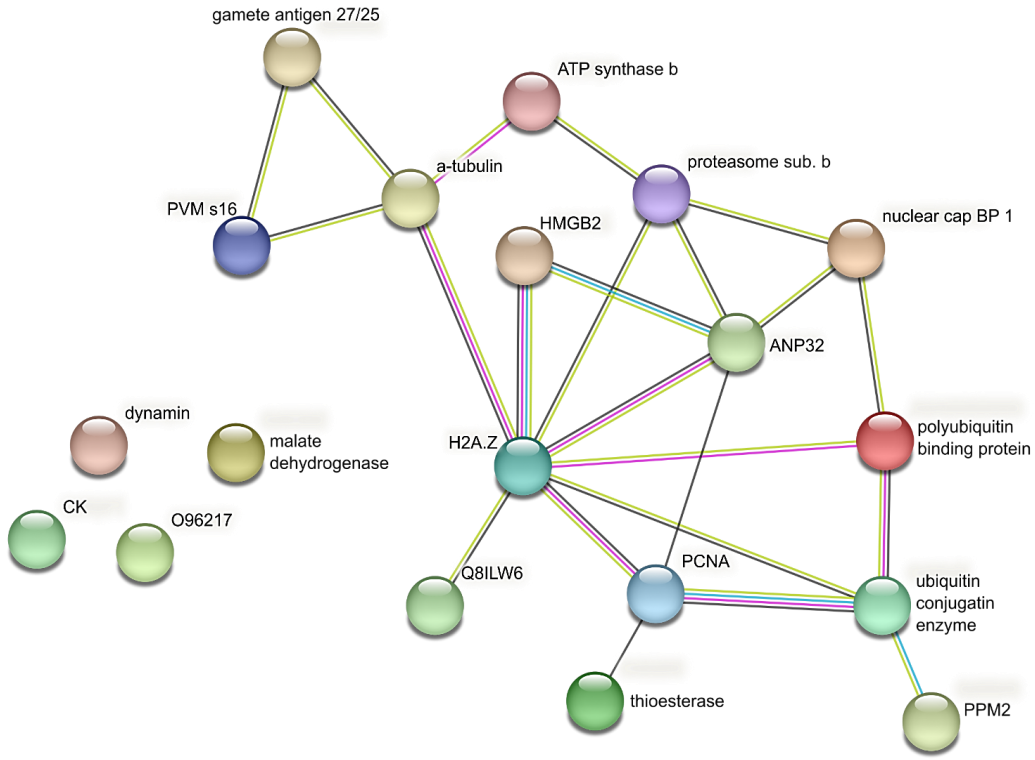


Figure 70. Interaction Network for Proteins Enriched in Nascent DNA. *The image illustrates a network analysis of proteins enriched on nascent DNA molecules (Table 5) created with STRING. The different colours of the connecting lines represent the types of evidence supporting each association: co-expression (brown), functional experiments (pink), databases (cyan), and text-mining (yellow).*

Next, I performed a Gene Ontology analysis of the list of the 19 proteins enriched in EdU vs thymidine (Table 5, Fig. 71). Indeed, the characterization of the "biological process" associated with the proteins of the dataset revealed an enrichment in terms associated to DNA replication such as leading strand elongation, DNA metabolic process, and regulation of DNA replication. In addition, terms associated with chromatin processing after replication were also enriched (post-replication repair, regulation of transcription, DNA synthesis involved in DNA repair, translesion synthesis and chromatin silencing), suggesting that, in *P. falciparum*, the mechanisms directing DNA repair might be happening almost synchronously as the DNA is being synthesized. In addition, there was a strong enrichment of the term "N-terminal protein lipidation", which involves the covalent attachment of a lipid group to the amino

terminus of a protein [362], [363]. This kind of protein post-translational modification could be indirectly related to active DNA replication since protein lipidation is implicated in regulation of phosphatase and kinase activities [364], which are essential players of activation of DNA replication.

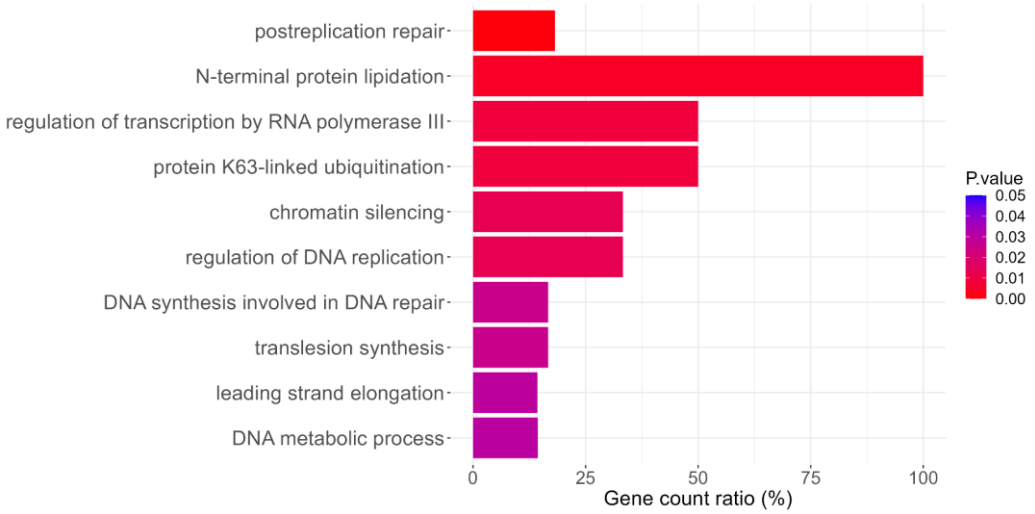


Figure 71. Gene Ontology analysis of the biological process of the proteins enriched in nascent DNA

These results suggest that the iPOND-MS screen successfully identified proteins associated with DNA replication, although the direct contribution to DNA replication of some the proteins enriched remains to be explored.

4. Work in progress

In addition to the successful experiments described in the previous sections, it is also important to acknowledge the valuable insights gained from experiments that did not yield the desired outcomes. Towards the goal of characterising the replicative complex of *P. falciparum* I attempted to generate additional epitope tagged lines that would have provided complementary information, but unfortunately were not successfully obtained.

In this section, I will present a comprehensive account of the experimental approaches undertaken, the challenges encountered, and the lessons learned from these attempts, shedding light on the complexities of working with *Plasmodium falciparum* parasites.

4.1. ORC subunits as baits

As previously introduced, the origin recognition complex plays a crucial role in the initiation of DNA replication by binding to the origins of replication. Due to their pivotal function, ORC subunits have

served as excellent bait proteins, allowing for the isolation and study of interacting replisome components in other systems [365], [366].

In addition to trying to co-IP *Pf*ORC2::HA and to generate the Turbo-ID line *Pf*ORC2::HA-BirA*, I also tried to verify direct interactions between the ORC proteins using a co-IP strategy in a double tagged line. To do that I used CRISPR/Cas9 aiming to insert a Ty epitope tag in the C-terminal end of either *Pf*ORC1 or *Pf*ORC5 in the parasite line carrying *Pf*ORC2::HA. I modified the existing C-terminal HA constructions and replaced the HA tag with a Ty tag (Fig. 72). Unfortunately, after 8 attempts using two different guide RNAs, I did not recover transgenic parasites.

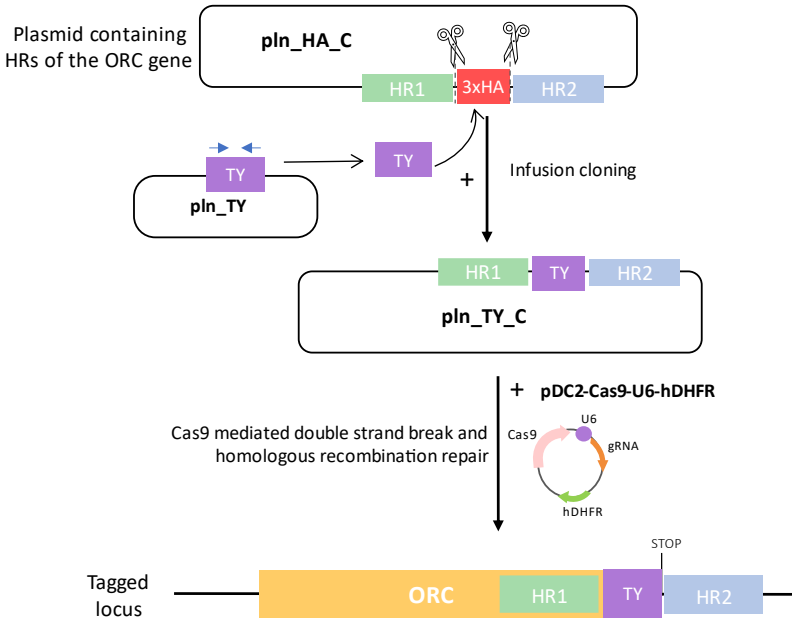


Figure 72. Strategy to generate the Ty tagged PfORC1 and 5 lines. *The Ty tagged was cloned by PCR and inserted by infusion cloning at the place of the HA tag in the pIn_HA_C vector.*

4.2. MCM subunits

In parallel, I also designed a strategy to exploit the *Pf*MCM proteins as baits to further characterise the members of the replicative machinery. I started by selecting *Pf*MCM2 and *Pf*MCM6 for the following reasons: *Pf*MCM2 is the only one of the MCM proteins that has a Nuclear Localization Signal (NLS) [247], and therefore potentially enters the nucleus independently from the other proteins, making it an interesting target for characterization and to use as bait to identify other interacting members of the replisome. On the other hand, *Pf*MCM6 has been shown to bind strongly to the chromatin [247], which made it a great target for: i) co-IP experiments and ii) additional ChIP-seq experiments at the onset of DNA replication to map origins of replication and compare this profile with that of the ORCs.

Previously characterised in *P. falciparum* [247], these MCM subunits show a high sequence identity with their human homologs (~40%). For this reason, I decided to generate endogenously tagged parasites by adding the triple hemagglutinin (HA_3) epitope tag in the same terminal sites as their human homologs: N-term for *Pf*MCM2, and C-term for *Pf*MCM6. To do this, the first approach was based on the use of CRISPR/Cas9 to insert the HA_3 tag in these genes (Fig. 73).

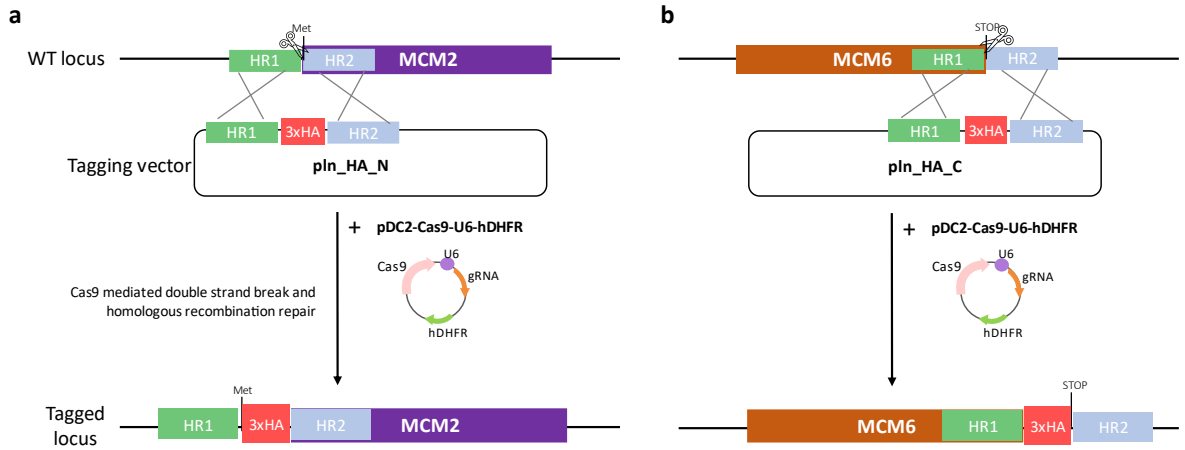


Figure 73. Strategy to generate parasites carrying a HA_3 tag in MCM2 or MCM6. *The HA_3 tag is inserted in the N-term of MCM2 (a) and C-term of MCM6 (b).*

I transfected the pLN_HA_MCM2 and the Cas9 vector into wild type *Pf*3D7 ring stage parasites in 8 different occasions and with two different guide RNAs but never obtained transgenic parasites. In the case of *Pf*MCM6, after 4 transfections, I obtained a mixed population of edited and non-edited parasites that survived the drug pressure and showed positive signal on a diagnostic PCR as well as on a preliminary western blot and IFA (Fig. 74). Unfortunately, I was unable to obtain pure clones.

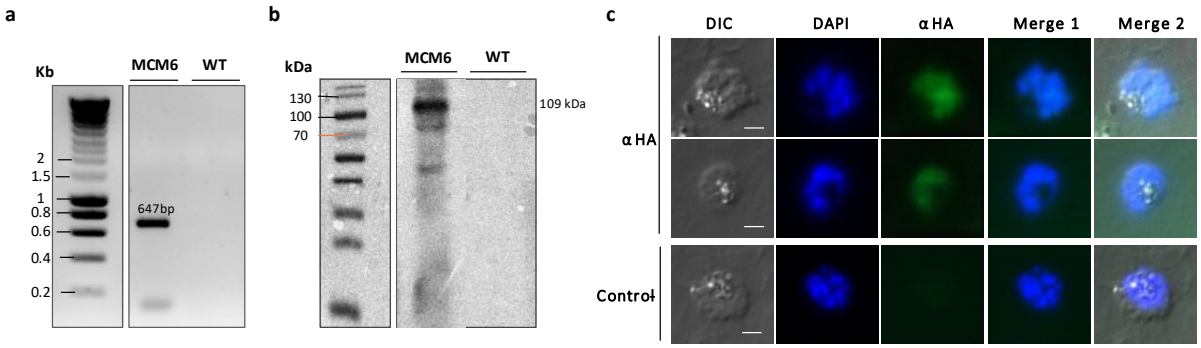


Figure 74. Detection of tag in MCM6::HA_3 mixed population*. a) Genotyping PCR of the integration of the tag at the 3' end. Expected size is indicated above the band. b) Detection of the HA_3 tagged MCM6 protein by immunoblotting. Expected size is indicated beside the blot. Equal amounts of proteins were loaded in each lane. c) Detection of the HA tagged PfMCM6 protein by immunofluorescence in schizont stage parasites. Scale bar represents 2 µm.*

In addition, I have also generated a targeting construct to obtain *Pf*MCM6::HA_3-BirA* parasites and perform Turbo-ID experiments (which would be informative even in a mixed population of edited and non-edited parasites) but following several transfections I did not obtain transgenic parasites.

4.3. *PCNA1*

The Proliferating Cell Nuclear Antigen (PCNA) protein has been shown to be a good marker of active DNA replication [2], [367]. During replication, PCNA tethers the DNA polymerases to the template DNA and forms a ring-shape complex around the DNA acting as a sliding clamp pushing the DNA polymerase progressing through the DNA molecule [214], [216]. Being a key component of the replisome machinery, I hypothesised that it could serve as an ideal candidate bait to perform proximity labelling in replicating *P. falciparum* parasites and hence identify the members of the replicative complex [367].

At first, I aimed to use a similar strategy as the one used for *Pf*ORC2 and *Pf*MCM6: CRISPR/Cas9 to insert an HA_3-BirA* tag at the C-terminal end of the *Pf*PCNA1 gene. However, after several unsuccessful attempts, I decided to clone the full length (FL) CDS region (835 bp long) into the pLN_HA_C plasmid, along with the promoter, and express it episomally (Fig. 75a).

I obtained resistant parasites carrying the episomal FL_PCNA1::HA_3-BirA* extra copy. This was confirmed by diagnostic PCR (Fig. 75b), and I validated the localization and expression of the tagged *Pf*PCNA1 inside the nuclei by immunofluorescence (Fig. 75c). However, the HA_3 tagged protein could not be detected by western blot and, after several attempts of incubating the parasites with biotin followed by streptavidin pulldown, no difference was seen between the samples incubated with biotin and without. This prevented me from proceeding to the proximity labelling experiment to analyse the biotinylated interactors by mass spectrometry.

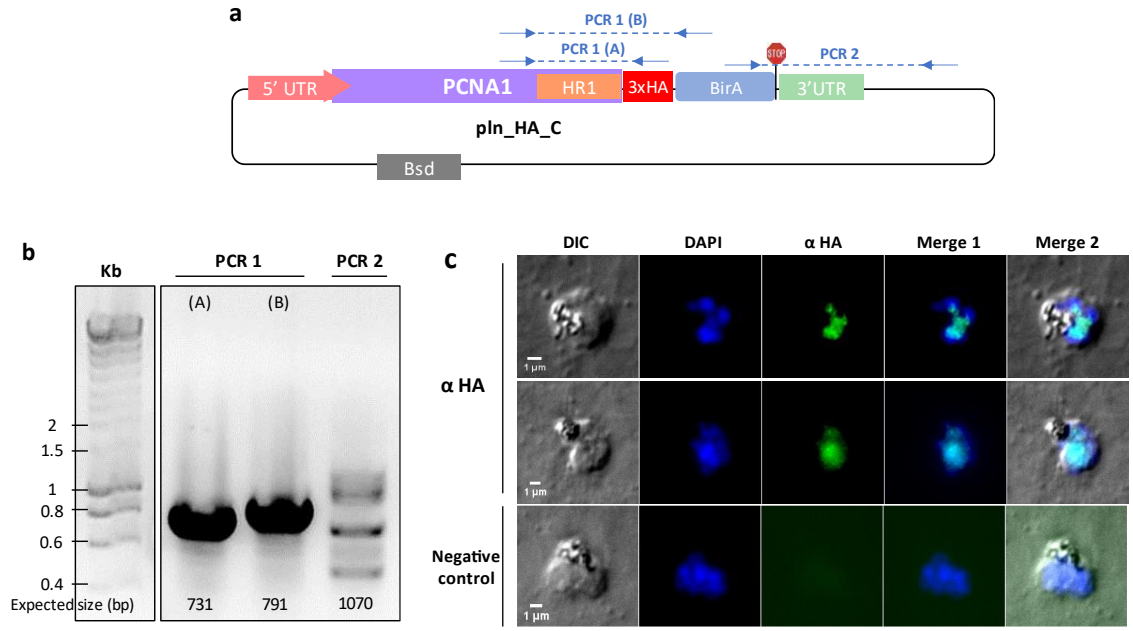


Figure 75. Full length PfPCNA1 tagged with HA_3-BirA* epitope tags. *a) Plasmid construction of the Full length PCNA1::HA_3-BirA. b) Genotyping PCR of the tagged PfPCNA1. Two different PCR reactions were performed to confirm the presence of both tags in the episomal vector: the HA_3 (PCR 1 – A) and the BirA* (PCR 1 – B). Expected sizes are indicated below. c) Detection of the HA_3 tagged PfPCNA1 protein (in green) by immunofluorescence in schizont stage parasites. DNA is labelled with DAPI in blue. Scale bar represents 1 µm.*

4.4. *CRK1*

Another protein potentially involved in DNA replication and/or cell cycle regulation is *Pf*Crk1 (cdc2-related protein kinase 1). The *P. berghei* homolog *Pb*Crk-1 was shown to be essential for the completion of the intraerythrocytic asexual cycle [368].

This made the *Pf*Crk-1 gene an interesting candidate to study its contribution to DNA replication in *P. falciparum* parasites. I first attempted to knock-out the *Pf*Crk-1 gene using CRISPR/Cas9 to remove the entire CDS (Fig. 76). I transfected this construction along with two different guide RNAs on several occasions, but never obtained mutant parasites.

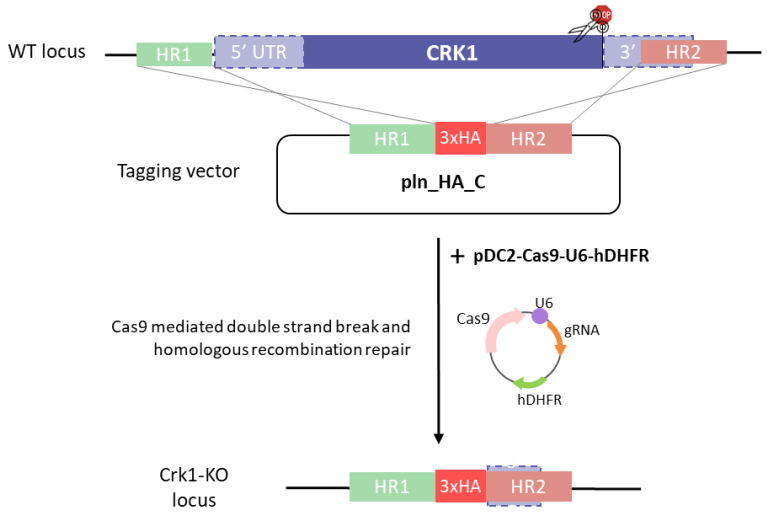


Figure 76. PfCrk-1 Knock-out strategy with CRISPR/Cas9.

5. Conclusion and perspectives

The second aim of my PhD project focused in dissecting the composition of the replicative complex of *Plasmodium falciparum* parasites. To achieve this, I have leveraged different technologies to identify and characterize the key members of the replisome machinery as well as their interactors. Although the initial approach to use proteins as baits and perform immunoprecipitation experiments was not successful, the iPOND strategy yielded moderately successful results. However, further optimisation is still necessary. This has allowed me to validate the presence of known members of this complex, such as *Pf*ORC5 or *Pf*PCNA1, but also to identify potential candidates that may play a role in *P. falciparum* replication. Further validation of the candidate proteins will be required in future studies.

Finally, the team will continue with our efforts to tag additional members of the replicative machinery, such as other subunits of the *Pf*MCM complex, aiming to characterise them and perform proximity labelling experiments to study their interactome. The results will be compared with the set of proteins identified with those from the iPOND experiments, strengthening our knowledge of the key players of DNA replication and potentially identifying additional unknown interactors.

The multiple approaches that I have taken to obtain a snapshot of the replicative complex have proven challenging and illustrate the complexities of working with *P. falciparum* parasites. However, they have allowed me to explore new strategies and gain expertise in plasmid design for the generation of transgenic parasite lines, and experiment planning and optimization of protocols for protein characterization.

General discussion

Throughout the development of my PhD project, I have explored different aspects of the initiation of DNA replication in *P. falciparum*, including the identification of origins of replication and characterization of their genetic landscape, the exploration of the dynamics of replication fork progression throughout the genome, and the study of the replisome components.

Most of the efforts to study DNA replication in eukaryotes have focused on model organisms from the Opisthokonta group, such as metazoans and yeast. Although a few recent studies have started to explore the dynamics of DNA replication in asexual intraerythrocytic *P. falciparum* parasites [1]–[4], the molecular mechanisms driving initiation of replication and the genomic determinants of replication origin specification remained largely elusive. Apicomplexan parasites are evolutionarily divergent from the opisthokonts and, as such, display significant differences in core biology and do not necessarily adhere to principles applied to canonical eukaryotic DNA replication. During their development inside host erythrocytes, the parasites replicate their genome multiple times within the same cell cycle, which allows them to increase their population by several orders of magnitude very quickly. Since this high multiplication rate is linked to their pathogenicity, treatments targeting this replicative process have the potential to be highly effective. The increasing emergence of widespread resistance to commonly used antimalarials enhances the need for developing new drugs that target the proliferative stage of this pathogen.

The goal of this study was to unravel the complexities of DNA replication in *P. falciparum.* To do so, I have combined three different approaches to map the plasmodial origins of DNA replication, as well as analysed the composition of the replicative complex through isolation of proteins bound to active replication forks. By adopting a multidisciplinary approach encompassing techniques such as genomic analysis, molecular biology, and proteomics, I sought to contribute to the growing body of knowledge surrounding this intricate process.

1. Genetic landscape of the origins of replication

First, by measuring the increase of DNA content through flow cytometry and timing the incorporation of a thymidine analogue into DNA, I concluded that about 30% of a semi-synchronous parasite culture initiated the first round of DNA synthesis at ~29 hours post infection (hpi). This was in concordance with the estimated timing of onset of S-phase in *Plasmodium* parasites previously described [369]. Then, the ChIP-seq experiment allowed me to map the binding sites of *Pf*ORC1 and *Pf*ORC2 proteins before the start of replication (25 hpi). Although the presence of specific ARS-like consensus sequences in *P. falciparum* had been previously suggested [328], I did not identify any origin-specific motif in the *Pf*ORC$_{1\text{-}2}$ binding sites dataset that supported this. Given the extremely high AT content of

the *P. falciparum* genome (>80%), *Pf*ORC$_{1-2}$ binding sites were surprisingly found in regions with relatively high GC content (27%) and were associated with the presence of G-quadruplex forming sequences (G4FS). G-quadruplexes have been suggested to play a role in origin specification in mammalian cells, as well as *Drosophila* and *Leishmania major* parasites [7]–[9], through excluding nucleosome occupancy and, thus, favouring ORC binding. In the case of *P. falciparum*, it is still unclear whether G-quadruplexes directly recruit ORC proteins, or if this secondary structure provides a region devoid of nucleosomes that in turn promote *Pf*ORC binding. Another feature known to influence ORC recruitment in metazoan organisms is the presence of specific histone post-translational modifications (PTMs) like H3K9 acetylation or H3K4 methylation [317], [330], [370], [371]. However, available *Pf*PTM datasets were not found to be significantly associated with *Pf*ORC$_{1-2}$ binding sites, suggesting that the presence of these epigenetic marks is not likely to be involved in *Pf*ORC$_{1-2}$ recruitment in *P. falciparum*.

In contrast, active transcription was found to be a strong determinant of origin specification, although TSS and TTS are depleted in *Pf*ORC$_{1-2}$ sites, contrary to human cells, where the presence of an active TSS is necessary and sufficient for ORC binding [317], [323]. This strong association with active transcription is also different from yeast ORC binding sites, where they are mainly found in intergenic regions, but also excluded from TSSs [176]. Interestingly, *Pf*ORC$_{1-2}$ were found to be enriched on the bodies of transcriptionally active genes, which was an unexpected finding since it could be a source of conflict between the replication and transcription machineries. However, the presence of the RNA polymerase in actively transcribed genes might in turn lead to a more relaxed status of the surrounding chromatin and lead to the exclusion of nucleosomes, favouring binding of the ORC to those genes. In *Drosophila* cells, for example, the ORC binding sites also lack a consensus origin motif and are enriched at genomic regions with an increased nucleosome turnover, such as the promoters of actively transcribed genes [371]. The extreme AT-richness of the intergenic regions and promoters (90-95%) of *P. falciparum* genes could be the reason of the depletion of *Pf*ORC$_{1-2}$ binding sites in TSSs, since the more balanced GC content of the internal regions of genes would provide a more stable platform for the assembly of the pre-replicative complex, hence, promoting recruitment here.

Another surprising feature was found in *var* genes, where promoters were significantly enriched in *Pf*ORC$_{1-2}$ binding sites, showing a completely different pattern of enrichment than in core genes or even in other heterochromatin multigene families like *rifin* and *stevor* genes, that also display an enrichment of *Pf*ORC$_{1-2}$ binding sites within their gene bodies. Promoters of *var* genes display also a significant enrichment in G4FS [337], which have been suggested to play a role in the transcriptional control of this gene family [372]. Whether it is the presence of G4FS in these GC-rich regions what attracts *Pf*ORC binding, or in contrast it is their direct implication in the transcriptional silencing of the

cognate *var* gene remains to be elucidated. The *Pf*ORC1 subunit has already been shown to have a specific role in *var* gene expression regulation through an association with the telomere-associated histone deacetylase *Pf*Sir2 [253], [336]. Our results suggest that this implication in *var* gene silencing is probably conserved in *Pf*ORC2, although the exact mechanism of the regulation of mono-allelic expression in *P. falciparum* parasites is still unresolved.

Furthermore, I identified 4796 active origins of replication in *P. falciparum* by sequencing short nascent strands (SNS-seq) at the beginning of schizogony, which is a number comparable to what has been reported in the human parasite *Leishmania major* with the same origin mapping method [8], and the genomes of both species are of a comparable size (33 Mb of *Lm* vs. 24 Mb of *Pf*). The set of active ORIs identified displayed a genome-wide distribution highly similar to the one of *Pf*ORC$_{1\text{-}2}$ binding sites, with most of the *Pf*ORC$_{1\text{-}2}$ binding sites being located within 2 kb of an active site of DNA replication initiation. Active ORIs are also associated with G4FS, even in a stronger manner, and display a similar TSS/TTS depletion and enrichment within gene bodies. Interestingly, our data suggests that strong transcription in *P. falciparum* is strongly associated with origin firing, since active ORIs were strongly enriched in genes from the top quartile of expression. To avoid conflicts between transcription and replication, the activation of origins in these highly expressed genes requires a spatial-temporal coordination mechanism that regulates the activity of the different polymerase complexes. In addition, parasites need to maintain a balance in the number of ORIs to activate replication from, since they need to activate enough origins to allow replication of the whole genome in the absence of genome instability, but also activate as few as possible to minimize the risk of replication-transcription conflicts.

Finally, the third method I used to map active origins of replication has proven to be highly efficient and accurate for an organism with a relatively small genome (23 Mb) like *P. falciparum*. With Oxford nanopore sequencing technology we can map the incorporation of nucleoside analogues into DNA from the specific electric profile signatures that they generate when the DNA strand translocates through the pore, allowing to detect active replication forks and infer the coordinates of replication initiation sites using NanoForkSpeed [5], [10]. The NFS ORIs dataset obtained was highly concordant with the mapped *Pf*ORC$_{1\text{-}2}$ binding sites and SNS-seq origins, illustrating the robustness of the NFS method and validating its efficacy to identify origins of replication throughout the genome of *Plasmodium falciparum*.

Our results also show the existence of replication origin clusters or “initiation zones” where most of the ORIs (~75%) could be found. These were defined as 10 kb-long genomic regions that contained 3 or more *Pf*ORC$_{1\text{-}2}$ binding sites, or 5 or more active ORIs mapped with SNS-seq or NFS. These origin

clusters come from population-based methods and therefore reflect a general preference for certain replication initiation regions in the genome, where each origin is not necessarily activated in the same cell, but the population tendency is to initiate DNA replication from these regions. Given the significant association between strong transcription and origin firing that our results show, having a restrictive sequence-based ORI specification system could lead to spatial and temporal conflicts between replication and transcription that could eventually hamper proper parasite growth and multiplication. In contrast, having a more flexible replication firing initiation zones approach and a mechanism that coordinates DNA replication initiation with active transcription ensures the efficiency of the replication program.

In other model organisms, such as *Drosophila,* human or mouse, the number of licensed origins (*i.e.,* ORC binding sites) is significantly higher than the number of origins where replication is actually triggered from. For instance, in *S. cerevisiae*, where origins are defined by the consensus ARS, the overall origin firing efficiency is less than 50%, with some origins being activated every cell cycle and the remaining inactive [199]. In contrast, human cells display an origin firing efficiency that ranges between 5-20% [200]. This excess of origin licensing has a critical role in the maintenance of genome stability, as the additional not activated (dormant) origins could eventually be activated and trigger replication if needed. This mechanism ensures the successful replication of the whole genome in the case of fork stalling events due to replicative stress. In contrast, the number of licensed origins I detected (*Pf*$ORC_{1\text{-}2}$ binding sites: 1861) was not higher than the number of active origins mapped with SNS-seq (4796) or NFS (t29h: 3786). This is probably the result of the stringent parameters I used for the identification of the *Pf*$ORC_{1\text{-}2}$ binding sites, since these sites had to be common to four different datasets (two different proteins, each with two different replicates), that in turn ensured a high confidence licensed origins dataset.

Moreover, NanoForkSpeed allows to investigate the pace at which replication forks progress through the DNA. I detected a median fork speed of ~1.5 kb/min that was altered and decreased significantly in specific regions of the genome with a highly packed chromatin state that might hamper replication fork progression, like centromeres and telomeres; but remained fairly constant regardless of the stage of schizogony, as there was only a difference of 46 bp/min (3%) between the two timepoints. Whether this slight decrease of replication fork velocity at 35 hpi is biologically relevant remains unclear. These results are consistent with previously published DNA combing data, where replication speed was reported to be 1.4 kb/min for 24-30 hpi parasites and 1.2 kb/min for 32-38 hpi parasites [341]. The use of techniques like DNA combing has allowed to make single cell measurements, instead of population averages, of replication fork speed and inter-origin distances (IODs). These two

measurements are variable between species, with an average replication fork speed of 2.1 kb/min in *S. cerevisiae* [5], for example, and 1.3 kb/min in human cells [223]. IODs are very large in trypanosomatid parasites such as *Trypanosoma brucei* (160 kb) and *Leishmania mexicana* (226 kb) [241], [373] or shorter in other systems, like 137 kb in mouse cells, 73 kb in Drosophila [163], or 46 kb in yeast [374]. In my case, long reads that harboured multiple origins allowed the estimation of single nuclei IOD, which ranged from 9 kb to 46 kb at 29 hpi and from 6 kb to 48 kb at 35 hpi. These IODs are slightly shorter than what was reported in the previously mentioned DNA combing study [341], where the IOD for 24-30 hpi parasites was 68.5 kb. However, this difference is probably caused by the limited range in length of the nanopore reads that contained more than one origin, which did not go over ~99kb, while DNA combing experiments allow for the stretching of DNA fragments reaching up to 500 kb long [341].

In addition, the single molecule measurements performed with NFS showed a significant correlation between IOD and the cognate replication fork speed, which was true for both timepoints. Longer IODs correlated with faster fork progression, while shorter IODs correlated with slower fork progression, akin to what has been observed in human cells [223], regardless of the advanced status of schizogony. Overall, these results indicate that asexual *Plasmodium falciparum* parasites can complete each round of DNA replication in 15 minutes, both at the beginning or middle of schizogony.

Altogether, this study provides insights into the replicative process that allows these parasites to massively proliferate within the human host by several orders of magnitude in a matter of days.

2. Replication machinery of *Plasmodium falciparum*

The second main objective of my PhD project was to identify the composition of the replicative complex of *Plasmodium falciparum*.

To achieve this, I initially used an approach that involved utilizing proteins key to DNA replication as bait molecules in immunoprecipitation and proximity labelling experiments to selectively capture the interacting proteins. Since these initial attempts were not successful I changed the approach and explored alternative methodologies that permitted the isolation of proteins specifically associated with nascent DNA at active replication forks - iPOND.

By leveraging the iPOND methodology, I was able to isolate proteins directly associated with the replicative complex during active DNA replication and identify a set of proteins enriched in nascent

DNA. Among these are known key components of the replisome, such as *Pf*PCNA1, but other known members were not detected, such as subunits of the ORC or DNA polymerases. The STRING analysis of protein-protein interactions between the 19 proteins enriched in active replication forks, hence potentially involved in DNA replication, revealed a high number of functional interactions within most members. Some of the identified proteins were predicted to be related to mature chromatin processes, like the high mobility group protein B2 (HMGB2), suggested to be involved in chromatin remodelling for transcriptional regulation after replication [360], or the acidic leucine-rich nuclear phosphoprotein 32-related protein (ANP32), for which a STRING individualized analysis showed interactions with a nucleosome assembly protein (Q8I2W3), a histone acetyltransferase (Q8III2), the DNA methyltransferase 1-associated protein 1 (C6KTC1) and an ATP-dependent helicase involved in chromatin remodelling (C0H4W3). The fact that they were enriched at active replication forks could suggest that, in *P. falciparum*, the mechanism of DNA replication is coupled almost simultaneously with the mechanism of chromatin processing that includes re-assembly of nucleosomes, re-establishment of PTMs in histone tails and chromatin remodelling for transcriptional regulation. In addition, this notion is strengthened by the results of the gene ontology analysis performed, where the biological processes enriched range from terms clearly related to DNA replication, such as leading strand elongation, DNA metabolic process, or regulation of DNA replication; to terms also associated with chromatin processing after replication, like DNA repair, translesion synthesis and chromatin silencing, suggesting that the mechanisms directing DNA repair or chromatin modifications may potentially be happening in a tightly synchronous manner to DNA replication. Likewise, the putative ubiquitin-conjugating enzyme E2 (O97241) and the putative polyubiquitin binding protein (A0A5K1K8W0) could have an indirect role in DNA replication by mediating the ubiquitination of key members of the replisome machinery, such as *Pf*PCNA1 (whose activity in human cells is regulated through poly- and de-ubiquitination [361]) or *Pf*ORC2 (in concordance to our WB results that show the presence of a ubiquitinated form of *Pf*ORC2, Fig. 25), which makes them interesting candidates to study their contribution to this process.

The identified enriched proteins can now be subjected to functional analyses to elucidate their specific roles in DNA replication and assess their potential as targets for antimalarial interventions. The expression and localization of the potential candidates will be characterized throughout the parasite's intraerythrocytic life cycle and their contribution to DNA replication will be analysed.

Altogether, though some of the attempted approaches did not yield the desired outcomes, the insights gained from these experiments serve as a foundation for future perspectives, highlighting the need for continued efforts to unravel the intricate mechanisms underlying DNA replication of this pathogen.

Appendices

Appendix 1. Primers and gRNAs used in this study

Primers used in this study

Primer name	Sequence (5'-3')	Description
argm0045	CTATGACGTACCAGACTATGC	triple HA tag genotyping
argm0046	GCATAGTCTGGTACGTCATAG	triple HA tag genotyping
argm0078	GTTGTGTGGAATTGTGAGCG	sequence guides in pDC2-Cas9-U6-hDHFR vector
argm0104	GAATGGATGGATACCTATTGAG	downstream of ORC2 HR2. 3' int. pcr
argm0106	CTATGATTACAAGACAACCAAC	downstream of ORC5 HR2. 3' int. pcr
argm0107	TATGTGCTTTGAATCTTTATTAAC	upstream of ORC5 HR1. 5' int. pcr
argm0108	GGTATGAATGTTGTTCATCCG	upstream of ORC1 HR1. 5' int. pcr
argm0109	AGAAATTAAACTTTTACTAGTACC	genotyping of ORC1::HA wt control
argm0111	GAGTTGATAATATAAAAAAATAAAC	genotyping of ORC5::HA wt control
argm0112	GGAGCTATAAAATATATAAAAGGG	genotyping of ORC2::HA wt control
argm0133	CTGCTTATATGGAATCGGATC	upstream of ORC2 HR1. 5' int. pcr
argm0216	GACAATACTGTGCCTCTGAAG	BirA tag genotyping
argm0217	CTTCAGAGGCACAGTATTGTC	BirA tag genotyping
argm0218	CATATGGATGTTTTCTTTTTACC	downstream of ORC1 HR2. 3' int. pcr
argm0220	AAATATATCACCTAGGCTGTGCGTCAGTTAGAATCG	MCM6 3xHA cloning primer HR1 fwd
argm0221	AGGGTATCCACCGCCGGCAAAATTGTCAATTTCTTCTTGAAAG	MCM6 3xHA cloning primer HR1 rev
argm0222	GGTGTGACCCCTTAAGATTGTAATATTATATATATCTATTATG	MCM6 3xHA cloning primer HR2 fwd
argm0223	AAGCTTGGGGGGATCCAAAATTGAAAAATACATTATTGTTACAAG	MCM6 3xHA cloning primer HR2 rev
argm0224	CGGCGGTGGAACTAGTGAGGTAATAAATAAATAAATAAATATA	MCM2 3xHA cloning primer HR2 fwd
argm0225	AAGCTTGGGGGGATCCTCTTGTTTTCTTTGCATCTGC	MCM2 3xHA cloning primer HR2 rev
argm0226	ATAAATAATTAAGATATCATATATATAAAGGGTTATTATATAAC	MCM2 3xHA cloning primer HR1 fwd
argm0227	AAGGGTACATCCTAGGTGTTTAATAAAGAATTATTTTATAATAT	MCM2 3xHA cloning primer HR1 rev
argm0228	GGTGTGACCCCTTAAGCCAATTAAAAAATAAAGTGATAAAAA	Crk1 KO cloning primer HR1 fwd
argm0229	AAGCTTGGGGGGATCCCGTATATACATATATATATTATCTT	Crk1 KO cloning primer HR1 rev
argm0230	AAATATATCACCTAGGAGAACATTGGAAATTTAAGATATGA	Crk1 KO cloning primer HR2 fwd
argm0231	AGGGTATCCACCGCCGGCTCTCTTTATTTTATTTATAAGGGGA	Crk1 KO cloning primer HR2 rev
argm0240	AATAGAAATATATCACCTAGGGATGTGATCTTAGAAAATGATGT	TY cloning primer for ORC5. fwd pcr1 (TY)
argm0241	CGGCGCGCCAAATAATCAACTCATCTAGAGGTATT	TY cloning primer for ORC5. rev pcr1 (TY)
argm0242	GTTGATTATTTGGCGCGCCGGCGGTGG	TY cloning primer for ORC5. fwd pcr2 (HR1)
argm0243	TGTATGGTATTTTTGCTTAAGTCACACCCCGTGGTCTAGA	TY cloning primer for ORC5. rev pcr2 (HR1)

argm0244	AATAGAAATATATCACCTAGGGTACTCTTTACCTTATTCGATTGGC	TY cloning primer for ORC1. fwd pcr1 (TY)
argm0245	CGGCGCGCCAGTAGAAGTTAAGTTTCTTAGAGC	TY cloning primer for ORC1. rev pcr1 (TY)
argm0246	TAACTTCTACTGGCGCGCCGGCGGTGG	TY cloning primer for ORC1. fwd pcr2 (HR1)
argm0247	TTTATTTTGATGGTACTTAAGTCACACCCCGTGGTCTAGA	TY cloning primer for ORC1. rev pcr2 (HR1)
argm0249	CGTGGTCTAGAGGGTCCTG	TY tag genotyping
argm0250	GTACATACTAACCAAGATCCAC	TY tag genotyping
argm0253	AGGTGCACAAAGATCTATGAG	upstream of MCM6 HR1. 5' int. pcr
argm0254	GAGAAAGAATTTATAAAATTATATAG	downstream of MCM6 HR2. 3' int. pcr
argm0259	CGAAAAGTGACTTAAGATTGTAATATTATATATATCTATTATG	MCM6 3xHA-BirA cloning primer HR2 fwd
argm0260	ATAAATAATTAAGATATCCTGTGCGTCAGTTAGAATCG	MCM6 3xHA-BirA cloning primer HR1 fwd
argm0261	CGGCCAATGCCCTAGGAAAATTGTCAATTTCTTCTTGAAAG	MCM6 3xHA-BirA cloning primer HR1 rev
argm0266	CGAAAAGTGACTTAAGAATATATAAAAGGGAAATTACAATT	ORC2 3xHA-BirA cloning primer HR2 fwd
argm0267	ATAAATAATTAAGATATCCAAAGAGTTAATGATTCTAATGTAC	ORC2 3xHA-BirA cloning primer HR1 fwd
argm0268	CGGCCAATGCCCTAGGTAGCTCCTCTGATATCCTTTTG	ORC2 3xHA-BirA cloning primer HR1 rev
argm0269	CGAAAAGTGACTTAAGGTACATATTTTCATATATATATTTATC	PCNA1 3xHA-BirA cloning primer HR2 fwd
argm0270	AAGCTTGGGGGGATCCTATTGGTGTGTGTATTCATATAAAAG	PCNA1 3xHA-BirA cloning primer HR2 rev
argm0271	ATAAATAATTAAGATATCGTTTTAGGTGTAAATATTGCATC	PCNA1 3xHA-BirA cloning primer HR1 fwd
argm0272	CGGCCAATGCCCTAGGATCTTTATTATCCATATCGTCATC	PCNA1 3xHA-BirA cloning primer HR1 rev
argm0293	CAAATCCTTCTTCACAATCTGG	PCNA1 genotyping primer
argm0296	GTAACATTTAAAAATTGTCAATTTC	genotyping of MCM6::HA wt control
argm0323	CTCATTATAGATGTGATCGTGA	sequencing primer PCNA1 full length
argm0350	TTCTTCAATGCCTAGGTGACAAAGAGTTAATGATTCTAATGTAC	fwd, cloning ORC2 HR1 (+ HA-BirA*) in pSLI vector
argm0351	TTCCTTCTCCGTCGACCTTTTCGGCAGACCGCAGAC	rev, cloning ORC2 HR1 (+ HA-BirA*) in pSLI
argm0352	TTCTTCAATGCCTAGGTGACTGTGCGTCAGTTAGAATCGC	fwd, cloning MCM6 HR1 (+ HA-BirA*) in pSLI
argm0353	TTCCTTCTCCGTCGACCACCCCGTGAGCATAATCCG	rev, cloning MCM6 HR1 (+HA-BirA*) in pSLI
argm0354	GCCTCACGCTACACTTGTG	NeoR+T2A in pSLI vector genotyping primer
argm0355	CGAACATTAAGCTGCCATATC	NeoR in pSLI vector genotyping primer

Guide RNAs used in this study

Name	Sequence (5'-3')	Description
g_argm0019	ATTGGATAAGGAATCTGGAGACAT	guide 1 sequence ORC1
g_argm0020	AAACATGTCTCCAGATTCCTTATC	guide 1 sequence ORC1
g_argm0021	ATTGAAATAACCAAAAAATTACTA	guide 2 sequence ORC1
g_argm0022	AAACTAGTAATTTTTTGGTTATTT	guide 2 sequence ORC1
g_argm0025	ATTGGGAGCTATAAAATATATAAA	guide sequence ORC2
g_argm0026	AAACTTTATATATTTTATAGCTCC	guide sequence ORC2
g_argm0023	ATTGTATATTATCAACTCATCTAG	guide sequence ORC5
g_argm0024	AAACCTAGATGAGTTGATAATATA	guide sequence ORC5
g_argm0062	ATTGATATATATAATATTACAATA	guide 1 sequence MCM6
g_argm0063	AAACTATTGTAATATTATATATAT	guide 1 sequence MCM6
g_argm0064	ATTGAATTATATAGTTCTTTAAAA	guide 2 sequence MCM6
g_argm0065	AAACTTTTAAAGAACTATATAATT	guide 2 sequence MCM6
g_argm0066	ATTGAATAATTCTTTATTAAACAA	guide 1 sequence MCM2
g_argm0067	AAACTTGTTTAATAAAGAATTATT	guide 1 sequence MCM2
g_argm0068	ATTGAATTCTTTATTAAACAATGG	guide 2 sequence MCM2
g_argm0069	AAACCCATTGTTTAATAAAGAATT	guide 2 sequence MCM2
g_argm0078	ATTGATAGGTATAGACCTGAAAAT	guide 1 sequence Crk1 KO
g_argm0079	AAACATTTTCAGGTCTATACCTAT	guide 1 sequence Crk1 KO
g_argm0080	ATTGGGATTCTTCAAAAAGGGAGG	guide 2 sequence Crk1 KO
g_argm0081	AAACCCTCCCTTTTTGAAGAATCC	guide 2 sequence Crk1 KO
g_argm0082	ATTGACTTTTATATGAATACACAC	guide 1 sequence PCNA1
g_argm0083	AAACGTGTGTATTCATATAAAAGT	guide 1 sequence PCNA1

Appendix 2. Replication timing (Fig. 21)

Immunofluorescence to measure DNA replication timing.

Fig 21b. Percentage of parasites in which DNA replication took place (*i.e.,* EdU positive) at 25, 29 and 35 hpi. At least one hundred parasites were counted for each timepoint.

Timepoint		**t25h**	**t29h**	**t35h**
EdU -	count	125	116	6
	%	98.4%	81.1%	3.4%
EdU +	count	2	27	168
	%	**1.6%**	**18.9%**	**96.6%**
Total cells		127	143	174

Fig 21c. Number of nuclei showing positive EdU signal per parasite. More than 40 parasites were counted for each timepoint.

EdU positive nuclei per cell	**t25h**	**t29h**	**t35h**
0	45	15	1
1	1	12	2
2	0	11	4
3	0	3	7
4	0	0	7
5	0	0	5
6	0	0	4
7	0	0	2
8	0	0	3
9	0	0	2
10	0	0	1
11	0	0	1
12	0	0	1
13	0	0	0
14	0	0	0
15	0	0	1
Average	**0**	**1.05**	**5.17**

Appendix 3. Binding sites of *Pf*ORC1, *Pf*ORC2 t25h

All sequencing data generated in this study can be made available upon request and will be deposited in the ENA database under the accession code PRJEB62206.

ChIP-seq results

Tables containing the narrow peaks coordinates obtained from MACS2 peak calling. Tables include Chromosome, Start, End, Name, Score, Fold enrichment, and -log10(q value).

- Sheet 1: *Pf*ORC1 intersection replicate 1 – replicate 2. Only narrow peaks coordinates common to both replicates were kept (before filtering). Fragments overlapping (if overlap was > 50 bp) with IgG dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 2: *Pf*ORC1 replicate 1. Fragments overlapping (if overlap was > 50 bp) with IgG dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 3: *Pf*ORC1 replicate 2. Fragments overlapping (if overlap was > 50 bp) with IgG dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 4: *Pf*ORC2 intersection replicate 1 – replicate 2. Only narrow peaks coordinates common to both replicates were kept (before filtering). Fragments overlapping (if overlap was > 50 bp) with IgG dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 5: *Pf*ORC2 replicate 1. Fragments overlapping (if overlap was > 50 bp) with IgG dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 6: *Pf*ORC2 replicate 2. Fragments overlapping (if overlap was > 50 bp) with IgG dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 7: *Pf*ORC$_{1\text{-}2}$. Only narrow peaks coordinates common to *Pf*ORC1 (intersection of rep. 1-2) and *Pf*ORC2 (intersection of rep. 1-2) were kept. Fragments overlapping (if overlap was > 50 bp) with IgG dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 8: *Pf*ORC$_{1\text{-}2}$ clusters. The genome was divided into 10 kb windows, with a sliding window of 5 kb. *Pf*ORC$_{1\text{-}2}$ sites were then overlapped with these windows. A cluster was then considered if a given 10 kb window contained 3 or more *Pf*ORC$_{1\text{-}2}$ sites. Table includes Chromosome, Start, End, Number of *Pf*ORC$_{1\text{-}2}$ sites found in that window, bp overlap between 10 kb window and *Pf*ORC$_{1\text{-}2}$ peaks, window size, and percentage (%) of window bp within peaks.

Appendix 4. SNS-seq origins

All sequencing data generated in this study can be made available upon request and will be deposited in the ENA database under the accession code PRJEB62206.

SNS-seq results

- Sheet 1: SNS-seq ORIs. Table containing the narrow peaks coordinates obtained from MACS2 peak calling. Table includes Chromosome, Start, End, Name, Score, Fold enrichment, and -log10(q value). Fragments overlapping (if overlap was > 50 bp) with the RNAse treated control dataset were subtracted, and peaks of length < 50 bp were removed.
- Sheet 2: SNS-seq clusters. The genome was divided into 10 kb windows, with a sliding window of 5 kb. SNS-seq ORIs were then overlapped with these windows. A cluster was then considered if a given 10 kb window contained 5 or more SNS-seq ORIs. Table includes Chromosome, Start, End, Number of SNS-seq ORIs found in that window, bp overlap between 10 kb window and SNS-seq peaks, window size, and percentage (%) of window bp within peaks.

Appendix 5. NanoForkSpeed origins

All sequencing data generated in this study can be made available upon request and will be deposited in the ENA database under the accession code PRJEB62206.

NFS results

Tables of ORIs include Chromosome, Start (start of the left-going fork), End (start of the right-going fork), Center (midpoint: (End-Start)/2), Strand.

- Sheet 1: t29h ORIs
- Sheet 2: t35h ORIs
- Sheet 3: t29h single molecule detected ORIs. Cases where more than one ORI was detected from the same sequencing read.
- Sheet 4: t35h single molecule detected ORIs. Cases where more than one ORI was detected from the same sequencing read.
- Sheet 5: t29h termination events. Calculated as the midpoint between two converging forks. Tables include Chromosome, Start (end of the right-going fork), End (end of the left-going fork), center (midpoint: (End-Start)/2), Strand. The end of the fork is considered as the end of the increasing slope of BrdU signal (*i.e.,* end of the pulse).
- Sheet 6: t35h termination events. Calculated as the midpoint between two converging forks. Tables include Chromosome, Start (end of the right-going fork), End (end of the left-going fork), center (midpoint: (End-Start)/2), Strand.
- Sheet 8: t29h forks. Table includes Chromosome, Start, End, Speed (calculated as the length of the labelled DNA region (end-start) divided by the duration of the BrdU labelling pulse (2 minutes)), Strand, Direction of the fork (L: left, R: right), and Type (Leading or Lagging).
- Sheet 9: t35h forks. Table includes Chromosome, Start, End, Speed (calculated as the length of the labelled DNA region (end-start) divided by the duration of the BrdU labelling pulse (2 minutes)), Strand, Direction of the fork (L: left, R: right), and Type (Leading or Lagging).
- Sheet 10: Single cell IOD vs fork speed. For those cases where more than one ORI was detected from the same sequencing read, the distance between the two ORIs was calculated, and only the speed of the upstream incoming replication fork was kept.
- Sheet 11: t29h ORIs clusters. The genome was divided into 10 kb windows, with a sliding window of 5 kb. NFS t29h ORIs were then overlapped with these windows. A cluster was then considered if a given 10 kb window contained 5 or more NFS t29h ORIs. Table includes

Chromosome, Start, End, Number of NFS t29h ORIs found in that window, bp overlap between 10 kb window and NFS t29h, window size, and percentage (%) of window bp within peaks.

References

[1] C. S. Simon, V. S. Stürmer, and J. Guizetti, “How Many Is Enough? - Challenges of Multinucleated Cell Division in Malaria Parasites,” *Front. Cell. Infect. Microbiol.*, vol. 11, May 2021, doi: 10.3389/FCIMB.2021.658616.

[2] S. Klaus *et al.*, “Asynchronous nuclear cycles in multinucleated Plasmodium falciparum facilitate rapid proliferation,” *Sci. Adv.*, vol. 8, no. 13, pp. 1–12, 2022, doi: 10.1126/sciadv.abj5362.

[3] F. I. G. Totañes, J. Gockel, S. E. Chapman, R. Bártfai, M. A. Boemo, and C. J. Merrick, “A genome-wide map of DNA replication at single-molecule resolution in the malaria parasite Plasmodium falciparum,” *Nucleic Acids Res.*, vol. 51, no. 6, p. 2709, Apr. 2023, doi: 10.1093/NAR/GKAD093.

[4] M. Ganter *et al.*, “Plasmodium falciparum CRK4 directs continuous rounds of DNA replication during schizogony,” *Nat. Microbiol. 2017 25*, vol. 2, no. 5, pp. 1–9, Feb. 2017, doi: 10.1038/nmicrobiol.2017.17.

[5] B. Theulot *et al.*, “Genome-wide mapping of individual replication fork velocities using nanopore sequencing,” *Nat. Commun. 2022 131*, vol. 13, no. 1, pp. 1–14, Jun. 2022, doi: 10.1038/s41467-022-31012-0.

[6] A. Fisher *et al.*, “A yeast chromosomal origin of DNA replication defined by multiple functional elements,” *Science (80-.).*, vol. 255, no. 5046, pp. 817–823, 1992, doi: 10.1126/SCIENCE.1536007.

[7] P. Prorok *et al.*, “Involvement of G-quadruplex regions in mammalian replication origin activity,” *Nat. Commun. 2019 101*, vol. 10, no. 1, pp. 1–16, Jul. 2019, doi: 10.1038/s41467-019-11104-0.

[8] R. Lombraña *et al.*, “Transcriptionally Driven DNA Replication Program of the Human Parasite Leishmania major,” *Cell Rep.*, vol. 16, no. 6, pp. 1774–1786, Aug. 2016, doi: 10.1016/j.celrep.2016.07.007.

[9] D. M. MacAlpine, H. K. Rodríguez, and S. P. Bell, “Coordination of replication and transcription along a Drosophila chromosome,” *Genes Dev.*, vol. 18, no. 24, p. 3094, Dec. 2004, doi: 10.1101/GAD.1246404.

[10] M. Hennion *et al.*, “FORK-seq: replication landscape of the Saccharomyces cerevisiae genome by nanopore sequencing,” *Genome Biol.*, vol. 21, p. 125, 2020, doi: 10.1186/s13059-020-02013-3.

[11] F. E. Cox, “History of the discovery of the malaria parasites and their vectors,” *Parasites and Vectors*, vol. 3, no. 1, pp. 1–9, Feb. 2010, doi: 10.1186/1756-3305-3-5/COMMENTS.

[12] R. Carter and K. N. Mendis, “Evolutionary and Historical Aspects of the Burden of Malaria,” *Clin. Microbiol. Rev.*, vol. 15, no. 4, p. 564, Oct. 2002, doi: 10.1128/CMR.15.4.564-594.2002.

[13] P. Schlagenhauf, “Malaria: from prehistory to present,” *Infect. Dis. Clin.*, vol. 18, no. 2, pp. 189–205, Jun. 2004, doi: 10.1016/J.IDC.2004.01.002.

[14] Geneva: World Health Organization (WHO), *World malaria report 2022*, Licence: I. 2022.

[15] D. A. Milner, “Malaria Pathogenesis,” *Cold Spring Harb. Perspect. Med.*, vol. 8, no. 1, p. a025569, Jan. 2018, doi: 10.1101/CSHPERSPECT.A025569.

[16] W. Stone, B. P. Gonçalves, T. Bousema, and C. Drakeley, “Assessing the infectious reservoir of falciparum malaria: past and future,” *Trends Parasitol.*, vol. 31, 2015, doi: 10.1016/j.pt.2015.04.004.

[17] K. S. Aly ASI, Vaughan AM, “Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host.,” *Annual Review of Microbiology*, 2009. https://sci-hub.se/https://doi.org/10.1146/annurev.micro.091208.073403 (accessed Feb. 15, 2022).

[18] J. Cox-Singh *et al.*, “Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening,” *Clin. Infect. Dis.*, vol. 46, no. 2, pp. 165–171, Jan. 2008, doi: 10.1086/524888/2/46-2-165-FIG002.GIF.

[19] A. Kiszewski, A. Mellinger, A. Spielman, P. Malaney, S. E. Sachs, and J. Sachs, “A GLOBAL INDEX REPRESENTING THE STABILITY OF MALARIA TRANSMISSION,” 2004.

[20] L. A. Baton and L. C. Ranford-Cartwright, “Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito,” *Trends Parasitol.*, vol. 21, no. 12, pp. 573–580, Dec. 2005, doi: 10.1016/J.PT.2005.09.012.

[21] T. Habtewold, Z. Groom, L. Duchateau, and G. K. Christophides, “Detection of viable plasmodium ookinetes in the midguts of anopheles coluzzi using PMA-qrtPCR,” *Parasites and Vectors*, vol. 8, no. 1, pp. 1–10, Dec. 2015, doi: 10.1186/S13071-015-1087-8/FIGURES/5.

[22] K. Matuschewski, “Getting infectious: formation and maturation of Plasmodium sporozoites in the Anopheles vector,” *Cell. Microbiol.*, vol. 8, no. 10, pp. 1547–1556, Oct. 2006, doi: 10.1111/J.1462-5822.2006.00778.X.

[23] J. P. Vanderberg and U. Frevert, “Intravital microscopy demonstrating antibody-mediated immobilisation of Plasmodium berghei sporozoites injected into skin by mosquitoes,” *Int. J. Parasitol.*, vol. 34, no. 9, pp. 991–996, Aug. 2004, doi: 10.1016/J.IJPARA.2004.05.005.

[24] L. M. Yamauchi Lucy M., A. Coppi, G. Snounou, and P. Sinnis, “Plasmodium sporozoites trickle out of the injection site,” *Cell. Microbiol.*, vol. 9, no. 5, p. 1215, May 2007, doi: 10.1111/J.1462-5822.2006.00861.X.

[25] M. Loubens, L. Vincensini, P. Fernandes, S. Briquet, C. Marinach, and O. Silvie, “Plasmodium sporozoites on the move: Switching from cell traversal to productive invasion of hepatocytes,” *Mol. Microbiol.*, vol. 115, no. 5, pp. 870–881, May 2021, doi: 10.1111/MMI.14645.

[26] A. M. Vaughan, A. S. I. Aly, and S. H. I. Kappe, “Malaria Parasite Pre-Erythrocytic Stage Infection: Gliding and Hiding,” *Cell Host Microbe*, vol. 4, no. 3, pp. 209–218, Sep. 2008, doi: 10.1016/J.CHOM.2008.08.010.

[27] P. C. Burda, R. Caldelari, and V. T. Heussler, “Manipulation of the Host Cell Membrane during Plasmodium Liver Stage Egress,” *MBio*, vol. 8, no. 2, Mar. 2017, doi: 10.1128/MBIO.00139-17.

[28] K. S. Griffith, L. S. Lewis, S. Mali, and M. E. Parise, “Treatment of Malaria in the United States: A Systematic Review,” *JAMA*, vol. 297, no. 20, pp. 2264–2277, May 2007, doi: 10.1001/JAMA.297.20.2264.

[29] J. F. Dubremetz, N. Garcia-Réguet, V. Conseil, and M. N. Fourmaux, “Invited review Apical organelles and host-cell invasion by Apicomplexa,” *Int. J. Parasitol.*, vol. 28, no. 7, pp. 1007–1013, Jul. 1998, doi: 10.1016/S0020-7519(98)00076-9.

[30] V. K. Goel, X. Li, H. Chen, S. C. Liu, A. H. Chishti, and S. S. Oh, “Band 3 is a host receptor binding merozoite surface protein 1 during the Plasmodium falciparum invasion of erythrocytes,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 9, p. 5164, Apr. 2003, doi: 10.1073/PNAS.0834959100.

[31] S. Singh, M. M. Alam, I. Pal-Bhowmick, J. A. Brzostowski, and C. E. Chitnis, “Distinct External Signals Trigger Sequential Release of Apical Organelles during Erythrocyte Invasion by Malaria Parasites,” *PLOS Pathog.*, vol. 6, no. 2, p. e1000746, Feb. 2010, doi: 10.1371/JOURNAL.PPAT.1000746.

[32] B. K. L. Sim, C. E. Chitnis, K. Wasniowska, T. J. Hadley, and L. H. Miller, “Receptor and Ligand Domains for Invasion of Erythrocytes by Plasmodium falciparum,” *Science (80-.).*, vol. 264, no. 5167, pp. 1941–1944, 1994, doi: 10.1126/SCIENCE.8009226.

[33] C. Crosnier *et al.*, “Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum,” *Nat. 2011 4807378*, vol. 480, no. 7378, pp. 534–537, Nov. 2011, doi: 10.1038/nature10606.

[34] M. Lamarque *et al.*, “The RON2-AMA1 Interaction is a Critical Step in Moving Junction-Dependent Invasion by Apicomplexan Parasites,” *PLoS Pathog.*, vol. 7, no. 2, p. 1001276, 2011, doi: 10.1371/JOURNAL.PPAT.1001276.

[35] M. L. Jones, E. L. Kitson, and J. C. Rayner, “Plasmodium falciparum erythrocyte invasion: A conserved myosin associated complex,” *Mol. Biochem. Parasitol.*, vol. 147, no. 1, pp. 74–84, May 2006, doi: 10.1016/J.MOLBIOPARA.2006.01.009.

[36] M. Koch and J. Baum, “The mechanics of malaria parasite invasion of the human erythrocyte – towards a reassessment of the host cell contribution,” *Cell. Microbiol.*, vol. 18, no. 3, pp. 319–329, Mar. 2016, doi: 10.1111/CMI.12557.

[37] J. Liu, E. S. Istvan, I. Y. Gluzman, J. Gross, and D. E. Goldberg, “Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 23, pp. 8840–8845, Jun. 2006, doi: 10.1073/PNAS.0601876103/SUPPL_FILE/01876FIG7.PDF.

[38] M. Marti, R. T. Good, M. Rug, E. Knuepfer, and A. F. Cowman, “Targeting malaria virulence and remodeling proteins to the host erythrocyte,” *Science (80-.).*, vol. 306, no. 5703, pp. 1930–1933, Dec. 2004, doi: 10.1126/SCIENCE.1102452/SUPPL_FILE/MARTI.SOM.PDF.

[39] M. Rug, S. W. Prescott, K. M. Fernandez, B. M. Cooke, and A. F. Cowman, “The role of KAHRP domains in knob formation and cytoadherence of P falciparum-infected human erythrocytes,” *Blood*, vol. 108, no. 1, pp. 370–378, Jul. 2006, doi: 10.1182/BLOOD-2005-11-4624.

[40] M. E. Wickham *et al.*, “Trafficking and assembly of the cytoadherence complex in Plasmodium falciparum-infected human erythrocytes,” *EMBO J.*, vol. 20, no. 20, p. 5636, Oct. 2001, doi: 10.1093/EMBOJ/20.20.5636.

[41] A. S. Paul *et al.*, “Co-option of Plasmodium falciparum PP1 for egress from host erythrocytes,” *Nat. Commun. 2020 111*, vol. 11, no. 1, pp. 1–13, Jul. 2020, doi: 10.1038/s41467-020-17306-1.

[42] S. Glushakova, V. Lizunov, P. S. Blank, K. Melikov, G. Humphrey, and J. Zimmerberg, “Cytoplasmic free Ca2+ is essential for multiple steps in malaria parasite egress from infected erythrocytes.,” *Malar. J.*, vol. 12, no. 1, pp. 1–12, Jan. 2013, doi: 10.1186/1475-2875-12-41/FIGURES/5.

[43] J. D. Dvorin *et al.*, “A plant-like kinase in plasmodium falciparum regulates parasite egress from erythrocytes,” *Science (80-.).*, vol. 328, no. 5980, pp. 910–912, May 2010, doi: 10.1126/SCIENCE.1188191/SUPPL_FILE/DVORIN_SOM.PDF.

[44] C. R. Collins *et al.*, “Malaria Parasite cGMP-dependent Protein Kinase Regulates Blood Stage Merozoite Secretory Organelle Discharge and Egress,” *PLoS Pathog.*, vol. 9, no. 5, p. 1003344, May 2013, doi: 10.1371/JOURNAL.PPAT.1003344.

[45] A. Ruecker *et al.*, “Proteolytic activation of the essential parasitophorous vacuole cysteine protease SERA6 accompanies malaria parasite egress from its host erythrocyte,” *J. Biol. Chem.*, vol. 287, no. 45, pp. 37949–37963, Nov. 2012, doi: 10.1074/JBC.M112.400820.

[46] J. A. Thomas *et al.*, “A protease cascade regulates release of the human malaria parasite Plasmodium falciparum from host red blood cells,” *Nat. Microbiol. 2018 34*, vol. 3, no. 4, pp. 447–455, Feb. 2018, doi: 10.1038/s41564-018-0111-0.

[47] V. L. Hale *et al.*, “Parasitophorous vacuole poration precedes its rupture and rapid host erythrocyte cytoskeleton collapse in Plasmodium falciparum egress,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 13, pp. 3439–3444, Mar. 2017, doi: 10.1073/PNAS.1619441114/-/DCSUPPLEMENTAL/PNAS.201619441SI.PDF.

[48] M. Abkarian, G. Massiera, L. Berry, M. Roques, and C. Braun-Breton, “A novel mechanism for egress of malarial parasites from red blood cells,” *Blood*, vol. 117, no. 15, pp. 4118–4124, Apr. 2011, doi: 10.1182/BLOOD-2010-08-299883.

[49] A. Sinha *et al.*, "A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium," *Nat. 2014 5077491*, vol. 507, no. 7491, pp. 253–257, Feb. 2014, doi: 10.1038/nature12970.

[50] B. F. C. Kafsack *et al.*, "A transcriptional switch underlies commitment to sexual development in malaria parasites," *Nat. 2014 5077491*, vol. 507, no. 7491, pp. 248–252, Feb. 2014, doi: 10.1038/nature12920.

[51] M. Filarsky *et al.*, "GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing," *Science (80-.).*, vol. 359, no. 6381, pp. 1259–1263, Mar. 2018, doi: 10.1126/SCIENCE.AAN6042/SUPPL_FILE/AAN6042_TABLES9.XLSX.

[52] A. Gomes *et al.*, "A transcriptional switch controls sex determination in Plasmodium falciparum," *Nature*, vol. (in press), no. January, 2022, doi: 10.1038/s41586-022-05509-z.

[53] N. J. Rogers, B. S. Hall, J. Obiero, G. A. T. Targett, and C. J. Sutherland, "A model for sequestration of the transmission stages of Plasmodium falciparum: adhesion of gametocyte-infected erythrocytes to human bone marrow cells.," *Infect. Immun.*, vol. 68, no. 6, pp. 3455–3462, Jun. 2000, doi: 10.1128/IAI.68.6.3455-3462.2000.

[54] A. Bartoloni and L. Zammarchi, "Clinical Aspects of Uncomplicated and Severe Malaria," *Mediterr. J. Hematol. Infect. Dis.*, vol. 4, no. 1, p. 201, 2012, doi: 10.4084/MJHID.2012.026.

[55] D. D. Laishram *et al.*, "The complexities of malaria disease manifestations with a focus on asymptomatic malaria," *Malar. J.*, vol. 11, no. 1, pp. 1–15, Jan. 2012, doi: 10.1186/1475-2875-11-29/TABLES/1.

[56] P. Daubersies *et al.*, "Rapid Turnover of Plasmodium falciparum Populations in Asymptomatic Individuals Living in a High Transmission Area," *Am. J. Trop. Med. Hyg.*, vol. 54, no. 1, pp. 18–26, Jan. 1996, doi: 10.4269/AJTMH.1996.54.18.

[57] C. Ruwende *et al.*, "Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. ," *Nature*, vol. 376, no. 6537, pp. 246–249, 1995.

[58] M. P. Grobusch and P. G. Kremsner, "Uncomplicated Malaria," *Curr. Top. Microbiol. Immunol.*, vol. 295, pp. 81–104, 2005, doi: 10.1007/3-540-29088-5_4.

[59] T. Hänscheid, M. P. Grobusch, J. Melo-Cristino, and B. G. Pinto, "Avoiding Misdiagnosis of Imported Malaria: Screening of Emergency Department Samples with Thrombocytopenia Detects Clinically Unsuspected Cases," *J. Travel Med.*, vol. 10, no. 3, pp. 155–159, Jun. 2003, doi: 10.2310/7060.2003.35735.

[60] P. Angchaisuksiri, "Coagulopathy in malaria," *Thromb. Res.*, vol. 133, no. 1, pp. 5–9, Jan. 2014, doi: 10.1016/J.THROMRES.2013.09.030.

[61] S. Gupta *et al.*, "Malaria and the Heart: JACC State-of-the-Art Review," *J. Am. Coll. Cardiol.*, vol. 77, no. 8, pp. 1110–1121, Mar. 2021, doi: 10.1016/J.JACC.2020.12.042/SUPPL_FILE/MMC1.DOCX.

[62] D. D. Brown, S. Solomon, D. Lerner, and M. Del Rio, "Malaria and acute kidney injury," *Pediatr. Nephrol.*, vol. 35, no. 4, pp. 603–608, Apr. 2020, doi: 10.1007/S00467-018-4191-0/FIGURES/1.

[63] A. R. Jensen, Y. Adams, and L. Hviid, "Cerebral Plasmodium falciparum malaria: The role of PfEMP1 in its pathogenesis and immunity, and PfEMP1-based vaccines to prevent it," *Immunol. Rev.*, vol. 293, no. 1, pp. 230–252, Jan. 2020, doi: 10.1111/IMR.12807.

[64] WHO, *Word Malaria Report 2021*. 2021.

[65] W. McGuire, A. V. S. Hill, C. E. M. Allsopp, B. M. Greenwood, and D. Kwjatkowski, "Variation in the TNF-α promoter region associated with susceptibility to cerebral malaria," *Nat. 1994 3716497*, vol. 371, no.

6497, pp. 508–511, 1994, doi: 10.1038/371508a0.

[66] B. A. Gyan *et al.*, "Allelic polymorphisms in the repeat and promoter regions of the interleukin-4 gene and malaria severity in Ghanaian children," *Clin. Exp. Immunol.*, vol. 138, no. 1, pp. 145–150, Aug. 2004, doi: 10.1111/J.1365-2249.2004.02590.X.

[67] M. Aidoo *et al.*, "Protective effects of the sickle cell gene against malaria morbidity and mortality," *Lancet*, vol. 359, no. 9314, pp. 1311–1312, Apr. 2002, doi: 10.1016/S0140-6736(02)08273-9.

[68] F. P. Mockenhaupt *et al.*, "α+-thalassemia protects African children from severe malaria," *Blood*, vol. 104, no. 7, pp. 2003–2006, Oct. 2004, doi: 10.1182/BLOOD-2003-11-4090.

[69] D. M. Langhi and J. O. Bordin, "Duffy blood group and malaria," *https://doi.org/10.1080/10245330500469841*, vol. 11, no. 5–6, pp. 389–398, Oct. 2013, doi: 10.1080/10245330500469841.

[70] D. Modiano *et al.*, "Haemoglobin C protects against clinical Plasmodium falciparum malaria.," *Nature*, vol. 414, no. 6861, pp. 305–308, Nov. 2001, doi: 10.1038/35104556.

[71] K. Chotivanich *et al.*, "Hemoglobin E: a balanced polymorphism protective against high parasitemias and thus severe P falciparum malaria," *Blood*, vol. 100, no. 4, pp. 1172–1176, Aug. 2002, doi: 10.1182/BLOOD.V100.4.1172.H81602001172_1172_1176.

[72] S. J. Allen *et al.*, "Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3," *Am. J. Trop. Med. Hyg.*, vol. 60, no. 6, pp. 1056–1060, 1999, doi: 10.4269/AJTMH.1999.60.1056.

[73] I. A. Cockburn *et al.*, "A human complement receptor 1 polymorphism that reduces Plasmodium falciparum rosetting confers protection against severe malaria," *Proc. Natl. Acad. Sci.*, vol. 101, no. 1, pp. 272–277, Jan. 2004, doi: 10.1073/PNAS.0305306101.

[74] K. Omi *et al.*, "CD36 Polymorphism Is Associated with Protection from Cerebral Malaria," *Am. J. Hum. Genet.*, vol. 72, no. 2, pp. 364–374, Feb. 2003, doi: 10.1086/346091.

[75] S. L. Pathirana *et al.*, "ABO-blood-group types and protection against severe, Plasmodium falciparum malaria," *http://dx.doi.org/10.1179/136485905X19946*, vol. 99, no. 2, pp. 119–124, Mar. 2013, doi: 10.1179/136485905X19946.

[76] I. Usynin, C. Klotz, and U. Frevert, "Malaria circumsporozoite protein inhibits the respiratory burst in Kupffer cells," *Cell. Microbiol.*, vol. 9, no. 11, pp. 2610–2628, Nov. 2007, doi: 10.1111/J.1462-5822.2007.00982.X.

[77] J. Tavares *et al.*, "Role of host cell traversal by the malaria sporozoite during liver infection," *J. Exp. Med.*, vol. 210, no. 5, pp. 905–915, May 2013, doi: 10.1084/JEM.20121130.

[78] A. Sturm *et al.*, "Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids," *Science*, vol. 313, no. 5791, pp. 1287–1290, Sep. 2006, doi: 10.1126/SCIENCE.1129720.

[79] R. E. Howes *et al.*, "Global Epidemiology of Plasmodium vivax," *Am. J. Trop. Med. Hyg.*, vol. 95, no. 6 Suppl, p. 15, 2016, doi: 10.4269/AJTMH.16-0141.

[80] X. Y. Yam and P. R. Preiser, "Host immune evasion strategies of malaria blood stage parasite," *Mol. Biosyst.*, vol. 13, no. 12, pp. 2498–2508, Nov. 2017, doi: 10.1039/C7MB00502D.

[81] C. Menéndez and C. Dobaño, "Malaria and immunity.," *Handb. Nutr. Immun.*, pp. 243–264, 2004.

[82] M. J. Ponsford *et al.*, "Sequestration and Microvascular Congestion Are Associated With Coma in Human

Cerebral Malaria," *J. Infect. Dis.*, vol. 205, no. 4, pp. 663–671, Feb. 2012, doi: 10.1093/INFDIS/JIR812.

[83] K. T. Andrews and M. Lanzer, "Maternal malaria: Plasmodium falciparum sequestration in the placenta," *Parasitol. Res. 2002 888*, vol. 88, no. 8, pp. 715–723, 2002, doi: 10.1007/S00436-002-0624-5.

[84] X. Y. Yam, M. Niang, K. G. Madnani, and P. R. Preiser, "Three Is a Crowd - New Insights into Rosetting in Plasmodium falciparum," *Trends Parasitol.*, vol. 33, no. 4, pp. 309–320, Apr. 2017, doi: 10.1016/J.PT.2016.12.012.

[85] "Global Malaria Programme." https://www.who.int/teams/global-malaria-programme/prevention/preventive-chemotherapies (accessed Jul. 19, 2023).

[86] E. A. Ashley and J. R. Poespoprodjo, "Treatment and prevention of malaria in children," *Lancet Child Adolesc. Heal.*, vol. 4, no. 10, pp. 775–789, 2020, doi: 10.1016/S2352-4642(20)30127-9.

[87] M. T. White *et al.*, "Immunogenicity of the RTS,S/AS01 malaria vaccine and implications for duration of vaccine efficacy: Secondary analysis of data from a phase 3 randomised controlled trial," *Lancet Infect. Dis.*, vol. 15, no. 12, pp. 1450–1458, Dec. 2015, doi: 10.1016/S1473-3099(15)00239-X.

[88] N. Arora, L. C. Anbalagan, and A. K. Pannu, "Towards Eradication of Malaria: Is the WHO's RTS,S/AS01 Vaccination Effective Enough?," 2021, doi: 10.2147/RMHP.S219294.

[89] R. Gosling and L. von Seidlein, "The Future of the RTS,S/AS01 Malaria Vaccine: An Alternative Development Plan," 2016, doi: 10.1371/journal.pmed.1001994.

[90] P. Venkatesan, "The future of malaria control in light of RTS,S," *The Lancet Microbe*, vol. 3, no. 4, p. e251, Apr. 2022, doi: 10.1016/s2666-5247(22)00070-2.

[91] M. S. Datoo *et al.*, "Efficacy and immunogenicity of R21/Matrix-M vaccine against clinical malaria after 2 years' follow-up in children in Burkina Faso: a phase 1/2b randomised controlled trial," *Lancet Infect. Dis.*, vol. 22, no. 12, pp. 1728–1736, Dec. 2022, doi: 10.1016/S1473-3099(22)00442-X.

[92] M. S. Datoo *et al.*, "Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial," *Lancet*, vol. 397, no. 10287, pp. 1809–1818, May 2021, doi: 10.1016/S0140-6736(21)00943-0.

[93] World Health Organization (WHO), "Guidelines for the Treatment of Malaria.," *Third Edition*, 2015. https://books.google.fr/books?hl=en&lr=&id=IVo0DgAAQBAJ&oi=fnd&pg=PP1&ots=9Umc-tQ7cQ&sig=_m9OVUjxxthVNQl1DyLi_Cjf1IQ&redir_esc=y#v=onepage&q&f=false (accessed Mar. 17, 2023).

[94] F. D. Krampa, Y. Aniweh, G. A. Awandare, and P. Kanyong, "Recent Progress in the Development of Diagnostic Tests for Malaria," *Diagnostics 2017, Vol. 7, Page 54*, vol. 7, no. 3, p. 54, Sep. 2017, doi: 10.3390/DIAGNOSTICS7030054.

[95] B. Blasco, Di. Leroy, and D. A. Fidock, "Antimalarial drug resistance: linking Plasmodium falciparum parasite biology to the clinic," *Nat. Med. 2017 238*, vol. 23, no. 8, pp. 917–928, Aug. 2017, doi: 10.1038/nm.4381.

[96] K. K. Dayananda, R. N. Achur, and D. C. Gowda, "Epidemiology, drug resistance, and pathophysiology of *Plasmodium vivax* malaria," *J. Vector Borne Dis.*, vol. 55, no. 1, p. 1, 2018, doi: 10.4103/0972-9062.234620.

[97] R. Perera, R. Wickremasinghe, G. Newby, A. Caldera, D. Fernando, and K. Mendis, "Malaria Control, Elimination, and Prevention as Components of Health Security: A Review," *Am. J. Trop. Med. Hyg.*, vol. 107, no. 4, pp. 747–753, Oct. 2022, doi: 10.4269/AJTMH.22-0038.

[98] B. M. Greenwood *et al.*, “Malaria: progress, perils, and prospects for eradication,” *J. Clin. Invest.*, vol. 118, no. 4, pp. 1266–1276, Apr. 2008, doi: 10.1172/JCI33996.

[99] K. J. Wicht, S. Mok, and D. A. Fidock, “Molecular Mechanisms of Drug Resistance in Plasmodium falciparum Malaria,” *https://doi.org/10.1146/annurev-micro-020518-115546*, vol. 74, pp. 431–454, Sep. 2020, doi: 10.1146/ANNUREV-MICRO-020518-115546.

[100] A. K. Pannu, “Malaria today: advances in management and control,” *Trop. Doct.*, vol. 49, no. 3, pp. 160–164, 2019, doi: 10.1177/0049475519846382.

[101] J. A. Watson, N. Nekkab, and M. White, “Tafenoquine for the prevention of Plasmodium vivax malaria relapse,” *The Lancet Microbe*, vol. 2, no. 5, pp. e175–e176, May 2021, doi: 10.1016/S2666-5247(21)00062-8.

[102] J. Achan *et al.*, “Quinine, an old anti-malarial drug in a modern world: Role in the treatment of malaria,” *Malar. J.*, vol. 10, no. 1, pp. 1–12, May 2011, doi: 10.1186/1475-2875-10-144/TABLES/2.

[103] S. N. Patel and K. C. Kain, “Atovaquone/proguanil for the prophylaxis and treatment of malaria,” *http://dx.doi.org/10.1586/14787210.3.6.849*, vol. 3, no. 6, pp. 849–861, Dec. 2014, doi: 10.1586/14787210.3.6.849.

[104] G. Pradel and M. Schlitzer, “Antibiotics in Malaria Therapy and their Effect on the Parasite Apicoplast,” *Curr. Mol. Med.*, vol. 10, no. 3, pp. 335–349, Apr. 2010, doi: 10.2174/156652410791065273.

[105] S. R. Meshnick, T. E. Taylor, and A. S. Kamchonwongpaisan, “Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy,” *Microbiol. Rev.*, vol. 60, no. 2, pp. 301–315, Jun. 1996, doi: 10.1128/MR.60.2.301-315.1996.

[106] L. Cui and X. Z. Su, “Discovery, mechanisms of action and combination therapy of artemisinin,” *http://dx.doi.org/10.1586/eri.09.68*, vol. 7, no. 8, pp. 999–1013, Oct. 2014, doi: 10.1586/ERI.09.68.

[107] N. Lang-Unnasch, M. E. Reith, J. Munholland, and J. R. Barta, “Plastids are widespread and ancient in parasites of the phylum Apicomplexa,” *Int. J. Parasitol.*, vol. 28, no. 11, pp. 1743–1754, Nov. 1998, doi: 10.1016/S0020-7519(98)00136-2.

[108] S. A. Ralph *et al.*, “Metabolic maps and functions of the Plasmodium falciparum apicoplast,” *Nat. Rev. Microbiol. 2004 23*, vol. 2, no. 3, pp. 203–216, Mar. 2004, doi: 10.1038/nrmicro843.

[109] M. W. White and E. S. Suvorova, “Apicomplexa Cell Cycles: Something Old, Borrowed, Lost, and New,” *Trends Parasitol.*, vol. 34, no. 9, pp. 759–771, Sep. 2018, doi: 10.1016/J.PT.2018.07.006.

[110] A. Flammersfeld, C. Lang, A. Flieger, and G. Pradel, “Phospholipases during membrane dynamics in malaria parasites,” *Int. J. Med. Microbiol.*, vol. 308, no. 1, pp. 129–141, Jan. 2018, doi: 10.1016/J.IJMM.2017.09.015.

[111] Y. W. Leong, B. Russell, B. Malleret, and L. Rénia, “Erythrocyte tropism of malarial parasites: The reticulocyte appeal,” *Front. Microbiol.*, vol. 13, p. 1022828, Oct. 2022, doi: 10.3389/FMICB.2022.1022828/BIBTEX.

[112] S. Antinori, L. Galimberti, L. Milazzo, and M. Corbellino, “Biology of Human Malaria Plasmodia Including Plasmodium Knowlesi,” *Mediterr. J. Hematol. Infect. Dis.*, vol. 4, no. 1, p. 201, 2012, doi: 10.4084/MJHID.2012.013.

[113] J. McDonald and C. J. Merrick, “DNA replication dynamics during erythrocytic schizogony in the malaria parasites Plasmodium falciparum and Plasmodium knowlesi,” *PLoS Pathog.*, vol. 18, no. 6, pp. 1–25, 2022, doi: 10.1371/journal.ppat.1010595.

[114] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, “Image analysis and machine learning for detecting malaria,” *Transl. Res.*, vol. 194, pp. 36–55, Apr. 2018, doi: 10.1016/J.TRSL.2017.12.004.

[115] C. Wongsrichanalai, M. J. Barcus, S. Muth, A. Sutamihardja, and W. H. Wernsdorfer, “A Review of Malaria Diagnostic Tools: Microscopy and Rapid Diagnostic Test (RDT),” 2007.

[116] X. Z. Su, K. D. Lane, L. Xia, J. M. Sá, and T. E. Wellems, “Plasmodium genomics and genetics: New insights into malaria pathogenesis, drug resistance, epidemiology, and evolution,” *Clin. Microbiol. Rev.*, vol. 32, no. 4, Oct. 2019, doi: 10.1128/CMR.00019-19/ASSET/2DF948B8-65A9-43ED-A513-BF0950E55D6B/ASSETS/GRAPHIC/CMR.00019-19-F0005.JPEG.

[117] M. J. Gardner *et al.*, “Genome sequence of the human malaria parasite Plasmodium falciparum,” *Nature*, vol. 419, no. 6906, pp. 498–511, 2002, doi: 10.1038/nature01097.

[118] D. J. Conway *et al.*, “Origin of Plasmodium falciparum malaria is traced by mitochondrial DNA,” *Mol. Biochem. Parasitol.*, vol. 111, no. 1, pp. 163–171, Nov. 2000, doi: 10.1016/S0166-6851(00)00313-3.

[119] R. J. M. (Iain. Wilson *et al.*, “Complete Gene Map of the Plastid-like DNA of the Malaria Parasite Plasmodium falciparum,” *J. Mol. Biol.*, vol. 261, no. 2, pp. 155–172, Aug. 1996, doi: 10.1006/JMBI.1996.0449.

[120] J. Carlton, “The Plasmodium vivax genome sequencing project,” *Trends Parasitol.*, vol. 19, no. 5, pp. 227–231, May 2003, doi: 10.1016/S1471-4922(03)00066-7.

[121] A. Pain *et al.*, “The genome of the simian and human malaria parasite Plasmodium knowlesi,” *Nature*, vol. 455, no. 7214, pp. 799–803, 2008, doi: 10.1038/nature07306.

[122] G. G. Rutledge *et al.*, “Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution,” *Nature*, vol. 542, no. 7639, pp. 101–104, 2017, doi: 10.1038/nature21038.

[123] W. L. Hamilton *et al.*, “Extreme mutation bias and high AT content in Plasmodium falciparum,” *Nucleic Acids Res.*, vol. 45, no. 4, pp. 1889–1901, Feb. 2017, doi: 10.1093/NAR/GKW1259.

[124] A. Goffeau *et al.*, “Life with 6000 Genes,” *Science (80-.).*, vol. 274, no. 5287, pp. 546–567, Oct. 1996, doi: 10.1126/SCIENCE.274.5287.546.

[125] V. Wood *et al.*, “The genome sequence of Schizosaccharomyces pombe,” *Nat. 2002 4156874*, vol. 415, no. 6874, pp. 871–880, Feb. 2002, doi: 10.1038/nature724.

[126] M. M. Zilversmit, S. K. Volkman, M. A. Depristo, D. F. Wirth, P. Awadalla, and D. L. Hartl, “Low-Complexity Regions in Plasmodium falciparum: Missing Links in the Evolution of an Extreme Genome,” *Mol. Biol. Evol.*, vol. 27, no. 9, pp. 2198–2209, Sep. 2010, doi: 10.1093/MOLBEV/MSQ108.

[127] C. Aurrecoechea *et al.*, “PlasmoDB: a functional genomic database for malaria parasites,” *Nucleic Acids Res.*, vol. 37, no. Database issue, p. D539, 2009, doi: 10.1093/NAR/GKN814.

[128] J. H. Gunderson *et al.*, “Structurally distinct, stage-specific ribosomes occur in plasmodium,” *Science (80-.).*, vol. 238, no. 4829, pp. 933–937, 1987, doi: 10.1126/SCIENCE.3672135.

[129] T. D. Otto *et al.*, “New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq,” *Mol. Microbiol.*, vol. 76, no. 1, pp. 12–24, Apr. 2010, doi: 10.1111/J.1365-2958.2009.07026.X.

[130] L. Chappell *et al.*, “Refining the transcriptome of the human malaria parasite Plasmodium falciparum using amplification-free RNA-seq,” *BMC Genomics*, vol. 21, no. 1, pp. 1–19, Jun. 2020, doi: 10.1186/S12864-020-06787-5/FIGURES/6.

[131] Z. Bozdech, M. Llinás, B. L. Pulliam, E. D. Wong, J. Zhu, and J. L. DeRisi, “The Transcriptome of the

Intraerythrocytic Developmental Cycle of Plasmodium falciparum," *PLoS Biol.*, vol. 1, no. 1, 2003, doi: 10.1371/JOURNAL.PBIO.0000005.

[132] Z. Zhang and F. S. Dietrich, "Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE," *Nucleic Acids Res.*, vol. 33, no. 9, pp. 2838–2851, May 2005, doi: 10.1093/NAR/GKI583.

[133] S. H. Adjalley, C. D. Chabbert, B. Klaus, V. Pelechano, and L. M. Steinmetz, "Landscape and dynamics of transcription initiation in the malaria parasite Plasmodium falciparum," *Cell Rep.*, vol. 14, no. 10, p. 2463, Mar. 2016, doi: 10.1016/J.CELREP.2016.02.025.

[134] P. R. Kensche *et al.*, "The nucleosome landscape of Plasmodium falciparum reveals chromatin architecture and dynamics of regulatory sequences," *Nucleic Acids Res.*, vol. 44, no. 5, pp. 2110–2124, Nov. 2016, doi: 10.1093/NAR/GKV1214.

[135] L. Cui and J. Miao, "Chromatin-Mediated epigenetic regulation in the malaria parasite Plasmodium falciparum," *Eukaryot. Cell*, vol. 9, no. 8, pp. 1138–1149, 2010, doi: 10.1128/EC.00036-10.

[136] A. Weiner *et al.*, "3D nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite Plasmodium falciparum," *Cell. Microbiol.*, vol. 13, no. 7, pp. 967–977, Jul. 2011, doi: 10.1111/J.1462-5822.2011.01592.X.

[137] R. Hernandez-Rivas, K. Pérez-Toledo, A. M. H. Solorio, D. M. Delgadillo, and M. Vargas, "Telomeric heterochromatin in plasmodium falciparum," *J. Biomed. Biotechnol.*, vol. 2010, 2010, doi: 10.1155/2010/290501.

[138] A. Scherf, L. M. Figueiredo, and L. H. Freitas-Junior, "Plasmodium telomeres: A pathogen's perspective," *Current Opinion in Microbiology*, vol. 4, no. 4. Elsevier Ltd, pp. 409–414, Aug. 01, 2001, doi: 10.1016/S1369-5274(00)00227-7.

[139] N. Ponts *et al.*, "Nucleosome landscape and control of transcription in the human malaria parasite," *Genome Res.*, vol. 20, no. 2, pp. 228–238, Feb. 2010, doi: 10.1101/GR.101063.109.

[140] W. A. M. Hoeijmakers *et al.*, "Plasmodium falciparum centromeres display a unique epigenetic makeup and cluster prior to and during schizogony," 2012, doi: 10.1111/j.1462-5822.2012.01803.x.

[141] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nat. 1997 3896648*, vol. 389, no. 6648, pp. 251–260, 1997, doi: 10.1038/38444.

[142] J. Miao, Q. Fan, L. Cui, J. Li, J. Li, and L. Cui, "The malaria parasite Plasmodium falciparum histones: Organization, expression, and acetylation," *Gene*, vol. 369, no. 1–2, pp. 53–65, Mar. 2006, doi: 10.1016/J.GENE.2005.10.022.

[143] J. Connacher, H. von Grüning, and L. Birkholtz, "Histone Modification Landscapes as a Roadmap for Malaria Parasite Development," *Front. Cell Dev. Biol.*, vol. 10, p. 617, Apr. 2022, doi: 10.3389/FCELL.2022.848797.

[144] A. M. Salcedo-Amaya *et al.*, "Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of Plasmodium falciparum," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 24, pp. 9655–9660, Jun. 2009, doi: 10.1073/pnas.0902515106.

[145] J. V. Torres-Perez, J. Irfan, M. R. Febrianto, S. Di Giovanni, and I. Nagy, "Histone post-translational modifications as potential therapeutic targets for pain management," *Trends Pharmacol. Sci.*, vol. 42, no. 11, pp. 897–911, Nov. 2021, doi: 10.1016/J.TIPS.2021.08.002/ATTACHMENT/EAFF29D8-2E30-41A6-897B-7541B0FF9C7F/MMC1.MP4.

[146] F. Ay, E. M. Bunnik, N. Varoquaux, J. P. Vert, W. S. Noble, and K. G. Le Roch, "Multiple dimensions of

epigenetic gene regulation in the malaria parasite Plasmodium falciparum," *BioEssays*, vol. 37, no. 2, pp. 182–194, Feb. 2015, doi: 10.1002/BIES.201400145.

[147] R. Bártfai *et al.*, "H2A.Z demarcates intergenic regions of the Plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3," *PLoS Pathog.*, vol. 6, no. 12, 2010, doi: 10.1371/journal.ppat.1001223.

[148] Z. Wang *et al.*, "Combinatorial patterns of histone acetylations and methylations in the human genome," *Nat. Genet. 2008 407*, vol. 40, no. 7, pp. 897–903, Jun. 2008, doi: 10.1038/ng.154.

[149] N. D. Heintzman *et al.*, "Histone modifications at human enhancers reflect global cell-type-specific gene expression," *Nat. 2009 4597243*, vol. 459, no. 7243, pp. 108–112, Mar. 2009, doi: 10.1038/nature07829.

[150] J. Tang, S. A. Chisholm, L. M. Yeoh, P. R. Gilson, and A. T. Papenfuss, "Histone modifications associated with gene expression and genome accessibility are dynamically enriched at Plasmodium falciparum regulatory sequences," *Epigenetics Chromatin*, pp. 1–25, 2020, doi: 10.1186/s13072-020-00365-5.

[151] S. A. Fraschka *et al.*, "Comparative Heterochromatin Profiling Reveals Conserved and Unique Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites," *Cell Host Microbe*, vol. 23, no. 3, pp. 407-420.e8, Mar. 2018, doi: 10.1016/j.chom.2018.01.008.

[152] J. J. Lopez-Rubio, L. Mancio-Silva, and A. Scherf, "Genome-wide Analysis of Heterochromatin Associates Clonally Variant Gene Regulation with Perinuclear Repressive Centers in Malaria Parasites," *Cell Host Microbe*, vol. 5, no. 2, pp. 179–190, Feb. 2009, doi: 10.1016/j.chom.2008.12.012.

[153] K. W. Deitsch and R. Dzikowski, "Variant Gene Expression and Antigenic Variation by Malaria Parasites," *https://doi.org/10.1146/annurev-micro-090816-093841*, vol. 71, pp. 625–641, Sep. 2017, doi: 10.1146/ANNUREV-MICRO-090816-093841.

[154] A. Scherf, J. J. Lopez-Rubio, and L. Riviere, "Antigenic Variation in Plasmodium falciparum," *https://doi.org/10.1146/annurev.micro.61.080706.093134*, vol. 62, pp. 445–470, Sep. 2008, doi: 10.1146/ANNUREV.MICRO.61.080706.093134.

[155] J. Spiegel, S. Adhikari, and S. Balasubramanian, "The Structure and Function of DNA G-Quadruplexes," *Trends Chem.*, vol. 2, no. 2, pp. 123–136, Feb. 2020, doi: 10.1016/J.TRECHM.2019.07.002.

[156] D. E. Gilbert and J. Feigon, "Multistranded DNA structures," *Curr. Opin. Struct. Biol.*, vol. 9, no. 3, pp. 305–314, Jun. 1999, doi: 10.1016/S0959-440X(99)80041-4.

[157] J. A. Capra, K. Paeschke, M. Singh, and V. A. Zakian, "G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces cerevisiae," *PLOS Comput. Biol.*, vol. 6, no. 7, p. e1000861, 2010, doi: 10.1371/JOURNAL.PCBI.1000861.

[158] Z. Y. Sun, X. N. Wang, S. Q. Cheng, X. X. Su, and T. M. Ou, "Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to Interfering with Nucleic Acid–Protein Interaction," *Mol. 2019, Vol. 24, Page 396*, vol. 24, no. 3, p. 396, Jan. 2019, doi: 10.3390/MOLECULES24030396.

[159] J. Lopes *et al.*, "G-quadruplex-induced instability during leading-strand replication," *EMBO J.*, vol. 30, no. 19, pp. 4033–4046, Oct. 2011, doi: 10.1038/EMBOJ.2011.316.

[160] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, and L. H. Hurley, "Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 18, pp. 11593–11598, Sep. 2002, doi: 10.1073/PNAS.182256799/SUPPL_FILE/2567SUPPTEXT.HTML.

[161] A. R. Langley, S. Gräf, J. C. Smith, and T. Krude, "Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq)," *Nucleic Acids Res.*, vol. 44, no. 21,

pp. 10230–10247, Dec. 2016, doi: 10.1093/nar/gkw760.

[162] D. Rhodes and H. J. Lipps, "G-quadruplexes and their regulatory roles in biology," *Nucleic Acids Res.*, vol. 43, no. 18, pp. 8627–8637, Oct. 2015, doi: 10.1093/NAR/GKV862.

[163] C. Cayrou *et al.*, "Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features," *Genome Res.*, vol. 21, no. 9, p. 1438, Sep. 2011, doi: 10.1101/GR.121830.111.

[164] C. Cayrou *et al.*, "New insights into replication origin characteristics in metazoans," *Cell Cycle*, vol. 11, no. 4, p. 658, Feb. 2012, doi: 10.4161/CC.11.4.19097.

[165] C. Cayrou *et al.*, "The chromatin environment shapes DNA replication origin organization and defines origin classes," *Genome Res.*, vol. 25, no. 12, pp. 1873–1885, Dec. 2015, doi: 10.1101/GR.192799.115.

[166] H. L. Gage and C. J. Merrick, "Conserved associations between G-quadruplex-forming DNA motifs and virulence gene families in malaria parasites," *BMC Genomics*, vol. 21, no. 1, pp. 1–18, Mar. 2020, doi: 10.1186/S12864-020-6625-X/FIGURES/9.

[167] E. Gazanion *et al.*, "Genome wide distribution of G-quadruplexes and their impact on gene expression in malaria parasites," *PLOS Genet.*, vol. 16, no. 7, p. e1008917, Jul. 2020, doi: 10.1371/JOURNAL.PGEN.1008917.

[168] O. Kikin, L. D'Antonio, and P. S. Bagga, "QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences," *Nucleic Acids Res.*, vol. 34, no. suppl_2, pp. W676–W682, Jul. 2006, doi: 10.1093/NAR/GKL253.

[169] V. Brázda *et al.*, "G4Hunter web application: a web server for G-quadruplex prediction," *Bioinformatics*, vol. 35, no. 18, pp. 3493–3495, Sep. 2019, doi: 10.1093/BIOINFORMATICS/BTZ087.

[170] A. De Cian *et al.*, "Plasmodium Telomeric Sequences: Structure, Stability and Quadruplex Targeting by Small Compounds," *ChemBioChem*, vol. 9, no. 16, pp. 2730–2739, Nov. 2008, doi: 10.1002/CBIC.200800330.

[171] E. L. Cunningham and J. M. Berger, "Unraveling the early steps of prokaryotic replication," *Curr. Opin. Struct. Biol.*, vol. 15, no. 1, pp. 68–76, Feb. 2005, doi: 10.1016/J.SBI.2005.01.003.

[172] N. P. Robinson and S. D. Bell, "Origins of DNA replication in the three domains of life," *FEBS J.*, vol. 272, no. 15, pp. 3757–3766, Aug. 2005, doi: 10.1111/J.1742-4658.2005.04768.X.

[173] A. C. Leonard and M. Méchali, "DNA Replication Origins," *Cold Spring Harb. Perspect. Biol.*, vol. 5, no. 10, p. a010116, Oct. 2013, doi: 10.1101/CSHPERSPECT.A010116.

[174] Y. Hu and B. Stillman, "Origins of DNA replication in eukaryotes," *Mol. Cell*, vol. 83, no. 3, pp. 352–372, Feb. 2023, doi: 10.1016/J.MOLCEL.2022.12.024.

[175] O. Ganier, P. Prorok, I. Akerman, and M. Méchali, "Metazoan DNA replication origins," *Curr. Opin. Cell Biol.*, vol. 58, pp. 134–141, Jun. 2019, doi: 10.1016/J.CEB.2019.03.003.

[176] J. J. Wyrick *et al.*, "Genome-wide distribution of ORC and MCM proteins in S. cerevisae: High-resolution mapping of replication origins," *Science (80-.).*, vol. 294, no. 5550, pp. 2357–2360, Dec. 2001, doi: 10.1126/SCIENCE.1066101/SUPPL_FILE/WEBTABLE5.XLS.

[177] C. A. Nieduszynski, Y. Knox, and A. D. Donaldson, "Genome-wide identification of replication origins in yeast by comparative genomics," *Genes Dev.*, vol. 20, no. 14, pp. 1874–1879, Jul. 2006, doi: 10.1101/GAD.385306.

[178] L. D. Mesner, V. Valsakumar, M. Cieslik, M. Pickin, J. L. Hamlin, and S. Bekiranov, "Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins," *Genome Res.*, vol. 23, no. 11, pp. 1774–1788, Nov. 2013, doi: 10.1101/GR.155218.113.

[179] E. Besnard *et al.*, "Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs," *Nat. Struct. Mol. Biol.*, vol. 19, no. 8, pp. 837–844, 2012, doi: 10.1038/nsmb.2339.

[180] N. Petryk *et al.*, "Replication landscape of the human genome," *Nat. Commun. 2016 71*, vol. 7, no. 1, pp. 1–13, Jan. 2016, doi: 10.1038/ncomms10208.

[181] F. Comoglio, T. Schlumpf, V. Schmid, R. Rohs, C. Beisel, and R. Paro, "High-Resolution Profiling of Drosophila Replication Start Sites Reveals a DNA Shape and Chromatin Signature of Metazoan Origins," *Cell Rep.*, vol. 11, no. 5, pp. 821–834, May 2015, doi: 10.1016/J.CELREP.2015.03.070.

[182] A. Costa and J. F. X. Diffley, "The Initiation of Eukaryotic DNA Replication," *https://doi.org/10.1146/annurev-biochem-072321-110228*, vol. 91, pp. 107–131, Jun. 2022, doi: 10.1146/ANNUREV-BIOCHEM-072321-110228.

[183] R. D. Klemm and S. P. Bell, "ATP bound to the origin recognition complex is important for preRC formation," *Proc. Natl. Acad. Sci.*, vol. 98, no. 15, pp. 8361–8367, Jul. 2001, doi: 10.1073/PNAS.131006898.

[184] A. F. Neuwald, L. Aravind, J. L. Spouge, and E. V. Koonin, "AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes," *Genome Res.*, vol. 9, no. 1, pp. 27–43, Jan. 1999, doi: 10.1101/GR.9.1.27.

[185] M. Makise, H. Takenaka, W. Kuwae, N. Takahashi, T. Tsuchiya, and T. Mizushima, "Kinetics of ATP binding to the origin recognition complex of Saccharomyces cerevisiae," *J. Biol. Chem.*, vol. 278, no. 47, pp. 46440–46445, Nov. 2003, doi: 10.1074/JBC.M307392200.

[186] J. C. W. Randell, J. L. Bowers, H. K. Rodríguez, and S. P. Bell, "Sequential ATP Hydrolysis by Cdc6 and ORC Directs Loading of the Mcm2-7 Helicase," *Mol. Cell*, vol. 21, no. 1, pp. 29–39, Jan. 2006, doi: 10.1016/J.MOLCEL.2005.11.023.

[187] J. L. Bowers, J. C. W. Randell, S. Chen, and S. P. Bell, "ATP Hydrolysis by ORC Catalyzes Reiterative Mcm2-7 Assembly at a Defined Origin of Replication," *Mol. Cell*, vol. 16, no. 6, pp. 967–978, Dec. 2004, doi: 10.1016/J.MOLCEL.2004.11.038.

[188] C. Speck, Z. Chen, H. Li, and B. Stillman, "ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA," *Nat. Struct. Mol. Biol. 2005 1211*, vol. 12, no. 11, pp. 965–971, Oct. 2005, doi: 10.1038/nsmb1002.

[189] S. Chen, M. A. De Vries, and S. P. Bell, "Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2–7 loading," *Genes Dev.*, vol. 21, no. 22, p. 2897, Nov. 2007, doi: 10.1101/GAD.1596807.

[190] B. Stillman, "The remarkable gymnastics of ORC.," *Elife*, vol. 11, Feb. 2022, doi: 10.7554/ELIFE.76475.

[191] S. Gupta, L. J. Friedman, J. Gelles, and S. P. Bell, "A Helicase-tethered ORC Flip Enables Bidirectional Helicase Loading," *Elife*, vol. 10, Dec. 2021, doi: 10.7554/ELIFE.74282.

[192] Z. Yuan and H. Li, "Molecular mechanisms of eukaryotic origin initiation, replication fork progression, and chromatin maintenance," *Biochem. J.*, vol. 477, no. 18, p. 3499, Sep. 2020, doi: 10.1042/BCJ20200065.

[193] R. A. Sclafani and T. M. Holzen, "Cell Cycle Regulation of DNA Replication," *Annu. Rev. Genet.*, vol. 41, no. 1, pp. 237–280, Dec. 2007, doi: 10.1146/annurev.genet.41.110306.130308.

[194] T. J. McGarry and M. W. Kirschner, “Geminin, an Inhibitor of DNA Replication, Is Degraded during Mitosis,” *Cell*, vol. 93, no. 6, pp. 1043–1053, Jun. 1998, doi: 10.1016/S0092-8674(00)81209-X.

[195] A. Ballabeni, R. Zamponi, J. K. Moore, K. Helin, and M. W. Kirschner, “Geminin deploys multiple mechanisms to regulate Cdt1 before cell division thus ensuring the proper execution of DNA replication,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 30, pp. E2848–E2853, Jul. 2013, doi: 10.1073/PNAS.1310677110/SUPPL_FILE/PNAS.201310677SI.PDF.

[196] J. G. Cook, D. A. D. Chasse, and J. R. Nevins, “The Regulated Association of Cdt1 with Minichromosome Maintenance Proteins and Cdc6 in Mammalian Cells,” *J. Biol. Chem.*, vol. 279, no. 10, pp. 9625–9633, Mar. 2004, doi: 10.1074/JBC.M311933200.

[197] G. Guilbaud, P. Murat, H. S. Wilkes, L. K. Lerner, J. E. Sale, and T. Krude, “Determination of human DNA replication origin position and efficiency reveals principles of initiation zone organisation,” *Nucleic Acids Res.*, vol. 50, no. 13, pp. 7436–7450, Jul. 2022, doi: 10.1093/NAR/GKAC555.

[198] M. Méchali, “Eukaryotic DNA replication origins: many choices for appropriate answers,” *Nat. Rev. Mol. Cell Biol. 2010 1110*, vol. 11, no. 10, pp. 728–738, Sep. 2010, doi: 10.1038/nrm2976.

[199] C. Heichinger, C. J. Penkett, J. Bähler, and P. Nurse, “Genome-wide characterization of fission yeast DNA replication origins,” *EMBO J.*, vol. 25, no. 21, pp. 5171–5179, Nov. 2006, doi: 10.1038/SJ.EMBOJ.7601390.

[200] R. Lebofsky, R. Heilig, M. Sonnleitner, J. Weissenbach, and A. Bensimon, “DNA replication origin interference increases the spacing between initiation events in human cells,” *Mol. Biol. Cell*, vol. 17, no. 12, pp. 5337–5345, Dec. 2006, doi: 10.1091/MBC.E06-04-0298/ASSET/IMAGES/LARGE/ZMK0120678710005.JPEG.

[201] I. Ilves, T. Petojevic, J. J. Pesavento, and M. R. Botchan, “Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins,” *Mol. Cell*, vol. 37, no. 2, pp. 247–258, Jan. 2010, doi: 10.1016/J.MOLCEL.2009.12.030.

[202] B. Stillman, “Origin recognition and the chromosome cycle,” *FEBS Lett.*, vol. 579, no. 4, pp. 877–884, Feb. 2005, doi: 10.1016/J.FEBSLET.2004.12.011.

[203] S. P. Bell and A. Dutta, “DNA Replication in Eukaryotic Cells,” *https://doi.org/10.1146/annurev.biochem.71.110601.135425*, vol. 71, pp. 333–374, Nov. 2003, doi: 10.1146/ANNUREV.BIOCHEM.71.110601.135425.

[204] L. Zou and B. Stillman, “Assembly of a Complex Containing Cdc45p, Replication Protein A, and Mcm2p at Replication Origins Controlled by S-Phase Cyclin-Dependent Kinases and Cdc7p-Dbf4p Kinase,” *Mol. Cell. Biol.*, vol. 20, no. 9, pp. 3086–3096, May 2000, doi: 10.1128/MCB.20.9.3086-3096.2000/ASSET/69040ED6-42BB-49EC-9C4F-7109F11FCEE3/ASSETS/GRAPHIC/MB0901620008.JPEG.

[205] H. Yabuuchi, Y. Yamada, T. Uchida, T. Sunathvanichkul, T. Nakagawa, and H. Masukata, “Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins,” *EMBO J.*, vol. 25, no. 19, pp. 4663–4674, Oct. 2006, doi: 10.1038/SJ.EMBOJ.7601347.

[206] M. E. Douglas and J. F. X. Diffley, “Recruitment of Mcm10 to sites of replication initiation requires direct binding to the minichromosome maintenance (MCM) complex,” *J. Biol. Chem.*, vol. 291, no. 11, pp. 5879–5888, Mar. 2016, doi: 10.1074/jbc.M115.707802.

[207] L. D. Langston and M. E. O’Donnell, “An explanation for origin unwinding in eukaryotes,” *Elife*, vol. 8, Jul. 2019, doi: 10.7554/ELIFE.46515.

[208] H. Li and M. E. O’Donnell, “The eukaryotic CMG helicase at the replication fork: Emerging architecture reveals an unexpected mechanism,” *Bioessays*, vol. 40, no. 3, Mar. 2018, doi: 10.1002/BIES.201700208.

[209] M. O'Donnell, L. Langston, and B. Stillman, "Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya," *Cold Spring Harb. Perspect. Biol.*, vol. 5, no. 7, Jul. 2013, doi: 10.1101/CSHPERSPECT.A010108.

[210] Y. Shiomi *et al.*, "ATP-dependent structural change of the eukaryotic clamp-loader protein, replication factor C," *Proc. Natl. Acad. Sci.*, vol. 97, no. 26, pp. 14127–14132, Dec. 2000, doi: 10.1073/PNAS.97.26.14127.

[211] N. Yao *et al.*, "Clamp loading, unloading and intrinsic stability of the PCNA, β and gp45 sliding clamps of human, E. coli and T4 replicases," *Genes to Cells*, vol. 1, no. 1, pp. 101–113, Jan. 1996, doi: 10.1046/J.1365-2443.1996.07007.X.

[212] V. P. Bermudez, A. Farina, V. Raghavan, I. Tappin, and J. Hurwitz, "Studies on human DNA polymerase ε and GINS Complex and their role in DNA replication," *J. Biol. Chem.*, vol. 286, no. 33, pp. 28963–28977, Aug. 2011, doi: 10.1074/jbc.M111.256289.

[213] Z. F. Pursell, I. Isoz, E. B. Lundström, E. Johansson, and T. A. Kunkel, "Yeast DNA polymerase ε participates in leading-strand DNA replication," *Science (80-.).*, vol. 317, no. 5834, pp. 127–130, Jul. 2007, doi: 10.1126/SCIENCE.1144067/SUPPL_FILE/PURSELL_SOM.PDF.

[214] J. L. Stodola and P. M. Burgers, "Resolving individual steps of Okazaki fragment maturation at msec time-scale," *Nat. Struct. Mol. Biol.*, vol. 23, no. 5, p. 402, May 2016, doi: 10.1038/NSMB.3207.

[215] X. Li, C. M. Stith, P. M. Burgers, and W. D. Heyer, "PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase δ," *Mol. Cell*, vol. 36, no. 4, p. 704, Nov. 2009, doi: 10.1016/J.MOLCEL.2009.09.036.

[216] C. M. Stith, J. Sterling, M. A. Resnick, D. A. Gordenin, and P. M. Burgers, "Flexibility of Eukaryotic Okazaki Fragment Maturation through Regulated Strand Displacement Synthesis," *J. Biol. Chem.*, vol. 283, no. 49, p. 34129, Dec. 2008, doi: 10.1074/JBC.M806668200.

[217] P. Garg, C. M. Stith, N. Sabouri, E. Johansson, and P. M. Burgers, "Idling by DNA polymerase δ maintains a ligatable nick during lagging-strand DNA replication," *Genes Dev.*, vol. 18, no. 22, pp. 2764–2773, Nov. 2004, doi: 10.1101/GAD.1252304.

[218] J. Qiu, Y. Qian, P. Frank, U. Wintersberger, and B. Shen, "Saccharomyces cerevisiae RNase H(35) Functions in RNA Primer Removal during Lagging-Strand DNA Synthesis, Most Efficiently in Cooperation with Rad27 Nuclease," *https://doi.org/10.1128/MCB.19.12.8361*, vol. 19, no. 12, pp. 8361–8371, Dec. 2023, doi: 10.1128/MCB.19.12.8361.

[219] D. P. Clark, N. J. Pazdernik, and M. R. McGehee, "Cell Division and DNA Replication," *Mol. Biol.*, pp. 296–331, Jan. 2019, doi: 10.1016/B978-0-12-813288-3.00010-0.

[220] L. Postow, N. J. Crisona, B. J. Peter, C. D. Hardy, and N. R. Cozzarelli, "Topological challenges to DNA replication: Conformations at the fork," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 15, pp. 8219–8226, Jul. 2001, doi: 10.1073/PNAS.111006998/ASSET/4D3B3FFC-31B4-460A-8A55-4788E1B0FCEF/ASSETS/GRAPHIC/PQ1110069005.JPEG.

[221] R. Bermejo *et al.*, "Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation," *Genes Dev.*, vol. 21, no. 15, pp. 1921–1936, Aug. 2007, doi: 10.1101/GAD.432107.

[222] J. Baxter, "'Breaking Up Is Hard to Do': The Formation and Resolution of Sister Chromatid Intertwines," *J. Mol. Biol.*, vol. 427, no. 3, pp. 590–607, Feb. 2015, doi: 10.1016/J.JMB.2014.08.022.

[223] C. Conti, B. Saccà, J. Herrick, C. Lalou, Y. Pommier, and A. Bensimon, "Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells," *Mol. Biol.*

Cell, vol. 18, no. 8, pp. 3059–3067, Aug. 2007, doi: 10.1091/MBC.E06-08-0689/ASSET/IMAGES/LARGE/ZMK0080781550004.JPEG.

[224] A. Azvolinsky, P. G. Giresi, J. D. Lieb, and V. A. Zakian, "Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae," *Mol. Cell*, vol. 34, no. 6, p. 722, Jun. 2009, doi: 10.1016/J.MOLCEL.2009.05.022.

[225] R. Yeung and D. J. Smith, "Determinants of Replication-Fork Pausing at tRNA Genes in Saccharomyces cerevisiae," *Genetics*, vol. 214, no. 4, pp. 825–838, Apr. 2020, doi: 10.1534/GENETICS.120.303092.

[226] E. V. Mirkin and S. M. Mirkin, "Replication fork stalling at natural impediments," *Microbiol. Mol. Biol. Rev.*, vol. 71, no. 1, pp. 13–35, Mar. 2007, doi: 10.1128/MMBR.00030-06.

[227] M. C. Gadaleta and E. Noguchi, "Regulation of DNA Replication through Natural Impediments in the Eukaryotic Genome," *Genes 2017, Vol. 8, Page 98*, vol. 8, no. 3, p. 98, Mar. 2017, doi: 10.3390/GENES8030098.

[228] A. Sfeir *et al.*, "Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication," *Cell*, vol. 138, no. 1, p. 90, Jul. 2009, doi: 10.1016/J.CELL.2009.06.021.

[229] P. C. Li, R. C. Petreaca, A. Jensen, J. P. Yuan, M. D. Green, and S. L. Forsburg, "Replication Fork Stability Is Essential for the Maintenance of Centromere Integrity in the Absence of Heterochromatin," *Cell Rep.*, vol. 3, no. 3, pp. 638–645, Mar. 2013, doi: 10.1016/J.CELREP.2013.02.007.

[230] E. Tsang and A. M. Carr, "Replication fork arrest, recombination and the maintenance of ribosomal DNA stability," *DNA Repair (Amst).*, vol. 7, no. 10, pp. 1613–1623, Oct. 2008, doi: 10.1016/J.DNAREP.2008.06.010.

[231] A. Serra-Cardona and Z. Zhang, "Replication-coupled nucleosome assembly as a passage of epigenetic information and cell identity," *Trends Biochem. Sci.*, vol. 43, no. 2, p. 136, Feb. 2018, doi: 10.1016/J.TIBS.2017.12.003.

[232] G. Almouzni and H. Cedar, "Maintenance of epigenetic information," *Cold Spring Harb. Perspect. Biol.*, vol. 8, no. 5, 2016, doi: 10.1101/cshperspect.a019372.

[233] J. M. Dewar and J. C. Walter, "Mechanisms of DNA replication termination," *Nat. Rev. Mol. Cell Biol. 2017 188*, vol. 18, no. 8, pp. 507–516, May 2017, doi: 10.1038/nrm.2017.42.

[234] D. Fachinetti *et al.*, "Replication Termination at Eukaryotic Chromosomes Is Mediated by Top2 and Occurs at Genomic Loci Containing Pausing Elements," *Mol. Cell*, vol. 39, no. 4, pp. 595–605, Aug. 2010, doi: 10.1016/J.MOLCEL.2010.07.024.

[235] J. M. Dewar, M. Budzowska, and J. C. Walter, "The mechanism of DNA replication termination in vertebrates," *Nature*, vol. 525, no. 7569, pp. 345–350, Sep. 2015, doi: 10.1038/NATURE14887.

[236] M. Maric, T. Maculins, G. De Piccoli, and K. Labib, "Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication," *Science (80-.).*, vol. 346, no. 6208, Oct. 2014, doi: 10.1126/SCIENCE.1253596/SUPPL_FILE/MARIC.SM.PDF.

[237] M. E. Francia and B. Striepen, "Cell division in apicomplexan parasites," *Nat. Rev. Microbiol. 2014 122*, vol. 12, no. 2, pp. 125–136, Jan. 2014, doi: 10.1038/nrmicro3184.

[238] M. Roques, A. Bindschedler, R. Beyeler, and V. T. Heussler, "Same, same but different: Exploring Plasmodium cell division during liver stage development," *PLOS Pathog.*, vol. 19, no. 3, p. e1011210, Mar. 2023, doi: 10.1371/JOURNAL.PPAT.1011210.

[239] N. Hall *et al.*, "A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and

proteomic analyses.," *Science*, vol. 307, no. 5706, pp. 82–6, Jan. 2005, doi: 10.1126/science.1103717.

[240] L. Borges-Pereira, B. K. M. Dias, M. K. Singh, and C. R. S. Garcia, "Malaria parasites and circadian rhythm: New insights into an old puzzle," *Curr. Res. Microb. Sci.*, vol. 2, Dec. 2021, doi: 10.1016/J.CRMICR.2020.100017.

[241] S. Stanojcic *et al.*, "Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes Leishmania and Trypanosoma brucei versus mammalian cells," *Sci. Rep.*, vol. 6, Mar. 2016, doi: 10.1038/SREP23142.

[242] B. M. Invergo, M. Brochet, L. Yu, J. Choudhary, P. Beltrao, and O. Billker, "Sub-minute Phosphoregulation of Cell Cycle Systems during Plasmodium Gamete Formation," *Cell Rep.*, vol. 21, no. 7, pp. 2017–2029, Nov. 2017, doi: 10.1016/J.CELREP.2017.10.071.

[243] H. Matthews, J. McDonald, F. I. G. Totañes, and C. J. Merrick, "Dynamics of DNA Replication during Male Gametogenesis in the Malaria Parasite Plasmodium Falciparum," *Cell. Microbiol.*, vol. 2022, 2022, doi: 10.1155/2022/2701868.

[244] H. Fang *et al.*, "Multiple short windows of calcium-dependent protein kinase 4 activity coordinate distinct cell cycle events during Plasmodium gametogenesis," *Elife*, vol. 6, May 2017, doi: 10.7554/ELIFE.26524.

[245] J. A. Vaughan, "Population dynamics of Plasmodium sporogony," *Trends Parasitol.*, vol. 23, no. 2, pp. 63–70, Feb. 2007, doi: 10.1016/J.PT.2006.12.009.

[246] S. Bennink, M. J. Kiesow, and G. Pradel, "The development of malaria parasites in the mosquito midgut," *Cell. Microbiol.*, vol. 18, no. 7, pp. 905–918, Jul. 2016, doi: 10.1111/CMI.12604.

[247] S. Patterson, C. Robert, C. Whittle, R. Chakrabarti, C. Doerig, and D. Chakrabarti, "Pre-replication complex organization in the atypical DNA replication cycle of Plasmodium falciparum: Characterization of the mini-chromosome maintenance (MCM) complex formation," *Mol. Biochem. Parasitol.*, vol. 145, no. 1, pp. 50–59, Jan. 2006, doi: 10.1016/j.molbiopara.2005.09.006.

[248] R. G. Ridley *et al.*, "DNA polymerase δ: gene sequences from Plasmodium falciparum indicate that this enzyme is more highly conserved than DNA polymerase α," *Nucleic Acids Res.*, vol. 19, no. 24, pp. 6731–6736, Dec. 1991, doi: 10.1093/NAR/19.24.6731.

[249] J. H. White *et al.*, "The gene encoding DNA polymerase α from Plasmodium falciparum," *Nucleic Acids Res.*, vol. 21, no. 16, pp. 3643–3646, Aug. 1993, doi: 10.1093/NAR/21.16.3643.

[250] P. Mitra, K. Banu, A. S. Deshmukh, N. Subbarao, and S. K. Dhar, "Functional dissection of proliferating-cell nuclear antigens (1 and 2) in human malarial parasite Plasmodium falciparum: Possible involvement in DNA replication and DNA damage response," *Biochem. J.*, vol. 470, no. 1, pp. 115–129, Aug. 2015, doi: 10.1042/BJ20150452.

[251] S. Patterson, C. Whittle, C. Robert, and D. Chakrabarti, "Molecular characterization and expression of an alternate proliferating cell nuclear antigen homologue, PfPCNA2, in Plasmodium falciparum," *Biochem. Biophys. Res. Commun.*, vol. 298, no. 3, pp. 371–376, Nov. 2002, doi: 10.1016/S0006-291X(02)02436-1.

[252] A. Gupta, P. Mehra, and S. K. Dhar, "Plasmodium falciparum origin recognition complex subunit 5: functional characterization and role in DNA replication foci formation," *Mol. Microbiol.*, vol. 69, no. 3, p. 646, Aug. 2008, doi: 10.1111/J.1365-2958.2008.06316.X.

[253] A. S. Deshmukh *et al.*, "The role of N-terminus of Plasmodium falciparum ORC1 in telomeric localization and var gene silencing," *Nucleic Acids Res.*, vol. 40, no. 12, pp. 5313–5331, Jul. 2012, doi: 10.1093/NAR/GKS202.

[254] P. Mehra, A. K. Biswas, A. Gupta, S. Gourinath, C. E. Chitnis, and S. K. Dhar, “Expression and characterization of human malaria parasite Plasmodium falciparum origin recognition complex subunit 1,” *Biochem. Biophys. Res. Commun.*, vol. 337, no. 3, pp. 955–966, Nov. 2005, doi: 10.1016/J.BBRC.2005.09.131.

[255] R. Sharma, B. Sharma, A. Gupta, and S. K. Dhar, “Identification of a novel trafficking pathway exporting a replication protein, Orc2 to nucleus via classical secretory pathway in Plasmodium falciparum,” *Biochim. Biophys. Acta - Mol. Cell Res.*, vol. 1865, no. 5, pp. 817–829, May 2018, doi: 10.1016/J.BBAMCR.2018.03.003.

[256] T. Iwanaga *et al.*, “Characterization of Plasmodium falciparum cdc2-related kinase and the effects of a CDK inhibitor on the parasites in erythrocytic schizogony,” *Parasitol. Int.*, vol. 62, no. 5, pp. 423–430, Oct. 2013, doi: 10.1016/J.PARINT.2013.05.003.

[257] C. Doerig *et al.*, “Pfcrk-1, a developmentally regulated cdc2-related protein kinase of Plasmodium falciparum,” *Mol. Biochem. Parasitol.*, vol. 70, no. 1–2, pp. 167–174, Mar. 1995, doi: 10.1016/0166-6851(95)00033-W.

[258] J. Kimmel *et al.*, “Gene-by-gene screen of the unknown proteins encoded on Plasmodium falciparum chromosome 3,” *Cell Syst.*, vol. 14, no. 1, pp. 9-23.e7, Jan. 2023, doi: 10.1016/J.CELS.2022.12.001.

[259] M. Zhang *et al.*, “Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis,” *Science (80-.).*, vol. 360, no. 6388, May 2018, doi: 10.1126/SCIENCE.AAP7847/SUPPL_FILE/AAP7847_ZHANG_SM_TABLE-S9.XLSX.

[260] E. Bushell *et al.*, “Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes,” *Cell*, vol. 170, no. 2, pp. 260-272.e8, Jul. 2017, doi: 10.1016/J.CELL.2017.06.030.

[261] A. S. Deshmukh *et al.*, “Regulation of Plasmodium falciparum Origin Recognition Complex subunit 1 (PfORC1) function through phosphorylation mediated by CDK-like kinase PK5,” *Mol. Microbiol.*, vol. 98, no. 1, pp. 17–33, Oct. 2015, doi: 10.1111/MMI.13099.

[262] D. Jirage *et al.*, “The malarial CDK Pfmrk and its effector PfMAT1 phosphorylate DNA replication proteins and co-localize in the nucleus,” *Mol. Biochem. Parasitol.*, vol. 172, no. 1, pp. 9–18, Jul. 2010, doi: 10.1016/J.MOLBIOPARA.2010.03.009.

[263] J. A. Robbins, S. Absalon, V. A. Streva, and J. D. Dvorin, “The Malaria Parasite Cyclin H Homolog PfCyc1 Is Required for Efficient Cytokinesis in Blood-Stage Plasmodium falciparum,” *MBio*, vol. 8, no. 3, May 2017, doi: 10.1128/MBIO.00605-17.

[264] B. Novák, J. C. Sible, and J. J. Tyson, “Checkpoints in the Cell Cycle,” *eLS*, no. February 2003, 2003, doi: 10.1038/npg.els.0001355.

[265] H. L. Smith, H. Southgate, D. A. Tweddle, and N. J. Curtin, “DNA damage checkpoint kinases in cancer,” *Expert Rev. Mol. Med.*, vol. 22, p. e2, 2020, doi: 10.1017/ERM.2020.3.

[266] C. Vaziri *et al.*, “A p53-Dependent Checkpoint Pathway Prevents Rereplication censes the chromatin for replication initiation once cyclin-dependent kinases become active at the onset of S phase. Besides initiating DNA replication at the G1-S transi,” *Mol. Cell*, vol. 11, pp. 997–1008, 2003.

[267] S. C. Fang, C. De Los Reyes, and J. G. Umen, “Cell Size Checkpoint Control by the Retinoblastoma Tumor Suppressor Pathway,” *PLOS Genet.*, vol. 2, no. 10, p. e167, Oct. 2006, doi: 10.1371/JOURNAL.PGEN.0020167.

[268] R. T. Abraham, “Cell cycle checkpoint signaling through the ATM and ATR kinases,” *Genes Dev.*, vol. 15, no. 17, pp. 2177–2196, Sep. 2001, doi: 10.1101/GAD.914401.

[269] R. van Biljon *et al.*, "Inducing controlled cell cycle arrest and re-entry during asexual proliferation of Plasmodium falciparum malaria parasites," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, 2018, doi: 10.1038/s41598-018-34964-w.

[270] S. E. Babbitt *et al.*, "Plasmodium falciparum responds to amino acid starvation by entering into a hibernatory state," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 47, p. E3278, Nov. 2012, doi: 10.1073/PNAS.1209823109/-/DCSUPPLEMENTAL.

[271] C. A. Alvarez and E. S. Suvorova, "Checkpoints of apicomplexan cell division identified in Toxoplasma gondii," *PLOS Pathog.*, vol. 13, no. 7, p. e1006483, Jul. 2017, doi: 10.1371/JOURNAL.PPAT.1006483.

[272] K. E. Baillie and P. C. Stirling, "Beyond Kinases: Targeting Replication Stress Proteins in Cancer Therapy," *Trends in Cancer*, vol. 7, no. 5, pp. 430–446, May 2021, doi: 10.1016/J.TRECAN.2020.10.010.

[273] U. Jo, Y. Murai, N. Takebe, A. Thomas, and Y. Pommier, "Precision Oncology with Drugs Targeting the Replication Stress, ATR, and Schlafen 11," *Cancers (Basel).*, vol. 13, no. 18, Sep. 2021, doi: 10.3390/CANCERS13184601.

[274] E. van Eijk, B. Wittekoek, E. J. Kuijper, and W. K. Smits, "DNA replication proteins as potential targets for antimicrobials in drug-resistant bacterial pathogens," *J. Antimicrob. Chemother.*, vol. 72, no. 5, p. 1275, 2017, doi: 10.1093/JAC/DKW548.

[275] A. Gregson and C. V. Plowe, "Mechanisms of resistance of malaria parasites to antifolates," *Pharmacol. Rev.*, vol. 57, no. 1, pp. 117–145, Mar. 2005, doi: 10.1124/PR.57.1.4.

[276] K. A. Gray *et al.*, "Correlation between Cyclin Dependent Kinases and Artemisinin-Induced Dormancy in Plasmodium falciparum In Vitro," *PLoS One*, vol. 11, no. 6, Jun. 2016, doi: 10.1371/JOURNAL.PONE.0157906.

[277] A. Barski and K. Zhao, "Genomic location analysis by ChIP-Seq," *J. Cell. Biochem.*, vol. 107, no. 1, pp. 11–18, May 2009, doi: 10.1002/JCB.22077.

[278] J. A. Belsky, H. K. Macalpine, Y. Lubelsky, A. J. Hartemink, and D. M. Macalpine, "Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly," *Genes Dev.*, vol. 29, no. 2, pp. 212–224, Jan. 2015, doi: 10.1101/GAD.247924.114.

[279] B. MA, "DNAscent v2: detecting replication forks in nanopore sequencing data with deep learning," *BMC Genomics*, vol. 22, no. 1, Dec. 2021, doi: 10.1186/S12864-021-07736-6.

[280] "How nanopore sequencing works." https://nanoporetech.com/support/how-it-works (accessed Jul. 18, 2023).

[281] P. A. Zhao, T. Sasaki, and D. M. Gilbert, "High-resolution Repli-Seq defines the temporal choreography of initiation, elongation and termination of replication in mammalian cells," *Genome Biol.*, vol. 21, no. 1, pp. 1–20, Mar. 2020, doi: 10.1186/S13059-020-01983-8/FIGURES/7.

[282] H. Miura *et al.*, "Mapping replication timing domains genome wide in single mammalian cells with single-cell DNA replication sequencing," *Nat. Protoc. 2020 1512*, vol. 15, no. 12, pp. 4058–4100, Nov. 2020, doi: 10.1038/s41596-020-0378-5.

[283] M. Blin *et al.*, "DNA molecular combing-based replication fork directionality profiling," *Nucleic Acids Res.*, vol. 49, no. 12, pp. e69–e69, Jul. 2021, doi: 10.1093/NAR/GKAB219.

[284] I. Ljungström, H. Perlmann, M. Schlichtherle, A. Scherf, and M. Wahlgren, "Methods in Malaria Research," *Evaluation*, pp. 26–27, 2004.

[285] C. Lambros and J. P. Vanderberg, "Synchronization of Plasmodium falciparum erythrocytic stages in

culture," *J. Parasitol.*, vol. 65, no. 3, pp. 418–420, 1979, doi: 10.2307/3280287.

[286] E. Knuepfer, M. Napiorkowska, C. Van Ooij, and A. A. Holder, "Generating conditional gene knockouts in Plasmodium - A toolkit to produce stable DiCre recombinase-expressing parasite lines using CRISPR/Cas9," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/s41598-017-03984-3.

[287] C. J. Merrick, "Transfection with thymidine kinase permits bromodeoxyuridine labelling of DNA replication in the human malaria parasite Plasmodium falciparum," *Malar. J. 2015 141*, vol. 14, no. 1, pp. 1–12, Dec. 2015, doi: 10.1186/S12936-015-1014-7.

[288] B. S. Crabb, A. F. Cowman, G. J. V Nossal, and W. Hall, "Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodium falciparum.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 14, p. 7289, Jul. 1996, doi: 10.1073/PNAS.93.14.7289.

[289] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with imageJ," *Biophotonics Int.*, vol. 11, no. 7, pp. 36–41, 2004, doi: 10.1201/9781420005615.ax4.

[290] F. P. C. and S. Bolte, "JACoP v2.0: improving the user experience with co-localization studies," *ImageJ User Dev. Conf. Novemb. 2009*, vol. 224, no. 3, pp. 213–232, 2006.

[291] "Microscopy Image Analysis Software - Imaris - Oxford Instruments." https://imaris.oxinst.com/ (accessed Jun. 30, 2023).

[292] J. M. Bryant *et al.*, "Exploring the virulence gene interactome with CRISPR/dCas9 in the human malaria parasite," *Mol. Syst. Biol.*, vol. 16, no. 8, p. e9569, Aug. 2020, doi: 10.15252/MSB.20209569.

[293] S. Andrews, "FastQC: a quality control tool for high throughput sequence data." Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010.

[294] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows–Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, Jul. 2009, doi: 10.1093/BIOINFORMATICS/BTP324.

[295] H. Li *et al.*, "The Sequence Alignment/Map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, Aug. 2009, doi: 10.1093/BIOINFORMATICS/BTP352.

[296] Y. Zhang *et al.*, "Model-based analysis of ChIP-Seq (MACS)," *Genome Biol.*, vol. 9, no. 9, pp. 1–9, Sep. 2008, doi: 10.1186/GB-2008-9-9-R137/FIGURES/3.

[297] A. R. Quinlan, "BEDTools: The Swiss-Army Tool for Genome Feature Analysis.," *Curr. Protoc. Bioinforma.*, vol. 47, pp. 11.12.1-34, Sep. 2014, doi: 10.1002/0471250953.BI1112S47.

[298] A. R. Quinlan and I. M. Hall, "BEDTools: A flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, no. 6, pp. 841–842, Jan. 2010, doi: 10.1093/bioinformatics/btq033.

[299] F. Ramírez *et al.*, "deepTools2: a next generation web server for deep-sequencing data analysis," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W160–W165, Jul. 2016, doi: 10.1093/NAR/GKW257.

[300] J. T. Robinson *et al.*, "Integrative Genomics Viewer," *Nat. Biotechnol.*, vol. 29, no. 1, p. 24, Jan. 2011, doi: 10.1038/NBT.1754.

[301] R. C. Team, "R: A language and environment for statistical computing.," *MSOR Connect.*, vol. 1, 2014.

[302] R. A. M. Villanueva and Z. J. Chen, "ggplot2: Elegant Graphics for Data Analysis (2nd ed.)," *https://doi.org/10.1080/15366367.2019.1565254*, vol. 17, no. 3, pp. 160–167, Jul. 2019, doi: 10.1080/15366367.2019.1565254.

[303] M. Wickham, Hadley; Vaughan, Davis and Girlich, "tidyr: Tidy Messy Data." 2023.

[304] G. Yu, L. G. Wang, and Q. Y. He, "ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization," *Bioinformatics*, vol. 31, no. 14, pp. 2382–2383, Jul. 2015, doi: 10.1093/BIOINFORMATICS/BTV145.

[305] Q. Wang *et al.*, "Exploring Epigenomic Datasets by ChIPseeker," *Curr. Protoc.*, vol. 2, no. 10, Oct. 2022, doi: 10.1002/CPZ1.585.

[306] K. Pérez-Toledo *et al.*, "Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes," *Nucleic Acids Res.*, vol. 37, no. 8, pp. 2596–2606, May 2009, doi: 10.1093/nar/gkp115.

[307] T. L. Bailey, J. Johnson, C. E. Grant, and W. S. Noble, "The MEME Suite," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W39–W49, Jul. 2015, doi: 10.1093/NAR/GKV416.

[308] P. Machanick and T. L. Bailey, "MEME-ChIP: motif analysis of large DNA datasets," *Bioinformatics*, vol. 27, no. 12, pp. 1696–1697, Jun. 2011, doi: 10.1093/BIOINFORMATICS/BTR189.

[309] A. Miles *et al.*, "Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum," *Genome Res.*, vol. 26, no. 9, pp. 1288–1299, 2016, doi: 10.1101/gr.203711.115.

[310] B. M. Sirbu, F. B. Couch, and D. Cortez, "Monitoring the spatiotemporal dynamics of proteins at replication forks and in assembled chromatin using Isolation of Proteins On Nascent DNA (iPOND)," *Nat. Protoc.*, vol. 7, no. 3, p. 594, Mar. 2012, doi: 10.1038/NPROT.2012.010.

[311] X. Zhang, A. H. Smits, G. B. A. Van Tilburg, H. Ovaa, W. Huber, and M. Vermeulen, "Proteome-wide identification of ubiquitin interactions using UbIA-MS," *Nat. Protoc. 2018 133*, vol. 13, no. 3, pp. 530–550, Feb. 2018, doi: 10.1038/nprot.2017.147.

[312] D. Szklarczyk *et al.*, "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D605–D612, Jan. 2021, doi: 10.1093/NAR/GKAA1074.

[313] H. Harashima, N. Dissmeyer, and A. Schnittger, "Cell cycle control across the eukaryotic kingdom," *Trends Cell Biol.*, vol. 23, no. 7, pp. 345–356, Jul. 2013, doi: 10.1016/J.TCB.2013.03.002/ATTACHMENT/67C9AE6E-84B9-4055-B237-CE220F5C781F/MMC1.XLS.

[314] J. A. Naughton and A. Bell, "Studies on cell-cycle synchronization in the asexual erythrocytic stages of Plasmodium falciparum," *Parasitology*, vol. 134, no. 3, pp. 331–337, Mar. 2007, doi: 10.1017/S0031182006001466.

[315] S. B. Buck, J. Bradford, K. R. Gee, B. J. Agnew, S. T. Clarke, and A. Salic, "Detection of S-phase cell cycle progression using 5-ethynyl-2'- deoxyuridine incorporation with click chemistry, an alternative to using 5-bromo-2'-deoxyuridine antibodies," *Biotechniques*, vol. 44, no. 7, pp. 927–929, Jun. 2008, doi: 10.2144/000112812/ASSET/IMAGES/LARGE/FIGURE2.JPEG.

[316] K. Masuda, C. Renard-Guillet, K. Shirahige, and T. Sutani, "Bioinformatical dissection of fission yeast DNA replication origins," *Open Biol.*, vol. 10, no. 7, p. 200052, Jul. 2020, doi: 10.1098/RSOB.200052.

[317] B. Miotto, Z. Ji, and K. Struhl, "Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers," *Proc. Natl. Acad. Sci.*, vol. 113, no. 33, pp. E4810–E4819, Aug. 2016, doi: 10.1073/pnas.1609060113.

[318] A. Gupta *et al.*, "Functional dissection of the catalytic carboxyl-terminal domain of origin recognition complex subunit 1 (PfORC1) of the human malaria parasite plasmodium falciparum," *Eukaryot. Cell*, vol. 8, no. 9, pp. 1341–1351, Sep. 2009, doi: 10.1128/EC.00170-09/SUPPL_FILE/SUPPLEMENTARY_DATA.PDF.

[319] M. Ghorbal, M. Gorman, C. R. MacPherson, R. M. Martins, A. Scherf, and J. J. Lopez-Rubio, "Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system," *Nat. Biotechnol. 2014 328*, vol. 32, no. 8, pp. 819–821, Jun. 2014, doi: 10.1038/nbt.2925.

[320] J. C. Wagner, R. J. Platt, S. J. Goldfless, F. Zhang, and J. C. Niles, "Efficient CRISPR-Cas9–mediated genome editing in Plasmodium falciparum," *Nat. Methods 2014 119*, vol. 11, no. 9, pp. 915–918, Aug. 2014, doi: 10.1038/nmeth.3063.

[321] A. H. Lee, L. S. Symington, and D. A. Fidock, "DNA Repair Mechanisms and Their Biological Roles in the Malaria Parasite Plasmodium falciparum," *Microbiol. Mol. Biol. Rev.*, vol. 78, no. 3, pp. 469–486, Sep. 2014, doi: 10.1128/MMBR.00059-13.

[322] S. G. Landt *et al.*, "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia," *Genome Res.*, vol. 22, no. 9, pp. 1813–1831, Sep. 2012, doi: 10.1101/GR.136184.111.

[323] G. I. Dellino *et al.*, "Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing," *Genome Res.*, vol. 23, no. 1, pp. 1–11, Jan. 2013, doi: 10.1101/gr.142331.112.

[324] W. Xu, J. G. Aparicio, O. M. Aparicio, and S. Tavaré, "Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae," *BMC Genomics*, vol. 7, no. 1, pp. 1–16, Oct. 2006, doi: 10.1186/1471-2164-7-276/FIGURES/7.

[325] G. A. Josling *et al.*, "Dissecting the role of PfAP2-G in malaria gametocytogenesis," *Nat. Commun.*, 2020, doi: 10.1038/s41467-020-15026-0.

[326] J. M. Santos *et al.*, "Red blood cell invasion by the malaria parasite is coordinated by the PfAP2-I transcription factor," *Cell Host Microbe*, vol. 21, no. 6, p. 731, Jun. 2017, doi: 10.1016/J.CHOM.2017.05.006.

[327] G. A. Josling *et al.*, "Dissecting the role of PfAP2-G in malaria gametocytogenesis," *Nat. Commun.*, vol. 11, no. 1, Dec. 2020, doi: 10.1038/S41467-020-15026-0.

[328] M. Agarwal, K. Bhowmick, K. Shah, A. Krishnamachari, and S. K. Dhar, "Identification and characterization of ARS-like sequences as putative origin(s) of replication in human malaria parasite Plasmodium falciparum," *FEBS J.*, vol. 284, no. 16, pp. 2674–2695, Aug. 2017, doi: 10.1111/FEBS.14150.

[329] A. Bedrat, L. Lacroix, and J.-L. L. Mergny, "Re-evaluation of G-quadruplex propensity with G4Hunter," *Nucleic Acids Res.*, vol. 44, no. 4, pp. 1746–59, Feb. 2016, doi: 10.1093/nar/gkw006.

[330] Q. Ding and D. M. MacAlpine, "Defining the replication program through the chromatin landscape," *Crit. Rev. Biochem. Mol. Biol.*, vol. 46, no. 2, pp. 165–179, Apr. 2011, doi: 10.3109/10409238.2011.560139.

[331] S. G. Prasanth, Z. Shen, K. V. Prasanth, and B. Stillman, "Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 34, pp. 15093–15098, Aug. 2010, doi: 10.1073/pnas.1009945107.

[332] Y. Tatsumi *et al.*, "Involvement of human ORC and TRF2 in pre-replication complex assembly at telomeres," *Genes to Cells*, vol. 13, no. 10, pp. 1045–1059, Oct. 2008, doi: 10.1111/J.1365-2443.2008.01224.X.

[333] S. G. Prasanth, K. V. Prasanth, K. Siddiqui, D. L. Spector, and B. Stillman, "Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance," *EMBO J.*, vol. 23, no. 13, pp. 2651–2663, Jul. 2004, doi: 10.1038/SJ.EMBOJ.7600255.

[334] M. M. Shareef, C. King, M. Damaj, R. Badagu, Da Wei Huang, and R. Kellum, "Drosophila heterochromatin protein 1 (HP1)/origin recognition complex (ORC) protein is associated with HP1 and ORC and functions

in heterochromatin-induced silencing," *Mol. Biol. Cell*, vol. 12, no. 6, pp. 1671–1685, Oct. 2001, doi: 10.1091/MBC.12.6.1671/ASSET/IMAGES/LARGE/MK0611521008.JPEG.

[335] D. T. S. Pak *et al.*, "Association of the Origin Recognition Complex with Heterochromatin and HP1 in Higher Eukaryotes," *Cell*, vol. 91, no. 3, pp. 311–323, Oct. 1997, doi: 10.1016/S0092-8674(00)80415-8.

[336] L. Mancio-Silva, A. P. Rojas-Meza, M. Vargas, A. Scherf, and R. Hernandez-Rivas, "Differential association of Orc1 and Sir2 proteins to telomeric domains in Plasmodium falciparum," *J. Cell Sci.*, vol. 121, no. 12, pp. 2046–2053, Jun. 2008, doi: 10.1242/JCS.026427.

[337] N. Smargiasso *et al.*, "Putative DNA G-quadruplex formation within the promoters of Plasmodium falciparum var genes," *BMC Genomics*, vol. 10, Aug. 2009, doi: 10.1186/1471-2164-10-362.

[338] K. J. Harvey and J. Newport, " CpG Methylation of DNA Restricts Prereplication Complex Assembly in Xenopus Egg Extracts ," *Mol. Cell. Biol.*, vol. 23, no. 19, pp. 6769–6779, Oct. 2003, doi: 10.1128/MCB.23.19.6769-6779.2003/ASSET/BA657CE0-97B3-444A-BCC6-705E2804F147/ASSETS/GRAPHIC/MB1930575006.JPEG.

[339] O. Hyrien, *How MCM loading and spreading specify eukaryotic DNA replication initiation sites*, vol. 5. 2016, p. 2063.

[340] O. Hyrien, "Peaks cloaked in the mist: the landscape of mammalian replication origins," *J. Cell Biol.*, vol. 208, no. 2, pp. 147–160, 2015, doi: 10.1083/JCB.201407004.

[341] S. Stanojcic, N. Kuk, I. Ullah, Y. Sterkers, and C. J. Merrick, "Single-molecule analysis reveals that DNA replication dynamics vary across the course of schizogony in the malaria parasite Plasmodium falciparum," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, doi: 10.1038/S41598-017-04407-Z.

[342] S. Makovets, I. Herskowitz, and E. H. Blackburn, "Anatomy and Dynamics of DNA Replication Fork Movement in Yeast Telomeric Regions," *https://doi.org/10.1128/MCB.24.9.4019-4031.2004*, vol. 24, no. 9, pp. 4019–4031, May 2023, doi: 10.1128/MCB.24.9.4019-4031.2004.

[343] T. J. Leach, H. L. Chotkowski, M. G. Wotring, R. L. Dilwith, and R. L. Glaser, "Replication of Heterochromatin and Structure of Polytene Chromosomes," *https://doi.org/10.1128/MCB.20.17.6308-6316.2000*, vol. 20, no. 17, pp. 6308–6316, Sep. 2023, doi: 10.1128/MCB.20.17.6308-6316.2000.

[344] Z. Hou, D. A. Bernstein, C. A. Fox, and J. L. Keck, "Structural basis of the Sir1–origin recognition complex interaction in transcriptional silencing," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 24, p. 8489, Jun. 2005, doi: 10.1073/PNAS.0503525102.

[345] T. Triolo and R. Sternglanz, "Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing," *Nature*, vol. 381, no. 6579, pp. 251–253, May 1996, doi: 10.1038/381251A0.

[346] E. G. Dastidar *et al.*, "Involvement of Plasmodium falciparum protein kinase CK2 in the chromatin assembly pathway," *BMC Biol.*, vol. 10, no. 1, pp. 1–16, Jan. 2012, doi: 10.1186/1741-7007-10-5/TABLES/2.

[347] A. C. Balestra *et al.*, "A divergent cyclin/cyclin-dependent kinase complex controls the atypical replication of a malaria parasite during gametogony and transmission," *Elife*, vol. 9, pp. 1–25, Jun. 2020, doi: 10.7554/ELIFE.56474.

[348] H. Fang *et al.*, "Epistasis studies reveal redundancy among calcium-dependent protein kinases in motility and invasion of malaria parasites," *Nat. Commun. 2018 91*, vol. 9, no. 1, pp. 1–14, Oct. 2018, doi: 10.1038/s41467-018-06733-w.

[349] T. C. Branon *et al.*, "Efficient proximity labeling in living cells and organisms with TurboID," *Nat. Biotechnol.*, vol. 36, no. 9, pp. 880–898, Oct. 2018, doi: 10.1038/NBT.4201.

[350] K. J. Roux, D. I. Kim, M. Raida, and B. Burke, "A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells," *J. Cell Biol.*, vol. 196, no. 6, p. 801, Mar. 2012, doi: 10.1083/JCB.201112098.

[351] E. Choi-Rhee, H. Schulman, and J. E. Cronan, "Promiscuous protein biotinylation by Escherichia coli biotin protein ligase," *Protein Sci.*, vol. 13, no. 11, pp. 3043–3050, Nov. 2004, doi: 10.1110/PS.04911804.

[352] J. E. Cronan, "Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase," *J. Nutr. Biochem.*, vol. 16, no. 7, pp. 416–418, Jul. 2005, doi: 10.1016/J.JNUTBIO.2005.03.017.

[353] D. G. May, K. L. Scott, A. R. Campos, and K. J. Roux, "Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation," *Cells 2020, Vol. 9, Page 1070*, vol. 9, no. 5, p. 1070, Apr. 2020, doi: 10.3390/CELLS9051070.

[354] J. Birnbaum *et al.*, "A genetic system to study Plasmodium falciparum protein function," *Nat. Methods*, vol. 14, no. 4, pp. 450–456, Mar. 2017, doi: 10.1038/NMETH.4223.

[355] A. L. Szymczak *et al.*, "Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide–based retroviral vector," *Nat. Biotechnol. 2004 225*, vol. 22, no. 5, pp. 589–594, Apr. 2004, doi: 10.1038/nbt957.

[356] B. M. Sirbu *et al.*, "Identification of proteins at active, stalled, and collapsed replication forks using isolation of proteins on nascent DNA (iPOND) coupled with mass spectrometry," *J. Biol. Chem.*, vol. 288, no. 44, pp. 31458–31467, Nov. 2013, doi: 10.1074/jbc.M113.511337.

[357] A. J. Lopez-Contreras *et al.*, "A Proteomic Characterization of Factors Enriched at Nascent DNA Molecules," *Cell Rep.*, vol. 3, no. 4, p. 1105, Apr. 2013, doi: 10.1016/J.CELREP.2013.03.009.

[358] H. Dungrawala and D. Cortez, "Purification of proteins on newly synthesized DNA using iPOND," *Methods Mol. Biol.*, vol. 1228, p. 123, 2015, doi: 10.1007/978-1-4939-1680-1_10.

[359] M. K. Bhattacharyya, S. Bhattacharyya Nee Deb, B. Jayabalasingham, and N. Kumar, "Characterization of kinetics of DNA strand-exchange and ATP hydrolysis activities of recombinant PfRad51, a Plasmodium falciparum recombinase," *Mol. Biochem. Parasitol.*, vol. 139, no. 1, pp. 33–39, Jan. 2005, doi: 10.1016/J.MOLBIOPARA.2004.09.007.

[360] S. Briquet *et al.*, "High-mobility-group box nuclear factors of Plasmodium falciparum," *Eukaryot. Cell*, vol. 5, no. 4, pp. 672–682, Apr. 2006, doi: 10.1128/EC.5.4.672-682.2006/SUPPL_FILE/SUPPLEMENTARY_EC00357_05_2.ZIP.

[361] K. Y. Lee, K. Yang, M. A. Cohn, N. Sikdar, A. D. D'Andrea, and K. Myung, "Human ELG1 Regulates the Level of Ubiquitinated Proliferating Cell Nuclear Antigen (PCNA) through Its Interactions with PCNA and USP1," *J. Biol. Chem.*, vol. 285, no. 14, p. 10362, Apr. 2010, doi: 10.1074/JBC.M109.092544.

[362] B. Chen, Y. Sun, J. Niu, G. K. Jarugumilli, and X. Wu, "Protein lipidation in cell signaling and diseases: function, regulation and therapeutic opportunities," *Cell Chem. Biol.*, vol. 25, no. 7, p. 817, Jul. 2018, doi: 10.1016/J.CHEMBIOL.2018.05.003.

[363] H. Jiang, X. Zhang, X. Chen, P. Aramsangtienchai, Z. Tong, and H. Lin, "Protein lipidation: Occurrence, mechanisms, biological functions, and enabling technologies," *Chem. Rev.*, vol. 118, no. 3, p. 919, Feb. 2018, doi: 10.1021/ACS.CHEMREV.6B00750.

[364] M. H. Wright *et al.*, "Validation of N-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach," *Nat. Chem.*, vol. 6, no. 2, p. 112, Feb. 2014, doi: 10.1038/NCHEM.1830.

[365] T. Takahashi, E. Ohara, H. Nishitani, and H. Masukata, “Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast,” *EMBO J.*, vol. 22, no. 4, pp. 964–974, Feb. 2003, doi: 10.1093/EMBOJ/CDG079.

[366] M. Weinreich, C. Liang, H. H. Chen, and B. Stillman, “Binding of cyclin-dependent kinases to ORC and Cdc6p regulates the chromosome replication cycle,” *Proc. Natl. Acad. Sci.*, vol. 98, no. 20, pp. 11211–11217, Sep. 2001, doi: 10.1073/PNAS.201387198.

[367] M. Srivastava *et al.*, “Replisome Dynamics and Their Functional Relevance upon DNA Damage through the PCNA Interactome,” *Cell Rep.*, vol. 25, no. 13, pp. 3869-3883.e4, Dec. 2018, doi: 10.1016/J.CELREP.2018.11.099.

[368] R. Rangarajan *et al.*, “Pbcrk-1, the Plasmodium berghei orthologue of P. falciparum cdc-2 related kinase-1 (Pfcrk-1), is essential for completion of the intraerythrocytic asexual cycle,” *Exp. Parasitol.*, vol. 112, no. 3, pp. 202–207, Mar. 2006, doi: 10.1016/J.EXPPARA.2005.11.002.

[369] J. Inselburg and H. S. Banyal, “Synthesis of DNA during the asexual cycle of Plasmodium falciparum in culture.,” *Mol. Biochem. Parasitol.*, vol. 10, no. 1, pp. 79–87, Jan. 1984.

[370] A. Unnikrishnan, P. R. Gafken, and T. Tsukiyama, “Dynamic changes in histone acetylation regulate origins of DNA replication,” *Nat. Struct. Mol. Biol.*, vol. 17, no. 4, p. 430, Apr. 2010, doi: 10.1038/NSMB.1780.

[371] H. K. MacAlpine, R. Gordân, S. K. Powell, A. J. Hartemink, and D. M. MacAlpine, “Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading.,” *Genome Res.*, vol. 20, no. 2, pp. 201–11, Feb. 2010, doi: 10.1101/gr.097873.109.

[372] A. Stanton, L. M. Harris, G. Graham, and C. J. Merrick, “Recombination events among virulence genes in malaria parasites are associated with G-quadruplex-forming DNA motifs,” *BMC Genomics*, vol. 17, no. 1, pp. 1–16, Nov. 2016, doi: 10.1186/S12864-016-3183-3/FIGURES/4.

[373] M. S. da Silva *et al.*, “Nuclear DNA Replication in Trypanosomatids: There Are No Easy Methods for Solving Difficult Problems,” *Trends Parasitol.*, vol. 33, no. 11, pp. 858–874, Nov. 2017, doi: 10.1016/J.PT.2017.08.002.

[374] A. Lengronne, P. Pasero, A. Bensimon, and E. Schwob, “Monitoring S phase progression globally and locally using BrdU incorporation in TK+ yeast strains,” *Nucleic Acids Res.*, vol. 29, no. 7, p. 1433, Apr. 2001, doi: 10.1093/NAR/29.7.1433.