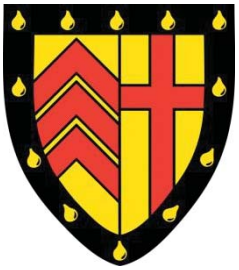
High-throughput reverse genetic screening in *Plasmodium berghei* using barcode sequencing

Ana Rita Batista Gomes



Clare College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy in Biological Science

September 2014

This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except where specified in the text. This dissertation is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other university. This dissertation does not exceed the prescribed limit of 60,000 words for the Degree Committee for the Faculty of Biology.

Ana Gomes

30th of September, 2014

Acknowledgements

My first thanks go to my PhD supervisors Oliver Billker and Christian Doerig for inviting me to be part of the EVIMalaR network, for guidance, and supervision during this PhD. Special thanks to Ellen Bushell and Mathieu Brochet for advice, discussion and assistance in all things.

Thank you Theo, Will, Kasia, Tom and Jaishree for being more than just lab buddies because a PhD is not just about working long hours in the lab. Caty, mon amie, pour son soutien sans lequel tout ce travail n'aurait été ni possible ni savoureux.

Susana and Leyla, my honorary mums, thank you for all the advice and helping me grow up during the past four years – you will be in my heart forever.

Alex and Kerstin, thank you so much for carefully reading this thesis despite not being biologists. Life in Cambridge would not have been this much fun without you. Thank you for everything.

Arthur, je te remercie profondément pour ton soutien au cours des derniers mois. J'ai hâte de savoir où vont nous conduire nos projets secrets.

Rita, minha companheira desde o dia em que nos conhecemos na GSLS induction. Nunca me hei-de esquecer de como nem foram precisas palavras para sabermos que ambas eramos portuguesas =) Foi um privilégio partilhar esta experiência contigo. Obrigada pela amizade, pelo companheirismo, por me arrastares para as danças de salão onde nos divertimos imenso, pelos jantares, por tudo... Nina e Vera, obrigada pelo vosso apoio durante esta última década. Fico feliz por saber que a nossa amizade resistiu a diferenças horárias de 12h. David, obrigada por estares sempre aí apesar dos nossos arrufos contantes. Tenho um feeling que um futuro grandioso te aguarda. Catarina e Iolanda muita força nesta recta final dos vossos doutoramentos e obrigada pelo vosso apoio.

To Gunnar and Céline, who have always inspired me, a sincere thank you for all the help and advice since the very beginning of my scientific journey.

Um beijo do tamanho do universo para a minha família. Obrigada pelo apoio incondicional e desculpem por não ter ido a casa o suficiente durante estes últimos quatro anos.

Last but not least, I would like to acknowledge EVIMalaR for awarding me the studentship, funded through the EU Seventh Framework Programme (FP7/2007-2013) and the Wellcome Trust Sanger Institute for funding my research and hosting me during my PhD.

Abstract

Malaria is a vector borne disease that causes one million deaths annually. The identification of novel drug targets is urgent but it requires a better understanding of the biology of *Plasmodium* parasites. Signature tagged mutagenesis (STM) has been used extensively in bacterial pathogens to identify virulence genes by parallel phenotyping of pools of individually tagged mutants.

Gene knock out (KO) vectors provided by a freely accessible resource produced at the Sanger Institute, *Plasmo*GEM, carry gene-specific barcodes that uniquely label parasites upon integration. This, together with an increased recombination frequency and strongly reduced incidence of episomes allowed me to establish a STM protocol for *Plasmodium berghei*, a rodent malaria parasite. Using this strategy, complex and defined pools of targeted KO mutants were reproducibly generated in a single mouse.

Vector-specific barcodes were amplified from daily blood samples by a polymerase chain reaction (PCR) and counted on a benchtop sequencer (MiSeq). This enabled the calculation of the relative growth rate of each population of mutants within a pool and how it changed during the infection. Each pool included a set of vectors that targeted genes that are only expressed in sexual and mosquito stages – *p25, p28, p230p* and *soap*. As these were known to be dispensable for asexual growth they were used as a normal growth reference for fitness cost analysis of the other mutants in the pool. Replicate experiments yielded nearly identical growth curves for each of the 48 populations of barcoded mutants. Southern hybridisation of separated chromosomes confirmed genomic integration events throughout the genome, many of which were further supported by PCR.

After this validation step, this technology was used to identify potential interaction pairs within the *P. berghei* kinome. A screen performed in six different mutant lines revealed multiple growth phenotypes that were recurrent in all backgrounds. Additionally, a severe growth defect was detected for a mutant lacking the *cdpk4* gene on a line expressing the resistant pkg^{T619Q} allele. This suggested the existence of an important genetic interaction between CDPK4 and PKG, which was further validated independently. In conclusion, this kind of high throughput genetic approach had no precedents in the malaria field and provides a promising basis for future screenings on large subsets of parasite genes.

Table of contents

List of Figures

List of tables

List of Appendices

Abbreviations

5-FC	5-fluorocytosine
5-FU	5-fluorouracil
ACT	Artemisinin combination therapies
AL	Adaptor ligation
aPK	Atypical protein kinases
AT	Adenosine thymidine
AT-content	Adenine and thymine content
ATP	Adenosine triphosphate
Bar-seq	Barcode analysis by sequencing
BKI	Bumped kinase inhibitors
bp	Base pairs
BWA	Burrows-Wheeler Aligner
CamK	Calcium/calmodulin-dependent kinases
cAMP	Cyclic adenosine monophosphate
CDK	Cyclin-dependent kinase
CDPKs	Calcium-dependent protein kinases
cGMP	Cyclic guanosine monophosphate
ChIP-Seq	Chromatin immune-precipitation
CK1	Casein-kinase 1
CLKs	CDK-like kinases
DA	Direct amplification
dAMP	Deoxyadenosine 5’-monophosphate
DNA	Deoxyribonucleic acid
ePK	Eukaryotic protein kinases
FBS	Fetal bovine serum
Gb	Giga base
gDNA	Genomic DNA
GFP	Green fluorescent protein
GOI	Gene of interest
GSK3	Glycogen synthase kinase
GW	Gateway Technology ®
h*dhfr*	Human dihydrofolate reductase
HRP	Horseradish peroxidase
i.p.	Intraperitoneal
i.v.	Intravenous
IHF	Integration host factor

Int	Integrase
Kb	Kilo base
KO	Knock out
MAP1	Mitogen-activated protein 1
MAP2	Mitogen-activated protein 2
MAPK	Mitogen-activated protein kinase
MAPKK	MAPK kinase
mRNA	Messenger RNA
ng	Nanogram
NGS	Next-generation sequencing
OI	organism of interest
OPK	Other protein kinases
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PF	Purity filter
PFGE	Pulsed-Field Gel Electrophoresis
PIKK	Phosphatidyl-inositol 3' kinase-related kinases
Plasmo GEM	*Plasmodium* genetic modification project
qPCR	Quantitative PCR
R&D	Research and development
RBCs	Red blood cells
RIO	Right open reading frame
RNA	Ribonucleic acid
RNAi	RNA interference
RNASeq	Transcriptome analysis
SBS	Sequencing by synthesis
SGA	Genetic array analysis
SNP	Single-nucleotide polymorphism
STM	Signature tagged mutagenesis
Tb	Tera base
TKL	Tyrosine-like kinases
TO mice	Theiler's Original mice
TyrK	Tyrosine kinases
UTR	Untranslated region
WBC	White blood cells
WGS	Whole genome sequencing
WT	Wild type
XA	Xanthurenic acid
Xis	Excisionase

Chapter 1

Introduction

The impact of malaria on human health is dramatic as financial constraints and wide-spread resistance to drugs hamper malaria control programmes. New, effective therapies are urgently required; their development however, relies on a better understanding of malaria parasite biology. The Evimalar network is one of the dedicated institutions created to promote malaria research as part of a global effort to eradicate this disease. As a PhD student of the EVIMalaR programme, I focused my efforts on the development of tools that enable large scale studies of the malaria parasite at the genetic level. To set the background for this work, this introduction outlines the disease burden, selected aspects of the malaria parasite biology, current high-throughput gene targeting strategies in model organisms and applications of such technologies to *Plasmodium* biology.

1.1 Malaria: A major global parasitic disease

Infectious diseases are still in the top 10 causes of death, having accounted for 18.4 % of total deaths worldwide in 2011 [1]. Although the widespread use of vaccines and drugs has dramatically decreased mortality from infectious diseases in developed countries, some them are beginning to emerge or re-emerge and are still prevalent in the developing countries [1].

Malaria is one of the oldest diseases known to mankind, the two appearing to have evolved together. Malaria literally means "bad air" (from the Italian words "*mal aria*"), named after the belief that this disease was caused by an unknown substance in the air arising from swamps [2]. Nowadays, it endangers half of the world's population (Fig.1.1). The WHO estimates that 219 million cases of malaria led to 660,000 deaths in 2012. Nearly 80 % of the cases and 90 % of the deaths are estimated to occur in sub-Saharan Africa, with children under the age of five and pregnant women being the most severely affected [3]. Malaria is a vector borne disease caused by parasites of the genus *Plasmodium*. The different species infect a wide variety of hosts including humans, monkeys, rodents, birds, and reptiles. The five human pathogens are: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. *P. falciparum* prevails in Africa and is the most deadly of these parasites, while *P. vivax* is less morbid but also widespread, and the other three species have a much lower incidence. The clinical symptoms of malaria include paroxysms (acute fever that is typically preceded by chills and rigor), vomiting, headache and anaemia. *P. falciparum*, in particular, can be responsible for cases of "severe malaria", a life-threatening condition that includes pronounced anaemia, disorders of the coagulation system and sequestration of the infected red

blood cells (RBCs) in the deep vasculature (e.g. brain, lungs and placenta in pregnant women) [4,5]. The neurological involvement may lead to a permanent coma and in some cases death. When a child recovers from a severe malaria episode, cognitive impairment is likely to occur thus reducing the child's lifelong potential [6]. The cyclic nature of the fever is a consequence of the strong inflammatory immune responses triggered by synchronous cycles of infection and release of parasites from circulating RBCs. This periodicity varies with the parasite; *P. malariae* causes fever every 72 hours, *P. knowlesi* every 24-28 hours and the remaining three species have 48-hour cycles. It is worth noting that *P. falciparum*, perhaps due to less synchronous growth, often causes an uninterrupted fever rather than periodic paroxysms [7].

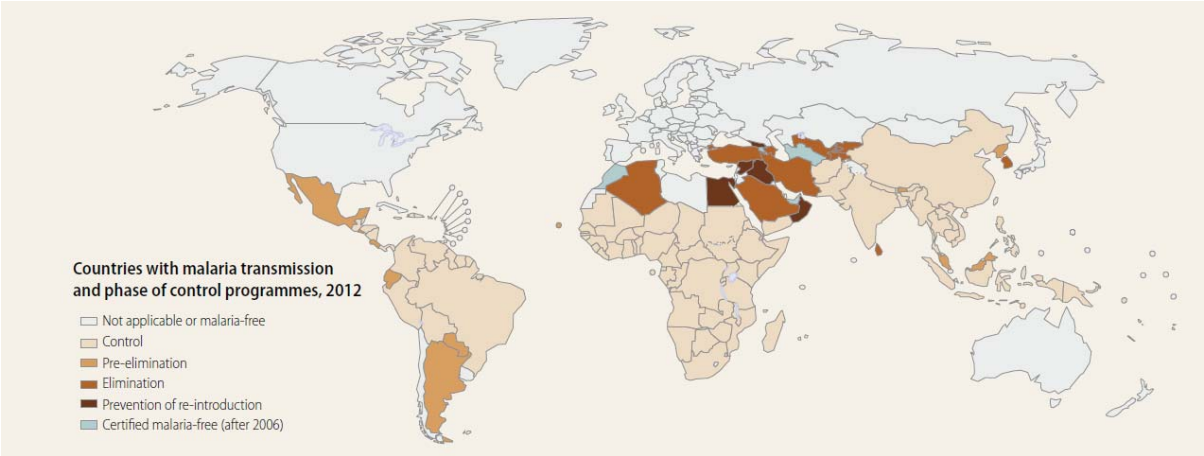


Fig. 1.1| Malaria endangers half of the world's population.
Malaria-free and malaria-endemic countries in phases of control, pre-elimination and elimination at the end of 2012. In "Pre-elimination" countries malaria test positivity rate is less than 5% during the malaria season (>5% "Control" phase); "Elimination" countries have zero incidence of locally transmitted infections; "Certified malaria free" countries have no locally transmitted infections for over a decade (Adapted from 2013 WHO report).

There is no natural acquisition of long lasting sterile immunity to malaria. In endemic areas, where exposure to infective mosquitoes is continuous, tolerance to infection may be seen in adults who, despite being asymptomatic, have a continuous low level of infection and thus act as reservoirs. Once these semi-immune individuals leave malaria endemic-areas, they lose their immunity within about six months [4]. This combined with genetic diversity, evolutionary plasticity and lifecycle complexity of *P. falciparum* complicate the development of malaria vaccines [8,9]. In the last two decades, over 40 vaccines designed to trigger an immune response against subunit components of liver or blood-stage parasites, or whole sporozoites (a mosquito stage form), have undergone clinical trials. Despite the promising results shown in pre-clinical and phase I-IIa trials, none accomplished full protection in the

field. Even the leading malaria vaccine candidate, RTS-S, demonstrated only modest protection against both clinical and severe malaria in young infants, in phase IIIb trials [10].

A range of malaria therapies exist. *P. falciparum* parasites have developed resistance against the majority of drugs introduced prior to the artemisinin combination therapies (ACT). For instance, chloroquine, once the drug of choice for prophylaxis and treatment, is no longer efficient in most areas where malaria is endemic [11]. Nowadays, it is only recommended for the treatment of malaria caused by *P. vivax* and *P. ovale* [12]. The ACT, the current first-line treatment for *P. falciparum* malaria combines the fast acting artemisinin-based compounds with a drug from a different class such as lumefantrine, mefloquine, piperaquine, among others, in order to reduce the chance of development and spread of resistance to either of the drugs. The second drug is chosen according to the local resistance patterns. Recent epidemiologic studies at the Thai-Cambodia border have, however, reported that the efficacy of ACT has decreased [8].

1.2 The life cycle of *Plasmodium* parasites

Malaria parasites belong to the phylum Apicomplexa, a large group of unicellular parasites that infect only animals. Other members of this phylum include *Cryptosporidium*, *Toxoplasma*, *Eimeria*, *Babesia* and *Theileria*. This phylum is defined by a specialised set of structures and secretory organelles at their apical tip, which is key to invasion and motility in –zoite stages [13]. This group of parasites features an unusual organelle named the apicoplast, which was acquired by secondary endosymbiosis between a free-living ancestor of these parasites and a red algae. The apicoplast is essential for parasite survival and contains biosynthetic pathways which have an equivalent in plants and bacteria, but not in animals such as type II pathway for *de novo* fatty acid synthesis [14,15].

Plasmodium species are obligate parasites with a complex life cycle that involves two different hosts: a vertebrate and a mosquito vector (Fig.1.2). Transmission to the vertebrate host is initiated by the bite of an infected female mosquito of the genus *Anopheles*, if the vertebrate is a mammal. During the blood meal *Plasmodium* sporozoites leave the mosquito salivary glands and enter the vertebrate's bloodstream (Fig. 1.2.1). The injected sporozoites migrate to the liver sinusoids where they traverse the vascular endothelium and invade hepatocytes (Fig. 1.2.2). There they multiply, giving rise to thousands of merozoites in 2-16 days, depending on the *Plasmodium* species (Fig. 1.2.3). Eventually merozoite-filled vesicles,

called merosomes, capable of infecting RBCs bud off from the hepatocyte into the liver sinusoids, thus starting the blood stage of the infection (Fig. 1.2.4) [16]. Inside the erythrocyte, a single merozoite replicates by schizogony and undergoes successive differentiations from ring through trophozoite stage eventually generating schizonts with 16-32 merozoites each. Finally the RBC ruptures and releases new merozoites, which in turn infect new erythrocytes (Fig. 1.2.5). While the initial stages of the infection (i.e. liver stages) are asymptomatic, repeated infection of erythrocytes by merozoites causes the symptoms and pathologies of malaria. During the asexual cycle in the blood a subset of parasites bypasses asexual multiplication and differentiates into sexually committed cells: the female and male gametocytes (Fig. 1.2.6). These forms are arrested in the G0 phase and only re-enter the cell cycle to produce gametes after being ingested by a mosquito (Fig. 1.2.7). In the mosquito midgut fertilisation of a female by a male gamete results in the formation of the zygote, the only diploid stage of an otherwise haploid parasite that develops into a motile and invasive ookinete. The ookinete crosses the midgut wall and forms an oocyst on the basolateral lamina (Fig. 1.2.8) where it will generate thousands of oocyst-derived sporozoites. When mature, these sporozoites migrate through the hemocoel to the mosquito's salivary glands (Fig. 1.2.9), making this mosquito infectious and hence completing the cycle (Fig 1.2.10) [17].

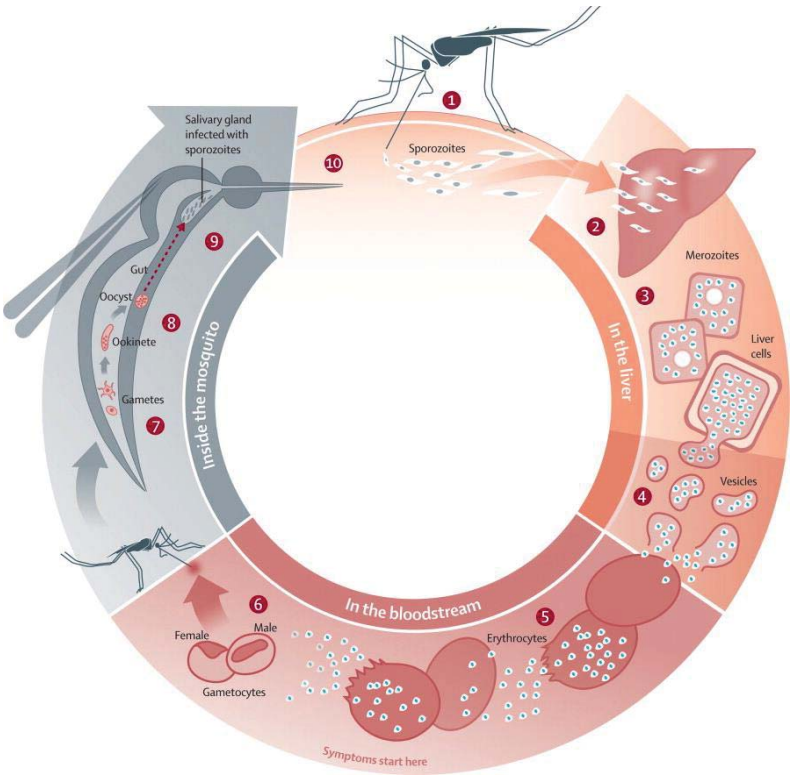


Fig. 1.2| Malaria life cycle.
The life cycle can be divided into three different stages: liver, blood and mosquito stages. The asexual cycle is initiated when the vertebrate host is bitten by an infected mosquito. The injected sporozoites migrate to the liver where they generate thousands of merozoites. After 2-16 days (63-72 hours in *P. berghei*, the rodent parasite in this study [18]) parasites leave the hepatocyte and invade erythrocytes, thus starting the symptomatic phase of the disease. Once gametocytes, the sexual precursors, are taken by another mosquito, fertilisation and mosquito colonisation occur. (Adapted from http://www.malariavaccine.org/malvac-lifecycle.php)

1.3 Next-generation sequencing technologies

DNA sequencing is the identification of the order of the four nucleotide bases adenine (A), guanine (G), cytosine (C), and thymine (T), in a molecule of DNA.

Until recently, DNA sequencing relied almost exclusively on Capillary/Sanger chemistry, a dideoxy chain termination method of sequencing [19,20]. Progress in technology across fields of microscopy, nucleotide chemistry, polymerase engineering, data storage and bioinformatics made next-generation sequencing (NGS) strategies possible.

Currently, the most commonly used NGS platforms are: 454 pyrosequencing (Roche/454 Life Sciences) [21], Illumina [22] and SOLiD (Applied Biosystems) [23]. These have produced an immense volume of accurate DNA sequence data at a fraction of the cost of the Sanger/Capillary method, which dramatically accelerated biological and biomedical research. Specifically, NGS have completely revolutionised several aspects of the field of genomics such as *de novo* genome sequencing, single-nucleotide polymorphism (SNP)

detection, chromatin immune-precipitation (ChIP-Seq) and transcriptome analysis (RNAseq). More recently, additional sequencing platforms such as Ion Torrent and Pacific Biosciences were introduced into the market [24].

NGS platforms differ in their sequencing biochemistry but their workflows are rather similar: DNA molecules are sheared into random short fragments which are then ligated *in vitro* to adaptor sequences at both ends to generate a so-called sequencing library.

A number of reviews have compared them to each other [25–28]. The advantages of NGS relative to Sanger sequencing are clear:

- *In vitro* construction and amplification of sequencing libraries;
- Higher degree of parallelism enabled by an array-based approach;
- Minimal volumes of reagents (picolitres or femtolitres) required since the array features are immobilized on a surface (flow cell).

Some disadvantages such as read-length and accuracy can nevertheless be listed; NGS reads are much shorter than Sanger sequencing and base-calls are, on average, at least tenfold less accurate. The latter is compensated by the huge number of reads generated, that together produce very accurate consensus sequences.

In this thesis I will focus solely on Illumina sequencing chemistry, as this was the only platform used for the present study.

1.3.1 Illumina sequencing

1.3.1.1 Illumina sequencing overview

The Illumina platform, sometimes still referred to as “the Solexa”, originated from work by Turcatti and colleagues [29,30]. It is optimised to generate large amounts of short DNA reads and is currently the cheapest sequencing technology per base of data. Read length has increased from 30 bp in 2008 to 300 bp in 2014, with the cost per bp also decreasing by several orders of magnitude in the same period. A recent study from the Sanger Institute estimated the error rate of Illumina reads to be below 0.4 % [27].

Table 1.1 summarises the main features of the different Illumina instruments currently on the market.

Table 1.1| Overview of current Illumina Instruments and their applications

	MiSeq	NextSeq 500	HiSeq 2500	Hi Seq X
Sequencing applications	Small genomes and amplicons	Genomes, exomes and transcriptomes	Production-scale genomes, exomes and transcriptomes	Population-scale human genomes
Output	0.3-15 Gb	20-39 Gb	10-180 Gb	1.6-1.8 Tb
Run time	5-55 hours	15-26 hours	7-40 hours	< 3 days
Reads per flow cell	25 Million	130 Million	300 Million	3 Billion
Maximum read length	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

1.3.1.2 Overview of library preparation procedures

Most library preparation protocols share the following workflow: DNA fragmentation, end repair, A-tailing and adaptor ligation, as illustrated in Figure 1.3. Illumina technology performs the best with DNA fragments that are 200-600 bp. Therefore long molecules of DNA need to be sheared prior to library preparation (except for some RNAseq protocols where the mRNA is sheared before reverse transcription [31]). This can be achieved by one of four different methods: enzymatic digestion, sonication (e.g. Covaris), nebulisation or hydrodynamic shearing (Fig.1.3, step 1). Next, end repair is used to generate blunt-ended, 5'-phosphorylated DNA ends compatible with the adaptor ligation strategy (Fig. 1.3, step 2). The enzymes involved in this step are T4 polynucleotide kinase and T4 DNA polymerase, both originally isolated from a bacteriophage. After end repair, fragments are "dA-tailed" by the Klenow fragment, a process by which a dAMP nucleotide is added onto the 3' end of blunted DNA fragments (Fig. 1.3, step 3). This step maximises ligation efficiency of adaptors carrying complementary dT-overhangs. Finally, T4 DNA ligase is used to catalyse the adaptor ligation step (Fig. 1.3, step 4). These are partially single stranded, forming a Y-shape, which allows each strand to have two different sequences added, one at each end (5' or 3') – crucial for the Illumina sequencing chemistry. A final clean-up step ensures removal of free library adaptors and adaptor dimers. This is critical as adapter-dimers are co-amplified with the adapter-ligated library fragments and reduce the sequencing capacity of the platform. After this step, the final libraries (Fig. 1.3, step 5) are quantified, usually by qPCR. Depending on their concentration they can either be directly used for sequencing, or amplified by PCR so that the desired concentration can be achieved.

Different libraries can be pooled and run together in the same lane in a process called multiplexing. This enables efficient use of the sequencing capacity of the instrument and only requires the incorporation of differently barcoded adaptors at the library preparation stage.

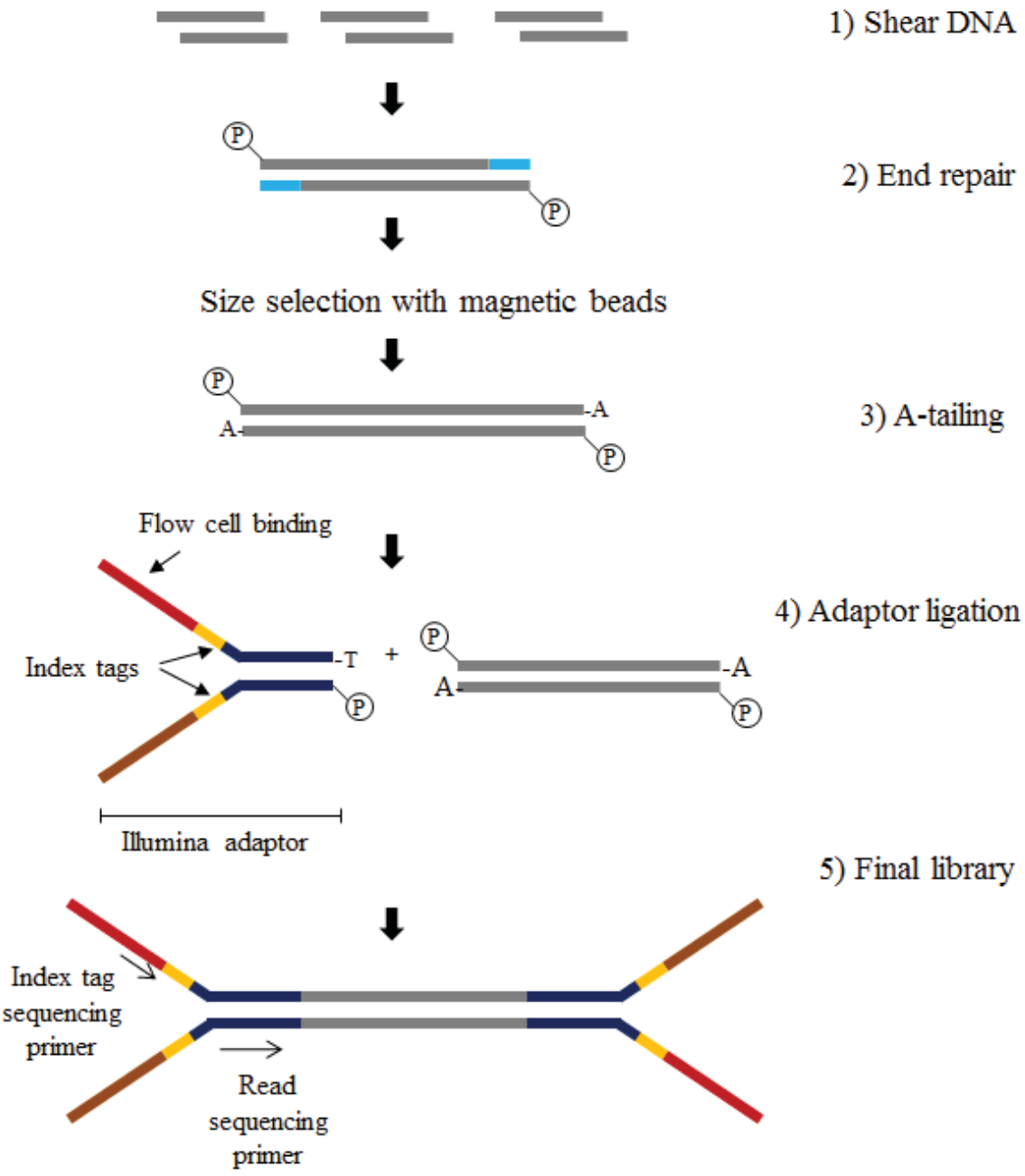


Fig. 1.3| Overview of a library preparation protocol.
Extracted gDNA is sheared in order to generate fragments of adequate size. Then end-repair and A-tailing follow. This enables the ligation of the Illumina adaptors. After this step the final libraries are ready to be quality controlled and sequenced.

1.3.1.3 Illumina sequencing chemistry

After careful quantification, a defined concentration (usually 1 to 4 nM) of each library is denatured and loaded into an Illumina flow cell. The latter is simply a glass surface coated with primers that are complementary to sequences present within the library adaptors. The Figure below illustrates, in nine steps, the Illumina sequencing workflow.

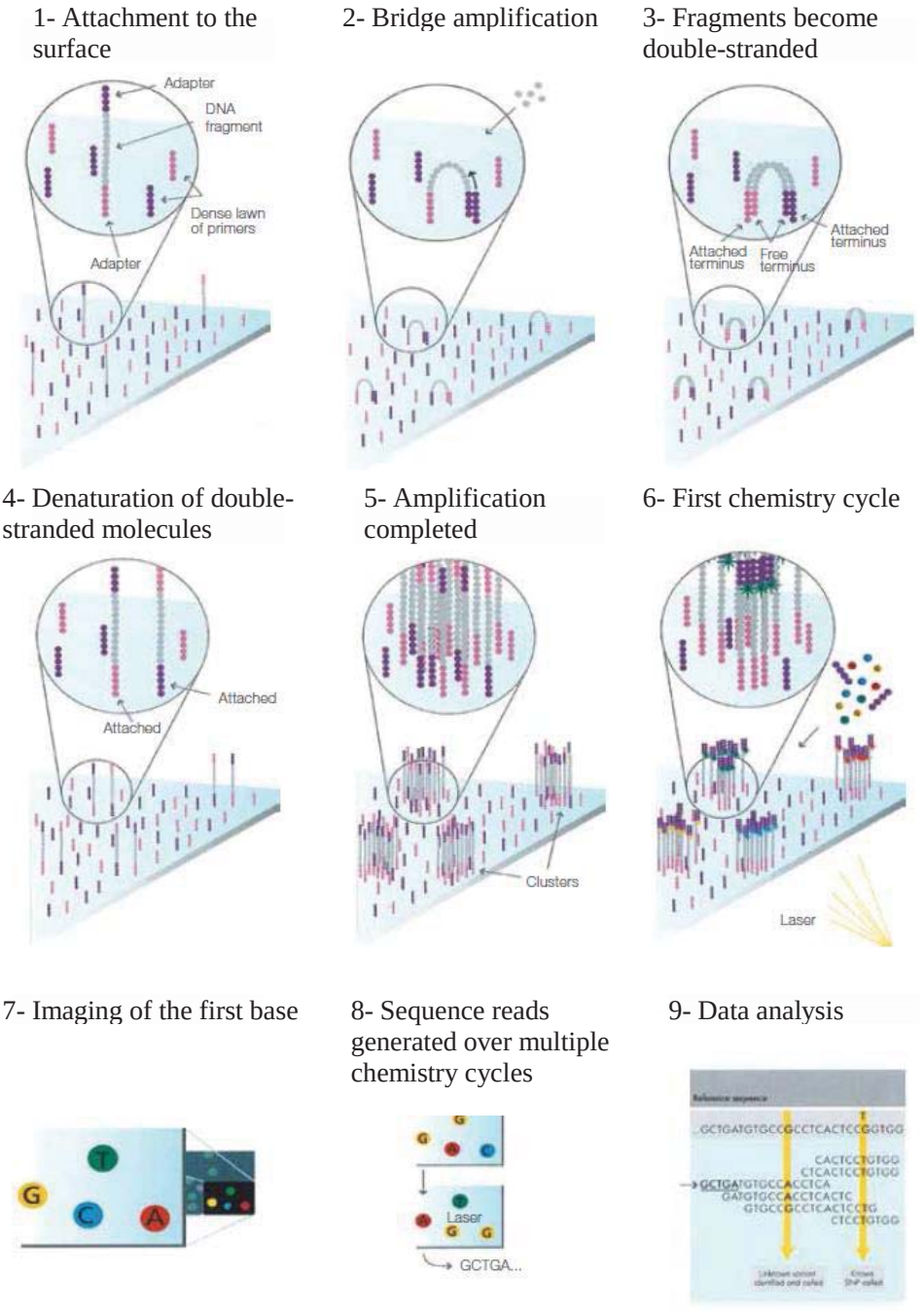


Fig. 1.4| Illumina sequencing chemistry.
1- Single stranded fragments randomly bind to complementary oligos on the surface of the flow cell. 2,3- unlabelled nucleotides and polymerase are added for solid-phase bridge amplification. At this stage the

fragments become double stranded. 4,5- After amplification is complete, a denaturation step results in millions of clusters of single-stranded molecules attached to the flow cell. 6- Determination of the first base; labelled reversible terminator nucleotides, primers and a polymerase are added to the flow cell thus initiating the first chemistry cycle. 7- The identity of the first base of each cluster is then recorded by imaging of the flow cell upon laser excitation. 8- After repeated cycles of chemistry the sequence of bases of a given library is determined. 9- Dedicated software converts the resulting image files into text files to allow data analysis by current bioinformatics tools. (Adapted from Illumina® Sequencing website)

After loading, the denatured, single-stranded library molecules are allowed to hybridise at one end with the primers on the flow cell (Fig. 1.4, step 1). This is followed by bridge amplification (Fig. 1.4, step 2), a process by which bound library molecules have their free adaptor end binding to a complementary primer sequence on the flow cell (forming a "bridge") and then act as template for the synthesis of the complementary strand of the duplex upon addition of the adequate reagents (Fig. 1.4, step 3). A subsequent denaturation step generates two single-stranded molecules that are both attached to the surface of the flow cell (Fig. 1.4, step 4). This "bridge amplification" is repeated until the desired cluster density is reached (Fig. 1.4, step 5).

After cluster formation sequencing is accomplished by SBS ("sequencing by synthesis") chemistry [22]. It is initiated with denaturation of the clusters and hybridisation of a sequencing primer complementary to the free end of the DNA molecules. Sequencing cycles follow, each consisting of the incorporation of a single modified nucleotide (Fig. 1.4, step 6) and subsequent high resolution imaging of the entire flow cell (Fig. 1.4, step 7). These nucleotides are modified in two ways: they are "reversible terminators" as they carry a chemically cleavable moiety at the 3' hydroxyl position that prevents incorporation of any other nucleotide in each cycle; and they have a chemically removable base-specific fluorophore that allows their identification by laser excitation in one of four different channels [30]. After imaging is complete, both the fluorescent dye and the terminator moiety are removed. Cycles of chemistry and imaging are repeated until the desired read length is reached (e.g. 150 bp). This is the end of a single-end read run. Paired-end sequencing requires subsequent removal of the first read products by denaturation followed by an equal number of cycles of SBS chemistry and imaging this time priming on the adaptor at the other end of the library. Positional information from the images of the flow cell links the two pieces of sequence data.

Computational analysis is then used to determine the base at each position (Fig. 1.4, step 8). In addition, a base-calling algorithm generates a quality value for each base call by quantifying the fluorescence signal from each cluster – any reads from mixed clusters are

filtered out. The percentage of reads that pass the "purity filter" constitute the purity filter (% PF) parameter that can be used for quality control purposes.

Mapping software can then be used to align the sequence data against a reference genome (Fig. 1.4, step 9). Reads produced by sequencing of genomic DNA are very useful to close gaps in genome annotations whereas reads from RNA-seq experiments provide information regarding gene expression and splicing.

1.4 Genetics of malaria parasites

The 24 Mb genome of *P. falciparum* was published in 2002, followed by rodent parasite species [32–35]. These data have shed light on the basic genome architecture and identified key structural elements, common metabolic and biosynthesis pathways and unique aspects that are shared among several *Plasmodium* parasites [34,36]. These genomes are organised in fourteen chromosomes and encode around 5400 genes, of which more than 60 % encode proteins with weak or no homology to other eukaryotes. Furthermore, their adenine and thymine (AT)-content is unusually high, i.e. close to 80 % in *P. falciparum*.

The human and rodent parasites share roughly 85 % of the genes at the level of content and order, and *P. falciparum*-specific genes that disrupt the conserved genome segments are predicted to play a role in host–parasite interactions [32–35]. This synteny strengthens the credibility of using the rodent model in many functional genomics studies [35].

1.5 Recombination in malaria parasites

Under ideal conditions, a DNA molecule could subsist for a maximum of 6.8 million years, although it would probably no longer be readable by polymerases after about 1.5 million years [37]. However, inside a living cell, DNA is constantly under the damaging effects of elements such as free radicals and other reactive species generated by metabolism, UV and ionizing radiation, including gamma rays and X-rays and its integrity relies on the constant action of repair systems. Additionally, normal progress of cell division and differentiation also generate DNA double-strand breaks (DSB) that might lead to transcriptional errors or chromosome rearrangements [38]. Two main repair mechanisms ensure repair of these breaks as faithfully as possible to prevent genome instability. These are non-homologous end joining (NHEJ) and homologous recombination (HR). The latter uses a

homologous template from which the lost DNA sequences are copied whereas in NHEJ the break is sealed in a template-independent fashion [39]. The choice of method varies with the organism – higher eukaryotes use predominantly NHEJ while prokaryotes prefer the high fidelity HR method [40,41]. A third method has recently been described named micro-homology-mediated end joining (MMEJ), where DNA breaks are repaired through re-connection of the broken ends at regions of micro-homology which are typically of 5 to 25 bp in length. This method always results in deletions and is highly associated with oncogenic chromosome rearrangements and genetic variation in humans [39].

In addition to the DSB caused by replication events *Plasmodium* DNA is continuously damaged by the immune response of the host [40]. Although haploid for most of their life cycle, malaria parasites are surprisingly thought to heavily rely on HR to repair DSBs. Accordingly, not only is the canonical machinery required for NHEJ, such as DNA ligase IV and KU 70/80 missing, but also *Plasmodium* parasites seem to be unable to close a linearized plasmid [32,40]. Given its role in closely related organisms like *Toxoplasma* parasites, it is thought that the loss of NHEJ pathways is a relatively recent event. This feature, i.e. predominance of HR repair mechanism, is particularly relevant for reverse genetic studies as it implies minimal off-target integration of targeting vectors.

In the absence of any template it has recently been shown that *Plasmodium* parasites are able to use a form of end-joining (EJ) repair method that resembles MMEJ and includes resection of single-stranded overhangs and insertion of templated short sequences at the site of the break [39]. This is consistent with the very limited set of products obtained in a study where DSBs were induced in unique areas of the genome [40]. This atypical EJ mechanism was however shown to be very infrequent and thus the prime mechanism of DSB repair in *Plasmodium* is resolutely HR [40].

1.6 The rodent model of malaria

Malaria research greatly relies on rodent parasite models, as maintaining the complete *P. falciparum* life cycle in the laboratory is still a difficult task. Rodents are the best non-primate models. Since their isolation from the wild between 1948 and 1974 in Central West Africa, rodent malaria parasites – *Plasmodium berghei* (used in this project), *P. chabaudi*, *P. vinckei* and *P. yoelii* – have allowed the study of several aspects of host-parasite-vector interactions, as we can easily access parasite stages from mosquitoes and from host's livers,

and evaluate potential interventions for malaria control [42,43]. Although none of the rodent *Plasmodium* species is the perfect model for *P. falciparum*, different species can be used to study different aspects of the infection. For instance, *P. chabaudi* is a model that best recapitulates the interplay between the parasite and the host immune response while *P. berghei* is a better model for studying the biology of transmission [42].

Apart from the fact that the whole parasite life cycle can be studied in laboratory conditions, rodent malaria parasites are more amenable to genetic modifications than *P. falciparum* [44]. In fact, it can take as little as two weeks to obtain a clonal population of *P. berghei* transgenic parasites, while for the human parasites this number is extended to at least two months. On the other hand, the availability of fluorescent rodent parasite lines and transgenic mice permits exploration of specific host parasite interactions that are not yet available for studies of human malaria [45].

1.7 Genetic engineering – new tools for reverse genetics

Our understanding of the molecular cues that drive *Plasmodium* development and differentiation has been hindered by the fact that only a small proportion of genes have been assigned functions experimentally. Gene function can be studied using forward or reverse genetics approaches. The former aims at the identification of gene(s) responsible for a particular phenotype. In malaria parasites such approaches are commonly mediated by transposon mutagenesis, a technology that has been used both in *P. berghei* [46] and in *P. falciparum* [47]. In contrast, reverse genetics involves alteration or disruption of a particular gene to enable the study of its biological function.

Reverse genetics is the favoured approach for analysing parasite gene functions. Usually, it comprises three main stages: assembly of the targeting vector, transfection of the parasites, and phenotypic analysis of the mutants. Several issues have delayed reverse genetics studies in malaria parasites. Although stable transfection technologies became available for *Plasmodium* parasites in the late 1990's [48–51] it was not until 2006 when efficient protocols were developed for *P. berghei*, the most genetically tractable of the malaria parasites [44]. The reason for this is the fact that *Plasmodium* parasites spend most of their life cycle intracellularly, inside a vacuole, with its genetic material enclosed within four membranes (i.e. erythrocyte membrane, parasitophorous vacuole, parasite membrane and nuclear envelope). Secondly, *Plasmodium* DNA is very AT-rich (~80 %) and unstable in *E. coli*, which makes preparation of targeting constructs difficult [32,33]. The reduced number

of selection markers has also been a limiting factor, especially for *P. berghei* studies where drug toxicity for the animals is an issue. Currently, the human dihydrofolate reductase (h*dhfr*) and the toxoplama *dhfr-ts* genes are used to select mutants with pyrimethamine. This drug inhibits DNA synthesis of the parasite by blocking the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, a crucial step for the biosynthesis of purines, pyrimidines, and certain amino acids [52]. The h*dhfr* also allows selection with the drug WR99210, but this requires subcutaneous injections unlike pyrimethamine which can be delivered orally. Until recently, the combination of both selection strategies was the only solution for performing a maximum of two sequential genetic modifications to *P. berghei* genomes, such as mutation of two non-consecutive genes or complementation of a KO. The recent development of a positive/negative recyclable selection cassette (*hdhfr/yfcu*) has made such studies more feasible**.** Negative selection is a process by which the loss of a marker is selected as its presence produces a toxic substance upon exposure to a suicide substrate (i.e. a prodrug). The system *hdhfr/yfcu* allows positive selection of mutant parasites with pyrimethamine and negative selection with 5-fluorocytosine (5-FC). After 5-FC treatment, this pharmacologically inactive compound is converted into the highly toxic form 5-fluorouracil (5-FU) by the enzyme coded by the *yfcu* gene – a bifunctional protein that combines yeast cytosine deaminase and uridyl phosphoribosyl transferase (UPRT)[53]. Exposure to this toxic metabolite forces mutant parasites to promote excision of the selection cassette through homologous recombination, facilitated by the presence of a directly repeated *Pbdhfr-ts* 3'UTR flanking this cassette [54].

Current technologies for the genetic manipulation of *Plasmodium* rely on the transfection of blood stages and subsequent drug selection, as described. Therefore, genes encoding products essential to blood stage development are not amenable to be disrupted in this manner. Conditional approaches are therefore required to study these genes.

RNA interference (RNAi) is widely used in model organisms [55,56] and parasites like trypanosomes [57], as a means to reversibly silence gene expression at the post-transcriptional level. However, *Plasmodium* lacks the enzymes required for RNAi-based approaches [58]. As a result, although less efficient, other conditional strategies have been developed such as the Tet-inducible system [59,60] and the destabilisation fusion domains strategy (for *P. falciparum* only) [61]. Conditional approaches for the study of essential genes in mosquito or liver stages are less complex and strategies like the Flp system and promoter swap can easily disrupt or silence genes beyond the blood stages [62–64].

So far, reverse genetics has only targeted 10 % of the *P. berghei* genes (http://www.pberghei.eu/). In order to increase the number of genes to which functions have been assigned experimentally it is necessary to move from a gene-by-gene approach towards high-throughput approaches.

1.7.1 The Gateway technology: DNA cloning using site-specific recombination

The bacteriophage lambda (λ) uses a site-specific recombination system when switching between the lytic and lysogenic pathways to promote its integration into the *E. coli* chromosome [65]. This recombination system has two major components: recombination sequences (*att* sites) and a set of proteins that mediate the recombination reaction (Integrase (Int), Integration host factor (IHF) and excisionase (Xis)). The lysogenic pathway is catalysed by Int and the *E. coli* IHF whereas the lytic pathway is mediated by the phage Int and Xis and the bacterial IHF. It is highly specific and conservative (i.e. no net gain or loss of nucleotides).

Hartley and colleagues exploited this system to devise a strategy by which these recombinases mediate the transfer of DNA fragments flanked by *att* sites into vectors also containing *att* sites [66]. This system has been commercialised by Invitrogen (currently Life technologies) since the late 1990s under the name of Gateway Technology ® (GW). It is used for the cloning and transferring of DNA fragments between different expression vectors in a high-throughput fashion while maintaining orientation and the reading frame of the fragment of interest [67].

This system carries out two reactions:

(1) *att*B × *att*P → *att*L + *att*R mediated by Int and IHF

(2) *att*L × *att*R → *att*B + *att*P mediated by Int, IHF and Xis.

The direction of the reactions is simply controlled by the enzyme cocktail and the available *att* sites. Each of these reaction mixes is commercially available as “BP clonase” (1) and “LR Clonase” (2). The original *att* sites have been mutated to ensure directionality and irreversibility of the *in vitro* *att*L x *att*R reaction. Specifically, *att*L1 sites react only with *att*R1 sites and *att*L2 sites react only with *att*R2 sites. GW technology was highly relevant for this project as it was part of a restriction enzyme-free cloning strategy to generate targeting vectors.

1.7.2 Recombineering, a homologous recombination based cloning strategy

Bacteriophages have been studied extensively as a means to understand the principles of homologous recombination. This process can be defined as a type of exchange of genetic material in which information is exchanged between two identical or nearly identical molecules of DNA, in a precise and accurate fashion. Recently, some of these phage recombination systems were explored as tools for genetic engineering of plasmids due to their precision and simplicity. In fact, their use at scale has greatly boosted functional genomics studies in model organisms [68].

The use of phage homologous recombination systems to carry out genetic engineering is termed recombineering [68]. This is a highly efficient recombination system that enables *in vivo* modification and subcloning of large fragments of DNA, without the need for restriction enzymes or error-prone PCR amplification [69]. Initially developed based on two proteins RecE/RecT from the Rac prophage, the system has also been developed for the analogous *red* operon of the λ phage (used in this project) [70]. The *red* operon includes three elements:

- Redα, a 5' to 3' exonuclease (functional equivalent of RecE);
- Redβ, an annealing protein (functionally equivalent of RecT);
- Redγ, an inhibitor of the major *E. coli* exonuclease and recombination complex (RecBCD, responsible for degradation of linear dsDNA).

The process is initiated after a double strand break, at which point Redα digests one of the DNA strands leaving the other strand as a 3' ended, single-stranded, DNA overhang. Then Redβ binds and coats the single strand and this complex aligns with the homologous DNA so that the 3' end can become a primer for DNA repair (Fig. 1.5). This is further assisted by Redγ, which inhibits the RecBCD exonuclease activity of *E. coli*. These enzymes are extremely efficient requiring only a minimal length of homology region of 42 bp to initiate recombination [71].

In practice, generating a targeting vector using recombineering involves two steps:

1) Flank the engineered DNA (e.g. resistance cassette, epitope tag) with short sequences homologous to the target where it is aimed to integrate. Due to their short size, the homology arms can be synthesised as primer overhangs and used to amplify the engineered DNA;

2) Transform the resulting PCR product into a bacterial host that carries both the Red system and the target DNA (either as an episome or within the chromosome). Provided that all elements have been provided, the recombinases promote the integration of the engineered

DNA into its target location and the final vector can easily be recovered by standard genomic DNA (gDNA) extraction.

In short, using recombineering we now can generate recombinant molecules without the need for unique or special sites, with greater precision and regardless of the size of the target molecules thus rendering traditional cloning methods obsolete for most applications.

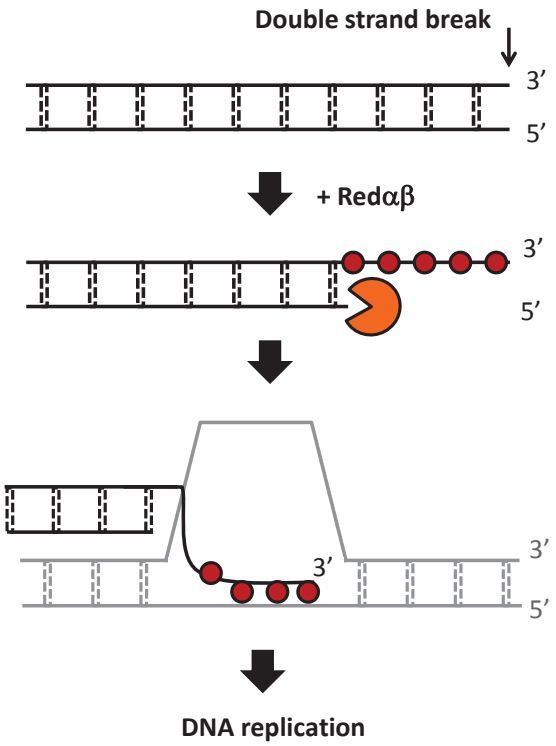


Fig. 1.5| Mechanism of Red recombination.
Upon a double strand break the 5'-3' exonuclease Redα (orange Pacman) digests one of the strands while Redβ (red circles) coats the resulting single strand. The homology regions from the engineered DNA (grey strand) are then used as templates for DNA repair.

1.7.3 Recombineering in *P. berghei* – the *Plasmo*GEM project

The recent generation of a *P. berghei* genomic DNA library using a low copy linear plasmid based on the bacteriophage N15 (pJAZZ, Lucigen) has now enabled the recombineering technology to be transferred to *Plasmodium* biology [72]. Currently, this library has 9113 clones with an average length of 9 kilobases (kb), and covers ~94 % of *P. berghei* genes. The high number of clones enables high coverage of most genes, for instance, 50 % of the genome is represented (at least partially) by at least five clones.

Together this library and the recombineering technology have set the basis for the development of the *Plasmodium* genetic modification project, *Plasmo*GEM. Launched in 2012 the *Plasmo*GEM project is based at the Wellcome Trust Sanger Institute (http://plasmogem.sanger.ac.uk/) and aims at producing a free community resource of genome-wide sets of genetic modification vectors for *P. berghei*. These vectors are generated through a highly efficient two-step recombineering pipeline, that has been scaled up to a 96-well plate format [73]. A brief description of each of the steps follows:

Step one: The gene of interest (GOI) present in a selected library clone is replaced *in vivo* by a positive/negative bacterial selection cassette *zeo-pheS* through recombineering (Fig. 1.6). This stage takes advantage of the short homology arms required to initiate recombination that are therefore synthetised as primer overhangs and used to amplify this cassette. At this point the Red/ET system is provided *in trans* as an inducible polycistronic unit in a temperature-sensitive, low copy number plasmid (pSC101-repA-BAD-gbaA). The pBAD promoter ensures that transcription of the recombinases (pSC101-repA-BAD-**gba**A) is activated only in the presence of arabinose, remaining otherwise inactive by action of the transcriptional repressor AraC. Two other layers of control have been added to the system: low copy number and thermo-sensitivity of the plasmid pSC101. The former is ensured by the origin of replication oriR101 and thermo-sensitivity is conferred by the protein RepA (pSC101-**repA**-BAD-gbaA), which is required for partitioning of plasmids to daughter cells at division [74]. This strategy therefore reduces the risk of undesired recombination events occurring as well as avoiding the recombineering plasmid to become a contaminant upon DNA extraction of the final product (*Plasmodium* vector) [68,75,76]. Finally, due to its role in general cellular integrity, recA (pSC101-repA-BAD-gba**A**) is also carried by the plasmid as the host cells are recA deficient in order to prevent recombination of the library clones [76]. After step one, the recombineered product is selected with the antibiotic zeocin, a broad-spectrum agent that induces double strand breaks of the DNA.

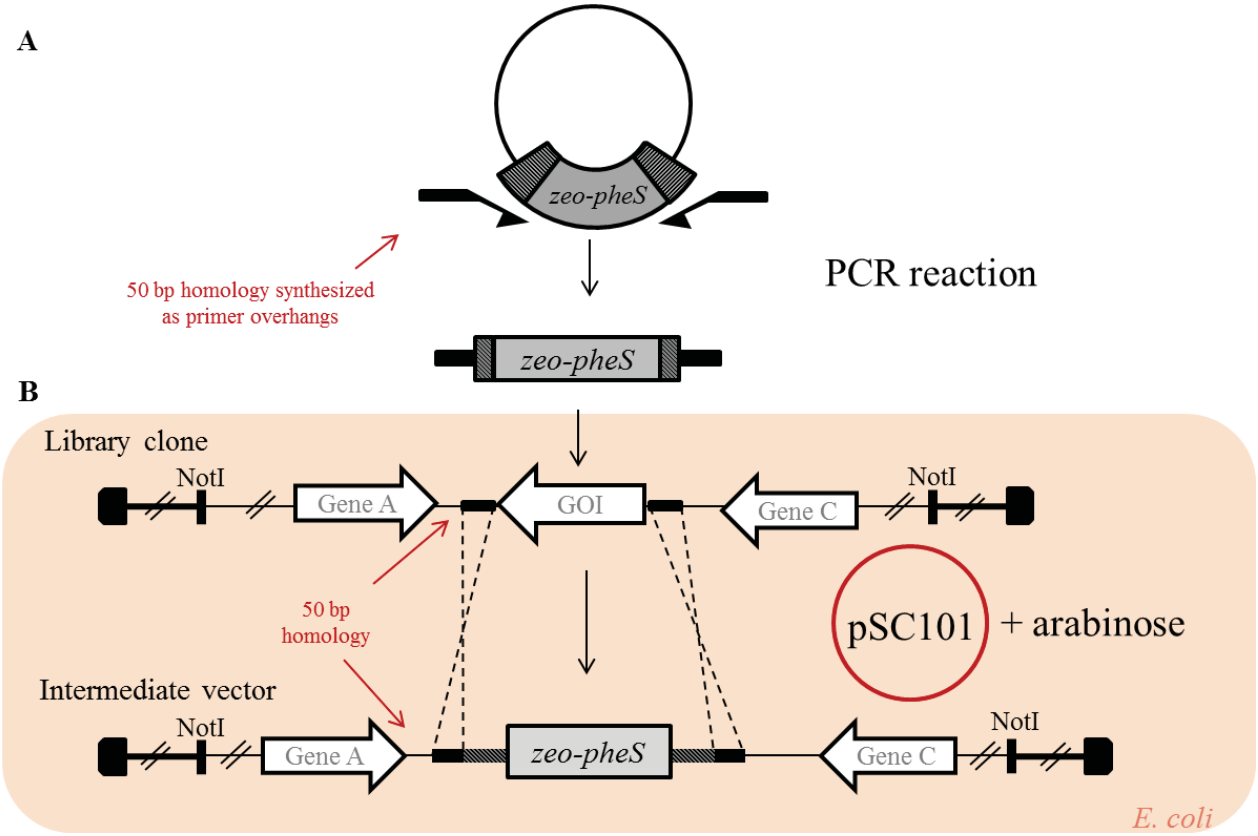


Fig. 1.6| First step – Recombineering reaction to generate a KO vector.
(A) A plasmid carrying the *zeo-pheS* cassette is used as template for a PCR reaction where the primers contain overhangs of 50 bp homologous to the flanking regions of the target gene. (B) This amplicon is electroporated into bacteria containing both the pSC101 plasmid and the library clone containing the target gene. Induction of the recombineering enzymes is accomplished with arabinose. Selection with zeocin is used to select the recombineered products. To generate a tagging vector, the approach is the same but the homology regions flank the stop codon in order to remove it.

Step two: the modified library clone is used as a substrate for an *in vitro att*L x *att*R GW reaction that replaces the *zeo-pheS* cassette by a parasite recyclable positive/negative cassette (*hdhfr-yfcu*) (Fig. 1.7) [54,73]. Negative selection against the *zeo-pheS* intermediate is then used to select the correct and final product – the *pheS* allele encodes a phenylalanyl-tRNA synthase *alpha* subunit that enables incorporation of the toxic phenylalanine analog p-chlorophenylalanine.

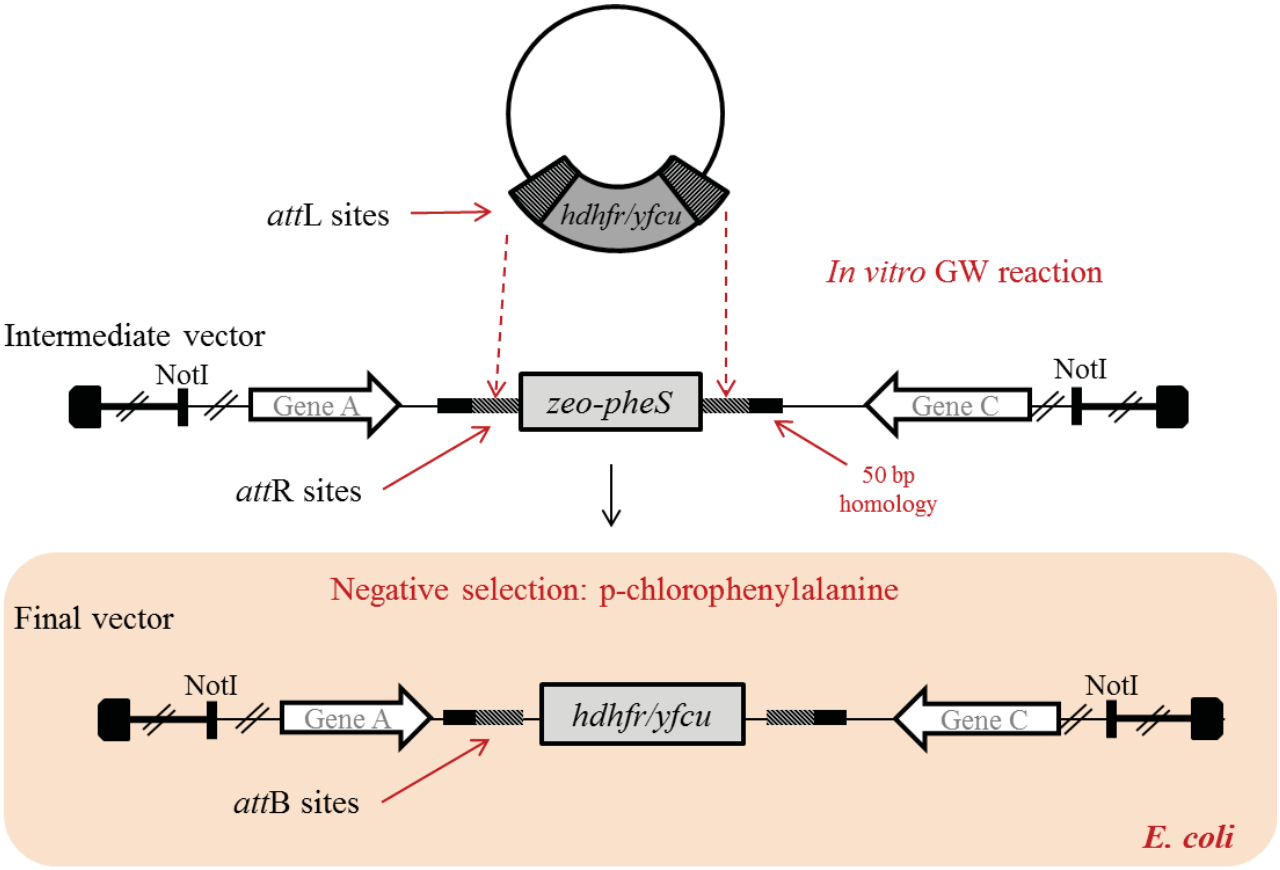


Fig. 1.7| Second step - Replacement of *zeo-pheS* cassette with *P. berghei* selection marker.
Gateway technology is used to replace the bacteria positive/negative selection cassette by the parasite selection marker. The reaction mix is then electroporated and the correct final product is isolated by negative selection.

This new approach has greatly reduced the time necessary for generation of large numbers of targeting vectors as its 96-well format has enabled the production of over 55 *Plasmo*GEM vectors in a period of less than two weeks, that only require *Not*I digestion to release the pJAZZ vector arms prior to transfection. Furthermore, as these vectors are generated from library clones they contain homology arms that are several kb long as opposed to the traditional 0.4-1 kb, which has greatly enhanced not only transfection efficiency but also transfection reproducibility [44,73,77]. Also, as these vectors are never circular, they are not expected to be kept in the cytosol as episomes.

1.8 High throughput reverse genetic screens

High throughput reverse genetics screenings have been vital for the understanding of model organisms at the molecular level. For instance, in *Saccharomyces cerevisiae* a library of a near-complete (96 %) collection of gene-deletion (KO) mutants has been generated [78], and around 9000 genes have already been targeted in highly germline-competent C57BL/6N mouse embryonic stem cells [79].

1.8.1 Signature tagged mutagenesis (STM)

Phenotypic analysis of mutants is the most laborious stage of any genetics screen, especially when *in vivo* infection models are required. Optimisation of plate-based bioassays has been accomplished, but only certain phenotypes can be analysed in this fashion, and although automated, each mutant is still screened individually. An efficient option to perform phenotypic analysis of large numbers of mutants is to analyse them in pools. Genetic footprinting was devised as a means to identify mutants within pools [80]. It involves the production of a large pool of mutants by transposon mediated mutagenesis, which will randomly include the disruption of the genes of interest, followed by inoculation of the pool under specific test conditions and the subsequent identification of the mutants that are present in the pool at the end of the experiment, by PCR. Some mutants, although present in the input inoculum will have been outgrown during the growth step due to the disruption of critical genes for the microbe's development. If, however, no detectable differences are observed between the pattern of PCR products before and after incubation it is likely that the disrupted gene is dispensable for survival under the tested conditions [80]. Genetic footprinting was first used to identify genes essential for the viability of *S. cerevisiae* under different growth conditions however, its low throughput encouraged the development of an alternative that would enable large-scale, parallel analysis of mutants [81]. This alternative was termed signature tagged mutagenesis (STM) and was presented by David Holden and co-workers in 1995 [82]. It is currently considered one of the most powerful and versatile large-scale genetic approaches to identify virulence determinants based on negative selection of attenuated mutants (i.e. mutants, which have lost the capacity to survive in a given host) [83]. Although very similar to genetic footprinting in its principle, the greatest advantage of STM is the ability to identify each mutant within a pool, not by individual PCRs, but by the presence of a

signature tag (i.e. unique short sequence of DNA) that is introduced into the microorganism upon integration of a transposon or targeting vector.

STM was first used to characterise virulence genes in *Salmonella typhimurium*. In this first report, a library of mutants was generated with a set of tagged transposons and then used to infect a rodent model of typhoid fever. Once infection was established, bacteria were recovered from the infected spleens and their tags were compared to the input pool using a dot-blot hybridisation approach. The tags included a 40 bp variable (and unique) central region that was flanked by constant annealing sites that also contained restriction sites. Such arrangement enabled not only the amplification of each tag by PCR and its subsequent cloning into the transposon, but also tag radio-labelling of the resulting pool prior to the hybridisation step [82].

STM has since been successfully applied to a wide range of microorganisms and yielded the identification of hundreds of new genes involved in virulence [84–89]. However, despite the preservation of the basic design (i.e. generation of tagged mutants followed by their propagation in pools and identification of the mutants present in the final pool through their tag), the variety of organisms to which this technique was applied led to the diversification of the mutagenesis method employed. For instance, *Neisseria meningitides* bacteria are not amenable to transposon mediated mutagenesis, therefore *in vitro* mutagenesis was used to generate mutants [90]. A summary of the mutagenesis methods is presented in Figure 1.8. These include *in vivo* and *in vitro* transposon mediated mutagenesis, shuttle mutagenesis, insertion–duplication mutagenesis by homologous recombination and gene replacement by homologous recombination [91].

The tag detection method was also optimised according to the available technologies and organisms. Hybridisation methods similar to the one used by Hensel *et al*. and variants that use digoxigenin or biotin labelled probes had as disadvantage the possibility of cross-hybridisation of probes that could lead to false positives. Although the construction of longer tags and pre-screen steps could help prevent such situations, alternative methods namely PCR-based approaches were developed. These included standard PCR reactions that relied on tag-specific primers and a flanking generic primer, less laborious multiplexed PCRs and also real-time PCR for a quantitative measurement of the abundance of each population of tags [91,92]. The yeast functional genetics field greatly benefited from STM. In 1996, shortly after Hensel's report, Shoemaker and co-workers implemented a new approach that enabled thousands of sequences to be analysed in parallel using high-density oligonucleotide arrays to detect the signature tags, termed barcodes, as they were gene specific [78]. This approach

allowed the measurement of the fitness cost of each gene deletion as differences in the intensities of the hybridisation signals reflected differences in the relative abundance of each population of mutants [93,94]. The libraries of barcoded mutants were generated by a gene-replacement strategy during which two 20-mer DNA barcodes – UPTAG and DOWNTAG, instead of one, were inserted for greater results confidence [94]. Despite being slower, the directed gene-replacement approach enabled prioritisation of targets and also made matching the presence of the barcode with the disrupted gene more straightforward [93]. Furthermore, the availability of libraries of single mutants simplified the validation of phenotypes seen in the pools as the original mutants were readily available.

Recently, NGS technologies were introduced as a means of barcode detection in STM experiments. This method, termed barcode analysis by sequencing or "Bar-seq" has been particularly relevant for the yeast functional genomics field and was shown to outperform the microarray detection methods (i.e. high-density oligonucleotide arrays) in terms of sensitivity, dynamic range, and limits of detection [95]. The Bar-seq strategy enables quantitative analysis of complex pools as the frequency at which a barcode is detected in the sequencing data is a measurement of the abundance of a given population of mutants within a pool [95,96].

In summary, STM is a very powerful and versatile technique that has revolutionised the functional genomics field as it can be applied to both *in vitro* and *in vivo* situations for a myriad of screens that include reverse and forward genetics and chemogenomic assays.

Mutagenesis method	
In vivo transposition	*Salmonella typhimurium, Mycobacterium tuberculosis, Vibrio cholerae, Yersinia enterocolitica, Legionella pneumophila, Brucella suis, Escherichia coli*
In vitro transposition	*Streptococcus pneumoniae, Neisseria meningitidis, Helicobacter pylori*
Shuttle mutagenesis	*Neisseria meningitidis*
Insertion-duplication mutagenesis	*Streptococcus pneumoniae*
Gene-replacement by homologous recombination	*Saccharomyces cerevisae*

In vivo transposition	1- Generation of a pool of tagged transposons. 2- Induce transposition into the organism of interest (OI). 3- Select clones with integrated barcoded transposons.
In vitro transposition	1- Transposition takes place *in vitro* (transposon activity independent of host factors) between DNA library of the OI and tagged transposons 2- Transformation of the OI with the mutagenized library. 3- Mutants are generated by homologous recombination.
Shuttle mutagenesis	1- Generation of a DNA library of the OI in a shuttle vector. 2- Transposition of barcoded transposons into this library takes place *in vivo* in a recipient organism. 3- Selected clones with the mutagenized library are introduced into the OI. 4- Mutants are generated by homologous recombination.
Insertion-duplication mutagenesis	1- Generation of targeting vectors by cloning DNA fragments from the OI. (synthetized by random or directed PCR) into barcoded vector. 2- Insertion of these vectors into the OI. 3- Mutants are generated by homologous recombination.
Gene-replacement by homologous recombination	1- Generation of a targeting vector that contains sequences that are homologous to the flanking regions of the gene of interest (GOI). 2- The vector should also contain a selection marker and a barcode. 3- Introduction of the vector into the target organism. 4- Homologous recombination generates a barcoded, resistant mutant.

Fig. 1.8| Summary of mutagenesis methods used for the generation of pools of mutants.
Top panel shows to which microorganisms the different methods were applied. Bottom set includes a brief description of each method. Adapted from ref [91].

1.8.2 Epistasis and genetic interactions

Genetic interactions can be described as biological phenomena that take place when the effect of a mutation depends on the genetic context in which it occurs. The same definition is given to the term epistasis as used in the population genetics field, although the classical definition by William Bateson restricts epistasis to a genetic interaction in which one mutation masks or suppresses the effects of another allele at another locus [97–99].

When a double mutant shows an expected, multiplicative phenotype compared to the corresponding single mutants it is very likely that these two genes do not interact either as parts of the same pathway or between pathways [100]. Genetic interactions can be classified as positive/alleviating or negative/aggravating (Fig. 1.9) [100,101]. The former refers to interactions where the simultaneous disruption of two genes yields a phenotype that is less severe than the phenotype expected from the sum of each independent mutation. Conversely, in a negative interaction the combined phenotype is more severe than expected, and in the most extreme cases (synthetic sick/lethal) the double mutation is lethal, unlike the single mutants [101]. A particular case of alleviating interaction is suppression. In suppression, the simultaneous perturbation (i.e. mutation or deletion) of two genes yields a wild type phenotype, despite the fact that each corresponding single mutant has an evident fitness loss. This is the case when one mutation counteracts the effects of another and is frequently associated with genes within the same pathway that also interact at the protein level [98]. Genetic interactions tend to occur among functionally related genes, although interactions of essential genes correspond to a broader functional range [100]. Although genetic interactions overlap with protein-protein interactions more often than expected by chance, such overlap is relatively rare, occurring at a frequency of less than 1% [102].

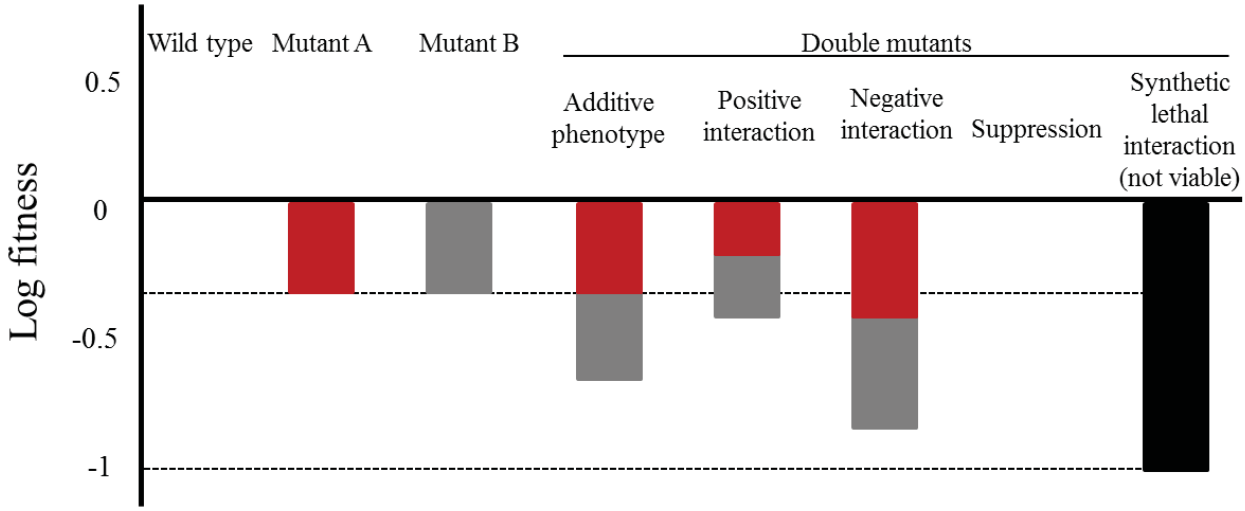


Fig. 1.9| Summary of genetic interactions.
The additive phenotype reflects the absence of genetic interactions. Positive or negative interactions take place when the phenotype of the double mutant is less or more severe than expected, respectively. Suppression happens when a second mutation counteracts the effect of a first mutation leading to a neutral phenotype. Synthetic lethal interaction is the most extreme case of a negative interaction and this double mutant is not viable.

Many recent insights into genetic interactions and networks have emerged from studies using *S. cerevisae.* Interestingly, under normal growth conditions, up to 80 % of *S. cerevisae* genes proved not to be essential for development, thus suggesting the presence of a high degree of interacting/compensatory pathways [94,103]. The availability of a collection of more than 6000 barcoded KO mutants allowed the test of 5.4 million gene-gene pairs for genetic interactions using a synthetic genetic array analysis (SGA) to screen the double mutants [104].

The highly complex life cycle of *Plasmodium* parasites suggests the existence of intricate signalling pathways. A systematic analysis of the *P. berghei* kinome suggested that only less than 35 % of the protein kinases were redundant for development unlike what was seen for the fission yeast (*Schizosaccharomyces pombe*) where 83 % of the protein kinases were amenable to deletion [105,106]. However, the scarcity of selectable markers available for *P. berghei* genetics has delayed genetic interaction studies as they require sequential modifications of the genome. In fact, only a very limited number of publications where sequential gene deletions were presented is available to date [107,108] and none involved signalling genes such as kinases, which due to their pivotal role in development and survival have been extensively targeted for genetic interaction studies in model organisms [109].

1.9 Protein kinases

A set of protein kinases was used throughout this dissertation to develop the screening method. For this reason a brief description of this family of proteins follows.

1.9.1 Eukaryotic protein kinases

Protein phosphorylation is the process by which kinases catalyse the transfer of phosphate groups from ATP to specific residues on their target proteins. It is a major regulatory mechanism that controls a myriad of cellular processes and is estimated to affect 30 % of the yeast proteome [110]. Protein kinases are one of the largest protein families, accounting for approximately 2 % of eukaryotic genomes [111]. For instance, the yeast genome encodes 127 protein kinases, while in humans the number increases to more than 500 [111,112]. The rapid and reversible nature of phosphorylation allows tight regulation of protein activity, localisation, stability, conformation and/or interaction with other proteins. Kinases themselves can be regulated in this fashion [113]. Kinase dysregulation is associated with a range of diseases, including vascular diseases, inflammatory disorders and cancers. For this reason kinases have been pursued as potential drug targets for the past three decades [114].

The catalytic domain of eukaryotic protein kinases (ePK) is characterized by highly conserved amino acids distributed in 11 subdomains (Fig. 1.10).

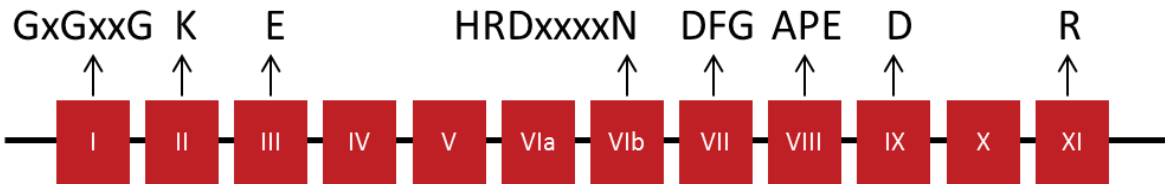


Fig. 1.10| Structure of ePKs catalytic domain.
The positions of amino-acid residues and motifs highly conserved throughout the ePK superfamily are indicated above the subdomains, using the single-letter amino-acid code with x being any amino-acid. The three glycine residues (GxGxxG) in subdomain I form a hairpin enclosing part of the ATP molecule; a lysine (K) in subdomain II, orientates the ATP molecules; a glutamate (E) in subdomain III forms a salt bridge with the former residue; in subdomain VIb aspartate (D) is thought to be the catalytic residue acting as a base acceptor; the aspartate in the DFG motif of subdomain VII, binds to the cation (Mg^{2+} or Mn^{2+}) associated with ATP; the glutamate (E) in subdomain VIII forms a salt bond with the arginine (R) in subdomain XI and provides structural stability of the C-terminal lobe; the aspartate in subdomain IX is involved in structural stability of the catalytic loop of subdomain VI through hydrogen bonding with the backbone. (Adapted from refs [115,116].)

According to their primary structure ePKs can be classified into seven major groups: ACG, CMGC, CaMK, CK1, STE, TKL and TyrK, that reflect broad functional categories [115,116]. Briefly:

1) **AGC**: includes the cyclic-nucleotide and calcium/phospholipid dependent kinases. E.g. PKA (cyclic-adenosine-monophosphate-dependent protein kinase), PKG (cyclic-guanosine-monophosphate-dependent protein kinase), PKC (protein kinase C) and related proteins;
2) **CMGC:** includes CDK (cyclin-dependent kinases), regulators of cell cycle progression, MAPK (mitogen-activated protein kinases), signal transducers that control effectors of cell cycle control and transcription, GSK3 (glycogen synthase kinase 3), also major regulators of cell proliferation and CLKs (CDK-like kinases) which play roles in RNA metabolism;
3) **CamK**: calcium/calmodulin-dependent kinases;
4) **CK1**: casein-kinase 1;
5) **STE**: includes PKs acting as regulators of MAPKs (STE stands for "sterile" and refers to the fact that these enzymes were first identified in a genetic screen of sterile yeast mutants);
6) **TKL:** tyrosine-kinase-like: related to tyrosine kinases but they are serine-threonine protein kinases;
7) **TyrK**: tyrosine kinases.

Atypical protein kinases (aPK) feature limited or no sequence similarity with ePKs however, they have demonstrated kinase catalytic activity experimentally. These can be divided in four groups: Alpha (e.g. myosin heavy chain kinase of *Dictyostelium*), PDHK (pyruvate dehydrogenase kinases), PIKK (phosphatidyl-inositol 3' kinase-related kinases) and RIO ('right open reading frame', as it was one of two adjacent genes that were found to be transcribed from a bidirectional promoter) [117,118].

A third group termed 'Other protein kinases' (OPK) allocates some ePKs that, although exhibiting some degree of sequence similarity to the main ePK groups, cannot be classified into any such group [117].

1.9.2 Protein kinases in *Plasmodium* parasites

Several independent groups have tried to assemble the kinomes (i.e. complete set of protein kinases) of *Plasmodium* parasites bioinformatically and according to the algorithm

used and the reference genomes available at the time, each study yielded slightly different results in terms of total numbers and classification [117–120]. The most recent analysis of *Plasmodium* was published by Miranda-Saavedra and co-workers and it compared twelve apicomplexan species using a validated computational tool, called Kinomer [117]. This kind of study relies on domain signature modelling of the conserved 11 subdomain structure of an ePK. Results showed that *Plasmodium* parasites have relatively small kinomes that represent between 1.2 % (*P. berghei* and *P. chabaudi*) and 1.6 % (*P. falciparum*) of their genomes [117].

The analysis of *P. falciparum* kinome yielded a total of 89 PKs with 65 ePKs, 19 FIKKs and 5 aPKs. Most of these enzymes were assigned to almost all major groups but some do not cluster with any group and some groups are not represented [117]. For instance, the CMGC group, which in other organisms gathers kinases involved in cell development and proliferation, contains the highest number of *Plasmodium* kinases (21 in *P. berghei* and 22 in *P. falciparum*). This is probably a requirement to meet the successive cycles of proliferation undergone as part of the parasitic life cycle. Examples include the CDKs, PfMAP-1/2, PfGSK3, PfCRK-1/3/4/5. The second most abundant group is the CamK (17 in *P. berghei* and 19 in *P. falciparum*) which evidences the importance of calcium signalling in *Plasmodium* development. Interestingly the CDPKs (calcium-dependent protein kinases) are part of this group which is only present in plants and other Apicomplexans – this topic is further developed in section 1.9.4. The CK1 group is the least represented group as in both *P. berghei* and *P. falciparum* only one *ck1* orthologue is found, unlike what has been found in other organisms where this group is expanded (e.g. 83/438 in *C. elegans* [121]). No *Plasmodium* PK clustered with either the STE or the TyrK groups. An important consequence of the former is the apparent absence of the canonical ERK1/2 pathways, despite the presence of MAP kinases homologues. In insects, plants, worms and mammals the TyrK group members are associated with hormone-response pathways. This suggests that this group of kinases arose as an adaptation to multicelullarity and therefore justifies its absence in both *Plasmodium* parasites and yeast [118]. Moreover, different mass spectrometry-based phosphoproteomic studies have confirmed that *P. falciparum* protein phosphorylation is involved in processes such as invasion and cytoadhesion, and also cell cycle control, DNA replication, transcription and translation [117,122].

Two main reasons for the seemingly low number of kinases found in *Plasmodium* genomes might be gene loss as an adaptation to the parasitic life cycle and the presence of *Plasmodium*-specific kinases that are too divergent to be found using the current methods that

are based on sequence similarity. Indeed, a set of *P. falciparum* PKs, named FIKK after their conserved four-residue motif in the kinase subdomain II (Phe-Ile-Lys-Lys) formed a tight cluster that despite containing most of the ePK conserved amino acids in the catalytic domain are not clearly related to any known ePK group [117,118]. So far, this family of kinases is restricted to Apicomplexa and is present as a single copy in *Plasmodium* parasites, except for *P. falciparum* and its closest relative *P. reichenowi* where it is expanded with 19 and six members, respectively [117,123]. All *P. falciparum* FIKK kinases have a variable extension N-terminal to their catalytic domain, which contain a PEXEL export signal motifs (required for exportation outside the parasitophorous vacuole), thus suggesting a role in host-parasite interaction for these enzymes [123].

Plasmodium parasites have also a small number of aPKs of the RIO and PIKK families. RIO proteins are widespread with at least two such enzymes present in organisms from Archaea to humans. In yeast these have been shown to be involved in rRNA processing and are essential for cell viability [117,124]. Similarly, both *P. falciparum* and *P. berghei* genomes have two orthologues and available data suggests that they are essential for the development of these parasites [105,125]. Enzymes belonging to the group PIKK (one in *P. berghei* and three in *P. falciparum*) are responsible for sensing DNA damage, nutrient-dependent signalling and nonsense-mediated RNA decay [117,126].

Altogether these studies showed that important divergences exist between malaria parasites and other eukaryotes such as the absence of the canonical ERK1/2 pathway components or the presence of Apicomplexa specific families of kinases. These differences mirror the obvious phylogenetic distance between Apicomplexa and Opisthokonta (lineage that includes *Homo sapiens*) and motivates the search for specific inhibitors as antimalarial drugs.

1.9.3 MAP kinases in *Plasmodium*

MAP kinases are serine/threonine PK responsible for regulation of cellular processes like mitosis, differentiation and cell survival that are activated through extracellular stimuli such as mitogens, osmotic stress or proinflammatory cytokines [127]. Two genes encoding atypical members of the MAP family, *map1* and *map2*, have been identified in both *P. falciparum* and *P. berghei*. However, while MAP1 seems to be related to the mammalian ERK7/8, MAP2, despite clearly belonging to the MAPK family, is somewhat divergent as it possesses a TSH motif in its activation loop motif instead of the canonical TxY (present in

MAP1) [119,128]. As previously mentioned, the STE group which includes a variety of kinases participating in MAPK signalling cascades is absent from the *Plasmodium* kinome. However, features of *Plasmodium* MAP kinases hint a different strategy for their activation. In the case of PfMAP-2, it has not only a TSH activation loop motif but also an insertion of about 26 amino acids at the N-terminal end. In addition, it seems to be phosphorylated and activated by the kinase PfNEK-1 which is not a member of the STE family but in this case acts as a MAPKK (MAPK kinase) equivalent. Moreover, the fact that all of these unusual features are present in other Apicomplexan parasites suggests a unique MAPK signalling mechanism [119].

Some functional differences have been observed between the MAP kinases of the rodent and the human parasites. While *Pfmap2* is likely to be essential for the erythrocytic cycle, *Pbmap2* can be deleted [129,130]. In the mutant, microgamete exflagellation is prevented, and hence parasite sexual reproduction and parasite transmission to the mosquito [130]. In contrast, the genetic inactivation of *map1* does not generate obvious phenotypic effects in either *P. berghei* or *P. falciparum* [129]. However, *Pfmap*2 is over-expressed in blood stages of the *Pfmap1* mutant [129] and *Pbmap2* transcripts are increased in liver stages of the *Pbmap1* mutant (Heussler, personal communication) suggesting that in both species the two MAP kinases may have partly overlapping functions and can partially complement each other's roles in development.

1.9.4 Calcium responsive kinases in *Plasmodium*

Signal transduction pathways allow cells to sense the environment and respond accordingly. This is a two-step process in which a first messenger (extracellular) binds a surface receptor followed by a second messenger that, within the cytoplasm, acts serving as chemical relay from the plasma membrane to the cytoplasm. Examples of second messengers are cyclic nucleotides (cGMP, cAMP) and calcium (Ca^{2+}). The latter is a major regulator of calcium-dependent protein kinases (CDPKs). These are abundant in plants (e.g. *Arabidopsis* has over 40 CDPKs) where they control a wide variety of processes including transcription and metabolism and also ion pumps and channels and the cytoskeleton [131]. Also present in apicomplexans (but not in animals), CDPKs seem to have a monophyletic origin as suggested by phylogenetic studies [132]. A CDPK contains three functional domains: a protein kinase catalytic domain, a carboxyl-terminal calmodulin-like domain with (usually) four EF-hands as

Ca^{2+} binding sites, and a junction domain between the kinase and the calmodulin-like domain [133].

In *Plasmodium* parasites calcium has been shown experimentally to regulate biological processes such as erythrocyte invasion by merozoites, motility and invasion by ookinetes and sporozoites, discharge of secretory organelles, and sexual differentiation in the mosquito vector [134–138]. In *P. falciparum* seven CDPKs have been identified (CDPK1-7) and found to be crucial for the development of the parasite. PfCDPK1 has been linked to motility and cell invasion mediated by microneme discharge and is thought to be essential for the intra-erythrocytic stages given the unsuccessful attempts to generate a KO mutant [139,140]. PfCDPK2 and PfCDPK3 are also likely to be essential during blood stage development although these have not been as explored [141]. Both PfCDPK3 and PfCDPK4 are thought to be involved in gametogenesis, while PfCDPK5 has been shown to have an essential role in the parasite egress from the erythrocytes [142–144]. Only PfCDPK4, PfCDPK6 and PfCDPK7 have been considered dispensable for parasite development at the blood stage [145].

The rodent parasites *P. berghei* have only six CDPKs, CDPK2 is missing, and unlike the human parasite four are not essential for completion of the intra-erythrocytic phase. The essential kinases include PbCDPK5 and PbCDPK7 and their function in the rodent parasites has not yet been elucidated [105]. Until recently, PbCDPK1 was considered to be essential during the asexual blood stages, therefore its function was best studied in the sexual stages where it was shown to activate translation of a subset of translationally repressed transcripts in the developing zygote stage and, its absence led to an arrest before the parasites could fully reach the ookinete stage [62,105]. Recent studies that include this project have, however, successfully deleted PbCDPK1 without an asexual growth phenotype despite its implication in invasion of *P. falciparum* parasites [146]. A function in ookinete's ability to invade the midgut epithelium has been proposed for PbCDPK3. Despite being able to glide through the blood meal in the mosquito gut the KO mutants are not capable of traversing the peritrophic membrane (i.e. protective layer formed from 12 h after a blood meal that protects the mosquito from pathogens) to reach the epithelial cells of the midgut [136]. Also not required for asexual stages, PbCDPK4 has a crucial role during male gametogenesis. Upon contact with xanthurenic acid (XA) in the mosquito gut, PbCDPK4 responds to a rapid increase in cytosolic calcium that occurs in gametocytes activating cell cycle progression, and promoting their differentiation into male gametes [138]. Finally, the absence of PbCDPK6 not only impairs sporozoite production but also renders the few produced sporozoites less infective to

hepatocytes likely due to a defect on the switch to the invasive phenotype, required for establishment of the liver infection [147].

Some CDPKs have their subcellular localisation determined by N-terminal acylation, which is thought to contribute to the specificity of such a small number of enzymes towards a ubiquitous secondary messenger [112,148]. Furthermore, the relatively small number of calcium effectors and the somewhat overlapping patterns of expression of the CDPKs at certain points suggest that not only do these enzymes interact with each other and perhaps other effectors but they might also have overlapping functions, thus conferring some redundancy to the system.

1.9.5 Defining the *Plasmodium* phospho-proteome

Fine-tuning of the intracellular machinery is achieved by several mechanisms that include transcriptional control, post-transcriptional control and post-translational modifications (PTMs) of proteins. PTMs are modifications of specific residues of proteins that often are reversible. Phosphorylation is one such alteration.

Advances in mass spectrometry based proteomic techniques have recently generated snap-shots of the phosphorylation status of organisms like *E.coli* [149], yeast [150] and even mice [151]. Since the apicomplexan parasites are quite divergent from most model organisms, dissecting *Plasmodium* signalling pathways is often a challenge as this cannot be fully achieved by bioinformatics predictions that use kinome and phospho-proteome data from other species [118]. For this reason in-depth phospho-proteome analyses of *Plasmodium* parasites have recently been performed [122,145,152,153]. Most phospho-proteome data has been produced through similar protocols and focused on the schizont stage as it is easily accessible in quantities that enable for mass-spectrometry based phospho-proteomics. Gene ontology (GO) analysis revealed that in schizonts the pospho-proteome is enriched for regulatory biological processes related to the basic transcription, translation and metabolic machinery which is in agreement with the schizogony process that these parasite undertake. Included in this dataset were at least 42 protein kinases [152].

The Treeck study [153] performed a comparison between the phospho-proteomes of *Toxoplasma gondii* and *P. falciparum* and showed that in these two related organisms the number of proteins implicated in secretory pathways that were phosphorylated suggests that these parasites might use phosphorylation as a means of regulating protein function outside their own boundaries. In *P. falciparum* parasites these included proteins located on the

parasitophorous vacuole and merozoite surface proteins such as MSP1 and MSP7. Secretion of the exonemes is required for merozoites egress in a process known to involve at least two kinases, CDPK5 and PKG. The latter was shown to be phosphorylated in its activation loop in a fashion consistent with autophosphorylation, thus, implying that PKG is not only regulated by changes in cGMP levels but also by phosphorylation. This phenomenon was shown to be present in at least 22 other protein kinases within the activation loop of the kinase, within one of the eleven kinase domains (e.g. CDPK1), or even outside the kinase domain (e.g. CDPK6)[145]. This suggests that the malaria protein kinases are organised in cascades, where the elements at the top are responsible for the phosphorylation, and therefore regulation, of their downstream partners.

Although we are still in the early stages of malaria phospho-proteomics research it is clear that phosphorylation is involved in most aspects of the parasite's biology and should thus be investigated to generate novel targets for pharmacological intervention.

1.10 Project aims

The major goal of this dissertation aims to overcome some of the limitations that currently prevent efficient reverse genetic screens in *Plasmodium* parasites. The specific aims of my work are:

(1) to explore the use of barcoded vectors with long homology arms as a means to carry out signature tagged mutagenesis in *P. berghei*;

(2) to develop barcode counting as a method for phenotyping the resulting complex mixtures of genetically modified parasites *in vivo*;

(3) to critically evaluate these new techniques by comparing their performance to conventional approaches using the parasite's protein kinases as a test case;

(4) to discover new kinase signalling pathways by conducting the first genetic interaction screen in *Plasmodium*.

The development of better therapies against malaria requires a deeper understanding of *Plasmodium* biology. Despite the establishment of methods for targeted genetic modification nearly two decades ago [51], only 10 % of the genes were assigned functions experimentally. As nearly half of the genome lacks any annotation, gene-by-gene strategies need to be scaled-up.

STM has been used extensively as a high-throughput method to identify microbial virulence genes by parallel phenotyping of pools of individually barcoded mutants [91]. These approaches encompass three different stages: (1) generation of pools of barcoded mutants, (2) propagation of the pool, and (3) detection of the mutants present in the final pool through their barcodes. The low transfection efficiency and high rate of false positives have prevented the development of such strategies for *Plasmodium* parasites. The recently developed *Plasmo*GEM vectors are linear and have improved integration efficiency due to long homology arms.

The first aim of this project was therefore to ask if STM technology could now be adapted to the rodent malaria parasite, *P. berghei*, and to use it to perform high throughput reverse genetic screens.

Chapter 3 describes the optimisation of the protocol. This involved maximisation of transfection efficiency, development of a parallel transfection strategy of pools of barcoded vectors and finally the development of a barcode sequencing approach, to read and count the barcodes of the pools of mutants generated. Chapter 4 presents the validation of the method

where a set of protein kinase genes was analysed by STM. In this chapter I also show how the barcode counting strategy enabled the analysis of the fitness of different mutants present in a pool.

As the STM technology enabled the accurate and reproducible measurement of fitness costs of single mutants growing within pools, another aim of this project was to detect genetic interactions through the analysis of the fitness of double mutants. Chapter 5 describes firstly how the presence of a recyclable selection cassette in each of the *Plasmo*GEM vectors enabled the generation of six selection marker free KO lines, and secondly, how these lines were used to perform genetic interaction screens. The latter involved measuring the fitness costs of the double mutants and comparing them to the fitness cost of the corresponding single mutants according to a multiplicative model for epistasis.

Chapter 2

Materials and Methods

All reagents were purchased from Sigma-Aldrich unless stated otherwise.

2.1 Parasitology

2.1.1 Rodents

The following rodents were used in this study:

Strain	Description
Theiler's Original (TO) mice	For routine work. Outbred, 6-8 weeks of age, male and female.
BALB/c mice	Used as recipients of STM experiments. Inbred, 6-8 weeks of age, male and female.
Wistar rats	Used to propagate parasites to be transfected in STM experiments. Outbred, 200 g – 250 g (~ 8weeks), female.

Animals were provided by Harlan Laboratories, UK. All animal research was conducted under licences issued by the UK Home Office using protocols reviewed by the ethics committee of the Wellcome Trust Sanger Institute.

2.1.2 Parasite lines

The following *P. berghei* parasite lines were used in this dissertation.

Line	Description
ANKA 2.34	WT line. For routine work and used as background for generation of transgenic parasites and as control for phenotypic analysis.
Clone RMgm-7	Selectable marker–free reporter strain expressing the mu3 variant of green fluorescent protein (GFP). This was generated in the *P. berghei* ANKA cl15cy1 background. Also used as background for generation of transgenic parasites and as control for phenotypic analysis.
gsk3 KO	Validation of STM results.
cdpk1 KO	Validation of STM results and used as background in the genetics interaction screen.
cdpk3 KO	Used as background in the genetics interaction screen.
cdpk4 KO	Used as background in the genetics interaction screen.
cdpk6 KO	Used as background in the genetics interaction screen.
pkg^{T619Q}	Used as background in the genetics interaction screen. (Provided by M.

	Brochet [154])
rio1 KO	Validation of STM results.
rio2 KO	Validation of STM results.
PBANKA_082960 KO	Validation of STM results.
tkl3 KO	Validation of STM results.

2.1.3 Parasite maintenance

All parasite lines were propagated in the above mentioned rodents. These were infected either by intraperitoneal (i.p.) or intravenous (i.v.) injection of *P. berghei* infected blood, using 30-gauge (G) needles. The latter route was administered in the tail vein and only used to infect the transfection recipients. Infections were monitored on thin blood smears made from one drop (~3 µL) of tail blood and stained with 10 % Giemsa stain solution modified diluted in water. Parasitaemia was counted under 1000x magnification and was expressed as the percentage of infected RBC per total number of RBC.

Parasite harvesting was always a terminal procedure. For that the rodents were anesthetised by i.p. injection of 10 mL/Kg of a solution containing 12 % ketamine and 0.16 % xylazine in phosphate-buffered saline (PBS). Blood was collected by cardiac puncture using syringes pre-loaded with 100 µL of heparin (30 U/mL).

2.1.4 Parasite cloning by limiting dilution

Limiting dilution was used to clone transgenic parasite populations. To this end, a TO mouse was infected with the mutant line. When parasitaemia reached 0.3-1 %, the mouse was sacrificed and the blood was diluted in RPMI1640 (Gibco) by successive dilutions to a final concentration of 1.5 parasites/100 µL. Next, 200 µL of this suspension, i.e. three parasites, were injected i.p. into each of a set of ten mice. Under *in vivo* conditions this procedure usually results in two to three positive infections. Parasitaemia was checked on day 9 post-infection and parasites were genotyped by PCR.

2.1.5 Selection marker recycling

All targeting vectors contained a recyclable positive/negative selection cassette (*hdhfr-yfcu*) [54]. After dilution cloning the selection marker was removed by negatively selecting

parasites with 5-FC (1 mg/mL) in drinking water, for seven days, from day 1 post-infection onwards. In the presence of negative selection, only parasites that lose the selection cassette as a result of homologous recombination between the direct repeats flanking the fusion gene are able to survive. On day 7-10 post-drug treatment the parasites obtained were genotyped and re-cloned.

2.1.6 Parasite phenotyping

2.1.6.1 Asexual growth curves of *P. berghei* parasites.

This method was used to perform pairwise comparisons of growth curves without using a barcode strategy. For each mutant, three mice were infected with 10^6 parasites i.p. Blood smears were taken daily to monitor parasitaemia. A total of 1500 RBCs were counted per replicate, per time-point to calculate growth curves for each mutant. These experiments were performed under a "severe" protocol according to our animal project licence.

2.1.6.2 Exflagellation assay

Approximately 5 µL of blood was taken by tail bleed and mixed with ookinete media (section 2.1.6.3.1). After 5 min, 10 µL of this mix was loaded on a haemocytometer, and the number of exflagellating microgametocytes per 1000 RBCs was counted. A minimum of five fields per replicate were counted.

2.1.6.3 Ookinete phenotyping

2.1.6.3.1 Ookinete culture

In order to induce reticulocytosis and hence gametocyte production [155], mice were injected i.p. with 0.2 mL phenylhydrazine (6 mg/mL) three days prior to infection with *P. berghei*. On day 4 post-infection, infected blood was collected and re-suspended in ookinete medium: RPMI-1640 Medium HEPES modification, with L-glutamine and 25 mM HEPES, without sodium bicarbonate (Gibco); 50 mg/L hypoxanthine; 50 units/mL of Pen/Strep; 100 µM xanthurenic acid in 6 mM $NaHCO_3$; filter-sterilised; final pH 7.4. Just before the blood was added, the medium was supplemented with 20 % FBS (gibco). The culture was incubated for 18 h at 19 °C. On the following day, cultures were checked for the presence of ookinetes on Giemsa stained smears.

2.1.6.3.2 Ookinete conversion rate calculation

To calculate the ookinete conversion rate, 18-20 hour ookinete cultures were sampled (50-100 µL) and spun for 1 min at 500 g. The pellets were stained with 13.1 antibody, an anti-P28 Cy3-labeled monoclonal antibody [156,157], diluted 1:500 in PBS, at room temperature for 10 min. The presence of 0.4 % Hoechst (bisBenzimide H33342 trihydrochloride) enabled the detection of the nuclei but this was not essential. Parasites were checked immediately on a Leica DM2500B and the ratio ookinetes : activated macrogametocytes, i.e. banana shaped : round forms, was calculated. Images were processed using ImageJ software v.1.44o.

2.1.6.3.3 Ookinete purification with magnetic beads

The ookinete cultures were centrifuged for 5 min at 500 g. Then, pellets were re-suspended in 8 mL of ookinete medium, in 15 mL Falcon tubes, and incubated with 5 µL of 1:50 dilution of an anti-P28 antibody coated with magnetic beads, for 5 min, on a rotating mixer. Next, tubes were placed on a magnetic Dynarack (Dynal Biotech) for 2 min or until the solution was clear. The culture medium was then removed and, without removing the tube from the stand, PBS was added to wash the bound ookinetes. After this, the tubes were removed from the magnetic stand and ookinetes were re-suspended in an appropriate volume of PBS.

2.1.6.4 Oocyst dissection and counts

Twelve days after infection, mosquitoes were anesthetised (10 min, at -20°C) and midguts were dissected out into PBS, using fine point dissection forceps and a 30 G needle. Dissections were carried out in sets of 10 midguts per strain, which were then covered with a coverslip. The number of oocysts present in each midgut was counted under phase contrast.

2.1.6.5 Sporozoite dissection and counts

Twenty-one days after infection, mosquitoes were anesthetised (10 min, at -20 °C) and salivary glands were dissected out into PBS, using fine point dissection forceps and a 30 G needle. Dissections were carried out in sets of 10 mosquitoes per strain. In the event that individual glands were damaged or lost during dissection, an equivalent number of extra mosquitoes would be dissected. Glands were transferred into a 1.5 mL tube and homogenised

with a micro-pestle in 100 µL. Next, the number of sporozoites was counted on a haemocytometer. For this a sample of 5 µL of the homogenate was diluted in 5 µL of PBS.

2.2 Generation of targeting vectors

2.2.1 Single tube protocol

All vectors not obtained from the *Plasmo*GEM resource (19 out of 54) were assembled using the small scale tube protocol as previously described [73]. Briefly, the *Pb*G library clone was chosen by the *Plasmo*GEM database as the optimal clone containing the gene of interest (GOI). This was amplified and electroporated (Bio-Rad Gene Pulser Xcell; parameters: 1800 V, 10 µF, 600 Ω) with the recombineering plasmid (pSC101-BAD-gbaA-tet). The transformed bacteria were incubated overnight at 30˚C in 4 mL of TB medium, with 30 µg/mL of kanamycin and 5 µg/mL of tetracycline. On the next day, this culture was diluted in fresh medium to an optical density (OD600) of 0.05. When the OD 600 reached 0.3-0.4, expression of the recombineering proteins was induced by addition of L-arabinose (0.2% w/v) and by a temperature switch to 37 ˚C. This allowed efficient expression of the proteins and concomitant loss of the plasmid which contained a thermo-sensitive origin of replication. After 40 min of induction, the cells were washed three times with ice-cold ultrapure water (3 min, 5000 g) and electroporated with 1 µg of a PCR product containing a *zeo-pheS* cassette flanked by Gateway attR1-attR2 sites [79]. This amplicon was generated with primers that included the 50 bp located immediately upstream and downstream of the start and stop codons, respectively. The sequence of these primers was provided by the *Plasmo*GEM database and the cycling programme was as follows: 95 °C 5’ // 95 °C 30’’ / 50 °C 30’’ / 68 °C 1’ (x30) // 68 °C 10’ //. The resulting culture was allowed to recover in 4 mL of TB medium for 70 min before zeocin (50 µg/mL) and kanamycin (30 µg/ mL) were added. This was incubated overnight at 37 ˚C. On the following morning, the recombineered library clone (i.e intermediate vector) was extracted using a mini-prep kit (Qiagen) according to the manufacturer’s instructions. This was followed by dialysis as high purity is crucial for the success of the gateway step. In the second step of this protocol, an *in vitro* Gateway reaction was used to replace the *zeo-pheS* cassette by the *Plasmodium* drug resistance cassette. The Gateway reaction was set up in 20 µL using 2 µL of LR clonase (Invitrogen), 100 ng Gateway Entry plasmid (pR6K-attL1-3xHA-hdhfr-yfcu-attL2), LR clonase buffer and 300 ng purified intermediate vector. The final vector was transformed into electrocompetent *E. coli* TSA and

plated on YEG-Cl agar with kanamycin (30 µg/ mL). Colonies were verified by PCR to ensure the presence of the cassette. Four of them were sequenced together with the vectors produced by the resource, to check the integrity of the homology arms and barcode sequences. Prior to transfection, constructs were digested with *Not*I to release the bacterial regulatory regions.

2.2.2 Vectors provided by the *Plasmo*GEM resource

The *Plasmo*GEM vectors were generated by recombinase mediated engineering in continuous liquid culture on 96-well plates as previously described [73], with the following modifications:

At the first step, the PCR product consisting of a *zeo-pheS* marker and 50 bp primer extensions homologous to the GOI was purified using the High Pure 96 UF Cleanup Kit (Roche) to improve Lambda Red recombination efficiency in *E. coli*. The resulting intermediate vectors were selected in liquid culture containing 30 µg/mL kanamycin and 50 µg/mL zeocin. After two rounds of antibiotic selection cultures were re-inoculated into fresh selective medium and incubated for no longer than 16 hours at 37 °C. Intermediate vector DNA was obtained using a Qiagen Plasmid Plus 96 Miniprep kit and eluted into 20 µL Tris-EDTA buffer.

At the second step, the Gateway reaction was set up in 20 µL using 2 µL of LR clonase (Invitrogen), 100 ng Gateway Entry plasmid (e.g. pR6K-attL1-3xHA-hdhfr-yfcu-attL2), LR clonase buffer and 300 ng purified intermediate vector. Gateway reactions were purified using the High Pure 96 UF Cleanup Kit. Electrocompetent *E. coli* TSA were transformed, plated on YEG-Cl agar and four colonies picked to verify the sequence of their homology arms.

2.3 Generation of mutant *P. berghei* parasites

2.3.1 Generation of single mutants

Single mutants were generated using a standard transfection protocol [44]. Briefly, one donor mouse was infected and monitored until blood parasitaemia reached 1-3 %. At this point the infected blood was harvested and used to prepare an overnight culture of schizonts. Each millilitre of blood was diluted in 50 mL of schizont medium (RPMI 1640 medium containing 25 mM glutamine and 25 mM HEPES (Gibco), 10 mM $NaHCO_3$, 100 U/mL

penicillin/streptomycin (Gibco) and supplemented with 25 % FBS), gassed for 90 seconds with malaria gas (1 % O_2, 3 % CO_2, 96 % N_2) and incubated at 36.5 °C with gentle shaking.

On the following morning mature schizonts were purified on a 55 % Nycodenz/PBS (v/v) cushion with low acceleration and brakes, for 20 min. The brown layer of schizonts was collected with a Pasteur pipette and washed in schizont medium (500 g, 2 min). The pellet of schizonts was then mixed with the transfection reagent (16 µL of P3 primary cell) and 2 µg of vector DNA (~5 µL), and electroporated using the 4D Nucleofector System (Lonza) and the pulse programme FI-115. Electroporated parasites were immediately injected i.v. into 6-8 week-old TO mice. Drug selection was initiated on the following day with pyrimethamine (70 mg/L in drinking water). Parasitaemia was monitored daily for five days from day 5 post-transfection.

In positive infections blood was harvested when it reached > 2 % parasitaemia and parasite gDNA was genotyped (sections 2.4.2, 2.4.3.2, 2.4.4).

*Plasmo*GEM identification numbers for vectors used in these experiments are listed in appendix I.

2.3.2 STM protocols

2.3.2.1 Parallel transfection

STM transfections were performed by electroporation of purified schizonts largely as described in section 2.3.1, with the following modifications: parasites for the schizont culture were propagated in female Wistar rats to achieve maximal transfection efficiency. Purified schizonts were washed in complete medium, mixed with 16 µL of the P3 transfection reagent (Lonza) and then with the pool of targeting vectors. This transfection mix was electroporated using the 4D Nucleofector System (Lonza) in 16-well strips according to the pulse program FI-115. DNA pools contained 100 ng of each of the *Plasmo*GEM vectors. Prior to transfection, the pool was digested with *Not*I to release the bacterial vector arms, precipitated with isopropanol and re-suspended in a volume not exceeding 6 µL.

Electroporated parasites were immediately injected i.v. into 6-8 week-old Balb/c inbred mice. Three batches of schizonts were transfected with the same vector pool. Resistant parasites were selected by pyrimethamine (70 mg/L in the drinking water). Infections were monitored daily by counting Giemsa stained thin blood films under 1000 x magnification.

*Plasmo*GEM identification numbers for vectors used in these experiments are listed in appendix I.

2.3.2.2 Collection and processing of STM time-points.

A small sample of blood (≤ 30 μL) was collected at exactly the same time each day from the tail vein on days 4, 5, 6, 7 and 8 post-transfection, in 200 μL of PBS. Blood was lysed by adding 1 mL of 1:10 dilution in dH_2O of pre-chilled lysis buffer (0.15 M NH_4Cl, 0.01 M $KHCO_3$, 1 mM Na_2EDTA; pH 7.4), followed by centrifugation for 3 min at 1000 g. Next, gDNA was extracted using the phenol-chloroform method described in section 2.4.3.1 and re-suspended in 50 μL of water (days 4, 5 and 6) or in 100 μL (days 7 and 8). Giemsa stained thin blood smears were performed daily to monitor infection.

2.4 DNA preparation and genotyping methods

2.4.1 White blood cell (WBC) removal

Collected blood was diluted 1:5 in PBS and then applied to a cellulose powder column previously equilibrated with PBS. More PBS was added to assist the blood moving through the column. The flow-through was collected into a 50 mL Falcon tube and then spun for 10 min at 1000 g. Supernatant was discarded and the parasite pellet was re-suspended in an appropriate volume that depended on the downstream application.

2.4.2 Blood lysis

After blood collection by tail bleed or cardiac puncture, RBCs were lysed using lysis buffer (0.15 M NH_4Cl, 0.01 M $KHCO_3$, 1 mM Na_2EDTA; pH 7.4). Blood was diluted in lysis solution at a 1:10 ratio in dH_2O and incubated on ice for 5 min. Then, the suspension was centrifuged at 1000 g and supernatant was discarded. The black pellet of parasites was re-suspended either in PBS or in TNE buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM EDTA) depending on the extraction method chosen.

2.4.3 DNA extraction methods

The extraction method was chosen according to the table below.

Table 2.1| Choice of DNA extraction method depending on starting material.

Stage	Method	White blood cell removal
Blood for PCR genotyping	DNeasy Blood & Tissue Kit (Qiagen)	No
Blood for Southern blot/ WGS/ or from STM time-points	Phenol-chloroform	Yes

2.4.3.1 Phenol-chloroform extraction of blood samples

Lysed parasite pellets were re-suspended in 600 µL of TNE buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM EDTA) containing 1 % SDS and 3 µL of RNAse A (20 mg/mL) and incubated at 37 °C, for 10 min. Next, 10 µL of proteinase K (20 mg/mL) were added and the mix was incubated at 37 °C for 45 min. At this point, 800 µL of buffered phenol:chloroform:isoamylalcohol (25:24:1) were added and samples were inverted several times and centrifuged at 10000 g for 5 min, at room temperature. The resulting aqueous upper phase was transferred to a new tube and 800 µL of chloroform:isoamylalcohol (24:1) were added, mixed and centrifuged in the same way. The new aqueous upper phase was transferred to a new tube and DNA was precipitated overnight at -20 ˚C with isopropanol at a v/v ratio of 1:1. On the following day, DNA was pelleted by centrifugation, air-dried and re-suspended in 50-100 µL of DNAse-free water.

2.4.3.2 DNeasy Blood & Tissue Kit

Blood was lysed according to section 2.4.2 and re-suspended in 200 µL of PBS. Then samples were processed according to the manufacturer's instruction and re-suspended in 100 µL of EB buffer.

2.4.4 PCR genotyping

Parasites were genotyped to check for correct integration of the *Plasmo*GEM vectors. Unless stated otherwise in the Figure, seven PCRs were performed on gDNA, for each line: (1) a positive control that targeted *rna polymerase II* (primers arg84/arg85); (2) a positive control for presence of the selection cassette (primers arg80/arg81); (3) 5' end of the WT locus; (4) 3' of the WT locus; (5) 5' end of the modified locus; (6) the 3' end of the modified locus, and (7) a long-range integration PCR. For reaction seven, a gene-specific primer annealed to the chromosome just outside of the vector's homology arm and paired with a primer annealing to the selection cassette within the targeting vector (either arg216 or arg218, depending on the orientation of the selection cassette relative to the first oligonucleotide).

All primers are listed in appendices II and III. PCR reaction mixes were prepared using Gotaq mix (Promega) according to manufacturer's guidelines. The cycling program was: 95 °C 5' // 95 °C 30'' / 55 °C 30'' / 68 °C 1'/kb (x30) // 68 °C 10' //. The elongation time varied according to the size of the longest amplicon, i.e. 1 min per kb.

2.4.5 Pulsed-Field Gel Electrophoresis (PFGE) of *P. berghei* chromosomes

Highly infected blood was collected by cardiac puncture and WBCs depleted (section 2.4.1) prior to lysis (section 2.4.2). The parasite pellets were subsequently re-suspended in 1 % low melting point agarose in TNE buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM EDTA) at 42 °C. While liquid, the mixture of parasites and agarose was loaded into plug moulds and left to settle for 20 min. These were then digested in 5 mL of SE buffer (0.5 M EDTA, pH 8; 1 % sarcosyl) with 100 µg/mL of proteinase K for 48 h at 37 °C.

A 1 % agarose gel was prepared in 0.5 x TBE and run in 0.5 x TBE, in a Chef Chiller DR III system from BioRad, according to the manufacturer's instructions and Pfander *et al* [73]. After the run, the gel was stained with ethidium bromide and imaged to assess chromosome migration. Then, the gel was depurinated in 0.25 M HCl for 20 min and denatured for another 20 min in denaturing buffer (1.5 M NaCl, 0.5 M NaOH). Finally, the DNA was transferred and hybridised according to section 2.4.7 to a Hybond XL membrane by capillarity and hybridised with a probe that targeted the *Pbdhfr* 3' UTR. This sequence was chosen as it naturally occurs in chromosome 7 and can therefore be used as positive control. Additionally, it is present in all *Plasmo*GEM vectors as a direct repeat, flanking the drug cassette. The probe was amplified from gDNA using primers arg496 and arg497 and the

following cycling programme: 95 °C 5' // 95 °C 30'' / 55 °C 30'' / 68 °C 30'' (x30) // 68 °C 10' //. Probe labelling and hybridisation were performed using a standard Southern blot protocol detailed in section 2.4.7.

2.4.6 DNA preparation for Southern blot analysis

DNA extracted from blood stage parasites was digested with restriction enzymes to provide suitable sized DNA fragments for analysis (Table 2.2). A minimum of 3 µg was digested for each hybridisation. All enzymes were purchased from NEB and reactions were prepared according to the manufacturer's guidelines.

Table 2.2| Summary of restriction enzymes used to digest parasite DNA for Southern blot analysis.

Parasite line	Enzymes	Expected sizes (WT/Mutant, bp)
***cdpk1* KO**	*Sna*BI + *Xba*I	3870/2771
***cdpk3* KO**	*Spe*I + *Eco*RI	5538/4098
***cdpk4* KO**	*Hind*III	5153/4083
***cdpk6* KO**	*Pst*I, + *Bgl*II	4706/2794
***gsk3* KO**	*Cla*I + *Xba*I	2707/ 821

After digestion, the DNA was separated on a 0.7 % agarose gel (Biorad laboratories), at 35V for 16 h. Once the run was finished, the gel was stained with ethidium bromide and imaged to assess DNA migration. This was followed by incubation in denaturing buffer (1.5 M NaCl, 0.5 M NaOH) for 20 min. Finally, the DNA was transferred onto a Hybond N^+ membrane (Amersham Biosciences) overnight by capillary action transfer. On the next day, the membrane was cross-linked by a UV cross-linker.

2.4.7 Southern blotting hybridisation

The membrane was pre-incubated in hybridisation buffer (0.25 M Na_2HPO_4, pH 7; 20% SDS; 0.25mM EDTA; 1 % BSA (w/v)) for 30 min, at 55 °C, while the probe was end-labelled with High Prime DNA labelling kit (Roche) and [^{32}P] γATP according to the manufacturer's guidelines. Next, the probe was purified using ProbeQuantTM G-50 Micro Columns according to the manufacturer's instructions (Amersham Biosciences) and denatured at 95 °C for 5 min. After denaturation, it was cooled down on ice for 5 min and added to the

hybridisation buffer. These were incubated at 55 °C, overnight, in a rotating oven. On the next morning, the buffer was removed and the membrane was washed three times in 2x SSC, for 20 min, at 55°C. The Southern blot was visualised by exposure to BioMaxTM MR film (Kodak) at -80°C.

2.5 Illumina sequencing

2.5.1 Library preparation

2.5.1.1 Adaptor ligation

2.5.1.1.1 Whole genome sequencing

Blood from three highly infected mice (parasitaemia >10 %) was collected and gDNA was prepared according to sections 2.4.1, 2.4.2 and 2.4.3.1. A minimum of 2 µg was sheared using a Covaris ultrasonicator (model E220) and settings suggested by the manufacturer to obtain ~ 500 bp fragments. Libraries were prepared using the TruSeq DNA PCR-Free LT Sample Prep Kit (Illumina) according to the manufacturer's guidelines. Prior to sequencing each library was quantified by qPCR using a Library Quantification kit (Kapa Biosystems – kk4834). Equimolar amounts of each library were pooled and the final library was diluted to 4 nM.

As the library preparation (i.e. steps from DNA fragmentation onwards, see 1.3.1.2) and quantification steps are available as a service provided by the Illumina C team at the Sanger institute, they were delegated, once I was fully trained on these protocols.

2.5.1.1.2 STM time-points

The initial library preparation protocol for STM samples was based on an adaptor ligation strategy. The kit NEBNext DNA Library Prep Master Mix Set for Illumina (NEB) was used for this purpose, according to the manufacturer's instructions, but starting from the "dA-Tailing of End Repaired DNA" step (see 1.3.1.2). A minimum of 500 ng of PCR amplicon sample was required to initiate this protocol.

Prior to sequencing each library was quantified by qPCR using a Library Quantification kit (Kapa Biosystems – kk4834). Equimolar amounts of each library were pooled (multiplexing) and the final library was diluted to 1 nM.

As the library preparation (i.e. from dA-Tailing onwards) and quantification steps are available as a service provided by the Illumina C team at the Sanger institute, they were delegated, once I was fully trained on these protocols.

2.5.1.2 Direct amplification

For reasons detailed in chapter 3, the direct amplification method became the standard method to prepare libraries to sequence the vector-specific barcodes. It relied on a nested PCR approach. Briefly, 1 μL of each DNA sample served as template for a PCR reaction using Advantage 2 Taq polymerase (Clontech) with primers arg444 and arg445 (95 °C 5' // 95 °C 30'' / 55 °C 20'' / 68 °C 8'' (x35) // 68 °C 10' //), which bind to constant annealing sites flanking each barcode. This generated a 167 bp amplicon that was further extended in a second PCR reaction using oligonucleotides that in their 5' extensions contain Illumina adaptors and sample-specific barcodes for multiplexing up to 32 samples in one run of a MiSeq instrument. For sample specific indexing, 5 μL of the first amplicon served as template for a further 10 amplification cycles (95 °C 5' // 95 °C 30'' / 68 °C 15''(x10) // 68 °C 5' //) using one generic oligonucleotide (PE1.0) and one of a set of 32 index oligonucleotides (listed in appendix IV). The size and quality of the resulting amplicons were verified on a 1.5 % agarose gel. Each library was purified using the MinElute PCR Purification Kit (Qiagen) and eluted in 50 μL of EB buffer. Then, these were quantified in a Qubit system and 100 ng of each library were pooled. The final pool was quantified and quality controlled by qPCR for the presence of sequencing adaptors using a Library Quantification kit (Kapa Biosystems- kk4834). Prior to loading, the final multiplexed library was diluted to 1 nM.

Quantification by qPCR and MiSeq loading steps were performed by the Illumina C team at the Sanger institute.

2.5.2 MiSeq run conditions

All samples were loaded and run on a MiSeq instrument by the low-throughput Illumina Bespoke team at the Sanger Institute.

2.5.2.1 Whole genome sequencing

These libraries were diluted to 4 nM and mixed with 1 % PhiX for internal control purposes and run at normal cluster density (8-10 x 10^5 clusters/mm^2). Reads were paired-end and 150 bp-long. Data were mapped with BWA, handled with SAMtools, and visualised in Artemis.

2.5.2.2 Barcode sequencing

Due to their low complexity, PCR amplicon libraries were diluted to 1 nM and then spiked with 40-50 % of PhiX before being loaded at low cluster density (4 x 10^5 clusters/mm^2). Reads were paired-end and 150 bp-long.

2.6 Data analysis

2.6.1 Fitness calculation by barcode counting

Using a Perl script written by Frank Schwach, barcode sequences were extracted from the sequencer output files (*fastq*) and counted for each gene, for every time point and input samples.

Then, the relative abundance of each barcode within the pool was determined using my own R script. The quantitation was considered reliable for barcodes accounting for at least 0.1 % of all counts.

The input sample, i.e. an aliquot of the pool of vectors used in the transfection, was used as positive control for the presence of vectors in the pool and also as an internal control for sample cross-contamination. Parasitaemia curves for these mutants were inferred by combining the relative abundance of each barcode with the observed total parasitaemia, determined by a Giemsa stained thin blood film. This was done by multiplying the daily proportion of each mutant by the corresponding overall parasitaemia.

The relative fitness (w) of a mutant represented by a barcode on a given day (d) was calculated according to Mani *et al* [101] by comparing the daily change in its relative abundance (A) to that of the reference genes (A_{Ri} ... A_{Rn}) with normal growth:

$$w_{gene\ d} = \frac{A_d}{A_{d-1}} : \frac{\sum_{i=1}^{n} \frac{A_{Ri\ d}}{A_{Ri\ d-1}}}{n}$$

where n=4 as the reference genes in these experiments were *soap*, *p25*, *p28* and *p230p*. This was done by diving the daily fold change of the relative abundance of each mutant (A_d / A_{d-1}, where *d-1* is the day of infection preceding day *d*) by the average of the equivalent daily fold change of the normal growth references $\frac{\sum_{i=1}^{n} \frac{A_{Ri\ d}}{A_{Ri\ d-1}}}{n}$. Statistical analyses compared the growth rate (G_r), i.e. the daily fold change, of each mutant against the normal-growth reference vectors as given by:

$$Gr_{gened} = \frac{A_d}{A_{d-1}} \text{ and } Gr_{Rd} = \frac{\sum_{i=1}^{n} \frac{A_{Ri\ d}}{A_{Ri\ d-1}}}{n}$$

using a two tailed T-test (unequal variance, p-values adjusted according to the false discovery rate method). A given mutant was considered viable when consistent growth of its barcode was observed for all time points in at least 2 of 3 replicates.

2.6.2 Genetic interaction coefficients

Genetic interaction coefficients were calculated for each day post-transfection as the difference between the observed and the expected fitness of a double mutant ($M_{1,2}$). The expected value was calculated from the observed fitness of the single mutants (M_1 and M_2) using a multiplicative model [101] as follows:

$$w_x(\text{obs}) * w_y(\text{obs}) = w_{xy}(\text{exp}),$$

where w_x(obs) is the observed fitness for mutant *x*; w_y(obs) is the observed fitness for mutant *y*, both calculated as described in section 2.6.1. and w_{xy}(exp) is the expected fitness of the double mutant *xy* assuming that genes *x* and *y* do not interact.

Uncertainty of the predicted fitness of the double mutant for a given day was determined from the observed standard deviations (SD) of the fitness values for the single mutants, according to the rules of error propagation for SD:

$$\sqrt{\left(\frac{SD_x}{w_x}\right)^2 + \left(\frac{SD_y}{w_y}\right)^2} \times w_{xy}(\text{exp}) .$$

2.7 Western blotting

Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using a NuPAGE Novex 4-12 % Bis-Tris pre-cast gel (Life Technologies), at a constant voltage of 200 V for 50 min, according to manufacturer's instructions. After electrophoresis was complete, the proteins were transferred to Immobilon®-P PVDF membranes (Sigma-Aldrich) using the XCell II™ blot module (Life Technologies) and 1X NuPAGE® transfer buffer (Life Technologies) (with 10 % methanol), at a constant voltage of 30 V for 1 hour, according to manufacturer's instructions. After the transfer was complete, the PVDF membrane was blocked overnight in 5 % milk (Marvel)/PBS, at 4 °C.

On the next day, the membrane was probed with primary antibodies for 1 hour at room temperature with rotation, followed by 3 washes with PBST. In this project I only used the anti-HA tag rabbit monoclonal antibody (Cell Signalling Technology), diluted 1:400, in

2 % milk/0.1% Tween-20/PBS (Sigma-Aldrich) (PBST). The membranes were then probed with horse radish peroxidase (HRP)-conjugated secondary antibodies (Cell Signalling Technology), diluted in PBST, for 1 hour at room temperature with rotation, followed by another 3 washes with PBST. The presence of the HRP-conjugated secondary antibodies on the membranes was detected using enhanced chemi-luminescence (Amersham ECL, GE Healthcare) and Biomax MR films (Kodak).

Chapter 3

Establishment of Signature tagged mutagenesis in *P. berghei*

Setting the scene

3.1 Introduction

The first aim of this project was to adapt the STM strategy to *P. berghei* parasites in order to enable large scale genetic screening.

STM screens have been designed in many ways, reflecting the diversity of genetic systems of different taxa. Common to all is a workflow that starts with (1) mutagenesis, i.e. generation of barcoded mutants, which is followed by (2) propagation of these mutants in pools and finally (3) identification of the mutants present after propagation through their barcode.

One main approach used to generate barcoded mutants in bacteria is *in vivo* transposition (Fig. 1.6). Fonager and colleagues have applied a *piggyBac* transposition system to *P. berghei* parasites, but fine tuning is yet to be achieved [46]. In yeast, directed gene-replacement has been the most used method to generate libraries of thousands of mutants that are then pooled and used in STM approaches [94]. Modifications that rely on homologous recombination are probably the most reliable method for genetic modification of *Plasmodium* parasites, although at different frequencies according to the species. However, the approach taken by the yeast field would be of very little use for *P. berghei* parasites as at least 12 mice are needed to generate a single clonal line.

Recently, a new generation of *P. berghei* targeting vectors was developed – the *Plasmo*GEM vectors [73]. These are linear vectors in which the length of homology arms is increased from 0.3 – 1.0 kb to several kb, to improve homologous integration frequencies. Additionally, they have not been reported to persist as episomes after transfection, which decreases the rate of false positives, as drug selection ensures elimination of the parasites where integration did not take place. A combination of improved integration with reduction of false positives made these vectors promising tools for a *P. berghei* adapted STM. To allow identification of mutants generated by these vectors within a pool, gene-specific barcodes were introduced into their basic design that labels mutants upon genome integration. These barcodes consisted of a 10-11 mer DNA sequence that was inserted into a ~100 bp-long module, located next to the B2 gateway site in all vectors. The length of the barcode permitted

that enough sequences with a hamming distance of four[1] (i.e. single error correction plus double error detection) [158], could be generated to cover the entire *Plasmodium* genome.

The barcode module was flanked by constant annealing sites, which enabled a bias-free amplification of all barcodes from a pool through a single PCR reaction.

Taken together these tools offered the opportunity to perform STM-like experiments in *P. berghei*. We hypothesised that the properties of the *Plasmo*GEM vectors would enable the generation of complex pools of mutants, thus circumventing the need to generate each mutant independently prior to parallel phenotyping. In other words, transfection of pools of barcoded *Plasmo*GEM vectors (mutagenesis step) would generate pools of barcoded mutants that after being expanded in a single mouse (propagation step) could be identified through their barcodes (identification step).

To test this hypothesis, various parameters needed to be optimised. These included transfection conditions, barcode detection, sequencing library preparation and sequencing run conditions.

3.2 Results

3.2.1 Optimisation of operating conditions for transfection

The ability to develop STM based screening in *Plasmodium* depends on the complexity of mutant pools that can be easily generated by co-transfecting multiple vectors. This in turn depends on the transfection efficiency that can be achieved. My first objective was therefore to identify the most suitable electroporator and the most adequate DNA concentration of each *Plasmo*GEM vector to use.

3.2.1.1 Choice of electroporator

One aspect that is critical for transfection efficiency is the type of the electroporation system used. The traditional Bio-Rad instruments have been surpassed by the Lonza electroporators, which are now the most efficient devices used to generate *P. berghei* transgenics.

[1] A Hamming distance of four enables single error correction plus double error detection, i.e. it takes four mutations, or sequencing/synthesis errors for one barcode to become another; one mutation can be corrected and two can be detected.

Two different Lonza electroporator systems were tested for their efficiency – Nucleofector II and 4D Nucleofector X unit (Fig. 3.1A). Using the same pool of schizonts cultured from two different mice, four different transfections were performed with each electroporator.

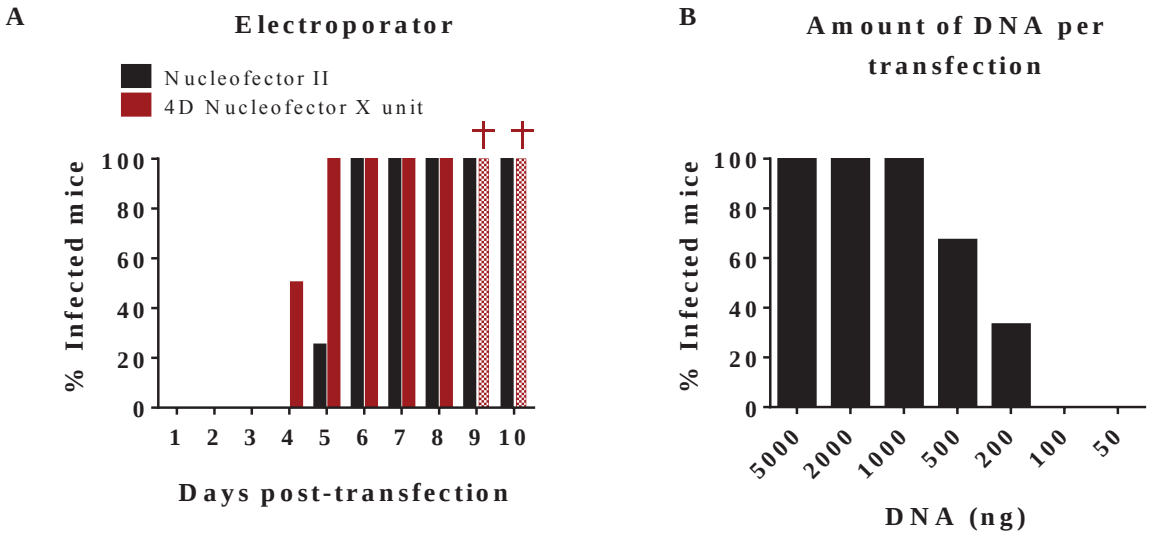


Fig. 3.1| The choice of electroporation system and DNA concentration are critical for maximum transfection efficiency.
(A) Impact of the choice of electroporator on patency. Four different transfections were performed using either the Nucleofector II or the 4D Nucleofector X unit electroporator to transfect 5 µg of a KO vector targeting *map1* gene. The graph shows the duration after transfection required for mice to develop parasitaemias visible on a Giemsa stained smear, which was used to infer transfection efficiency. † culled mice due to high infection. (B) Assessment of how little DNA is required for a successful transfection using the 4D Nucleofector X unit system (n=3 per concentration). All mice were injected intra-venously (i.v.) in the tail vein.

Transgenic parasites were obtained for all replicates. However, the 4D Nucleofector system proved to be more efficient since parasites were visible in the blood of two out four mice by day 4 post-transfection, whereas with the Nucleofector II system only one mouse had visible parasitaemia by day 5 post-transfection. All eight mice were diagnosed as infected on day 6.

Transfection efficiency was determined according to Janse *et al* [159] by comparing the number of surviving parasites in mice before and after pyrimethamine selection, based on the daily 10x multiplication rate of *P. berghei* in mice [160]. Mathematically this is defined as: $(n2/n1) \times 1/10^{d}$. To this end, the parasitaemia after injection of the transfected parasites (*n1*) was determined by counting Giemsa stained thin blood films at ~24h post-transfection, just before the start of the selection with pyrimethamine. Later, usually four to seven days (*d*) after

transfection, the number of drug-resistant parasites (*n2*) was determined from the parasitaemia counted on day *d*, when infection was patent.

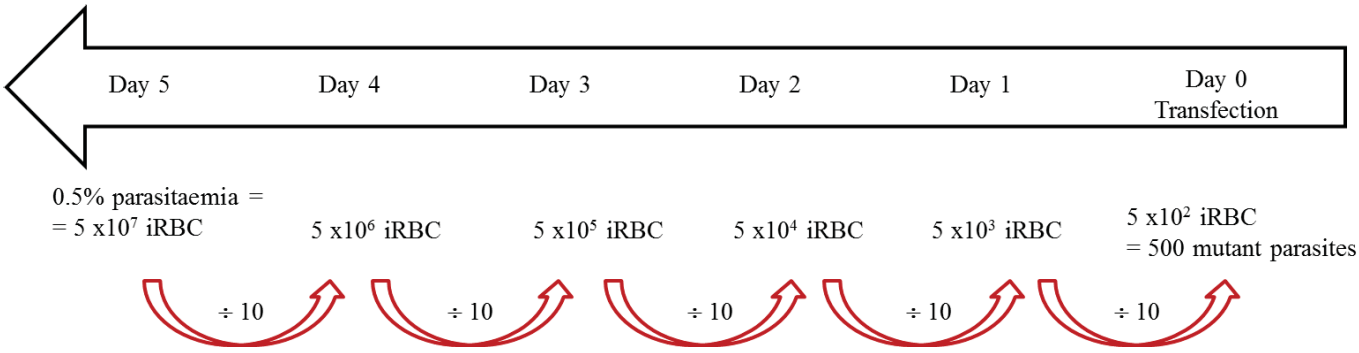
These results translated into a transfection efficiency of ~ $2x10^{-5}$ and ~$5x10^{-4}$ for the Nucleofector II and 4D, respectively.

3.2.1.2 Optimal DNA concentration

As transfection of pools of vectors was envisaged I explored what the minimal DNA concentration of each vector was to generate transgenic parasites, and whether very high concentrations would be more efficient.

For this, a dilution series of a KO vector for the *map1* gene (PbGEM-036210) was prepared to generate the following range: 5000 ng, 2000 ng, 1000 ng, 500 ng, 200 ng, 100 ng and 50 ng. Each of them was transfected in triplicates using the 4D Nucleofector X unit and the percentage of mice that became infected by day 10 post-transfection was registered. As expected, 2000 ng (the standard concentration for a *P. berghei* transfection) generated transgenic parasites for all replicates as did 5000 ng and 1000 ng, whereas concentrations below 200 ng were not successful. The transfection of 200 ng and 500 ng generated one and two infections, respectively (Fig. 3.1B). In addition, there was no difference in patency day for the highest concentrations, hence suggesting that above 2000 ng the system is probably close to saturation.

On day 5 post-transfection of the experiments where 1 µg of DNA was transfected, the observed parasitaemia was around 0.5 %. Based on the daily 10x multiplication rate of *P. berghei* parasites [160] during the exponential phase of growth this translates into approximately 500 independent integration events as the number of circulating RBCs in a mouse[2] is on average $1x10^{10}$, as detailed below:



[2] This RBC concentration applies to an average 6-8week-old, 30 g mouse.

Since transfecting 1 µg of DNA generates around 500 independent integration events while 0.1 µg produced none, I concluded that efficient transfection might be a threshold phenomenon, as is electroporation [161,162], and not linearly related to the amount of vector DNA used. I therefore predicted that even smaller quantities of vector DNA should reliably give rise to mutants when delivered as part of a vector pool as long as the total amount of DNA was at least 1 µg. To this end, I transfected two different pools (n=3) of equimolar amounts of 10 vectors with a total DNA concentration of 1000 ng (100 ng/vector) and 2000 ng (200 ng/vector). These pools included the KO vector used in previous experiments, as a positive control (*map1* KO). The choice of number of vectors had the objective of assessing the limit suggested by the previous experiment (i.e. no transgenic parasites < 200 ng/vector). All transfections were positive by day 6 post-transfection and integration for the control vector was detectable in all experiments by PCR.

The detection of *map1* KO vector in both conditions strongly indicated that the overall DNA concentration is a more important variable than the concentration of each vector for transfections of pools in the context of STM experiments. However, the same phenomenon that enables the transfection of small amounts of each vector and therefore simplifies the preparation of each pool prior to transfection (i.e. one miniprep yields enough DNA for several experiments) also raised the concern of integration of multiple vectors in the same genome. Mathematically, assuming a transfection efficiency of $5x10^{-4}$, as determined previously for the experiments shown in Figure 3.1A, and assuming that the integration of two different vectors are independent events, then the likelihood of a double integration event should be the product of the individual likelihoods, i.e. close to $2.5x10^{-7}$.

However, in reality integration events are almost certainly not independent since the efficiency with which DNA is delivered will vary between individual schizonts. I addressed this question experimentally by asking whether vectors lacking selection markers (i.e. intermediate vectors) could become "passengers" of vectors with selection marker.

KO vectors for known targetable genes were transfected with a 20-fold excess of intermediate vectors for different other targetable genes (Fig. 3.2A) that lacked a selection marker for *P. berghei* and could therefore only replicate when integrated into genomes that carried a second insertion of a final KO vector. PCR analysis of resistant parasites on day 9 post-transfection failed to detect such events (Fig. 3.2B), suggesting that double integration events are rare and that the large majority of the transgenic parasites are single mutants.

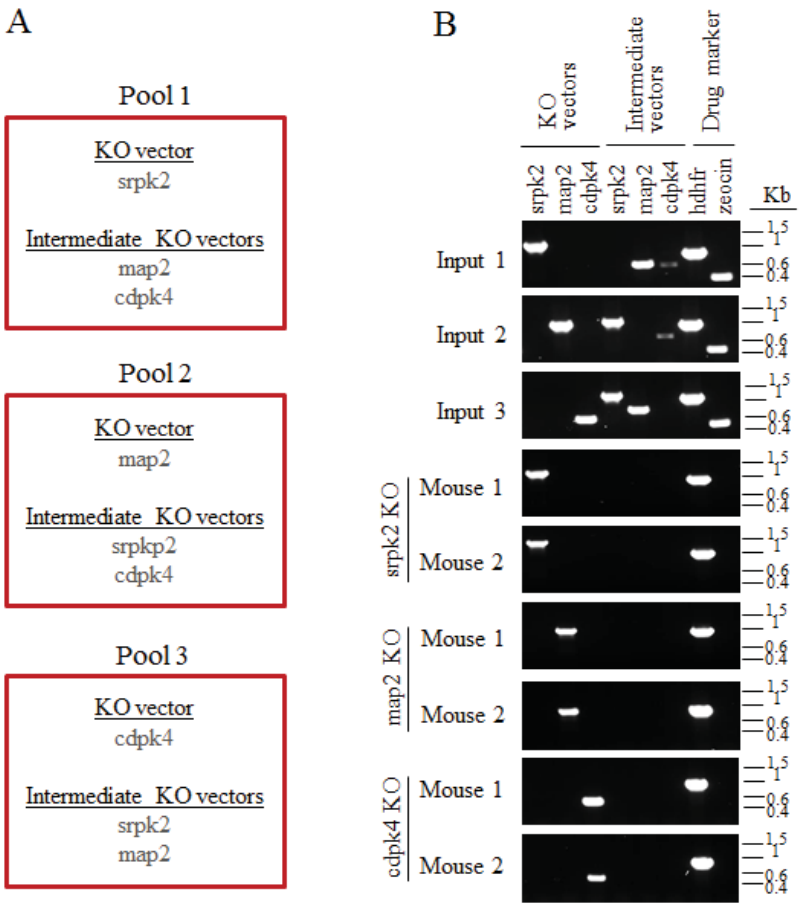


Fig. 3.2| Absence of passenger vectors lacking a selection cassette.
(A) Three different pools consisting of one final KO vector in the presence of a 20-fold excess of intermediate vectors (10 µg total DNA per transfection), which have the same homology arms but a zeocin resistance cassette that cannot be selected in *P. berghei* parasites were transfected in duplicates. (B) PCR genotyping performed on gDNA samples, from day 9 post-transfection from each of the six infected mice, failed to detect the presence of intermediate vectors that could only be selected if integrating into a genome where a KO vector was present.

3.2.2 Optimisation of barcode detection using Illumina sequencing

Before any attempt to perform STM experiments with *P. berghei* parasites, I investigated whether a "bar-seq" strategy (section 1.7.1) using Illumina sequencing would be feasible to read the *Plasmo*GEM barcodes. To this end, several pools of barcoded vectors were prepared and sequenced.

Pools 1 and 2 (Fig. 3.3A) were devised to look at the sequencing accuracy across two orders of magnitude of different ratios of each of 19 vectors that varied in abundance, while trying to reproduce realistic concentrations of vectors in real experiments. In some cases, as little as 5 ng (0.05 ng/µL) of vector was added (vectors 4, 15 and 17) to verify the sensitivity of the PCR reaction for low abundant barcodes in an attempt to mimic a pool of parasites containing a less fit (and therefore less abundant) population of mutants. Conversely, some

vectors (8 and 14) were more abundant (5.00 ng/μL) than the rest so that the impact of having a dominant barcode in the pool could be assessed. Some vectors were used as negative controls and were therefore not added to the pool but their barcode was searched for in the sequencing data (vectors 1 and 11) and in other cases a vector would be present in one of the pools but not the other (vectors 5 and 17).

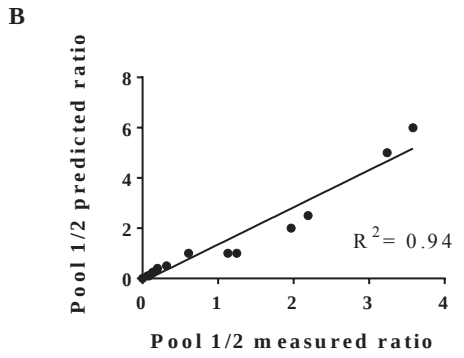
A correlation analysis between the predicted and the measured ratios yielded a high correlation coefficient (R^2) of 0.94 and is depicted in Fig. 3.3B. For instance, for vectors 2, 4, and 6 the predicted ratio was 2.00, 0.10, and 1.00 and the corresponding measured ratios were 1.97, 0.09, and 1.13, respectively. The negative controls were not detectable in either of the pools (1 and 11) and the same was seen for vectors 5 and 17 in pools 2 and 1, respectively.

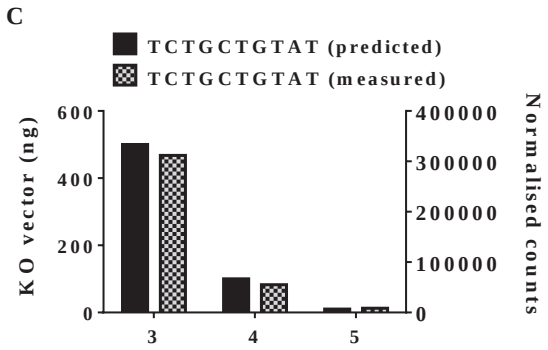
In parallel, three other less complex pools were prepared where the concentration of each of three vectors was adjusted to mimic three different hypothetical outcomes of real experiments: decline (Fig. 3.3C), steady maintenance (Fig. 3.3D) and increase (Fig. 3.3E) in abundance over time. The obtained patterns greatly resembled parasite growth and again very similar numbers between measured read counts and their corresponding prediction were obtained, clearly showing that less abundant vectors yield fewer read counts and vice-versa.

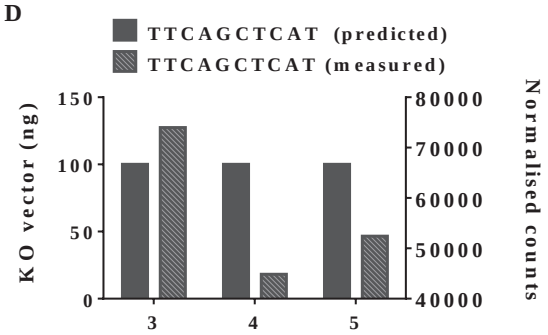
Together these data showed that barcode counting provides quantitative measurements that are sufficiently accurate to measure differences in barcode abundance within pools.

A

Barcode	pool 1 (ng)	pool 2 (ng)	Ratio pool 1/2
1)TGACCACTAT	0	0	-
2)TCGAAGTGAT	200	100	2.00
3)TGTACCTTAT	50	200	0.25
4)ACAGTGGTAT	5	50	0.10
5)TCTCGGTTAT	50	0	-
6)TGGCTCAGAT	300	300	1.00
7)TAGCTTGTAT	50	100	0.50
8)TGATACGTAT	100	500	0.20
9)TCCAGTCGAT	300	50	6.00
10)TCTGCTGTAT	100	300	0.33
11)TTAGGCATAT	0	0	-
12)TCCTCAATAT	100	100	1.00
13)TCGAGCGTAT	100	100	1.00
14)TAAGTTCGAT	200	500	0.40
15)GCCAATGTAT	5	50	0.10
16)TTCAGCTCAT	500	10	50.00
17)CGATGTTTAT	0	5	0.00
18)TTCTGTGTAT	50	10	5.00
19)TCAGATTCAT	500	200	2.50







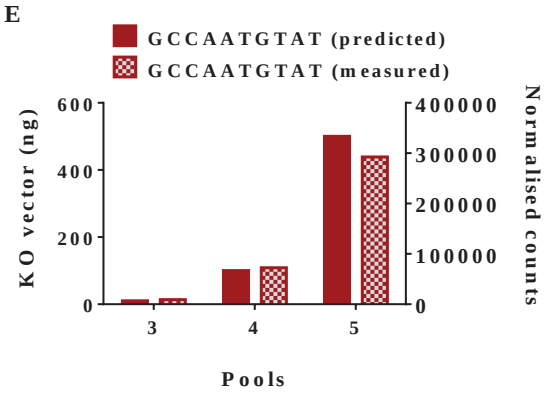


Fig. 3.3| The barcodes within the *Plasmo*GEM vectors are compatible with a bar-seq strategy.
(A) Known concentrations of the same set of vectors were used to prepare two different pools. The final volume of each pool was 100 µL. (B) Both pools were sequenced and the ratio pool 1/2 was calculated from the sequencing reads generated. A Pearson correlation analysis was performed to compare the measured with the predicted ratios of vectors. This analysis excluded all vectors for which a ratio could not be calculated. (C-E) Three other smaller pools (3-5) were prepared to look at feasibility of measuring growth patterns (decline, steady maintenance and increase) through time. The normalised counts patterns were very similar to the predicted ones.

3.2.3 Optimisation of Illumina library preparation

The *Plasmo*GEM barcodes are flanked by constant annealing sites. This ensures that barcodes can blindly be amplified in a single PCR reaction and eliminates any multi-template PCR bias [163]. This is particularly relevant for an STM approach as the pool of barcodes is expected to change during infection.

Amplification of barcodes from purified *Plasmo*GEM vectors was routinely done in 25 cycles. However, the same number of cycles was not enough to reach saturation in samples originated from infected blood. In order to reach the minimal concentration required for Illumina library preparation, either ten different reactions needed to be pooled for each sample or the number of cycles needed to be increased to 35. To exclude any data bias that this increase might induce, the impact on the quality of the data generated after either 25 or 35 cycles was compared (Fig. 3.4).

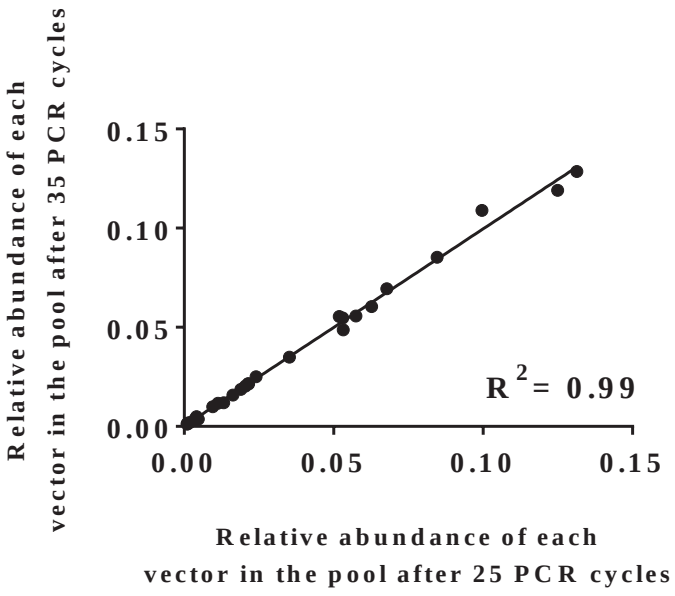


Fig. 3.4| Increasing the number of PCR cycles does not have a high impact on data quality.
The impact on the relative abundance of each vector within the pool after 25 or 35 PCR cycles at the library preparation stage was assessed. A Spearman correlation analysis between the two datasets yielded a high correlation coefficient (R^2= 0.994). Each library had 28 different vectors.

The number of cycles had very little impact on the data as shown by the high correlation coefficient (R^2=0.99) between samples where 28 different vectors were present at different abundances. As a result, 35 cycles were used to amplify all samples in this project since these yielded a more suitable concentration of amplicon.

Two different strategies can be used to prepare sequencing libraries from PCR amplicons: adaptor ligation (AL) or direct amplification (DA).

The AL method is by far the most common library preparation method used to generate next-generation sequencing libraries. Essentially, it requires the ligation of specific adaptor oligos to the pre-processed fragments of the DNA to be sequenced as illustrated in Figure 3.5 (left panel). As the STM amplicons are only 100 bp in length, the original protocol was simplified as shown (Fig. 3.5, right panel). Briefly, 500-1000 ng of the purified barcode PCR product were dA-tailed and then used to ligate Illumina adaptors. When needed, a PCR amplification step priming on the adaptor sequences could be used to boost the library yield.

The AL method, although robust, required high amounts of DNA and was very time-consuming for large numbers of samples. The second method, DA, relied exclusively on two rounds of PCR followed by one purification step (Fig. 3.6A). The first PCR amplified the barcode as previously and introduced priming sites for the second reaction at which stage Illumina adaptors were incorporated. A final clean-up step removed primer/adaptor dimers ensuring that the libraries were ready for loading and sequencing.

In order to choose the best approach, both methods were analysed in terms of preparation time, cost, input material needed and data generated.

On average, preparing a set of 32 samples using AL method took five days, cost £42 per sample and required at least 500 ng of purified PCR amplicon. Conversely, the DA method proved to be substantially faster as only two days were required to process the same set, more affordable (£18/sample) and used less sample (<10 ng of unpurified primary PCR product). The most important parameter, data generated, was not influenced by the library preparation method. High correlation coefficients were obtained for the analysis of six libraries that were prepared by both methods (Fig. 3.6 B, C).

Given that the DA method proved to be a faster and more cost-effective alternative to AL, it was chosen as the preferred method to generate samples throughout this project.

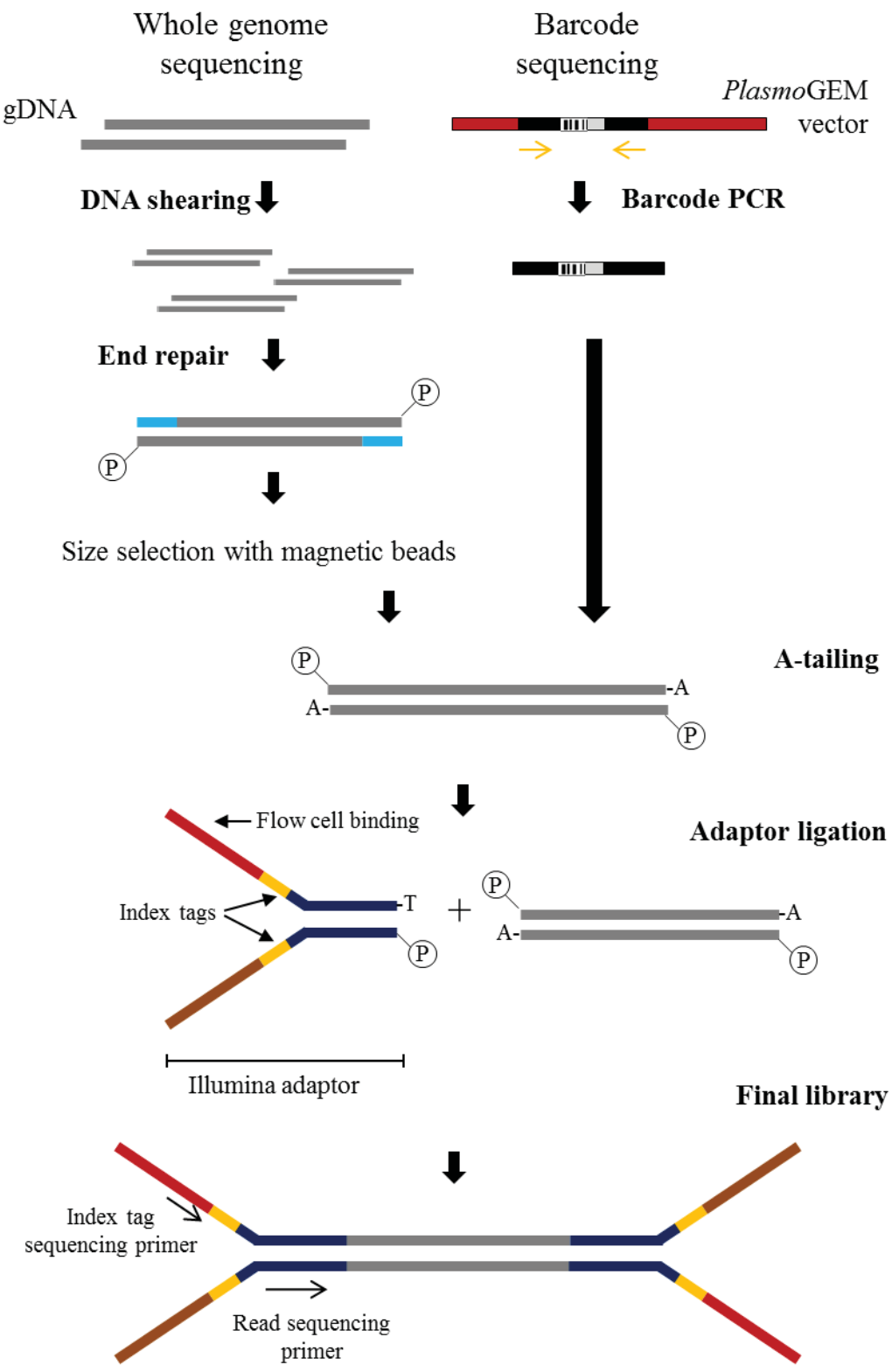
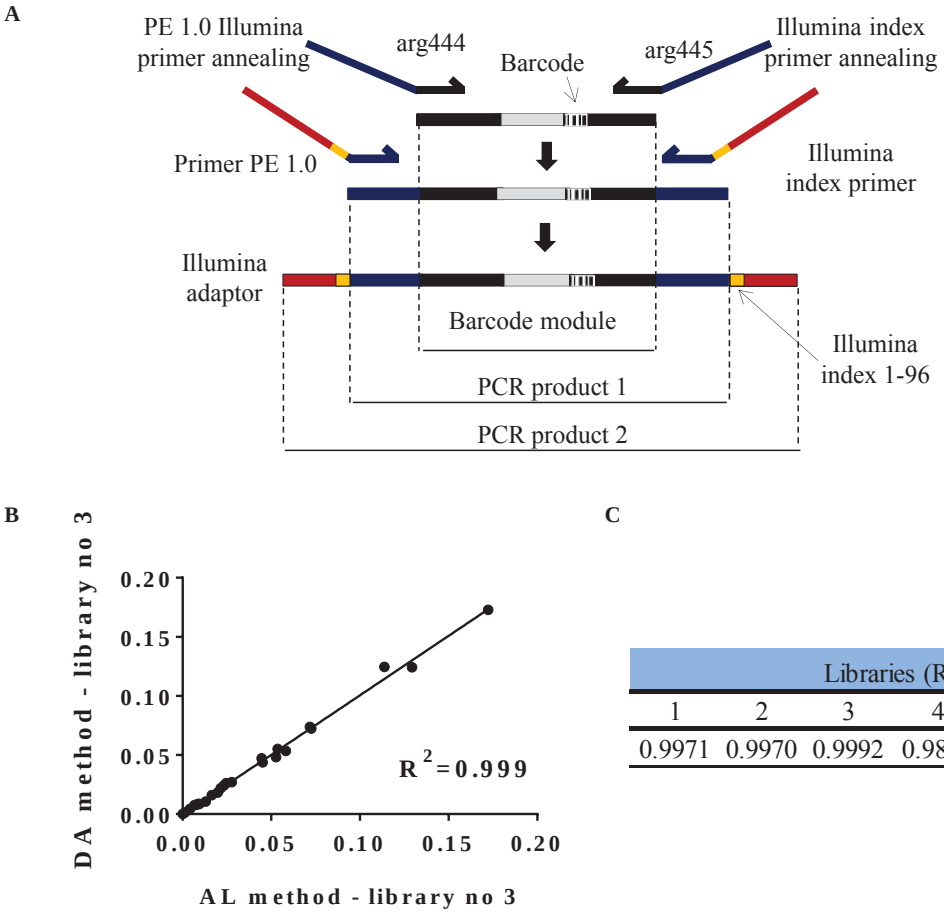


Fig. 3.5| Adaptor ligation method overview.
Typically, an adaptor ligation protocol for WGS samples involves the following steps: DNA shearing, end repair, a-tailing and adaptor ligation. In the case of the STM samples, no shearing or end repair steps are required. Instead, barcodes are amplified by PCR using the flanking annealing sites that are common to all *Plasmo*GEM vectors. Next a-tailing and subsequent adaptor ligation follow. The structure of the final library is depicted at the bottom.



Libraries (R^2)					
1	2	3	4	5	6
0.9971	0.9970	0.9992	0.9898	0.9979	0.9979

Fig. 3.6| Comparison between AL and DA library preparation methods.
(A) DA method overview. A nested-PCR approach amplified the barcodes (PCR1) and generated Illumina compatible libraries (PCR2). The priming sites that enabled the barcode amplification (black regions in primers arg444 and arg445) were unchanged from the AL method but they carry, as overhangs, priming sites for the next round of PCR. The second PCR reaction introduces the Illumina adaptors shown in red and yellow. (B, C) Spearman correlation analysis for six different libraries that were prepared by both AL and DA methods. Very high coefficients were obtained for all of them discarding the possibility of bias introduced by the DA method. (B) Regression analysis plot for library number three.

3.2.4 Optimisation of Illumina MiSeq run conditions

Whole genome sequencing libraries undergo a shearing step that generates random fragments of DNA. This creates a high degree of diversity at the sequencing level as the distribution of bases throughout the flow cell in each sequencing cycle becomes random. STM PCR amplicon libraries are of very low complexity as the order of each base is the same for every cluster except at the barcode region. Also, shearing is not recommended as they already are very short. This makes differentiation of the individual clusters at the imaging level a very difficult task. As very uniform clusters tend to be highly error-prone due to imaging limitations of the platform, it is essential that enough diversity is present to ensure accurate base-calling.

The PhiX spike-in is a base-balanced DNA library derived from the genome of a PhiX 174 bacteriophage. It is commonly used as a spike-in to generate diversity in the flow cell or as a control lane to validate the quality of each run.

The density of clusters on the flow cell is of vital importance to the throughput of the instrument as overload will impede accurate imaging of individual clusters. Normal MiSeq run conditions for high complexity libraries are ~ 8 x10^5 clusters/mm^2 and up to 5 % of PhiX. In order to optimise the run conditions for the STM libraries I tested different conditions of PhiX spike-in concentration and cluster density and determined their impact on the purity filter (%PF). Anticipating problems with cluster density due to the low complexity, the first tests were run at low density (< 4 x10^5 clusters/mm^2). In these conditions, a high correlation (R^2= 0.97) was observed between the spike-in abundance and the %PF of six different runs that were run at similar cluster densities (3 – 3.8 x10^5 clusters/mm^2) (Fig. 3.7 A). The absence of PhiX was highly detrimental to the run quality at both normal and low cluster densities (Fig. 3.7 B and C - Runs 8396 and 10499, respectively); neither of these runs achieved the minimal QC threshold. The addition of 5 % of PhiX increased the %PF of runs 10666 and 10600, 56.7 % and 56.5 %, respectively, at both low and normal density (2.9 and 6.5 x10^5 clusters/mm^2, respectively). However, only when the concentration of PhiX was increased to > 50 % did the %PF reach acceptable levels (at both densities) as shown in Figures 3.7 B and C, run IDs 10343 and 10537.

From these results I established that the ideal run conditions for STM libraries to achieve a %PF of at least 85 % were: a spike-in concentration of 40-50 % (as given by the linear regression shown in Fig. 3.7A) and cluster density of 4-6 x10^5 clusters/mm^2.

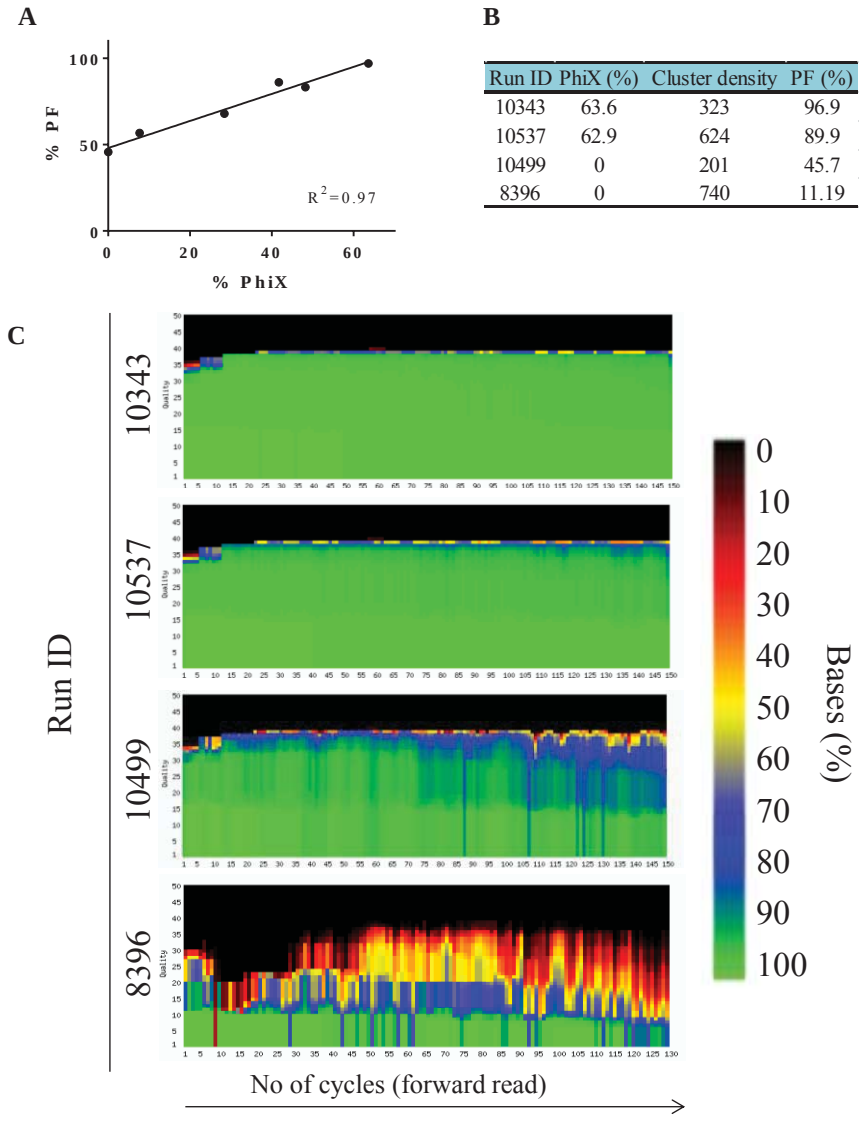


Fig. 3.7| Miseq run quality analysis.
(A) Comparison of the %PF of six different libraries run at similar cluster densities with different proportions of PhiX. The quality of the run highly correlated with the levels of PhiX. (B) Impact on %PF of different cluster densities in the presence and absence of PhiX. (C) Quality histograms of the runs shown in B. The x axis shows the number of sequencing cycles of each run while the y axis shows the quality of base calling: the colour gradient represents the percentage of bases that on a given cycle reached a given quality. Quality values higher than 30 (Q30) are optimal.

Finally, data reproducibility within and between runs for the same sample was verified. For this, two libraries were prepared from the same STM sample with different index adaptors. These were multiplexed and run in the same conditions twice. High correlation was obtained for data generated within runs ($R^2_{(10572)} = 0.9987$ and $R^2_{(10537)} = 0.9999$) and between runs ($R^2 = 0.9995$) thus suggesting that any variation between biological replicates should be seen as real variation (Fig. 3.8).

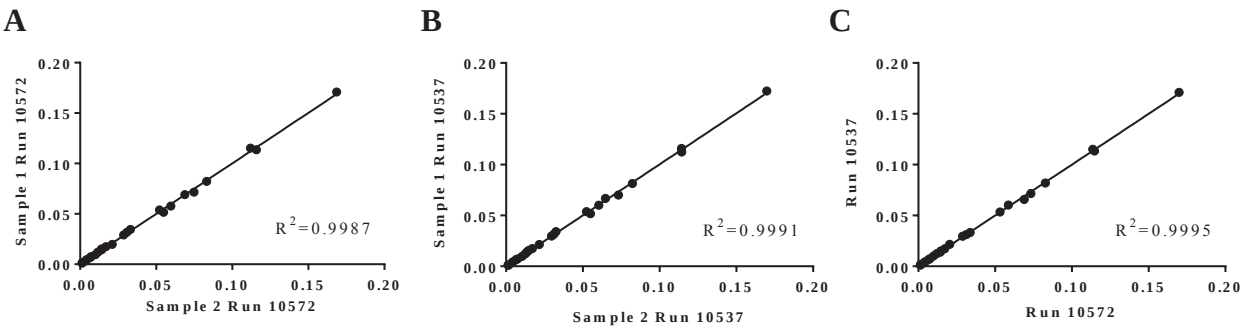


Fig. 3.8| Sequencing reproducibility within and between runs.
(A and B) Pearson correlation analyses of the relative abundances of each barcode within the pool between replicates run in the same lane. (C) Pearson correlation analysis between the averages of each run.
Run IDS: 10537 and 10572, both run at ~6x10^{5}clusters/mm^2; PhiX was spiked in at ~ 60 % and PF values were 89.9 % and 95.9 %, respectively.

3.3 Discussion

In this chapter I have shown that adequate choice of electroporation system can boost transfection efficiency by more than one order of magnitude, which is crucial for the development of an STM approach. In addition, as low as 100 ng of a single *Plasmo*GEM vector is enough to generate transgenic parasites, provided that they are part of a larger pool of at least 1 µg of DNA. This will, in theory, enable the generation of highly complex pools of transgenics in a single transfection.

The library preparation method (DA) optimised here will enable rapid processing of high numbers of samples for multiplex sequencing. However, one consequence of its simplicity is the generation of low complexity libraries. This was overcome by the optimisation of the run conditions, which included low cluster density (4-6 x10^5 clusters/mm^2) and the presence of a base-balanced spike-in (PhiX).

As the *Plasmo*GEM vectors have not been reported to be maintained as episomes and double integration events in the same transfection were shown to be too rare to be detected, if at all existent, I expect that each mutant will carry only one barcode, integrated in the genome.

Barcode sequencing proved to be a reliable method to count barcodes since the abundance of sequencing reads reflected the abundance of the barcodes in the samples. Applied to real STM samples this means that it will be possible to analyse the relative abundance of each barcode, i.e. mutant, and how it changes during infection.

Taken together these data indicated that from the technical point of view *P. berghei* STM approaches are feasible.

Chapter 4

STM analysis of protein kinase genes in *P. berghei*

4.1 Introduction

Following the development of an optimised method to count the relative abundance of barcodes of the *Plasmo*GEM vectors present in mixed pools, this chapter puts the new protocols to the test in a reverse genetics screen of eukaryotic protein kinases (ePK) genes of *P. berghei*. Protein kinases were chosen as test genes for these pilot experiments because they have recently been analysed systematically using conventional KO techniques [105]. According to the latter study, which served as a benchmark for the current analysis, out of a total of 66 ePKs, 23 are redundant for development of asexual stages, while 43 were identified as possibly essential as they could not be targeted.

The experimental setup of this pilot experiment included three stages (Fig. 4.1). Firstly, pools of *Plasmo*GEM targeting vectors were used to co-transfect triplicate pools of schizonts originally propagated for one passage in a Wistar rat[3].

Secondly, the resulting transfected parasites were injected into BALB/c mice as their inbred status reduced host variation to a minimum. Each of the three transfections gave rise to drug resistant parasites four days later, three days after drug selection was initiated. To monitor the composition of the parasite population as it developed in the host, small blood samples were collected from the tail vein, each day from day 4 to day 8 post-transfection, i.e. from the moment when parasitaemia became detectable on a Giemsa-stained thin blood film until the mice reached a critical health state. Additionally, our Home Office animal license did not permit daily blood sampling for longer than five consecutive days. Therefore, a total of five blood samples was collected from each mouse. Next, parasite gDNA was extracted using a phenol-chloroform method as maximal recovery was crucial.

Finally, these samples were used to amplify the vector specific barcodes by PCR, which were subsequently sequenced on a MiSeq instrument, using the optimised conditions described in Chapter 3.

An attempt to generate *Plasmo*GEM KO vectors for 63 ePK genes[4] by recombinase mediated engineering yielded a total of 41 barcoded vectors[5]. These included the CDPK group, MAP kinases, CRKs, RIO kinases, among others. Additionally, seven other KO vectors were prepared that served as additional references and controls. Four were vectors for

[3] The use of rats to boost transfection efficiency is documented in [159].
[4] Three ePK genes were not covered by the *Pb*G libraries.
[5] A list of these vectors can be found in appendix I

genes that were only expressed in sexual and mosquito stages: *p25*, *p28*, *soap* and *p230p*. The first three have well established functions in the ookinete: *p25* and *p28* are surface proteins, and *soap* is a secreted ookinete adhesive protein [164,165]. The last one is p230p gene, a redundant gene in *P. berghei* that is frequently used for transgene insertion [166]. A C-terminal tagging vector for this gene was used instead of a deletion vector because the latter did not integrate efficiently. For all these genes absence of a blood stage phenotype is well established and these mutants could therefore serve as references for normal growth against which fitness for all the other mutants could be calculated.

The other three constructs targeted metabolic enzymes, and were chosen as positive controls for detection of attenuated, i.e. slow, growth phenotypes. These were *plasmepsin IV*, an aspartic protease involved in haemoglobin degradation [167]; PBANKA_140160, a putative methyl transferase of unknown function[6] and PBANKA_110420, the E1β subunit of apicoplast branched chain α-ketoacid dehydrogenase [168].

Vectors were prepared as triplicate pools consisting of 100 ng of each of the 48 vectors.

[6] The selection of this target originated from personal observation in a parallel project.

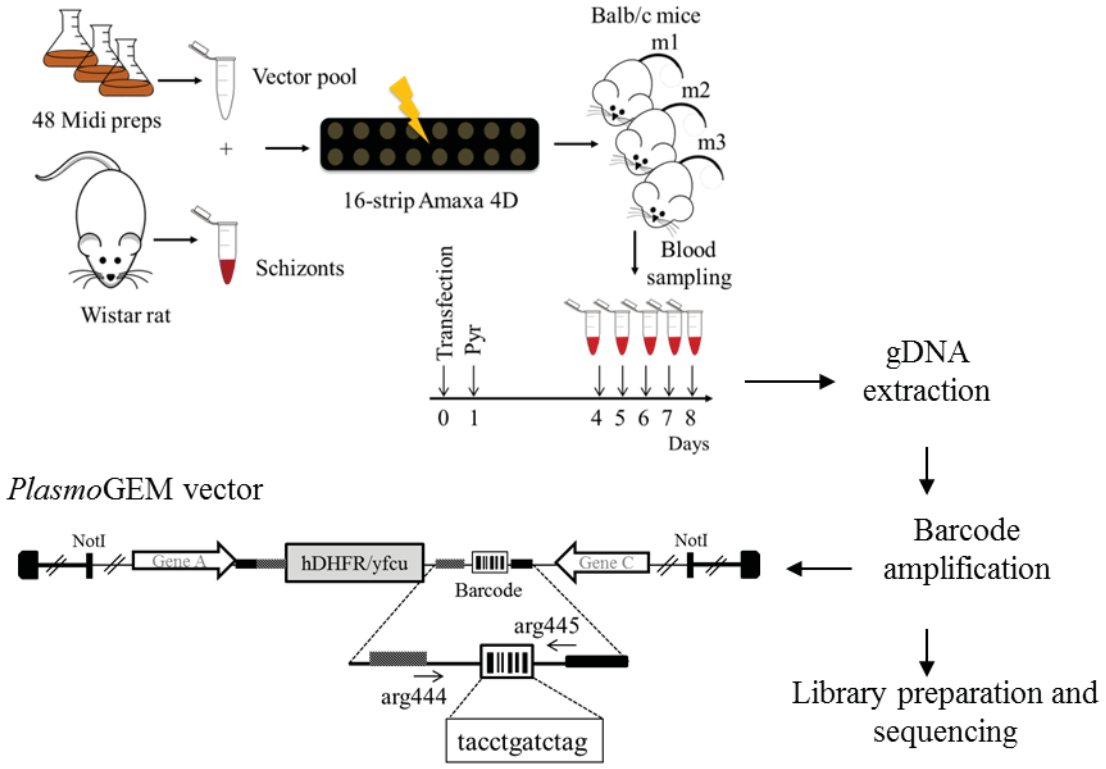


Fig. 4.1| Proposed experimental design for the *P. berghei* STM-Bar-seq experiments.
Midipreps for all vectors were prepared in order to ensure that enough material was available for all STM experiments presented in this thesis. Pools of vectors were transfected into schizonts using the 4D-Amaxa system and injected into three different mice. Drug selection with pyrimethamine started on day 1 post-transfection. Samples were collected on days 4-8 post-transfection and barcodes amplified and sequenced. Each time-point corresponded to an independent library. All libraries for each experiment were multiplexed in the same lane.

The main aims of this set of pilot experiments were: to determine if complex pools of mutants can be (1) generated through co-transfection of vectors and (2) followed over time; (3) to understand if the detection of barcodes corresponds to the detection of viable mutants and consequently if barcodes of essential genes are absent in the final pool; (4) how reproducibly is the generation of mutants from the same pool of vectors; (5) can the bar-seq strategy be used to measure growth rates during the infection, and (6) how can fitness costs be assessed.

4.2 Results

4.2.1 Barcode counting in *P. berghei* allows parallel phenotyping of mutants in a single mouse

The chosen workflow was successful, in that the overall parasitaemia of each mouse evolved as expected for a *P. berghei* infection, increasing rapidly during exponential phase at

ten-fold rate and then stabilised as it reached a growth plateau, with minimal variation across replicates (Fig. 4.2A). Five samples were collected from each of the three mice and a total of 16 were multiplexed and sequenced in the same MiSeq lane. The last sample, here termed the "input", was collected from the electroporation cuvette after the transfection. The data from this sample allowed detection and accurate measurement of the proportion of each vector at the moment of transfection.

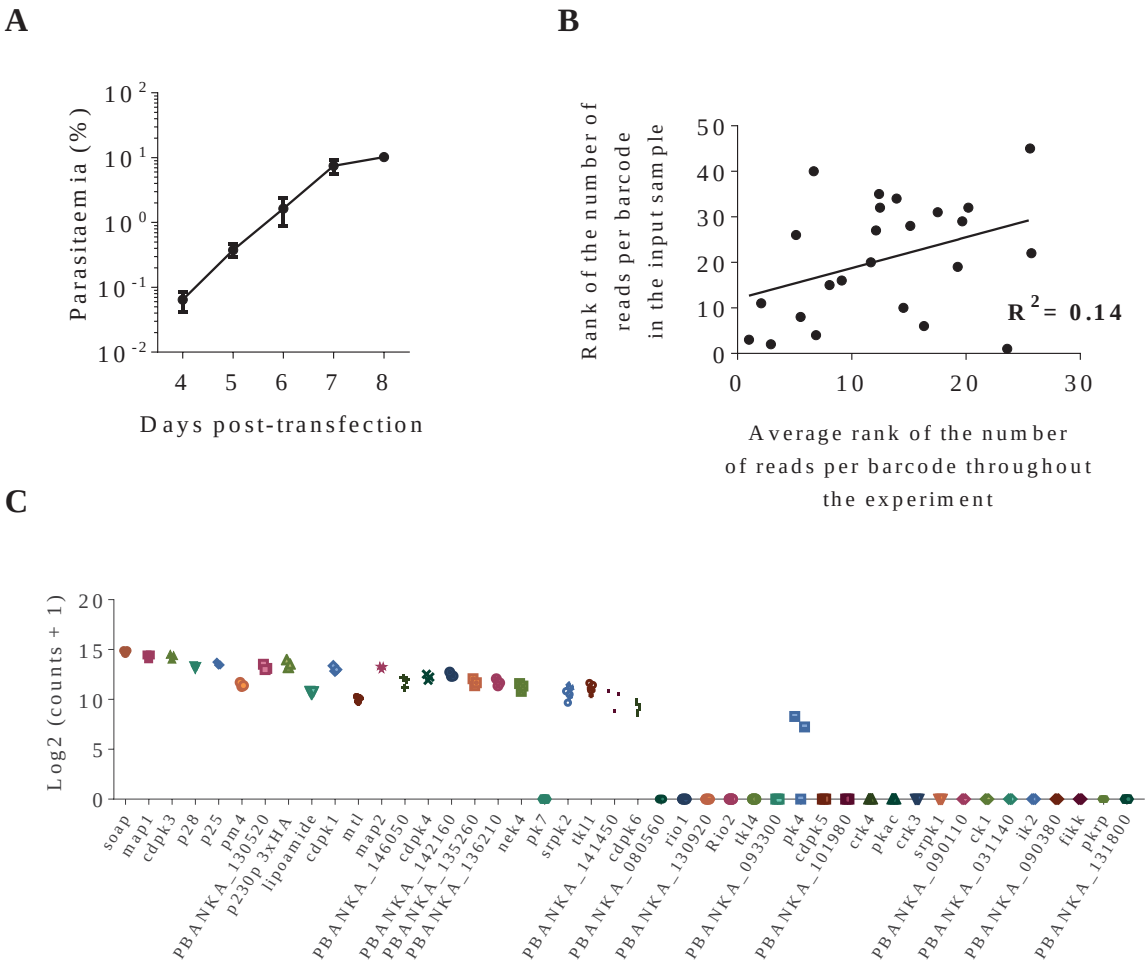


Fig. 4.2| Distribution of barcode counts for each gene and comparison with the input sample.
(A) Overall parasitaemia of the infections throughout the sample collection period showed very little variation between replicates. Error bars show standard deviations of the mean (n=3). (B) Pearson correlation analysis between the rank of the abundance of each barcode in the input sample and the average rank of the abundance of the corresponding barcode throughout the experiment. The low coefficient (R^2=0.14) indicated that the abundance of each vector in the transfection pool is not a major determinant of its subsequent abundance in the pool of mutants. (C) Distribution of the number of counts per barcode on day 7 post-transfection showed a range of abundances for the different barcodes within the pool.

Barcodes were extracted from the sequencing reads and counts per gene/barcode were calculated using a perl script developed by Frank Schwach. On average, three million reads

per lane (excluding PhiX reads) generated between 100 and 300 thousand barcode counts per library (time-point). The counting script was designed to be very conservative to reduce false positives to a minimum by counting only exact matches of a barcode flanked by at least five nucleotides of the barcode module.

The number of counts per barcode in the experimental libraries greatly varied from each other, unlike what was observed for the input sample (Fig. 4.2B, C). This suggested that a range of different abundances of each population of mutants (i.e. barcodes) had been generated despite the normalised quantity of DNA vector.

At this stage two different parameters were used to characterise each barcode: their abundance within the pool and their daily fold change. While the former reflects the integration efficiency and is also a consequence of the abundance of the respective vector in the DNA pool, the latter measures the growth rate, i.e. how the abundance of a given barcode changes over time.

Separate growth curves for all barcodes were calculated from their relative abundance and the overall parasitaemia over time (Fig. 4.3A).

As anticipated, the "normal growth" references (shown in green) were dominant throughout the experiment with highest growth rates (given by the slope of the curve). On the other hand, the attenuated references (shown in light orange) were less competitive and despite having started at similar numbers as the green references, their relative abundance decreased over time, also as expected.

Interestingly, the starting point of the growth curves was spread across at least two orders of magnitude despite the equivalent amount of DNA and parasites. No obvious correlation was found between the size of the homology arms of these vectors and the starting point of the growth curves (data not shown).

For the barcode counting to be used as a detection method of mutants it was crucial to investigate if detection of barcodes was equivalent to detection of integration events. Being linear, *Plasmo*GEM vectors are not expected to persist as episomes, which makes false positives very unlikely to start with. An analysis by Southern hybridisation of separated chromosomes (PFGE) showed genomic integration events throughout the genome. Interestingly, in the pool there were three vectors that targeted genes in chromosome eight[7] (PBANKA_080560, PBANKA_080800/*crk-4*, and PBANKA_083560/*pka*), but no band was present in the PFGE for this chromosome, which suggested that in the pool there were no

[7] The first two digits of the accession number indicate the chromosome number.

mutants with a gene disrupted in chromosome eight. Indeed, previously published data considered all three genes as possibly essential since integration of a KO vector in those loci was never obtained [105]. Further integration evidence was provided by PCR products showing integration of individual targeting vectors (Fig 4.3 C). Taken together, these data suggested that co-transfection of vectors can be used to generate pools of mutants.

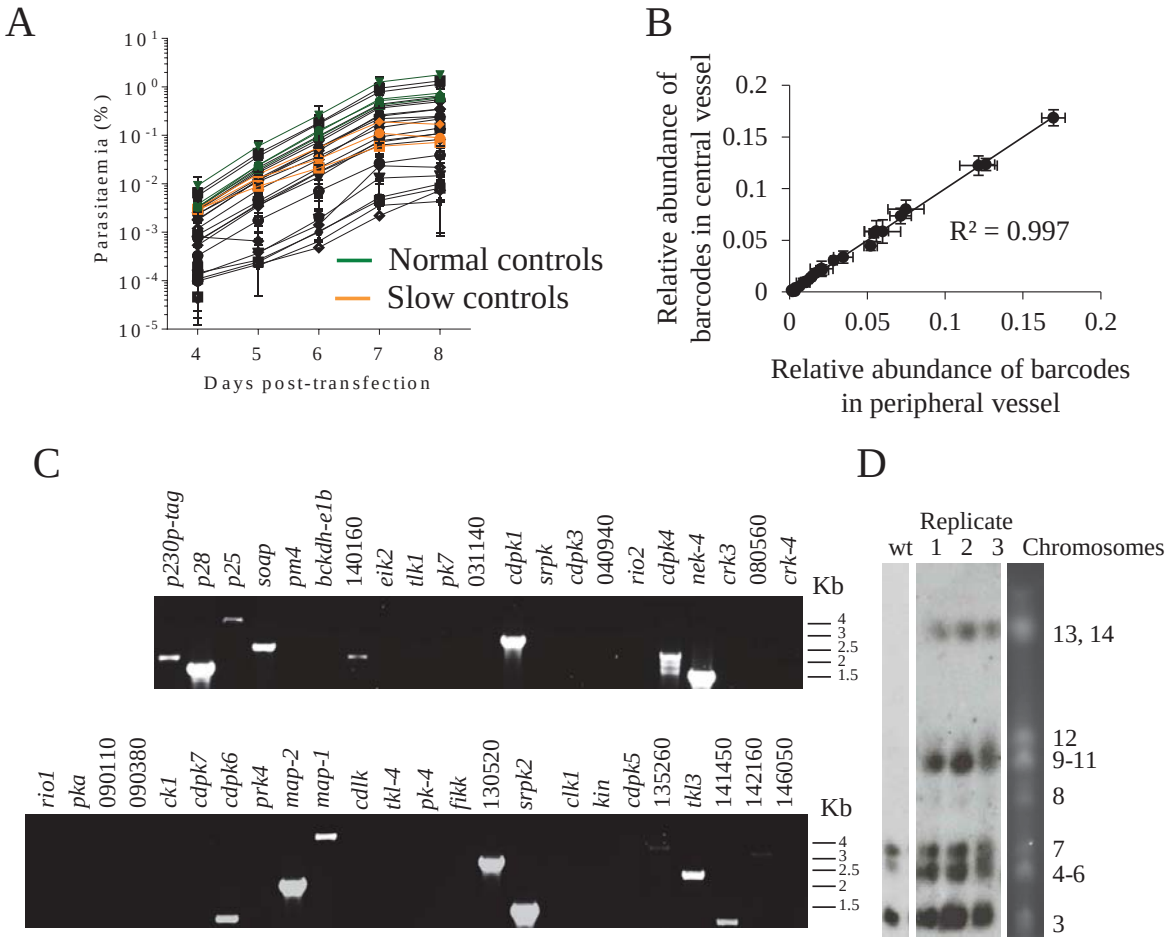


Fig. 4.3| Parallel transfection of pooled KO vectors generated pools of mutants

(A) Growth curves of the different mutants calculated from the overall parasitaemia and the relative abundance of each barcode. The "normal controls" were labelled in green and corresponded to the KOs of the following genes: *soap*, *p28*, *p25*, *p230p*-3xHA; while the "slow controls", labelled light orange, corresponded to KOs of *plasmepsinIV*, PBANKA_140160 and PBANKA_110420 (in descending order of abundance). The black lines represent all the other viable mutants present in the experiment. (B) Pearson correlation analysis between data generated from samples collected either from a peripheral or a central vessel. The high coefficient for these samples suggested that the route by which sampling was done was not a critical parameter. (C) PCR products across the predicted integration sites from the same experiment as in A and D, supporting genomic integration of many vectors whose barcodes were detected. The long homology arms of *Plasmo*GEM vectors can give rise to false negative results from PCR genotyping for some genes. (D) Southern hybridisation of chromosomes separated by PFGE with a ~500 bp probe for the 3' UTR of Pb*dhfr-ts*. This probe should target every integrated *Plasmo*GEM vector as it is present twice, flanking the drug resistance cassette and it should also detect the

endogenous *dhfr-ts*, located in chromosome 7 (positive control). Note that there is an exogenous site in on chromosome 3 in the background strain used due to the presence of a *gfp* gene that carried the same 3' UTR. Error bars show standard deviations from the mean (n=3).

Blood samples from days 4 – 7 were collected from tail blood (peripheral vessel), but on day 8 these were usually collected under terminal anaesthesia by cardiac puncture (central vessel) as larger samples were required for subsequent genotyping purposes. To ensure that the route by which sampling was performed would not impact on the distribution of parasites a correlation analysis was performed between the two types of samples collected on day 8 post-transfection. Very high correlation (R^2=0.997) between both samples was obtained suggesting that type of vessel used for sampling has no impact on the distribution of the parasites (Fig. 4.3B).

Growth curves were analysed using two parameters: the relative abundance of each barcode within the pool, and the relative fitness of each mutant, defined as the rate at which its abundance changed each day when compared to the average of the normal growth references. A second, independent experiment, also with triplicates, reproduced the data generated by experiment number one with remarkable accuracy. The parameters that were compared were the relative abundance of each population, the corresponding relative fitness, and the overall parasitaemia (Fig. 4.4A-C).

While both the relative abundance and the fitness parameters were highly reproducible between technical and biological replicates (Fig. 4.4A, B), only the latter, i.e. the shape of a growth curve, provided a quantitative measure for the fitness of a mutant. The relative abundance of a mutant within a pool was less informative since it was influenced by factors such as any local variation in recombination rates.

Fitness values of all four normal growth references were very similar and centred around the value one, i.e. their average, throughout time (Fig. 4.4D). Conversely, the average fitness of the attenuated references PBANKA_110420, PBANKA_140160, and *plasmepsin IV* was 0.65, 0.60, and 0.73, respectively, varying between 0.46 and 0.79 across all time points (Fig. 4.4 G-I). These significantly reduced fitness values (p-values < 0.05 for most time points, appendix V), confirmed my choice of attenuated reference mutants as adequate. As a result, these seven references were included in all STM experiments performed from this point onwards.

A closer look at the fitness of ePK mutants such as the CDPKs (Fig. 4.4E) revealed that the *cdpk3* mutants had an average fitness of 1.01 while *cdpk4* mutants were 10 % less fit. The

latter had not been detected by conventional methods where parasitaemias were counted manually [138]. Similarly, *cdpk6* mutants also had an average fitness of 0.88 but unlike *cdpk4* this decrease in fitness was significant on days 6 and 7 (appendix V). These findings reinforced the idea of a dynamic interplay between parasites and the host during infection that generates a multitude of phenotypes that can change daily, i.e. with every cycle. Despite being present in the input sample, no barcode counts were detected for either *cdpk5* or *cdpk7* in any of the replicates, an outcome that matched previously described data [105] where the disruption of these genes did not yield viable mutants. The same result, i.e. high number of counts in the input sample and complete absence in the experimental samples in this or other reverse genetics studies, was observed for other targets such as *fikk* and *crk4*, clearly strengthening the credibility of this approach. Quite surprising was the detection of the *cdpk1* KO mutant that, at the time when I performed these experiments, was still considered likely essential for asexual stage development [62]. In accordance with previously published data [129] the *map* kinase mutants (*map1* and *map2* KO) had growth curves very similar to the normal growth references and average fitness values of 1.02 and 1.00, respectively (Fig. 4.4F).

Importantly, barcode counting yielded reproducible fitness measurements that were independent of the relative abundance of a vector in the transfection pool as shown in Figure 4.4J. In this Figure it is also shown that the distribution of fitness values varies between 1.00 and around 0.50. Presumably lower fitness values could not be measured *in vivo* where fitter parasites will have outcompeted very slow growing mutants by the time the infection becomes patent.

For the analysis presented here, mutants were considered viable when counts for a given barcode were consistently present in every time-point for at least two out of three replicates.

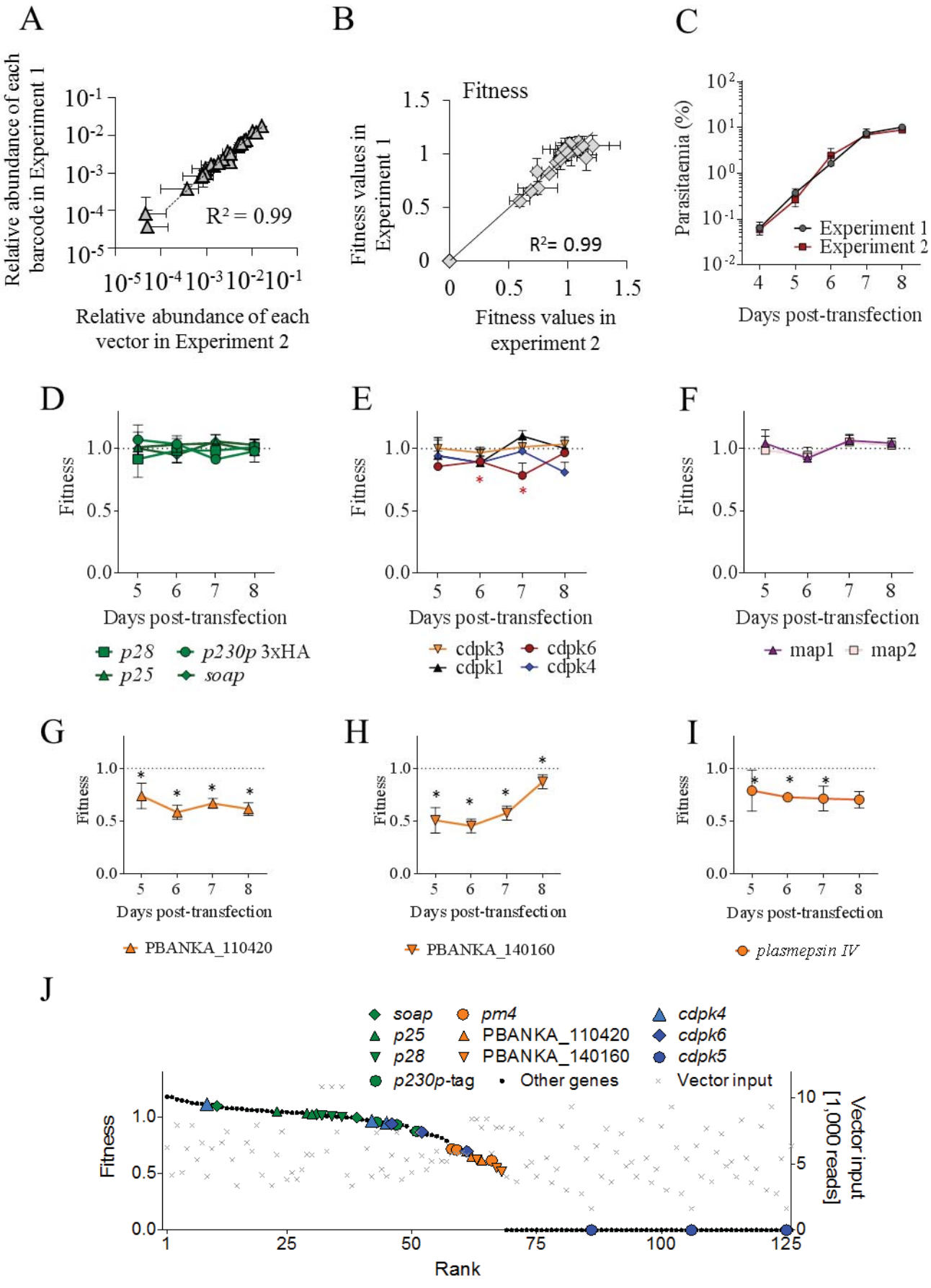

A
Relative abundance of each barcode in Experiment 1
R² = 0.99
Relative abundance of each vector in Experiment 2
B
Fitness
Fitness values in Experiment 1
R²= 0.99
Fitness values in experiment 2
C
Parasitaemia (%)
Experiment 1
Experiment 2
Days post-transfection
D
Fitness
Days post-transfection
p28
p230p 3xHA
p25
soap
E
Fitness
Days post-transfection
cdpk3
cdpk6
cdpk1
cdpk4
F
Fitness
Days post-transfection
map1
map2
G
Fitness
Days post-transfection
PBANKA_110420
H
Fitness
Days post-transfection
PBANKA_140160
I
Fitness
Days post-transfection
plasmepsin IV
J
soap
pm4
cdpk4
p25
PBANKA_110420
cdpk6
p28
PBANKA_140160
cdpk5
p230p-tag
Other genes
Vector input
Fitness
Vector input [1,000 reads]
Rank

Fig. 4.4| STM revealed a range of growth phenotypes.

(A) Pearson correlation analysis of average abundance values for each barcode, on day 7 post-transfection, of two independent STM experiments. (B) Pearson correlation analysis of average fitness values for each barcode, on day 7 post-transfection, of two independent STM experiments. Fitness was calculated from the replication rate of each barcode from day 6 to day 7 post-transfection relative to the average of the reference genes on the same period of time. (C) Overall average parasitaemia of the same two independent STM experiments. Note that the curves nearly overlap. (D) Fitness over time of the normal growth reference genes. Their consistency over time was striking, hence confirming that their choice as references was adequate. Fitness of reference mutants averages 1 by definition. (E, F) Fitness over time of selected mutants as labelled. The *cdpk1* gene had previously been considered essential for asexual development but it was not significantly less fit than the normal growth references; unlike the *cdpk6* KO parasites. (G-I) Fitness over time of the attenuated references. Statistical analysis indicated that these were significantly less fit than the normal growth references at nearly every time-point.
Error bars show standard deviations of the mean (n=3). * Different from reference mutants as determined by a two sided T-test corrected for multiple testing ($p<0.05$). (J) Distribution plot generated from a ranked list of day-6 fitness values measured for each gene in Experiment 1 (left axis). The relative abundance of a targeting vector in the electroporation cuvette at the moment of transfection, given by the input sample (grey crosses, right axis) did not predict whether a mutant could be obtained.

4.2.2 Comparison between barcode counting and a conventional deletion analysis

Different STM experiments yielded data for 46 ePKs for which there were published data available [105]. Barcode counting agreed with the previous study for 35 out of 47 ePK genes (76%), of which 15 were targetable and 20 were classified as likely essential ePKs genes. In the Tewari *et al* study [105], used as benchmark for this study, the authors targeted two ePKs that the STM approach failed to generate: *uis1* and *pk7*. It is uncertain whether such technical failures resulted from faults in individual vectors or from low recombinogenicity of the target loci. Excluding these two false negatives, the bar-seq strategy provided evidence for targetability for 25 genes, while the conventional approach targeted only 15 (plus *usi1* and *pk7*) [105]. The ten genes that had previously been considered possibly essential for erythrocytic development but were now targeted were *cdpk1*, *gsk-3*, *rio1*, *rio2*, PBANKA_082960, *tkl3*, PBANKA_141450, PBANKA_142160, *tkl1* and PBANKA_130520 [62,105]. As this corresponded to a rate of false positives of 34.5 %, I attempted to validate the first six by independently generating and carefully genotyping the mutants. This is further detailed in section 4.2.3. A table summarising this comparison between studies and including the daily fitness values for each mutant as well a description of the type of genotyping performed in each of the validation cases is presented in appendix V. The statistical analysis was also performed for each of the mutants comparing their growth rate to that of the normal growth references was also included in the above mentioned appendix.

4.2.3 Validation of false positives

If barcode counting detected non-integrated or incorrectly integrated targeting vectors, this would be a major disadvantage of the new method since it would result in false positives in STM screens. Therefore, it was important to validate mutants that were not previously considered viable. The generation and genotyping of KO clones was attempted for six of the ten new deletions detected by barcode sequencing. These were: *cdpk1*, *gsk-3*, *rio1*, *rio2*, *tkl3*, and PBANKA_082960 (Fig. 4.5). *cdpk1*, *gsk3*, PBANKA_082960 KO and *tkl3* parasites were easily generated and PFGE analysis showed clear integration of the vector in the expected chromosomes, i.e. 3, 4, 8 and 13, respectively (Fig. 4.5A). At this points *gsk3* and *cdpk1* KO parasites were cloned and further genotyping by PCR was performed (Fig 4.5 B,C) as these were two potential drug targets in *P. falciparum* [169,170]. Shortly after the generation of the *cdpk1* mutant, Jebiwott S. *et al*. reported a similar KO [146] and confirmed the phenotype at the ookinete stage, which previous work had described using a stage specific KO [62]. A short phenotypic analysis was carried out for the *gsk3* KO mutants (section 4.2.3.1). Cloned mutants for *tkl3* and PBANKA_082960 were generated as well but for genotyping purposes only. PCR data and gene maps are depicted in appendix VIII.

Kinases of the RIO family are atypical protein kinases involved in cell cycle progression and ribosome biogenesis [171–173] and have been considered essential in yeast and human cells [171–173]. These kinases have also been considered as possibly essential in both *P. berghei* [105] and *P. falciparum* [125]. However, barcode counts for both kinases, although sometimes at levels lower than the noise threshold, were consistently detected in different experiments.

The obtained RIO kinase mutants were harder to genotype and interpret. Unlike the other mutants, where harvesting took place on day 8 post-transfection, drug resistant parasites took until day 10 post-transfection to reach a parasitaemia of 2 % after transfection with a *rio1* KO vector. When targeting the *rio2* gene, parasites emerged even later, on day 14 post-transfection. PFGE results showed integration of the vector into chromosome 5 for *rio2*, as expected, but for the *rio1* KO mutant population, the intensity of the expected band on chromosome 14 was very weak (Fig. 4.5A). Since the 3'UTR of *Pbdhfr* was used as probe, the two copies present in the vector that flank the resistance cassette should have yielded twice the intensity of the endogenous sequence (internal control) on chromosome 7, which was clearly not the case, suggesting that not all parasites had integrated the vector. PCR

genotyping (Fig. 4.5D, E) showed the presence of the *rio* KO vectors (lanes 401/216, 403/218 and 396/218, 398/216)[8], but remained inconclusive since no integration PCR product could be obtained (lanes 405/216 and 400/218). In addition, both *rio* KO populations were still PCR positive for their respective WT sequence. Four attempts – two for each strain – at dilution cloning single parasites failed, and reducing the dilution factor to five parasites per mouse produced what appeared to be mixed populations of mutant and WT (not shown). Taken together these findings suggested that deletion of *rio1* or *rio2* produced mutants that were too unfit to produce infections as clones on their own. To characterise the target loci of both mutants further, non-clonal populations resulting from two different infections following dilution cloning for each mutant where subjected to whole genome sequencing (WGS).

Paired-end reads for the four libraries were mapped to the *P. berghei* ANKA reference using the Burrows-Wheeler Aligner (BWA) algorithm, manipulated with SAMtools and viewed with Artemis. The resulting alignments are depicted in Figure 4.6 for the *rio1* and Figure 4.8 for the *rio2* mutants. Targeting of *rio1* (lines blue and red) led to a substantial drop in coverage that corresponded exactly to the deletion region expected by the KO vector (Fig. 4.7A). The *rio2* populations served as controls. This indicated that the gene was deleted in all but a small number of WT parasites, thus confirming all other genotyping data.

The KO vector for *rio2* was expected to target the entire ORF of the gene (Fig.4.7B). However, a comparison of the coverage levels between all libraries at this locus (Fig. 4.8), revealed a ~29.7 kb duplication on either side of the *rio2* gene; coverage for the *rio2* itself was down to the reference levels. These results are consistent with a scenario in which a regional duplication including *rio2* had occurred in the parental parasite, and after transfection, the KO vector deleted only one of the two copies. These data are consistent with PFGE results, i.e. the vector integrated into the right chromosome, but also with the PCR data, where persistence of the WT locus was shown. Given these results, I speculated that, unlike *rio1*, *rio2* is likely to be essential and the integration of the KO vector in this locus was only possible due to the duplication. This also explained the delayed patency of these mutants as this duplication was, most likely, a very rare event.

[8] Gene maps illustrating binding sites of the different primers are available in appendix VI.

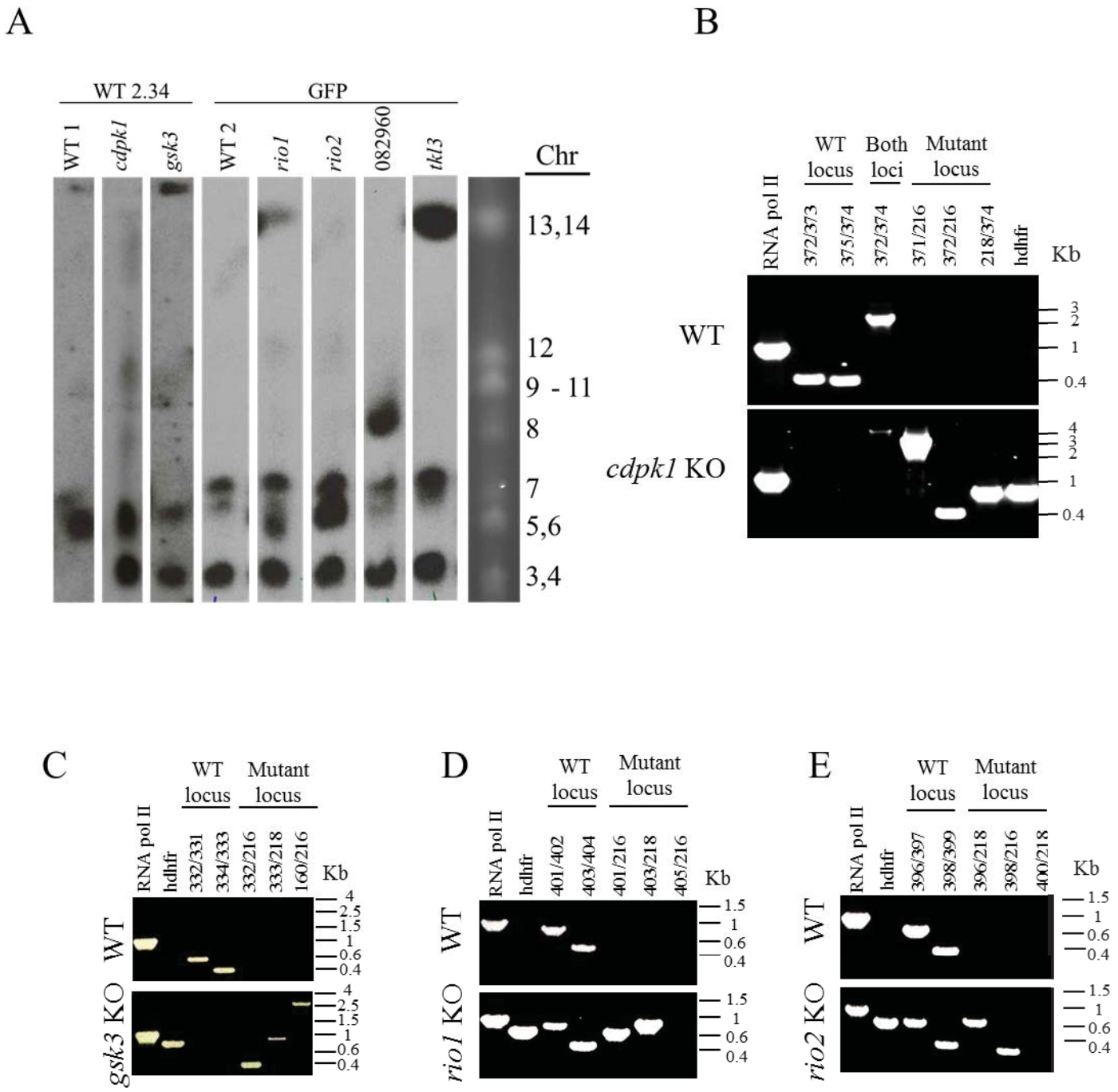


Fig. 4.5| Genotyping of the newly obtained mutants *cdpk1* KO, *gsk3* KO, PBANKA_08296 KO, *tkl3* KO, *rio1* KO and *rio2* KO.

(A) Southern hybridisation of chromosomes separated by PFGE with a ~ 500 bp probe for the 3' UTR of Pb*dhfr-ts*, as before. Note that there is an exogenous site in on chromosome 3 in the background strain labelled WT2 (GFP parasites). (B-E) PCR genotyping for *cdpk1* KO, *gsk3* KO, *rio1* KO and *rio2* KO to support the PFGE data shown in A. The first lane targeted *rna polymerase II* and was used as a positive control; the second lane targeted the resistance cassette and therefore should always be positive in the mutant but not in the WT lanes. The "WT locus" PCRs span either the 3' or 5' end of the genes. "Mutant locus" PCRs primed both inside the drug resistance cassette and outside the ORF. Maps showing the annealing sites of the genotyping primers are present in appendices VI and VII for the *rio* and *gsk3* KOs and in Figure 5.1A for *cdpk1* KO. PCR genotyping evidence as well as gene maps for PBANKA_082960 KO and *tkl3* KO can be found in appendix VIII.

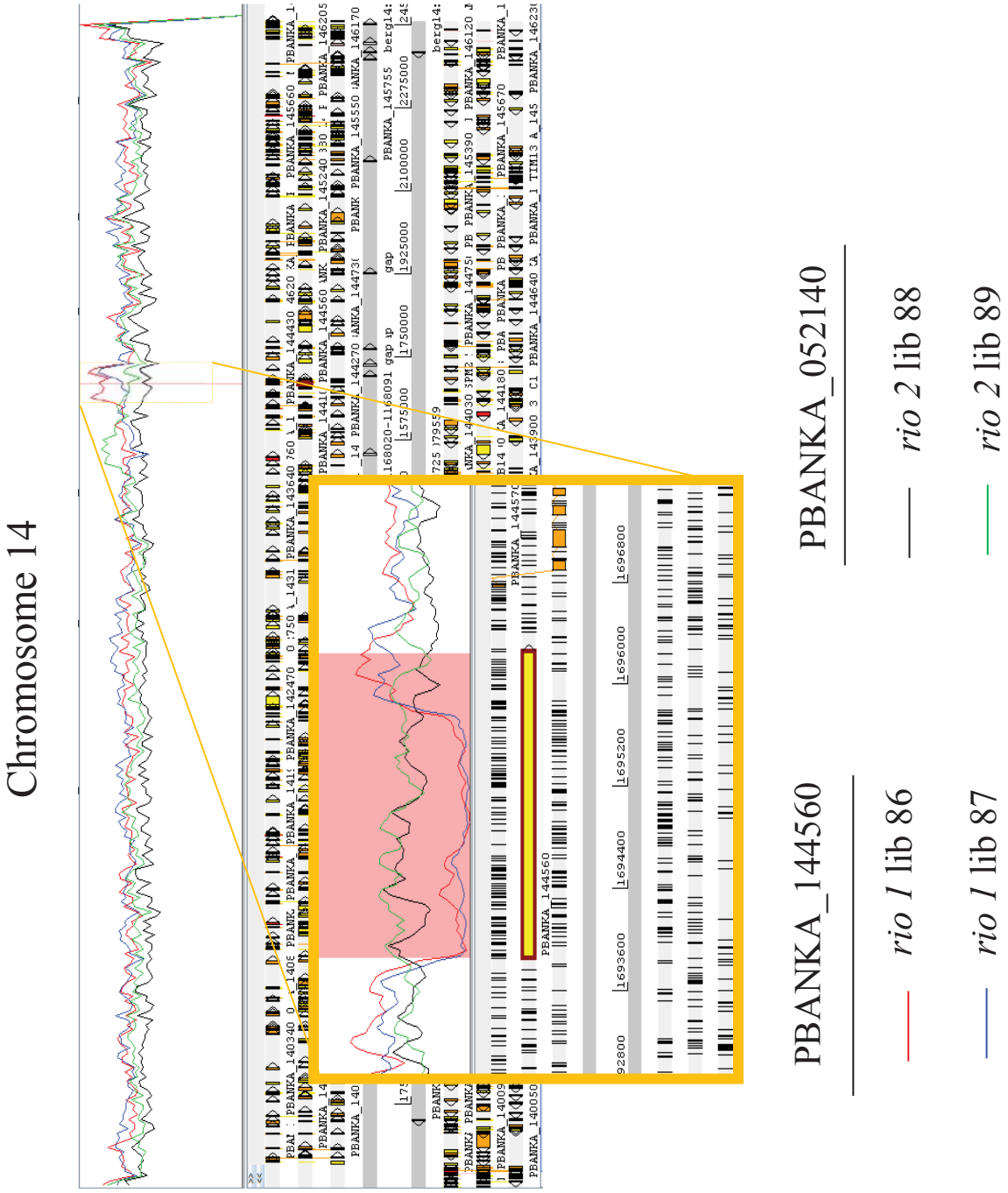
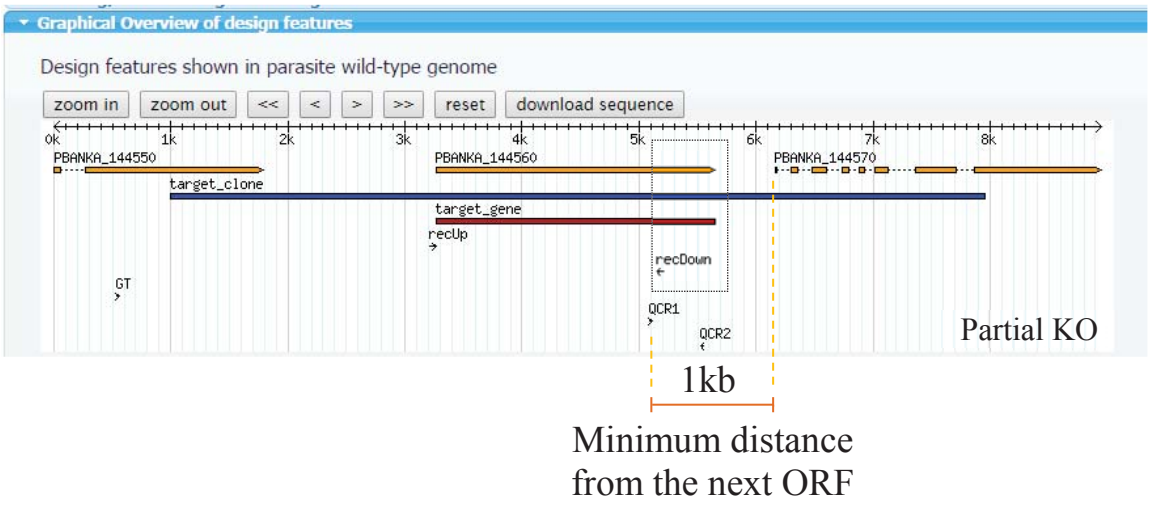


Fig. 4.6| WGS of RIO kinases - *rio1* locus is disrupted in the *rio1* mutants.
Mapped reads were visualised with the Artemis software. The Figure shows the coverage plots for the entire chromosome 14. Each line corresponds to each of the sequenced mutants (individual libraries – 86 to 89). Coverage across the chromosome was very even except for the KO region where the drop for libraries 86 and 87 indicated that *rio1* gene had been deleted. The rise in coverage at the last 500 bp of the gene is a result of the vector design, i.e. the vector did not target this region (please see Figure 4.7A). The pink highlighted region in the zoomed rectangle spans the *rio1* ORF (PBANKA_144560).
The numbers that are not preceded by "PBANKA_" are chromosome coordinates.

A

rio1 KO vector design



B

rio2 KO vector design

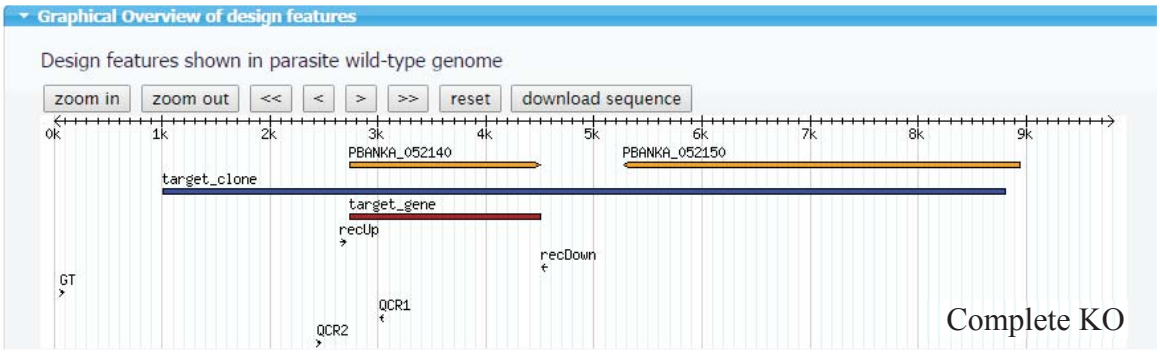


Fig. 4.7| RIO kinases KO vector designs.
Vector designs generated by the *Plasmo*GEM database for *rio1* and *rio2* genes (A and B, respectively). The standard *Plasmo*GEM KO designs aim to target the entire ORF. A few exceptions were implemented to include cases when: (1) There is not a library clone that covers the entire ORF; (2) One of the homology arms is less than 1 kb in length, in which case the design is adjusted to prioritise the homology arm; (3) There is not a minimum distance of 1 kb from the neighbour gene in which case the design is adjusted to avoid interfering with the other ORF's expression. This distance could be reduced to 0.8 kb if the downstream gene was in the reverse orientation, i.e. 1 kb from the next 5'UTR and 0.8 kb from the next 3'UTR. "recUp" and "recDown" represent the 50 bp homology regions between which the drug resistance cassette was inserted. (A) The design for *rio1* KO was included in the third category of exceptions and, as a result, the last ~500 bp were not included in the deletion region. (B) Design for *rio2* KO vector, showing the entire ORF was targeted for removal.
GT, QCR1 and QCR2 are quality control primers, automatically designed by the software.

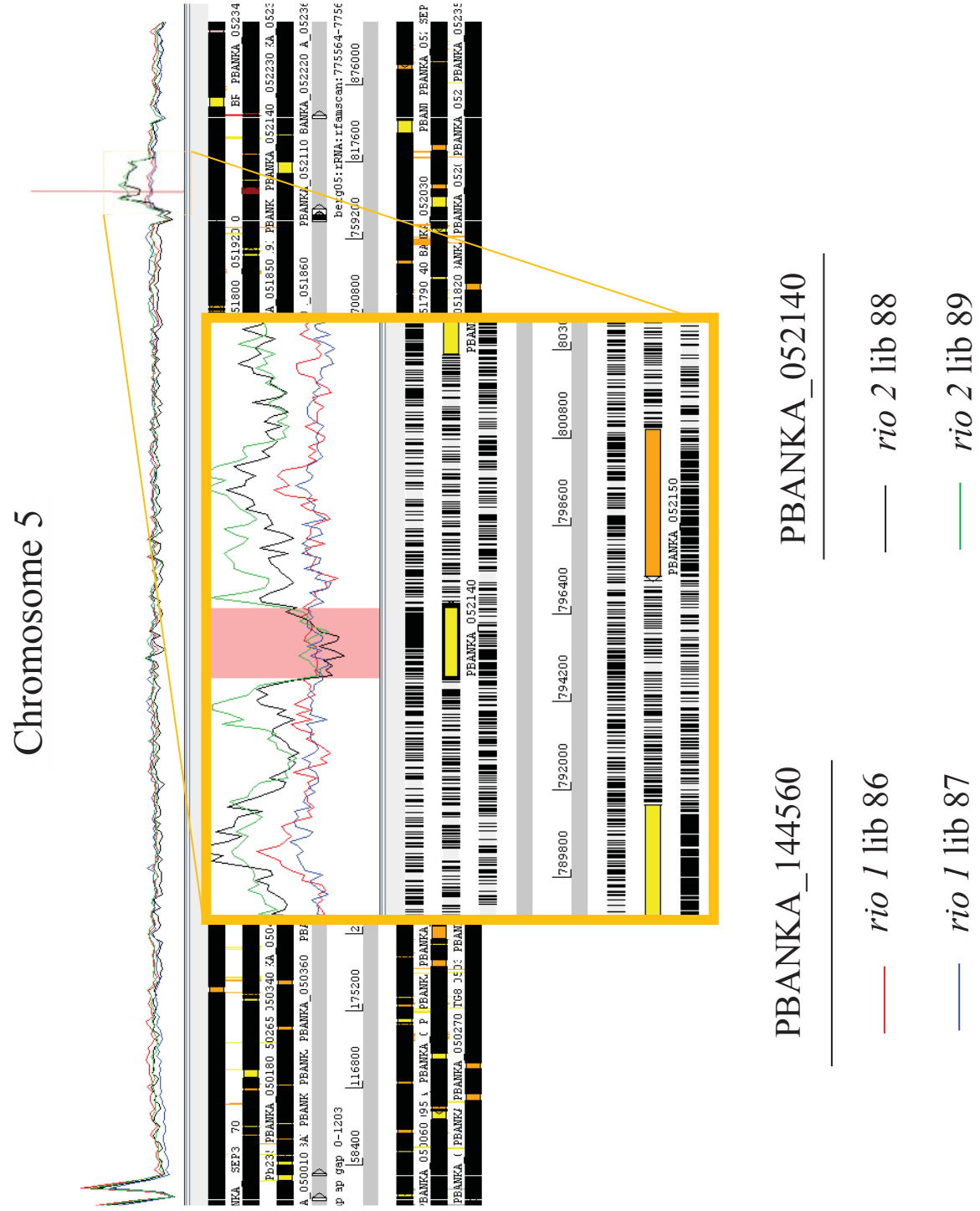


Fig. 4.8| WGS of RIO kinases - *rio2* locus.
Mapped reads for all mutant lines were visualised with the Artemis software. The Figure shows the coverage plots for the entire chromosome 5. Each line corresponds to each of the sequenced mutants (individual libraries – 86 to 89). Coverage across the chromosome was very even except for a ~29.7 kb region spanning the *rio2* gene, where a duplication was detected. This only affected the *rio2* libraries (lines green and black). This duplication did not include the *rio2* gene as the coverage for the ORF was down to the reference levels which meant that the targeting vector deleted one copy but not the other. The pink highlighted region in the zoomed rectangle spans the *rio2* ORF (PBANKA_052140).
The numbers that are not preceded by "PBANKA_" are chromosome coordinates.

4.2.3.1 *gsk3* KO phenotyping

As the *gsk3* gene had been considered essential for erythrocytic development by others [105], careful genotyping by PFGE (Fig. 4.5A), PCR (Fig. 4.5C) and Southern blot (Fig. 4.9A) confirmed that the *gsk3* gene had been deleted.

Barcode counting determined the average relative fitness of this mutant to be 1.01 for days 5-8 post-transfection (Fig. 4.9B). Interestingly, when comparing the growth pattern of a *gsk3* KO clone to the WT parasites for a longer period of time, a striking phenotype was revealed (Fig. 4.9C). On days 1-5, roughly the equivalent window of days 5-8 post-transfection in the screen in terms of overall parasitaemia, both curves seem to evolve evenly. However, beyond this time-point the growth pattern of the *gsk3* KO clone completely diverged from the WT. While the WT population increased over time until day 7 post-infection, the opposite was seen for the mutants. Only from day 8 onwards did their population increase.

Observation of Giemsa stained thin blood films indicated that the mutant parasites could only persist in reticulocytes. Figure 4.9D shows a snapshot of those smears, where blue arrows point at infected normocytes and black arrow at reticulocytes. Interestingly, the increase in parasitaemia registered from day 8 post-infection was accompanied by an increase in reticulocytaemia (~7 % on day 9 post-infection), presumably derived from the anaemia induced by the long-term infection.

The mutant parasites persisted in the mice for 16 days. At this point, severe hepatomegaly and splenomegaly, and severe anaemia became critical. Liver and spleen enlargement are some of the hallmarks of malaria pathology, but in these mice this was exacerbated, perhaps due to the unusual length of the infection (Fig 4.9E).

No significant differences were found between the *gsk* KO mutants and the WT for the number of exflagellation centres per 1000 RBCs (Fig. 4.9F), ookinete conversion rate (Fig. 4.9G), number of oocysts per midgut (Fig. 4.9H). However, the number of sporozoites per mosquito (Fig. 4.9I) was much reduced (p-value= 0.00022). Sufficient numbers of sporozoites reached the salivary glands and no delay in patency (day 4, for WT and *gsk3*, data not shown) was observed. These experiments were performed using a single clone and will therefore need to be reproduced with a second independent mutant.

Localisation experiments were performed by immunofluorescence of *gsk3*-3xHA tagged parasites but only a very diffuse and inconclusive signal was detected in the cytosol of schizonts, trophozoites, gametocytes and ookinetes (data not shown).

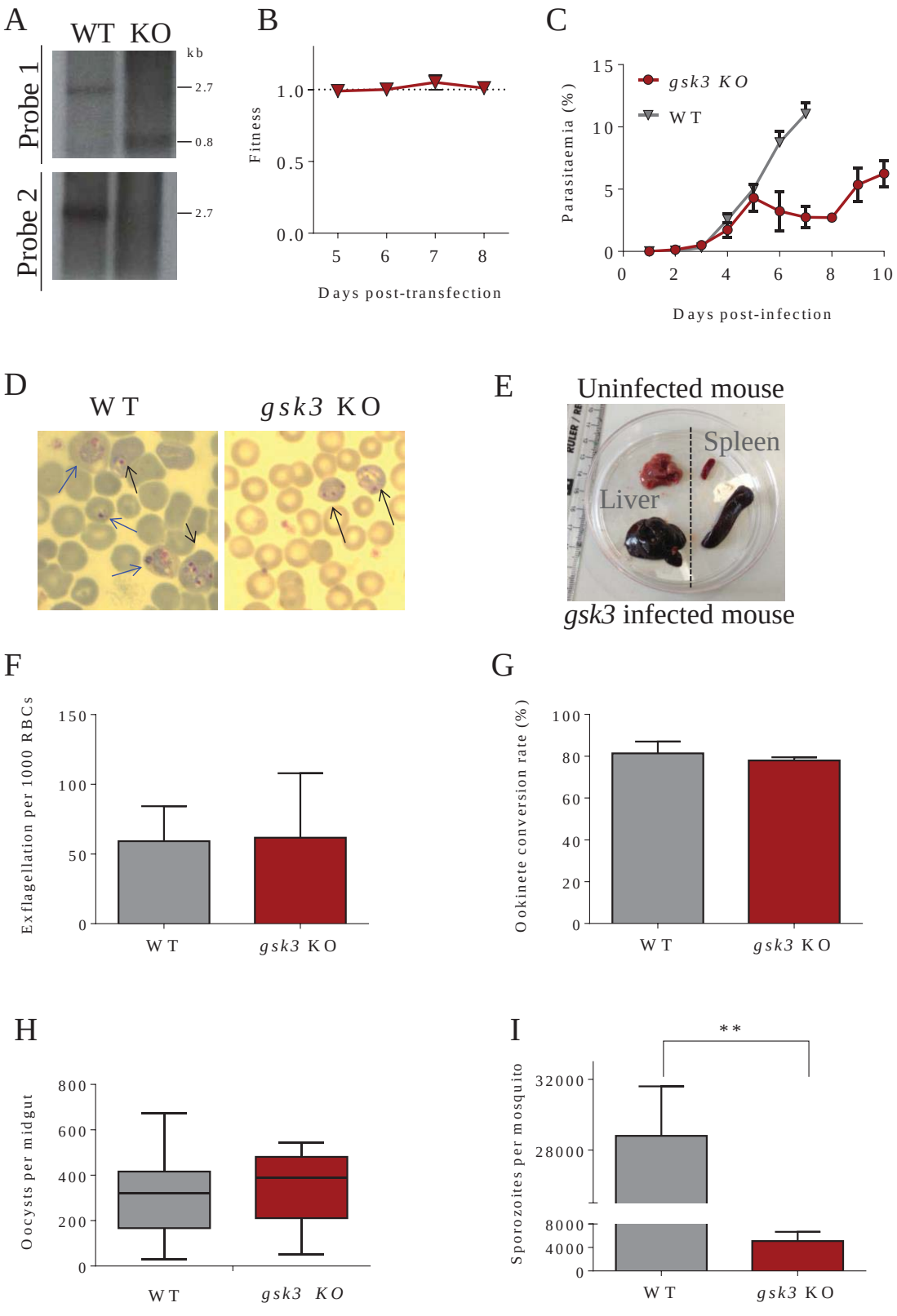


Fig. 4.9| *gsk3* KO genotyping and phenotypic analysis.
(A) Southern blot analysis of the *gsk3* locus to confirm gene disruption. Two probes were used: probe 1 to detect a difference in size between the WT (2.7 kb) and modified (0.8 kb) loci and probe 2 to hybridise only with WT-

specific sequences. A schematic representation of the locus with restriction sites used can be found in appendix VII. (B) Fitness over time for *gsk3* KO mutants obtained from an STM experiment showed that for early/low infections there was no fitness disadvantage. (C) Growth curves for the independently generated mutant, calculated from Giemsa stained thin blood smears prepared daily. Each mouse was injected with 10^6 infected RBCs. Error bars show standard deviations of the mean (n=3). During the first five days of this experiment where the parasitaemia was equivalent to the one in the STM experiments shown in B, the mutant and the WT growth curves were very similar. The attenuated phenotype was only detected beyond this time window. (D) Giemsa stained smears showing infected normocytes (blue arrows – the three left most arrows) and infected reticulocytes (black arrows). Reticulocytes are slightly larger than normocytes and stain purple due to a high content of nucleic acids. (E) Illustration of necrosis and enlargement of the liver and spleen of a *gsk3* KO infected mouse on day 16 post-infection. The other two replicates had a similar size and colouration. Imaging was performed *post-mortem*. (F-I) Phenotypic analysis of *gsk3* KO sexual stages: exflagellation centres per 1000 RBCs (F); ookinete conversion rate (n=3 cultures) (G); number of oocysts per mosquito midgut, dissected on day 12 post-feed (n=10 mosquitoes) (H); and number of sporozoites per mosquito, dissected on day 21 post-feed (n=10 mosquitoes) (I). * Different from the WT as determined by a two sided T-test (p-value= 0.00022).

4.3 Discussion

In this chapter I have shown that the *P. berghei* adapted STM strategy proposed in the previous chapter was fully validated. Parallel transfection of barcoded vectors proved to be a powerful strategy to screen large numbers of genes in *P. berghei*. Mutant barcoding enabled rapid and reliable measurement of the relative abundance of mutants in pools across three orders of magnitude and how it changed during infection. Fitness is a phenotype measurement central to many large-scale genetic studies and was originally measured in terms of population allele frequencies [101]. In this study, it was successfully calculated from the growth rates of each mutant line, as previously described [101], and used to assess the cost of each gene deletion.

The screen presented here, permitted the identification of nine kinases that are dispensable for blood stage development and had not been targeted before. This was likely due to a combination of factors that include (1) a higher integration efficiency of the *Plasmo*GEM vectors, (2) better transfection operating conditions and (3) higher sensitivity of the detection method. These discrepancies reinforce the idea that the absence of evidence for genomic integration is insufficient proof for the essentiality of a gene. Additional experiments such as gene tagging can supply supplementary information about the locus accessibility.

Five of the newly targeted genes were validated with generation of the independent mutants which included *cdpk1* and *gsk3* kinases, current targets of drug development research [143,169]. Data from these validation experiments confirmed the ability of barcode sequencing to identify new mutants and failed to detect evidence that STM screens will generate large numbers of false positives, as might be expected if *Plasmo*GEM vectors replicated as episomes or integrated by non-homologous recombination.

Eukaryotic ribosomes or 80 S ribosomes are so called due to their sedimentation coefficient (S). They are comprised of two unequal subunits, a small subunit (40 S) and a large subunit (60 S). Each of them contains multiple ribosomal proteins embedded on a scaffold of ribosomal RNA (rRNA). The small subunit is further assembled from 18 S rRNA and 32 ribosomal proteins [174]. The RIO (right open reading frame) family of atypical protein kinases plays an important role in ribosome biogenesis and is subdivided into three subfamilies Rio1, Rio2, and Rio3. Every organism from archaea to humans has both *rio1* and *rio2* kinases, while *rio3* genes are only found in multicellular eukaryotes [124,171]. In *S. cerevisiae* there are two *rio* kinases, *rio1* and *rio2*. Both were identified as necessary for processing of 20 S pre-rRNA into mature 18 S rRNA. Depletion of either of *rio* kinases leads to accumulation of 20 S pre-rRNA and cell cycle arrest [172]. In addition, *rio1* plays an important role in cell cycle, during G1 to S transition and is involved in the control of the onset of anaphase. On the other hand, *rio2* has been proposed to promote nuclear export of pre-40 S subunits [171,175]. The fact that deletion of either *rio1* or *rio2* is lethal for the cells, suggests that they perform distinct functions [103].

In malaria parasites these kinases have previously been considered essential for blood stage development [105,125]. However, the current study suggested that the *rio1* gene can be targeted. Despite the failure in obtaining a pure population of deletion mutants, WGS data clearly showed that the vast majority of the parasites had that gene deleted.

The *rio2* gene, on the other hand, is likely to be essential in *P. berghei* parasites as it is in other organisms. Interestingly, a random duplication of a segment of gDNA that included the *rio2* gene enabled the integration of the targeting vector into one copy of the duplicated loci. This allowed parasites to become resistant to pyrimethamine while still carrying a functional copy of the *rio2* gene. Random amplification of large regions of DNA characterises a remarkable phenomenon that has recently been shown to occur randomly amongst individual parasites. This is part of a two-step strategy, identified to be responsible for drug resistance development [176]. Step one involves random amplifications that include genes that confer resistance to a given antimalarial and therefore allow the parasite to survive. Then, to ensure tolerance of increasing drug pressure, additional copies of this amplification are generated. While the original copy of a given gene ensures the native role is still carried out, each of the other copies that generate a pseudo-ploidy state might carry mutations that enhance drug resistance or increase parasite fitness in a given context. Loss of these

amplification regions takes place when pressure is alleviated or when its fitness cost is no longer advantageous [176].

This is a very rare event that drives development of drug resistance [176] and, in this case, was the cause for the only confirmed false positive in the STM screen.

GSK3 was first described as a protein kinase that phosphorylates glycogen synthase [177]. Since then it has been implicated in a number of processes like metabolic control, embryonic development, cell proliferation and adhesion, Alzheimer's disease, circadian rhythm, and oncogenesis [178]. Due to its implication in several human diseases this multifunctional enzyme has been targeted for drug development research. Likewise, its likely essential role in *Plasmodium* development [105,145] has prompted the search for specific inhibitors [169]. In fact, treatment of *P. berghei* infections with LiCl led to a reduction in parasitaemia relatively comparable to treatment with chloroquine [179], a drug once widely used to treat *P. falciparum* infections but currently only recommended for the treatment of malaria caused by *P. vivax*, *P. malariae* and *P. ovale*.

Co-localisation experiments in *P. falciparum* parasites have shown the presence of GSK-3 in Maurer's clefts, within the erythrocyte cytoplasm, as vesicle-like structures [169]. Maurer's clefts are membranous structures that are present in *P. falciparum* infected RBCs [180]. These have a crucial role in virulence as they mediate protein sorting and export and have important implications on (1) the adherence properties of the infected RBC and consequently on splenic clearance, and (2) on the RBC permeability as a means to allow the acquisition of nutrients [181]. This observation regarding GSK3 localisation prompted the hypothesis of it being involved in protein trafficking. Localisation experiments with an epitope tagged allele of (Pb*gsk3*-3xHA) failed to reproduce such pattern (data not shown). At the time of these experiments this was not surprising since presence of Maurer's clefts-like structures had not been reported in *P. berghei*. However, recently biochemical and microscopic data revealed the presence of intra-erythrocytic membranous structures termed *P. berghei*-induced structures (IBIS). Although morphologically different from the Maurer's clefts it indicates that the rodent parasite also creates an intracellular network in infected RBCs [182,183]. In light of these recent developments, follow-up experiments on the localisation of PbGSK-3 will include a replacement of the HA epitope tag with a fluorophore such as mCherry to allow live imaging of this protein throughout the lifecycle.

Barcode counting experiments revealed that unlike what was previously suggested [105], *gsk3* can be disrupted in *P. berghei*. Generation of a thoroughly genotyped, cloned KO

mutant validated such results. Interestingly, a brief phenotypic analysis suggested that these mutants have a strong preference for reticulocytes.

Reticulocytes encompass a minor though heterogeneous population of RBCs precursors that can be grouped according to their levels of the transferrin receptor (CD71) expression. They mature into normocytes (normal RBCs) over a period of 72 hours during which these cells undergo profound changes such as loss of surface receptors and reticular matter. Some *Plasmodium* species have a preference for these young RBCs over normocytes. For instance, in the presence of equal numbers of erythrocytes and reticulocytes, *P. berghei* is ~150 times more likely to infect a reticulocyte [184]. Among the species infecting humans, *P. falciparum* is able to invade RBCs of all ages while *P. vivax* is restricted to reticulocytes. This tropism towards different sets of RBC fractions has a major impact on the course of infection and pathology which is why *P. falciparum* infections are life-threatening.

Preliminary phenotypic experiments suggested that the disruption of the *gsk3* gene exacerbated the reticulocyte preference of *P. berghei* parasites. Supporting this finding was the observation on Giemsa stained thin blood smears of (1) reticulocytes as the only infected RBCs and (2) a drop in parasitaemia between days 5 and 8 post-infection, i.e. moment when reticulocytes in circulation become limiting. It was unclear whether the *gsk3* mutants could not fully develop in normocytes or not invade them at all. The drop in parasitaemia, on the other hand, was attributed to the suppression of the erythropoietic response. The latter is a phenomenon that has been reported to occur during malaria infections, presumably as a host protective mechanism [184,185]. This response has not been seen during non-malaria anaemias and is only alleviated once haemoglobin levels drop below critical levels, which correlates with the increase in parasitaemia from day 8 post-infection onwards. As a consequence of this reticulocyte preference, *gsk3* mutants could persist for at least twice as long in the rodent bloodstream when compared to the WT *P. berghei* strains. Perhaps as a consequence of this prolonged infection, the liver and spleen of these mice showed extensive morphological signs of failure. A histological study of these organs would be required to best characterise this finding.

Furthermore, experiments that would help clarifying the nature of this phenotype would include complementation of the modified locus and validation of the reticulocyte preference phenotype through invasion assays with normal and reticulocyte enriched blood. The drop in parasitaemia from day 5 post-infection onwards is also consistent with adaptive immune clearance. It would therefore be interesting to access parasite growth in RAG mutant mice, where the disruption of the *rag* gene, involved in V(D)J recombination, prevents B and T cell

differentiation [186]. On the other hand, immunological profiling of cytokines during early and late stages of infection could also be quite informative.

In mammalian cells, GSK3 plays a central role in the control of proliferation through the repression of transcription factors such as CREB, NFκB and AP-1 [187]. Thus, transcriptome analysis of *gsk3* KO parasites could potentially reveal a similar function for this kinase in *Plasmodium* parasites.

Chapter 5

A genetic interaction screen reveals a new signalling pathway

5.1 Introduction

The clear increase in scale that parallel phenotyping by barcode sequencing provided offers new opportunities for reverse genetic screening that were previously unachievable for malaria parasites. One of these is the study of genetic interactions at scale. The last aim of this project was therefore to demonstrate that such studies are now possible in *P. berghei* parasites.

Genetic interactions result from the interplay between genes that have a function either within the same pathway or between interconnected pathways. In model organisms, systematic screens for genetic interactions have provided deep insights into biological pathways [100,102].

Since I have demonstrated in Chapter 3 that in *P. berghei* the likelihood of manipulating two loci simultaneously is negligible, this kind of study requires sequential manipulations of the genome. This has become more feasible in *P. berghei* parasites since the development of a positive-negative recyclable selection cassette [54]. As this cassette is part of all *Plasmo*GEM vectors, I generated a panel of marker free mutants that were used as background lines to repeat the ePK deletion screen described in Chapter 4. This approach allowed me to look at the fitness of dozens of double mutants in the same experiment and revealed a case of epistasic genetic interaction in the *P. berghei* kinome.

5.2 Results

5.2.1 Choice of the genetic backgrounds

CDPKs are protein kinases that are found in apicomplexan parasites and plants. This makes them very attractive targets for antimalarial drug development [139,170,188]. *P. berghei* parasites have six CDPKs. This family of protein kinases has been shown to control important signalling processes in these parasites such as motility, development and egress [136,138,144,189]. Due to their pivotal role in so many aspects of the parasite's life cycle it is surprising that four out of six are dispensable for blood stage development. In fact, *cdpk1* and *cdpk4*, were even reported to be expressed at the protein level in blood stages [34,62]. We therefore hypothesised that this seeming genetic redundancy among the CDPKs was, in fact, the result of genetic buffering, i.e. the existence of compensatory effects amongst

them. For this reason the targetable CDPKs were chosen as the genetic backgrounds against which genetic interactions were searched for. Additionally, two other genetic backgrounds were selected: a resistant allele of the *pkg* gene (*pkg*^{T619Q})[9], and a double KO mutant for the MAP kinases.

The *pkg* gene encodes a protein kinase that is activated by cGMP. The binding of this secondary messenger promotes the phosphorylation of a number of biologically important targets in higher eukaryotes, being implicated in the regulation of smooth muscle relaxation, platelet function, sperm metabolism, cell division, and nucleic acid synthesis [190].

In *Plasmodium*, *pkg* is an essential gene for blood stage development. Therefore, its function has been studied through a chemical genetics approach. According to this strategy, a mutation of the amino-acid that controls the access of small molecules to the ATP-binding pocket, the so-called gatekeeper residue, generates a mutant line that is selectively resistant to an inhibitory compound without compromising the functionality of the enzyme [191]. Parasite growth is strongly impaired in the presence of Compound 1 (a trisubstituted pyrrole 4-[2-(4-fluorophenyl)-5-(1-methylpiperidine-4-yl)-1H-pyrrol-3-yl] pyridine (C1)), a reversible but potent and selective inhibitor of PKG. A single point mutation in the *pkg* gene to replace the small threonine residue with a bulky glutamine one (T619Q in *P. berghei* and T618Q in *P. falciparum*) is sufficient to render parasites resistant to this chemical without affecting parasite growth [192,193].

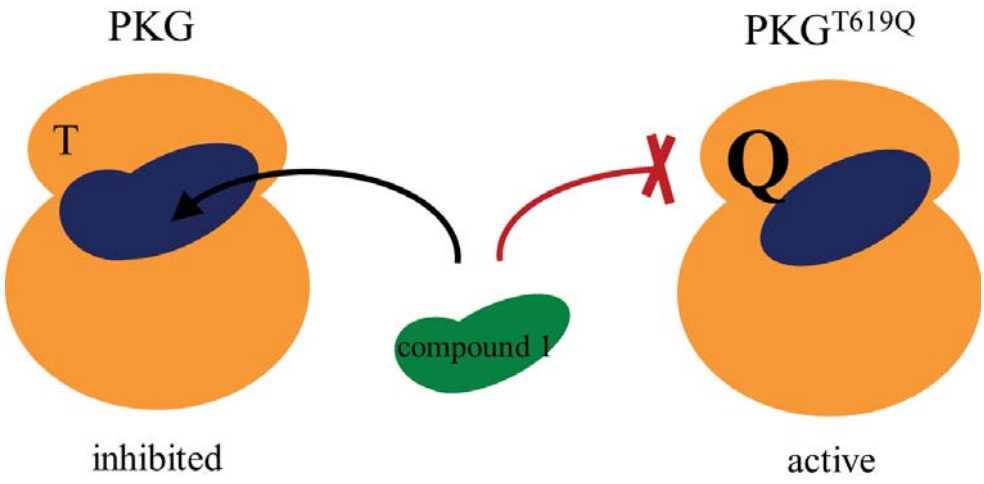


Fig. 5.1| The mutation of the gatekeeper amino-acid renders PKG resistant to Compound 1.
PKG has a small threonine (T619 in *P. berghei*) gatekeeper residue that when mutated into a glutamine (Q) residue, prevents the binding of the compound. This compound selectively inhibits the parasite PKG and not the host since the metazoan PKG has a large gatekeeper.

[9] This line was kindly provided by Mathieu Brochet [154].

Comparative analyses of the effect of this compound in WT and resistant lines revealed that PKG plays crucial roles throughout development. For instance, transient inhibition of PKG impairs schizogony [192] and gametogenesis [194] of *P. falciparum* parasites. Furthermore, it has also been shown that in *P. berghei* gametocytes, stimulation with XA is followed by mobilisation of intracellular calcium through the activation of PKG and possibly PI-PLC [154]. The observation that PKG activation is required for regulation of cytosolic levels of calcium, which in turn regulate CDPKs, led us to include the resistant *pkg*^{T619Q} allele in the screen.

The MAP cascade is highly conserved in eukaryotes. MAP kinases are responsible for regulation of important cellular processes, like mitosis, differentiation, and cell survival. They are activated through extracellular stimuli, such as mitogens, osmotic stress or pro-inflammatory cytokines [127]. However, very little is known about its role in *Plasmodium* parasites as they lack the canonical upstream regulators and downstream effectors of these kinases [119]. The overexpression of Pb*map2* transcripts in mutants lacking Pb*map1* (Volker Heussler, personal communication) suggested the existence of a compensatory effect if not an interaction between these two genes. Given the relevance of these kinases in other organisms and our gap in knowledge about this pathway in *Plasmodium* parasites, I attempted to generate a double KO (dKO) mutant for both *map* kinases (see section 5.2.2.1). As this mutant was unexpectedly viable, indicating that perhaps there is a third kinase with overlapping functions, this double mutant was used as a genetic background in the genetic interaction screen.

5.2.2 Generation of selection marker free backgrounds

A total of six different genetic backgrounds were produced to look at genetic interactions between ePKs. Firstly, single KO mutants for each of the chosen genes (*cdpk1*, *cdpk3*, *cdpk4*, *cdpk6*, *pkg*^{T619Q} and *map1*) were generated. These lines were then cloned and their selection cassette was excised [54] with negative selection. A final dilution cloning step ensured that these marker free lines were clonal and suitable to be used as genetic backgrounds for the genetic interaction screen. This workflow was repeated twice to generate a double KO for the *map* kinase genes. Figure 5.1 A-D shows Southern blot and PCR genotyping results for each of the *cdpk* mutants. For each of the Southern blots I used two different probes: one that targeted a sequence present in both the WT and the mutant and another probe that only targeted the gene sequence. The former should display the difference

in the size of the locus after removal of the gene while the latter ensured that very rare cases of gene duplications (e.g. *rio2* KO mutants) would not go unnoticed.

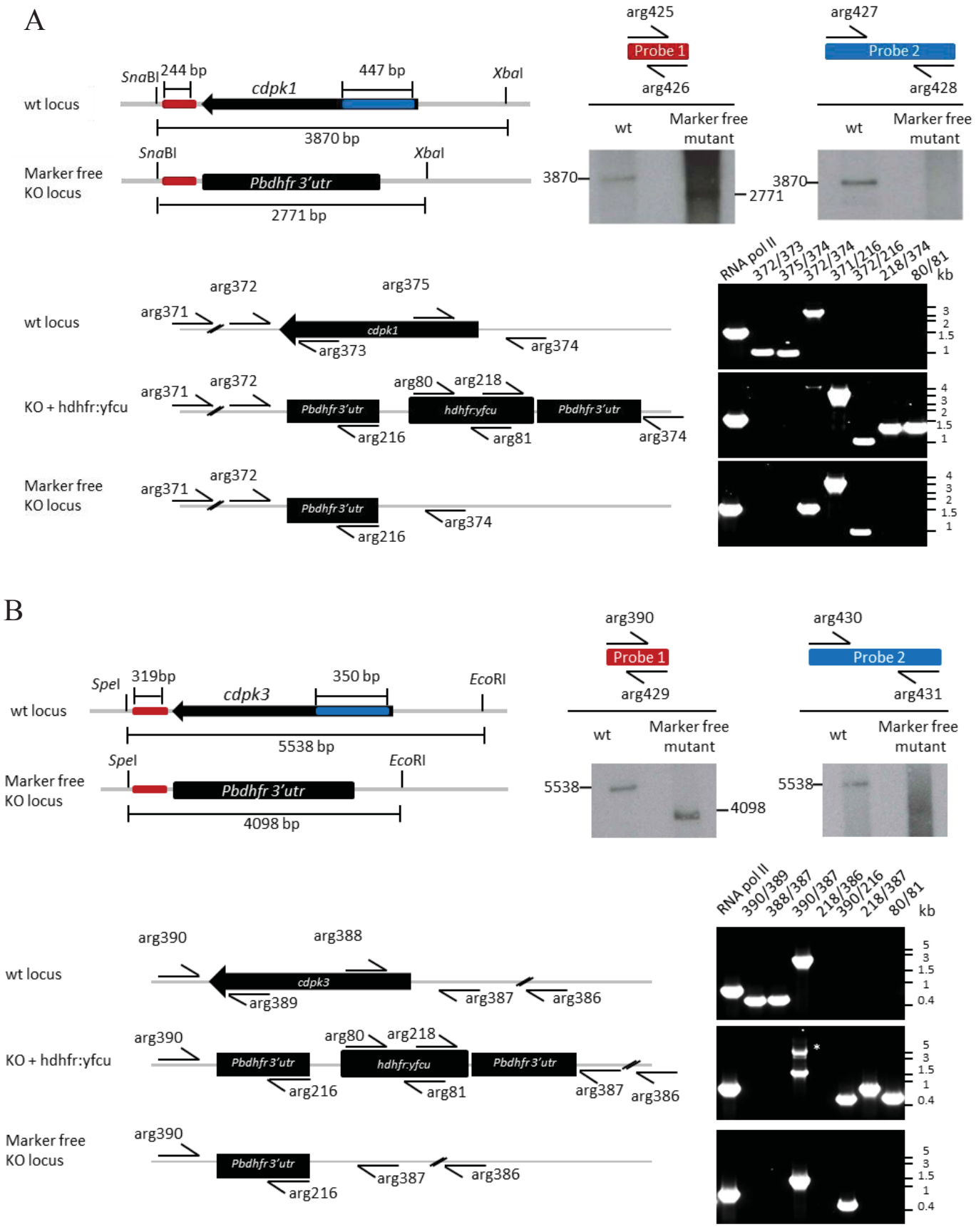
All mutants were successfully obtained, and genotyping was unambiguous.

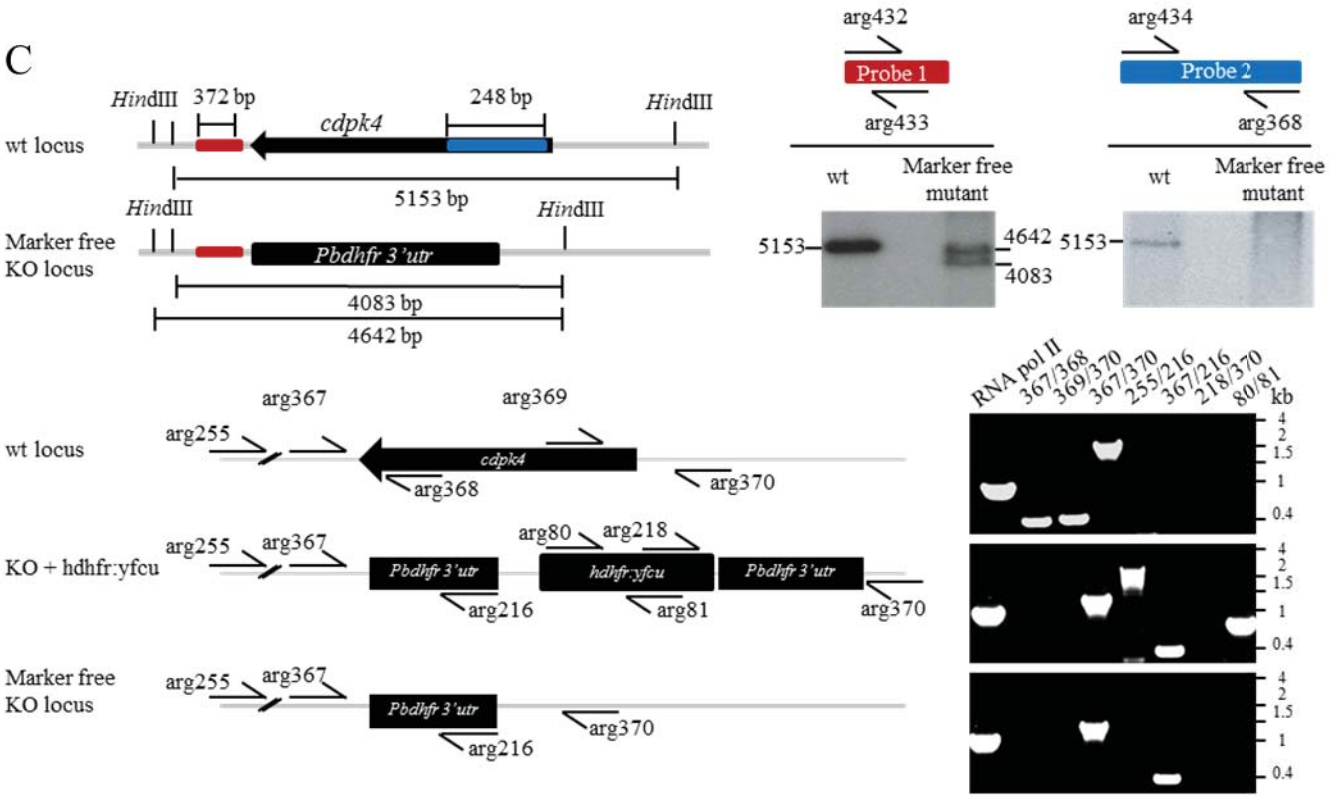
5.2.2.1 Phenotypic analysis of the *map* double KO revealed a *map2* KO-like phenotype

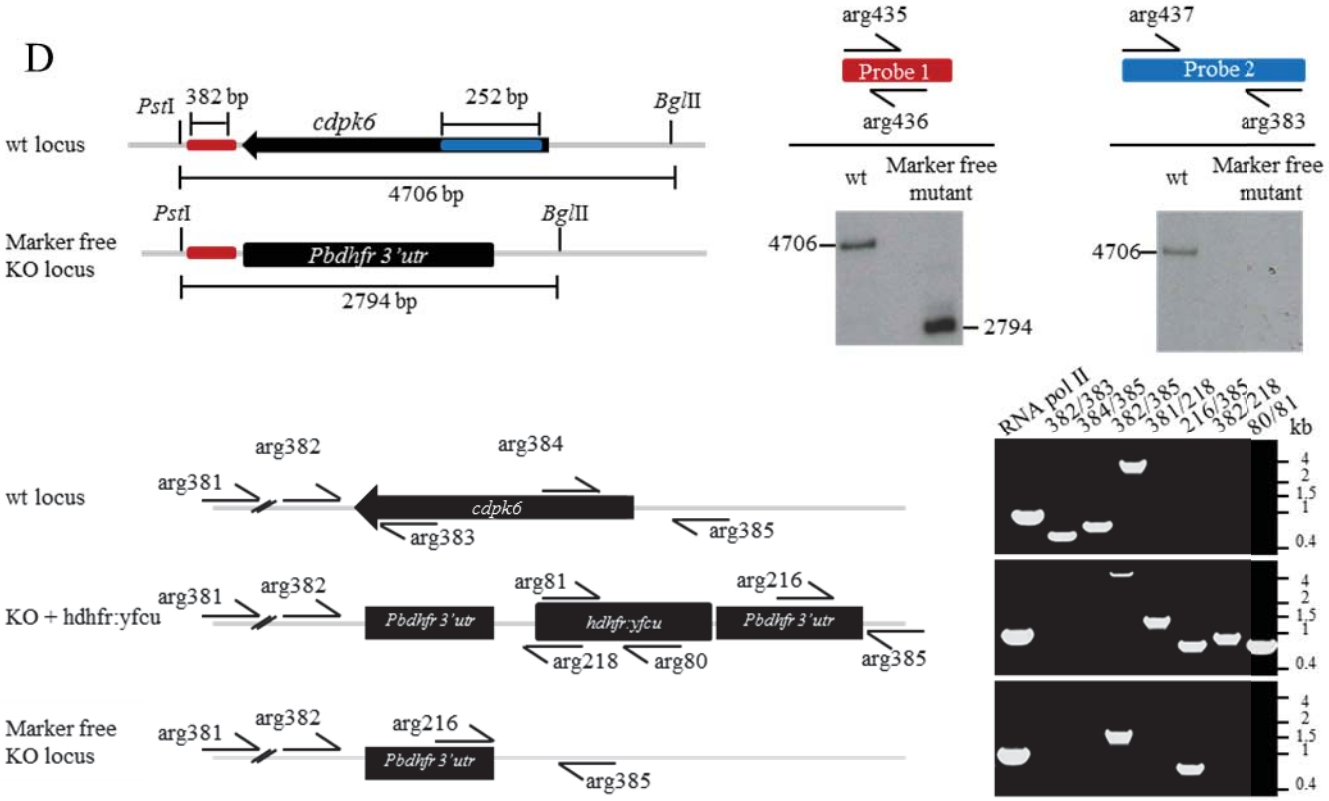
While the *map1* gene can be disrupted in both *P. berghei* and *P. falciparum* parasites, *map2* is required for the asexual development in *P. falciparum* but not in *P. berghei* [129]. In *P. berghei*, *map1* seems completely dispensable, while *map2* is required for microgametogenesis, specifically genome replication and cytoskeletal rearrangements associated with mitosis and assembly of axonemes [138]. However, despite this crucial role during gamete formation the *Pbmap2* gene can be targeted in blood stages. In fact, both *map* genes could be disrupted simultaneously in *P. berghei* (Fig 5.3 A, B). The phenotype of this double mutant was not noticeably different from the WT in the asexual blood stages (Fig. 5.3C), but it showed a sexual phenotype similar to the *map2* KO alone. Figure 5.3D and E show the highly reduced ookinete conversion rate which was a result of the blockade in microgamete exflagellation, confirming the previous analysis of the *map2* single mutant [138].

As both kinases were successfully disrupted it is possible that there is a third kinase that is up-regulated in the absence of the other two.

Fig. 5.2| Genotyping of CDPK KO genetic backgrounds (next two pages)
The genotype of each of the CDPK mutants was confirmed by Southern blot of restricted gDNA (top panel of each Figure) and PCR (bottom panel). Two different probes were used to hybridise the DNA fragments on the Southern blots. One probe should detect both loci, but show a difference in the size as consequence of gene removal, and another that should only hybridise with the WT gene sequence and therefore not with the mutant DNA. The restriction digest strategy is illustrated next to each blot. For the PCR genotyping eight reactions were performed to distinguish three different gDNAs: the WT, the mutant after the first round of cloning and the cloned mutant after resistance cassette removal. These included one positive control (*rna polymerase II*), two reactions targeting the WT gene (5' and 3' end), three others targeting the modified locus resistance cassette (two detected the presence of the vector and the third detected correct integration), one reaction across the gene that should generate differently sized bands for each of the three gDNAs (this reaction was particularly important to identify excision of the resistance cassette), and one last reaction that detected the parasite's resistance cassette. (A-D) Genotyping of *cdpk1* KO, *cdpk3* KO, *cdpk4* KO, and *cdpk6* KO mutants, respectively.
Note that the white asterisk (*) in the middle panel of the gel picture in Figure B (*cdpk3* KO genotyping) denotes the band with the correct size. In Figure C, the presence of two bands on the left panel of the Southern blot is the result of incomplete digestion by *Hind*III enzyme that is present twice downstream the *cdpk4* gene, depicted on the Southern blot map on the left.


A
wt locus
SnaBI
244 bp
cdpk1
447 bp
XbaI
3870 bp
Marker free
KO locus
Pbdhfr 3'utr
2771 bp
arg425
Probe 1
arg426
arg427
Probe 2
arg428
wt
Marker free
mutant
3870
2771
RNA pol II
372/373
375/374
372/374
371/216
372/216
218/374
80/81
kb
arg371
arg372
arg375
arg373
arg374
KO + hdhfr:yfcu
arg80
arg218
hdhfr:yfcu
arg216
arg81
B
SpeI
319bp
cdpk3
350 bp
EcoRI
5538 bp
4098 bp
arg390
arg388
arg389
arg387
arg386
arg429
arg430
Probe 1
Probe 2
arg431
5538
4098
390/389
388/387
390/387
218/386
390/216
218/387
80/81


C
HindIII
372 bp
cdpk4
248 bp
HindIII
wt locus
5153 bp
HindIII
HindIII
Marker free
KO locus
Pbdhfr 3'utr
4083 bp
4642 bp
arg432
Probe 1
arg433
arg434
Probe 2
arg368
wt
Marker free
mutant
5153
4642
4083
arg367
arg369
arg255
wt locus
cdpk4
arg368
arg370
arg80
arg218
KO + hdhfr:yfcu
Pbdhfr 3'utr
hdhfr:yfcu
Pbdhfr 3'utr
arg216
arg81
Marker free
KO locus
RNA pol II
367/368
369/370
367/370
255/216
367/216
218/370
80/81
kb
4
2
1.5
1
0.4


D
PstI
382 bp
cdpk6
252 bp
BglII
wt locus
4706 bp
PstI
BglII
Marker free
KO locus
Pbdhfr 3'utr
2794 bp
arg435
Probe 1
arg436
arg437
Probe 2
arg383
wt
Marker free
mutant
4706
2794
arg382
arg384
arg381
wt locus
cdpk6
arg383
arg385
arg81
arg216
KO + hdhfr:yfcu
Pbdhfr 3'utr
hdhfr:yfcu
Pbdhfr 3'utr
arg218
arg80
Marker free
KO locus
RNA pol II
382/383
384/385
382/385
381/218
216/385
382/218
80/81
kb
4
2
1.5
1
0.4

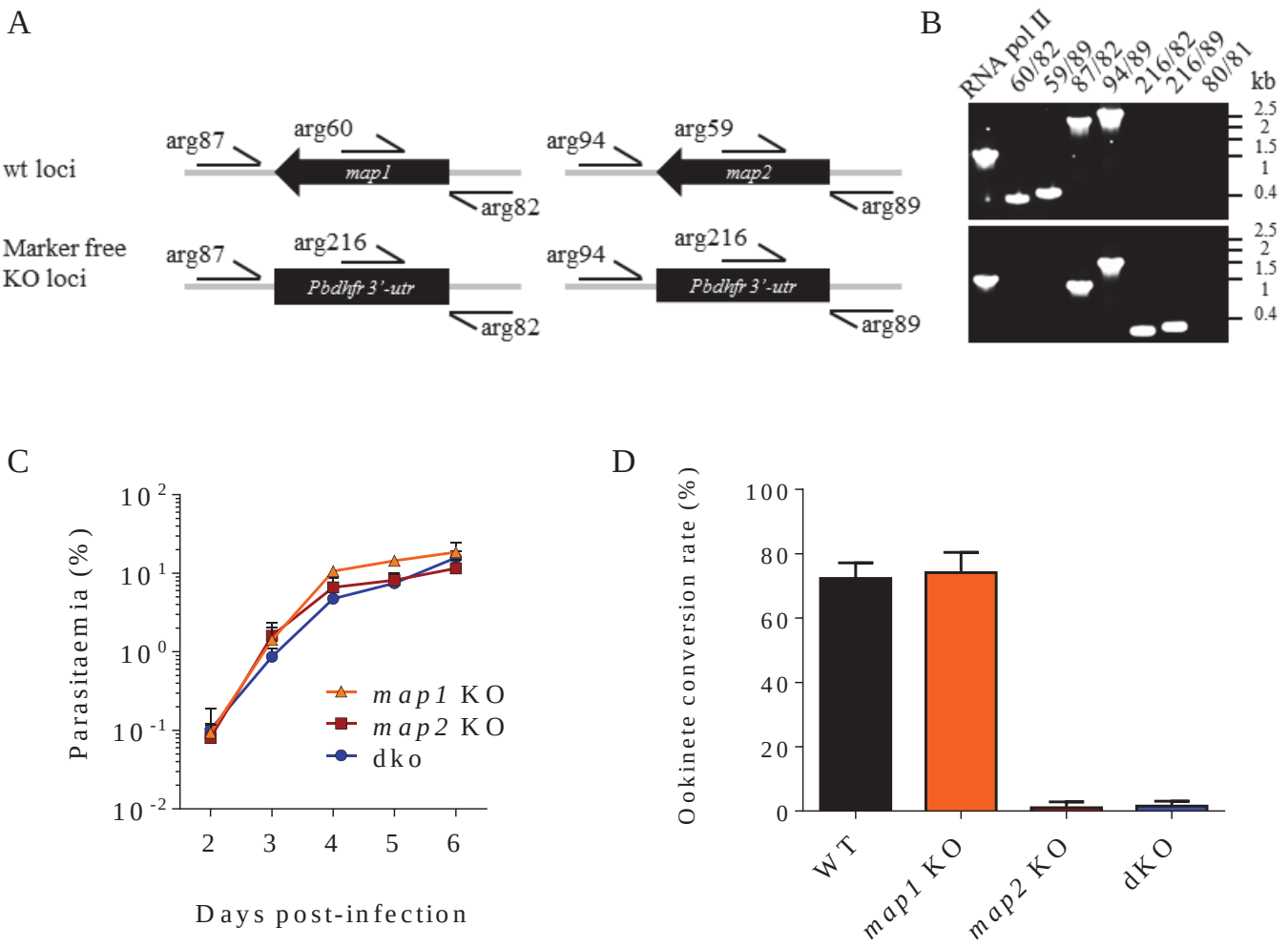


Fig. 5.3| Genotyping and phenotyping of the double KO mutant *map1*$^{-}$/*map2*$^{-}$ (dKO).
(A,B) PCR confirmation of the genotype of the dKO mutant. (A) Illustration of the primer annealing sites. (B) Eight reactions were performed to genotype the final mutant. One positive control (RNA polymerase II), two reactions to detect either *map1* or *map2* genes, two others that spanned each of the gene's ORF to detect excision of the drug cassette, two more reactions to detect the 3' end of each of the modified loci and one last PCR to detect the presence of the complete resistance cassette. (C-D) Phenotyping of the dKO mutant. (C) Comparison of the growth patterns of the cloned single *map* mutants (no detectable fitness cost) to the dKO mutant. (D) Ookinete conversion rate was greatly impaired in the dKO mutant as it was in the single *map2* KO. Three different cultures were sampled twice for each parasite line. (E) Imaging of one of the ookinete cultures used to calculate the conversion rate in (D). Parasites were stained with anti-P28 Cy3-labeled monoclonal antibody and

Hoechst to label the DNA. Most parasites in the double KO culture were non-fertilised female gametes (i.e. round forms) rather than banana-shaped ookinetes as the ones found in the WT culture (top panel). Scale bar = 5μm.

5.2.3 Revealing epistasis in the *P. berghei* kinome

The experiments performed on the WT background (Chapter 4) became an important control for this interaction screen as I used the same vector stocks to generate the DNA pools. Each pool lacked the targeting vector that had been used to generate the corresponding background but no other changes were implemented.

One immediate consequence of using a KO line as the genetic background in an STM experiment is the presence of two barcodes in each mutant parasite, i.e. one from the background and another one from the DNA transfected during the screen. This number is increased to three barcodes in the case of the dKO, where each background carries two barcodes that together should account for $^2/_3$ of the total barcode counts, per time-point.

At the moment of transfection, the genetic background accounts for 100 % of the parasite barcodes. As the new set of barcodes integrates and the drug selection clears the non-transfected parasites, this percentage should decrease to 50 % (or 33.3% in case the where the background has two barcodes).

Indeed, the analysis of the counts for the background barcodes over time showed that they stabilised on ~50 % and ~33% as expected (Fig. 5.4). However this only happened, from day 6 post-transfection onwards, which suggested that it takes approximately four to five days for the drug selection (and the spleen clearance) to clear all the parasites that did not integrate a new vector. Day 4 post-transfection was the time-point with the most excess of background counts relatively to the newly transfected barcode.

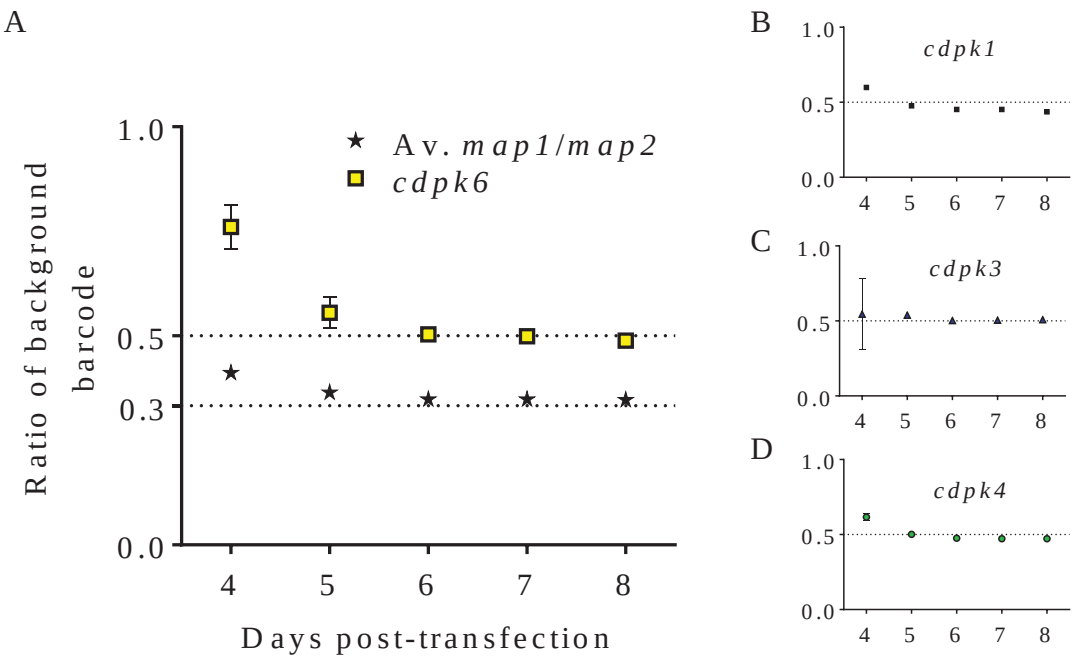


Fig. 5.4| Analysis of the genetic background barcode counts.
The pool of vectors studied in chapter 4 was then re-transfected in each of these background lines without the vector(s) targeting the already deleted loci in each particular case. The proportion of the barcode corresponding to the genetic background for each of the parasite lines was calculated for (A) *cdpk6* KO and dKO; (B) *cdpk1* KO; (C) *cdpk3* KO and (D) *cdpk4* KO. Note that in (A) it is the average of the proportion of barcode counts for *map1* and *map2*.

Six parallel transfections yielded fitness data for 258 double or triple mutants. Using a multiplicative model of epistasis [101], interaction coefficients were determined from the relative fitness phenotypes of the individual mutations and the fitness of the double (or triple) mutants (Fig. 5.5A,B).

Two putative interactions, one positive and one negative were detected and are highlighted in green and red, respectively.

A severe growth defect was detected for a mutant lacking the otherwise redundant *cdpk4* gene on a line expressing the resistant *pkg*^{T619Q}3xHA allele [154] (red circle). This suggested the existence of an important genetic interaction between *cdpk4* and *pkg* and was therefore validated with independently generated mutants, as described in the next section.

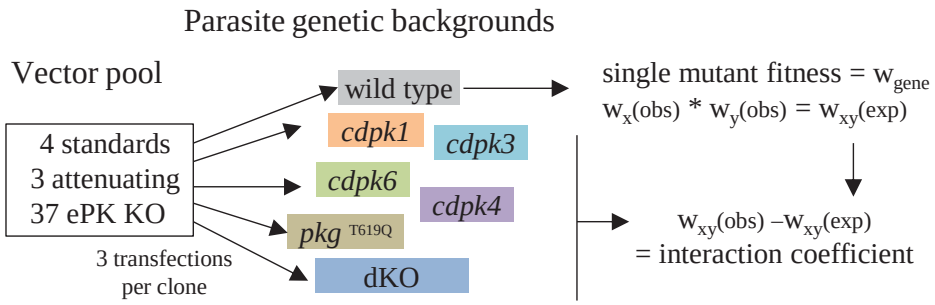
The positive interaction was detected for the pair *cdpk3* x PBANKA_146050 and was characterised by an increased fitness of the double mutant compared to the individual mutants (1.0 and 0.9, respectively). This proved to be specific to the pair *cdpk3* x PBANKA_146050.

PBANKA_146050 KO mutants fail to differentiate into ookinetes efficiently. The few that establish a mosquito infection do not produce sporozoites. This gene is *gak* (cyclin G-associated kinase) and its orthologue has been implicated in cell cycle progression due to a role in clathrin-mediated membrane trafficking [195]. *cdpk3*, on the other hand, has been associated with ookinete motility triggered by calcium signaling. Its disruption greatly decreases the ookinete ability to infect the mosquito midgut [136].

The link between these two genes is unknown and due to time constraints I did not pursue this putative interaction further.

All grey circular symbols in Figure 5.5B show data points in which targeting vectors integrated with borderline efficiency and fitness calculations were based on too few read counts to be considered reliable. These were hits that included the genes *srpk* and PBANKA_040940, both of which considered technical fails of the respective vectors for unknown reasons as evidence from others showed that they are targetable [105].

A


Parasite genetic backgrounds
Vector pool
4 standards
3 attenuating
37 ePK KO
3 transfections per clone
wild type
cdpk1
cdpk3
cdpk6
cdpk4
pkg T619Q
dKO
single mutant fitness = w_{gene}
$w_x(obs) * w_y(obs) = w_{xy}(exp)$
$w_{xy}(obs) - w_{xy}(exp)$
= interaction coefficient

B

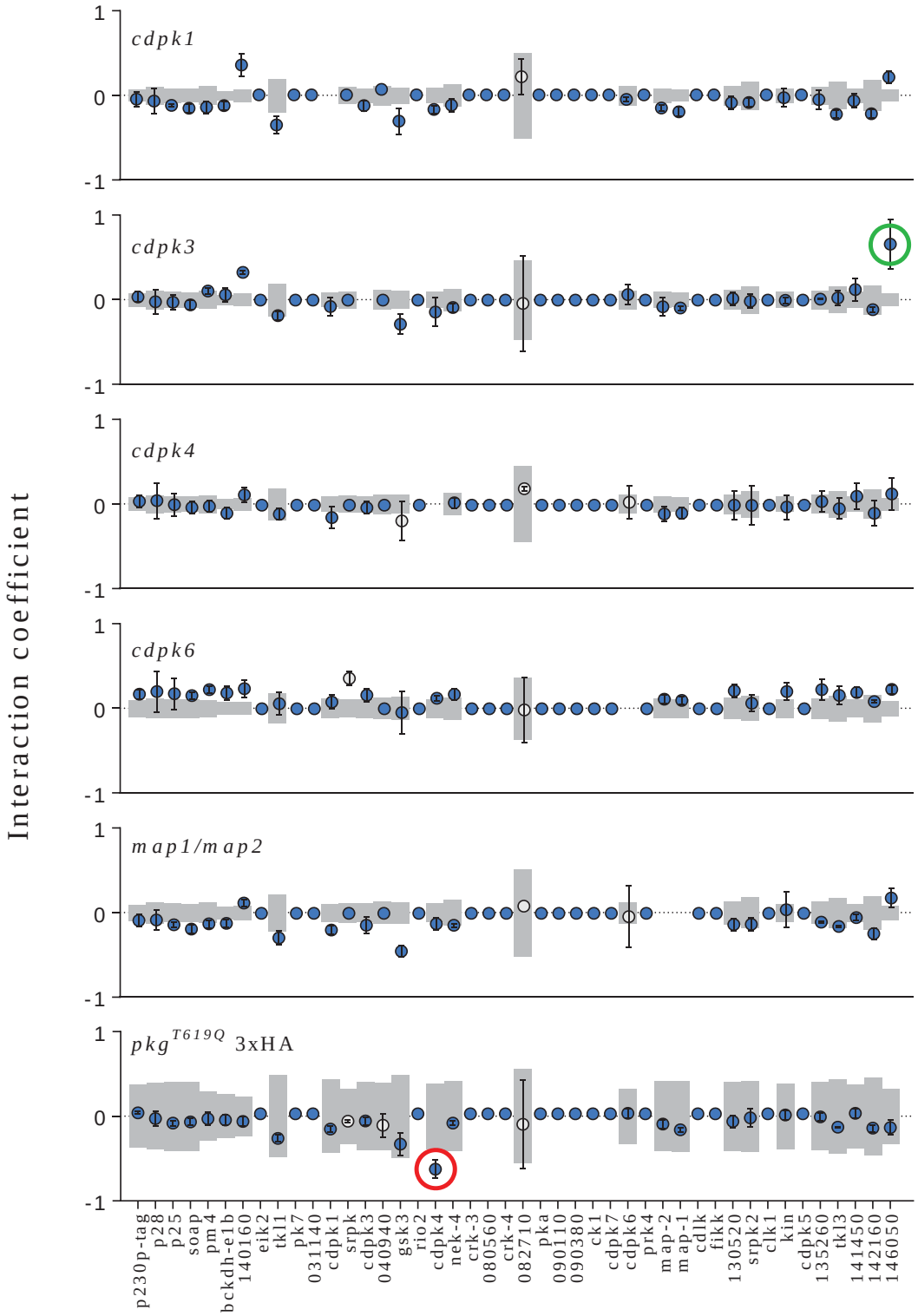

Interaction coefficient
cdpk1
cdpk3
cdpk4
cdpk6
map1/map2
pkg T619Q 3xHA
1
0
-1
p230p-tag
p28
p25
soap
pm4
bckdh-e1b
140160
eik2
tkl1
pk7
031140
cdpk1
srpk
cdpk3
040940
gsk3
rio2
cdpk4
nek-4
crk-3
080560
crk-4
082710
pka
090110
090380
ck1
cdpk7
cdpk6
prk4
map-2
map-1
cdlk
fikk
130520
srpk2
clk1
kin
cdpk5
135260
tkl3
141450
142160
146050

Fig. 5.5| Interaction coefficients for 258 double and triple mutants (next page).
(A) Schematic illustration of the interaction study. The interaction coefficient between two genes is defined as the difference between the observed fitness (W_{obs}) of a double mutant and its expected fitness (W_{exp}). The latter assumes that there is no interaction between the genes and that the fitness of the double mutant is the result of a cumulative effect of both mutations. (B) Genetic interaction coefficients on day 7 post-transfection (blue symbols). The "no interaction" area has been delimited by grey bars and reflects the combined uncertainty (two standard deviations) of the combined fitness measurements. In some cases, low barcode counts (< 0.25 % of the total) led to potentially inaccurate coefficient calculations. These are therefore shaded grey.
Green and red circles highlight putative positive and negative interactions, respectively. Mutants not viable were included in the graph as having a coefficient of zero. The absence of the grey bar was the result of no fitness values (i.e. likely essential gene) for the corresponding single mutant. Note that because the *pkg*^{T619Q} mutant does not carry a barcode, its fitness was calculated from parasitaemia growth curves. This slightly less accurate method of calculating the single mutant fitness is the reason for the uncertainty grey boxes being larger than in the other mutants.
Error bars show standard deviations of the mean of W(obs) (n=3).

5.2.4 Validation of *cdpk4*-*pkg* interaction

The work presented in this section was mostly performed in collaboration with Mathieu Brochet and is shown here due to its relevance for the project.

The analysis of the interaction coefficients suggested the existence of an interaction between *pkg* and *cdpk4*. Statistical analysis of the fitness of the double mutant indicated that it was significantly different from the normal growth references, throughout infection (Fig. 5.6A).

In order to validate this putative interaction, a double mutant was generated independently, by disruption of the *cdpk4* gene in the same strain used for the screen (*pkg*^{T619Q}). Careful examination of the growth pattern of the double mutant in comparison with the WT and the single mutant *cdpk4* KO was also indicative of a strong fitness cost caused by the double mutation (Fig. 5.6B).

Western blot analysis of schizont extracts from a parasite strain expressing *cdpk4*-3xHA but defective for production of gametocytes [196] showed that *cdpk4* is not only expressed in gametocytes as previously shown [138], but also in asexual stages as is *pkg* (Fig 5.6C), in agreement with proteomic analyses previously published [34].

Next, complementation of the double mutant was performed by either re-introducing the *cdpk4* gene or by repairing the gatekeeper mutation. This complementation strategy generated the following genotypes: *pkg*^{T619Q}3xHA/*cdpk4*-3xHA and *pkg*-3xHA/*cdpk4* KO, respectively and their growth patterns are depicted in Figure 5.6D. Although successful, neither of the strategies restored growth of the complemented mutants to WT levels. These results suggested

that the epitope tag and / or the generic 3'UTR used in both complementation vectors imposed a moderate fitness cost and generated hypomorphic alleles.[10] A complementation strategy with vectors that do not carry a 3xHA epitope tag has been devised and work is ongoing.

Altogether, these findings support the existence of a genetic interaction between *pkg* and *cdpk4*.

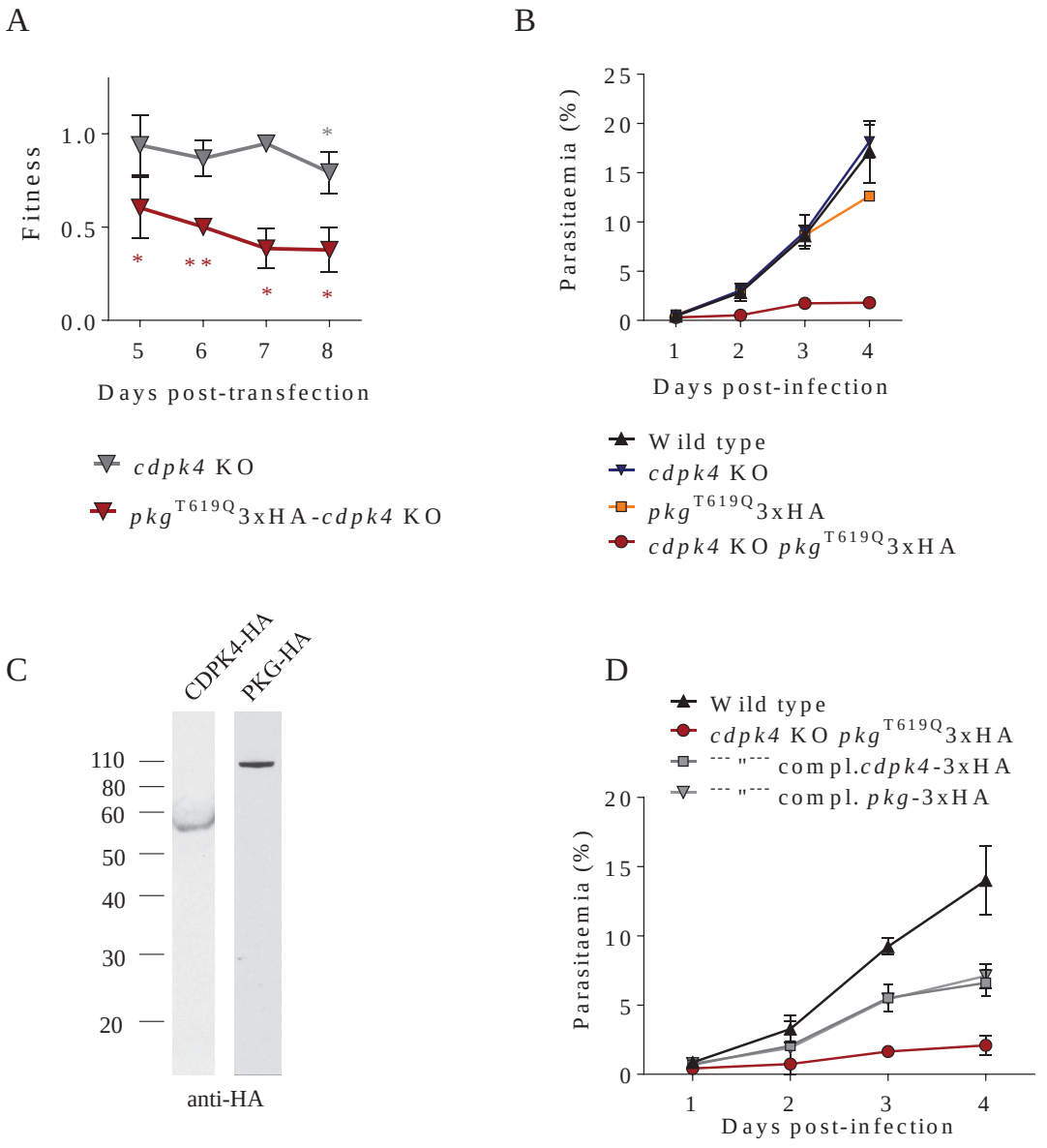


Fig. 5.6| The genetic interaction between *cdpk4* and *pkg* was also detected in independently generated mutants. (A) Bar-seq data showing a fitness comparison between the highly attenuated double mutant pkg^{T619Q}3xHA and *cdpk4* KO. Error bars show standard deviations from the mean (n=3). *Different from the WT reference mutants (included in all experiments) as determined by a two sided T-test corrected for multiple testing; *p < 0.05, **p < 0.01. (B) Growth curves of independently generated mutants over four days of infection determined from

[10] A hypomorphic allele is an allele with reduced levels of gene activity.

Giemsa stained smears. (C) Western blot performed on total protein extracts from *cdpk4*-3xHA and *pkg*-3xHA schizonts showed that both proteins are expressed during the asexual stages. The membrane was probed with an anti-HA antibody. The expected sizes were ~60 KDa and ~100 KDa, respectively. (D) Growth pattern of the complemented mutant lines and respective parental line. Each mutation, *cdpk4* KO and pkg^{T619Q}, was complemented independently. The grey squares represent the mutant where the *cdpk4* locus was restored and the inverted grey triangles show the mutant where the *pkg* point mutation was corrected.

5.3 Discussion

Redundancy between pathways ensures robustness of biological processes. This is the reason why in most organisms a remarkable proportion of genes can be disrupted without compromising viability. For instance, only 20 % of yeast genes are required for haploid development in optimal culturing conditions [100]. This network buffering complicates the analysis of traditional reverse genetics studies.

Systematic analyses in *S. cerevisiae* have shed light on important properties of genetic networks and their components. These have shown that synthetic interactions are highly biased toward genes that have related functions [102] and have connected genes that despite performing related functions lacked a direct functional relationship. For instance, the *rfc5* gene which encodes a subunit of the replication factor C and is required for the checkpoint that responds to replication block and DNA damage [197] was shown to interact with *nse1* and *smc6*, both members of a complex responsible for structural maintenance of chromosomes [198].

The screening approach presented here shows for the first time that synthetic interactions can be revealed through a genetic screen in *Plasmodium*. In this chapter I have shown that deletion screening by barcode counting on different genetic backgrounds enables growth phenotyping of dozens of double or triple mutants in the same mouse, in the timeframe of a normal transfection.

The screen revealed a negative interaction between *pkg* and *cdpk4*. The double mutant had a marked growth disadvantage that was validated by independent generation of the mutant. A function for *cdpk4* in blood stages was surprising as CDPK4 has been shown to be a major sexual stage regulator [138,189]. PKG, on the other hand has been shown to control the intracellular levels of calcium by regulating the biosynthesis of phosphoinositides [154]. A direct genetic interaction between *pkg* and *cdpk4* had never been documented.

Calcium plays a critical role in merozoite egress. During the last hour of *P. falciparum* erythrocytic cycle, merozoite egress is initiated by the release of calcium from the endoplasmic reticulum (ER) stores in a PKG dependent fashion [154]. This event promotes: (1) activation of CDPK5 and CDPK4, (2) induction of calcium-dependent vacuole swelling (3) destabilization of the RBC cytoskeleton [199,200]. Activation of PKG also triggers the secretion of organelles called micronemes and the discharge of the protease PfSUB1 into the parasitophorous vacuole that is responsible for modifying merozoites surface proteins as part of the egress process [201].

Our knowledge of the exact role of CDPKs in egress is still very rudimentary, but a link between *pkg* and *cdpk5* and their function in blood stage merozoites egress has been reported in *P. falciparum* [201]. Dvorin and colleagues showed that the depletion of CDPK5 prevents mature schizonts from egressing despite normal maturation of egress factors. Physical disruption of these schizonts released infectious merozoites [144]. In this study the depletion of CDPK5 was achieved with a destabilisation domain (DD) strategy. A transcriptome analysis of these arrested mutants was subsequently performed 48 hours after removal of shield. As this dataset is publicly available, I looked at the effect of CDPK5 depletion on *cdpk4* and found that the *cdpk4* transcript was ranked in the top 1 % of down-regulated genes. Unfortunately, the lack of replicates and technical details prevented statistical analysis of this dataset. Nevertheless, this is an indication that CDPK5 and CDPK4 might be part of the same pathway. In light of these data it is tempting to propose the following hypothetical model of parasite egress (Fig. 5.7):

(A) In WT parasites, activation of a fully functional PKG triggers microneme secretion and calcium release. This stimulus activates CDPK5 and CDPK4 that cooperatively promote merozoites egress upon microneme and exoneme secretion.

(B) A less efficient egress could be the cause for the reduced growth rate of the double mutant parasites. It has recently been shown that PKG^{T619Q} mutants, although viable throughout the life cycle [154], are less efficient at triggering calcium release (Mathieu Brochet, personal communication). This decreased level of intracellular calcium which does not seem to affect activation of the egress pathways in single mutants becomes critical in the absence of CDPK4, perhaps because CDPK5 on its own is unable to efficiently trigger merozoites release.

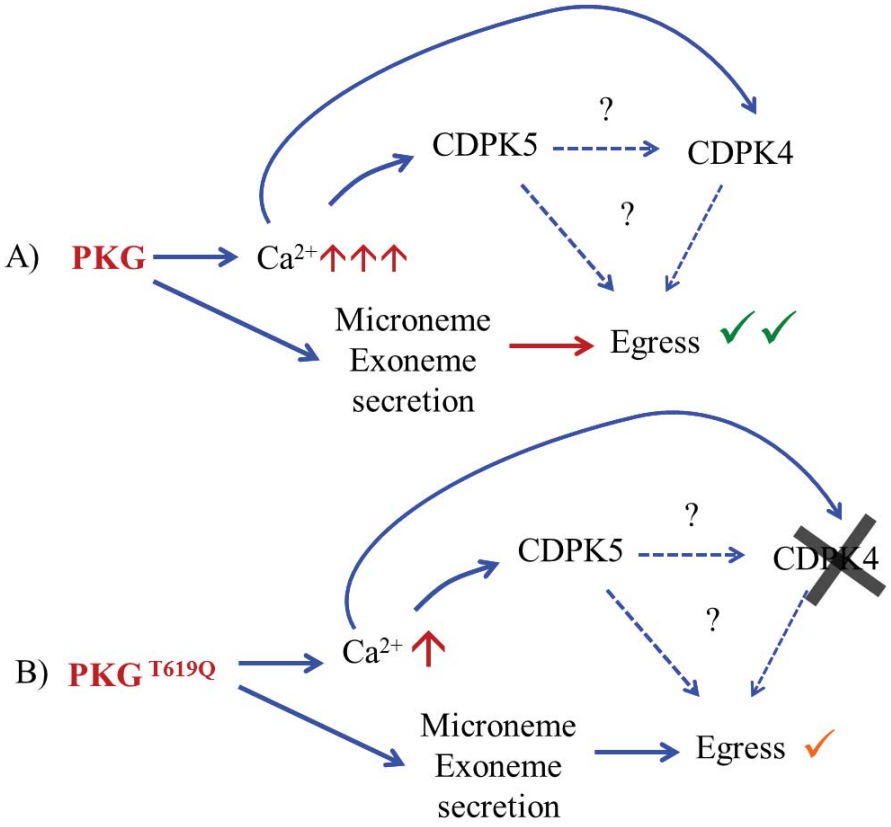


Fig. 5.7| A role for CDPK4 in merozoites egress – a possible model.
(A) Activation of PKG leads to the increase of intracellular calcium that in turn activates CDPK5 and CDPK4. In addition, PKG is also responsible for the discharge of proteases and microneme / exoneme secretion which lead to merozoites egress in a CDPK5-dependent fashion, potentially in cooperation with CDPK4. (B) As the mutated allele of PKG is less efficient at promoting the release of intracellular calcium, the absence of CDPK4 greatly impairs merozoites egress.

The use of hypomorphic alleles enables genetic studies that would not otherwise be possible. The fact that this severe phenotype is not revealed in mutants that only lack the *cdpk4* gene could also mean that a fully functional PKG, but not a hypomorphic allele, is able to divert this signal to an equivalent effector, such as another CDPK.

Follow-up on this matter would involve the inspection of the number of merozoites per schizont to exclude the unlikely possibility of this phenotype being caused by a reduced number of cells per cycle. If this hypothesis is excluded, a transcriptome analysis of *cdpk4* KO and the double mutant, as well as comparative invasion assays, would be crucial to confirm that the decreased fitness is a result of a less efficient egress phenotype. The transcriptome analysis would help clarify whether this phenotype is caused by gene down-regulation of a key factor. It would also be interesting to know whether over-expression of CDPK5 decreases the severity of this phenotype or even if disruption of CDPK5 is possible if CDPK4 is over-expressed.

In addition, as *cdpk4* has a similar role in both *P. berghei* and *P. falciparum*, i.e. it can be deleted in asexual blood stages but blocks microgamete exflagellation [138,143] and a Pb*cdpk4* mutant can be complemented with the Pf*cdpk4* gene [188], it would be interesting to check if the same applies to the interaction with *pkg*. Generation of transgenic parasites is a much longer process in *P. falciparum* than in *P. berghei*, but alternative tools would allow this hypothesis to be tested almost immediately, namely the existence of an equivalent *pkg* point mutation strain (pkg^{T618Q}) [192] and *cdpk4* inhibitors [188].

Chapter 6

General discussion

A better understanding of the biology of *Plasmodium* parasites is crucial for the identification of novel drug targets to treat malaria. Our understanding of the molecular cues that drive *Plasmodium* development and differentiation, and how it can be interfered with has been limited by the fact that only a small proportion of genes have been assigned a function experimentally.

Signature tagged mutagenesis has been used extensively in bacterial pathogens to identify virulence genes by parallel phenotyping of pools of individually tagged mutants [91]. The main aim of this dissertation was the development of such strategies for the study of *P. berghei* biology.

A three-step strategy was developed that involved (1) parallel transfection of barcoded *Plasmo*GEM vectors, followed by (2) a propagation step during which small blood samples were collected from day 4 to 8 post-transfection, and finally (3) calculation of the fitness of each population of mutants within the pool, through sequencing of their barcodes. Optimisation of each step was followed by validation of the method with protein kinases chosen as test genes. Using this strategy, complex and defined pools of KO mutants were reproducibly generated in a single mouse. Barcode sequencing enabled the measurement of the daily fitness of each mutant and how it evolved on the course of infection. For the first time in malaria research, hundreds of mutants can now be screened for quantitative growth phenotypes, which is a much more powerful approach than the traditional bimodal essential/redundant classification.

One way of using STM is in drop out screens, where disruption of essential genes produces lethal phenotypes and its consequent loss from the study pool. However, *in vivo* systems have a high selection pressure. While this promotes competition and reduces false positives it also prevented the survival of mutants with a fitness lower than 0.45 of the WT-like references. Despite the absence of fitness data for these highly attenuated or non-viable mutants, knowing which genes are essential for parasite development is highly valuable for drug discovery research. Coupling the STM/bar-seq strategy with the newly developed conditional systems such as DiCre mediated gene excision [202], FLP/FRT mediated conditional mutagenesis [64] or protein destabilisation domain strategies [61] will be crucial to validate targetability of this group of highly attenuated or non-viable mutants.

In some pilot experiments, I asked whether mutants for which barcodes were either missing at the end of the experiments (i.e. where likely essential genes were disrupted), or were inconsistent, could be observed more abundantly when mutants with normal values of fitness were excluded from the transfection. After this, some of the cases that had previously

been considered “inconsistent”, because no growth measurements had been obtained for at least two out of three replicates, were rescued. However, the absence of growth references prevented the calculation of fitness values for these mutants.

Although the throughput and sensitivity of the STM approach are unparalleled in *Plasmodium*, some caveats need to be considered. As with any gene knock out experiment if a targeting vector fails to integrate, the consequent absence of a given mutant in the final pool might be mistaken as gene essentiality. Furthermore, while genetic screens can generate new hypotheses, validation and follow-up experiments still require the independent generation of clonal lines. Importantly, this technology is of limited utility to study mosquito stages. Random fertilisation between the different genotypes present in a pool generates heterozygous (i.e. double mutant) zygotes that are not detectable by the barcoding method. The fact that barcodes are amplified through the same PCR reaction makes it impossible to pinpoint the origin of each barcode. One approach to overcome this limitation is to generate pools of mutants in a background where production of one of the gametes is impaired, for instance a male defective mutant like the *map2* KO, and then supply the cultures with the missing set of gametes. This would ensure that the “male” barcode is always the same. However, these experiments would have to be analysed as genetic interaction experiments. Beyond the zygote stage, meiosis complicates this issue further due to unpredictable cross-over events that may either repair disrupted loci and generate WT alleles or create mutants with several genes deleted.

It will be possible, however, to use barcode counting for the identification of genes required for gametocyte maturation and fertility by asking which barcodes are present among the asexual population in an infection, but absent from the gametocyte population. Finally, it is important to remember that the STM study design makes it impossible to detect essential genes whose function is compensated by other mutants in the pool. However, this is unlikely to be a frequent event.

The new parallel transfection approach coupled with bar-seq detection of the mutants within a pool proved to be more efficient and sensitive than previous strategies and allowed the detection of nine new mutants including *gsk3*, a gene previously targeted for drug development [169]. This finding was validated with a cloned mutant and its phenotypic analysis revealed a strong reticulocyte preference. In normal culturing conditions only normocytes are used as hosts for the parasites. If the *gsk3* KO phenotype seen in *P. berghei*

extends to *P. falciparum*, the absence of reticulocytes in culture would be the reason why the mutant could not be generated. Understanding what determines this host cell specificity (i.e. normocytes versus reticulocyte) would be of great relevance for drug development research. For instance, if drugs are directed against factors like GSK3, although this would be highly effective in culture, it would be of little use *in vivo* since *P. falciparum* parasites can also invade reticulocytes.

The properties of the STM approach have enabled large scale genetic screens with a dramatic reduction in workload and also a sharp decrease in the animal usage. Previously, in order to check a set of genes for redundancy one would need at least as many animals as genes. With the strategy presented here, I was able to screen a list of 50 genes in a single mouse. For a new project that will systematically screen 1000 metabolic enzymes, the vector pool size has been scaled up to ~100 vectors. Such large scale screens are now possible because data from each experiment are directly comparable and can be analysed together. This clear increase in scale offers new opportunities for reverse genetic screening that were previously unachievable for malaria parasites. I have demonstrated this here for the study of genetic interactions at scale by screening 41 ePK genes in six different mutant lines for growth phenotypes. This screen generated fitness data for 258 double or triple mutants and revealed an important interaction between *cdpk4* and *pkg*.

Biological circuits are protected by a certain degree of redundancy to ensure robustness of biological processes. However, some genes are central to several networks and cannot be disrupted. These are referred to as "network hubs" [100]. The study of these genes often requires the use of hypomorphic alleles or conditional strategies. The negative interaction between *pkg* and *cdpk4* was exactly one such example revealed by the use of a hypomorphic allele of the hub *pkg*.

Chemical genetics is another potential application for the barcoding approach. This is performed by testing barcoded mutants for sensitivity to sub-lethal concentrations of chemicals. These mutants can be the result of loss-of-function, point mutations (resistant or hypomorphic alleles) or even over-expression of genes. The principle applied here is similar to that of genetic interactions and greatly relies on the detection of unexpected fitness phenotypes of mutants in the presence and absence of chemicals. In yeast, hierarchical clustering of data generated by such screens revealed a number of genes associated with multidrug resistance when overexpressed. These genes, if disrupted, made the strains sensitive

to diverse compounds [203]. This kind of approach has also the potential to group together genes of unknown function based on the likely mode of action of a chemical. This would be highly relevant for malaria parasites as half of the protein coding genes lack annotation.

Progress in recent years in the field of genetics has suggested that gene essentiality is contextual, i.e. that a phenotype results from the interplay between the genotype and the environment in which it is expressed. The availability of a high throughput approach now offers the possibility of investigating not only gene-gene interactions, but also gene-environment interactions. For instance, in collaboration with Maria Mota and Liliana Mâncio (IMM, Lisbon) we performed some preliminary experiments that looked at the impact of the nutritional status of the host in the fitness of a set of mutants.

Other examples of environment-gene interactions include for example the use of different strains of receptor mice where genes controlling, for instance, cell receptors or immunological factors, could be either disrupted or over-expressed. Such experiments would be highly relevant for the study of host-pathogen interactions.

Plasmodium species diverged from their last common ancestor 67.8 million years ago [204]. It can therefore be expected that significant differences have emerged between the rodent parasites and those infecting humans. In fact, in some aspects the differences are obvious, for example the duration of the asexual cycle is 24 h in *P. berghei* and 48 h in *P. falciparum*, and the time required for the liver stage development is 48 h for the first and close to two weeks for the second. As a result, direct translation of findings between species can be limited in some cases. For this reason transfer of the STM approach to *P. falciparum* would be of great clinical relevance. The recent development of efficient site specific genome editing technologies for *P. falciparum* parasites such as zinc finger nucleases (ZFNs) [205] and CRISPR/Cas9 [206] have substantially increased the efficiency and ease with which genes can be targeted in these parasites. On the other hand, *P. knowlesi* parasites have recently been adapted to culturing conditions and arise now as a promising model to study human malaria [207]. These parasites are 1,000 fold more amenable to genetic manipulation than *P. falciparum* and, similarly to *P. berghei* parasites can be transfected with linear DNA [207].

As a result, the same parallel transfection of barcoded vectors could potentially be applied to these parasites. Furthermore, the fact that these parasites can be propagated in culture would certainly bring a new dimension to malaria reverse genetics. The disadvantage

of this system is the absence of host selection pressure and natural conditions such as reticulocytes as mentioned above. This will likely influence the rate of false positives, for example due to the absence of clearance of degenerated parasites that carry a barcode, and false negatives, i.e. if in natural conditions a given mutant would be viable, which will need to be controlled for. In my opinion this will be the next great challenge in the *Plasmodium* biology field.

In conclusion, the kind of high throughput genetic approach described here provides the basis for future screenings on large subsets of parasite genes. These will greatly contribute to the understanding of the molecular biology of *Plasmodium* parasites and will certainly take malaria research closer to eradication of this disease.

References

1. World Health Organization. WHO methods and data sources for global causes of death 2000-2011. [Internet]. Geneva, Switzerland: WHO; 2013. Available: http://www.who.int/healthinfo/statistics/GlobalCOD_method.pdf

2. Reiter P. From Shakespeare to Defoe: malaria in England in the Little Ice Age. Emerg Infect Dis. 2000;6: 1–11. doi:10.3201/eid0601.000101

3. World Health Organization. World malaria report 2012. WHO. 2012;

4. Ansari MT, Saify ZS, Sultana N, Ahmad I, Saeed-Ul-Hassan S, Tariq I, et al. Malaria and artemisinin derivatives: an updated review. Mini Rev Med Chem. 2013;13: 1879–902. Available: http://www.ncbi.nlm.nih.gov/pubmed/24070206

5. Crawley J, Chu C, Mtove G, Nosten F. Malaria in children. Lancet. Elsevier Ltd; 2010;375: 1468–81. doi:10.1016/S0140-6736(10)60447-3

6. Fernando SD, Rodrigo C, Rajapakse S. The "hidden" burden of malaria: cognitive impairment following infection. Malar J. 2010;9: 366. doi:10.1186/1475-2875-9-366

7. Karunaweera ND, Wijesekera SK, Wanasekera D, Mendis KN, Carter R. The paroxysm of Plasmodium vivax malaria. Trends Parasitol. 2003;19: 188–93. Available: http://www.ncbi.nlm.nih.gov/pubmed/12689650

8. Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, et al. Artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med. 2009;361: 455–67. doi:10.1056/NEJMoa0808859

9. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature. Nature Publishing Group; 2012;487: 375–9. doi:10.1038/nature11174

10. Wykes MN. Why haven't we made an efficacious vaccine for malaria? EMBO Rep. Nature Publishing Group; 2013;14: 661. doi:10.1038/embor.2013.103

11. Goldberg DE, Siliciano RF, Jacobs WR. Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. Cell. Elsevier Inc.; 2012;148: 1271–83. doi:10.1016/j.cell.2012.02.021

12. World Health Organization. WHO | Guidelines for the treatment of malaria. Second edition. WHO. World Health Organization; 2010; Available: http://www.who.int/malaria/publications/atoz/9789241547925/en/

13. Sibley LD. Intracellular parasite invasion strategies. Science. 2004;304: 248–53. doi:10.1126/science.1094717

14. Baumeister S, Winterberg M, Przyborski JM, Lingelbach K. The malaria parasite Plasmodium falciparum: cell biological peculiarities and nutritional consequences. Protoplasma. 2010;240: 3–12. doi:10.1007/s00709-009-0090-3

15. Waller RF, Ralph SA, Reed MB, Su V, Douglas JD, Minnikin DE, et al. A Type II Pathway for Fatty Acid Biosynthesis Presents Drug Targets in Plasmodium falciparum. Antimicrob Agents Chemother. 2003;47: 297–301. doi:10.1128/AAC.47.1.297-301.2003

16. Prudêncio M, Rodriguez A, Mota MM. The silent path to thousands of merozoites: the Plasmodium liver stage. Nat Rev Microbiol. 2006;4: 849–56. doi:10.1038/nrmicro1529

17. Khan SM, Waters AP. Malaria parasite transmission stages: an update. Trends Parasitol. 2004;20: 575–80. doi:10.1016/j.pt.2004.10.001

18. Sturm A, Amino R, van de Sand C, Regen T, Retzlaff S, Rennenberg A, et al. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. Science. 2006;313: 1287–90. doi:10.1126/science.1129720

19. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. Nature. 1977;265: 687–95. Available: http://www.ncbi.nlm.nih.gov/pubmed/870828

20. Swerdlow H, Wu SL, Harke H, Dovichi NJ. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. J Chromatogr. 1990;516: 61–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/2286629

21. Droege M, Hill B. The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets. J Biotechnol. 2008;136: 3–10. doi:10.1016/j.jbiotec.2008.03.021

22. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456: 53–9. doi:10.1038/nature07517

23. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005;309: 1728–32. doi:10.1126/science.1117389

24. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol. 2013;14: 405. doi:10.1186/gb-2013-14-6-405

25. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. Nature Publishing Group; 2010;11: 31–46. Available: http://www.ncbi.nlm.nih.gov/pubmed/19997069

26. Moorthie S, Mattocks CJ, Wright CF. Review of massively parallel DNA sequencing technologies. Hugo J. 2011;5: 1–12. doi:10.1007/s11568-011-9156-3

27. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13: 341. doi:10.1186/1471-2164-13-341

28. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26: 1135–45. doi:10.1038/nbt1486

29. Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Res. 2006;34: e22. doi:10.1093/nar/gnj023

30. Turcatti G, Romieu A, Fedurco M, Tairi A-P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic Acids Res. 2007;36: e25–e25. doi:10.1093/nar/gkn021

31. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5: 621–8. doi:10.1038/nmeth.1226

32. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature. 2002;419: 498–511. doi:10.1038/nature01097

33. Carlton JM, Angiuoli S V, Suh BB, Kooij TW, Pertea M, Silva JC, et al. Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature. 2002;419: 512–9. doi:10.1038/nature01099

34. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW a, Berriman M, et al. A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science. 2005;307: 82–6. doi:10.1126/science.1103717

35. Kooij TW a, Carlton JM, Bidwell SL, Hall N, Ramesar J, Janse CJ, et al. A Plasmodium whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. PLoS Pathog. 2005;1: e44. doi:10.1371/journal.ppat.0010044

36. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. Nat Rev Genet. Nature Publishing Group; 2012;13: 315–28. doi:10.1038/nrg3187

37. Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proc Biol Sci. 2012;279: 4724–33. doi:10.1098/rspb.2012.1745

38. Branzei D, Foiani M. Regulation of DNA repair throughout the cell cycle. Nat Rev Mol Cell Biol. 2008;9: 297–308. doi:10.1038/nrm2351

39. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 2008;24: 529–38. doi:10.1016/j.tig.2008.08.007

40. Kirkman LA, Lawrence EA, Deitsch KW. Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. Nucleic Acids Res. 2014;42: 370–9. doi:10.1093/nar/gkt881

41. Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. Nat Rev Mol Cell Biol. 2010;11: 196–207. doi:10.1038/nrm2851

42. Ramiro RS, Reece SE, Obbard DJ. Molecular evolution and phylogenetics of rodent malaria parasites. BMC Evol Biol. 2012;12: 219. doi:10.1186/1471-2148-12-219

43. Limenitakis J, Soldati-Favre D. Functional genetics in Apicomplexa: potentials and limits. FEBS Lett. Federation of European Biochemical Societies; 2011;585: 1579–88. doi:10.1016/j.febslet.2011.05.002

44. Janse CJ, Ramesar J, Waters AP. High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite Plasmodium berghei. Nat Protoc. 2006;1: 346–56. doi:10.1038/nprot.2006.53

45. Amino R, Thiberge S, Blazquez S, Baldacci P, Renaud O, Shorte S, et al. Imaging malaria sporozoites in the dermis of the mammalian host. Nat Protoc. 2007;2: 1705–12. doi:10.1038/nprot.2007.120

46. Fonager J, Franke-Fayard BMD, Adams JH, Ramesar J, Klop O, Khan SM, et al. Development of the piggyBac transposable system for Plasmodium berghei and its application for random mutagenesis in malaria parasites. BMC Genomics. 2011;12: 155. doi:10.1186/1471-2164-12-155

47. Balu B, Chauhan C, Maher SP, Shoue DA, Kissinger JC, Fraser MJ, et al. piggyBac is an effective tool for functional analysis of the Plasmodium falciparum genome. BMC Microbiol. 2009;9: 83. doi:10.1186/1471-2180-9-83

48. Goonewardene R, Daily J, Kaslow D, Sullivan TJ, Duffy P, Carter R, et al. Transfection of the malaria parasite and expression of firefly luciferase. Proc Natl Acad Sci U S A. 1993;90: 5234–6. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=46690&tool=pmcentrez&rendertype=abstract

49. Wu Y, Kirkman LA, Wellems TE. Transformation of Plasmodium falciparum malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. Proc Natl Acad Sci U S A. 1996;93: 1130–4. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=40043&tool=pmcentrez&rendertype=abstract

50. Van der Wel AM, Tomás AM, Kocken CH, Malhotra P, Janse CJ, Waters AP, et al. Transfection of the primate malaria parasite Plasmodium knowlesi using entirely heterologous constructs. J Exp Med. 1997;185: 1499–503. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2196274&tool=pmcentrez&rendertype=abstract

51. Van Dijk MR, Waters AP, Janse CJ. Stable transfection of malaria parasite blood stages. Science. 1995;268: 1358–62. Available: http://www.ncbi.nlm.nih.gov/pubmed/7761856

52. Ferone R, Burchall JJ, Hitchings GH. Plasmodium berghei dihydrofolate reductase. Isolation, properties, and inhibition by antifolates. Mol Pharmacol. 1969;5: 49–59. Available: http://www.ncbi.nlm.nih.gov/pubmed/4392112

53. Orr RY, Philip N, Waters AP. Improved negative selection protocol for Plasmodium berghei in the rodent malarial model. Malar J. 2012;11: 103. doi:10.1186/1475-2875-11-103

54. Braks JAM, Franke-Fayard B, Kroeze H, Janse CJ, Waters AP. Development and application of a positive-negative selectable marker system for use in reverse genetics in Plasmodium. Nucleic Acids Res. 2006;34: e39. doi:10.1093/nar/gnj033

55. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, et al. RNAi in budding yeast. Science. 2009;326: 544–50. doi:10.1126/science.1176945

56. Clemens JC, Worby CA, Simonson-Leff N, Muda M, Maehama T, Hemmings BA, et al. Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proc Natl Acad Sci U S A. 2000;97: 6499–503. doi:10.1073/pnas.110149597

57. Ngô H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proc Natl Acad Sci U S A. 1998;95: 14687–92. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24510&tool=pmcentrez&rendertype=abstract

58. Baum J, Papenfuss AT, Mair GR, Janse CJ, Vlachou D, Waters AP, et al. Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. Nucleic Acids Res. 2009;37: 3788–98. doi:10.1093/nar/gkp239

59. Meissner M, Krejany E, Gilson PR, de Koning-Ward TF, Soldati D, Crabb BS. Tetracycline analogue-regulated transgene expression in Plasmodium falciparum blood stages using Toxoplasma gondii transactivators. Proc Natl Acad Sci U S A. 2005;102: 2980–5. doi:10.1073/pnas.0500112102

60. Pino P, Sebastian S, Kim EA, Bush E, Brochet M, Volkmann K, et al. A tetracycline-repressible transactivator system to study essential genes in malaria parasites. Cell Host Microbe. 2012;12: 824–34. doi:10.1016/j.chom.2012.10.016

61. Armstrong CM, Goldberg DE. An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. Nat Methods. 2007;4: 1007–9. doi:10.1038/nmeth1132

62. Sebastian S, Brochet M, Collins MOO, Schwach F, Jones MLL, Goulding D, et al. A Plasmodium Calcium-Dependent Protein Kinase Controls Zygote Development and Transmission by Translationally Activating Repressed mRNAs. Cell Host Microbe. 2012;12: 9–19. doi:10.1016/j.chom.2012.05.014

63. Suarez C, Volkmann K, Gomes AR, Billker O, Blackman MJ. The malarial serine protease SUB1 plays an essential role in parasite liver stage development. PLoS Pathog. 2013;9: e1003811. doi:10.1371/journal.ppat.1003811

64. Lacroix C, Giovannini D, Combe A, Bargieri DY, Späth S, Panchal D, et al. FLP/FRT-mediated conditional mutagenesis in pre-erythrocytic stages of Plasmodium berghei. Nat Protoc. 2011;6: 1412–28. doi:10.1038/nprot.2011.363

65. Landy A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annu Rev Biochem. 1989;58: 913–49. doi:10.1146/annurev.bi.58.070189.004405

66. Hartley JL, Temple GF, Brasch MA. DNA cloning using in vitro site-specific recombination. Genome Res. 2000;10: 1788–95. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=310948&tool=pmcentrez&rendertype=abstract

67. Katzen F. Gateway(®) recombinational cloning: a biological operating system. Expert Opin Drug Discov. 2007;2: 571–89. doi:10.1517/17460441.2.4.571

68. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet. 2001;2: 769–79. doi:10.1038/35093556

69. Murphy KC. Phage recombinases and their applications. [Internet]. 1st ed. Advances in virus research. Elsevier Inc.; 2012. pp. 367–414. doi:10.1016/B978-0-12-394438-2.00008-6

70. Muyrers JP, Zhang Y, Testa G, Stewart AF. Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. 1999;27: 1555–7. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=148353&tool=pmcentrez&rendertype=abstract

71. Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet. 1998;20: 123–8. doi:10.1038/2417

72. Ravin V, Ravin N, Casjens S, Ford ME, Hatfull GF, Hendrix RW. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J Mol Biol. 2000;299: 53–73. doi:10.1006/jmbi.2000.3731

73. Pfander C, Anar B, Schwach F, Otto TD, Brochet M, Volkmann K, et al. A scalable pipeline for highly effective genetic modification of a malaria parasite. Nat Methods. 2011;8: 1078–82. doi:10.1038/nmeth.1742

74. Miller CA, Ingmer H, Cohen SN. Boundaries of the pSC101 minimal replicon are conditional. J Bacteriol. 1995;177: 4865–71. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=177259&tool=pmcentrez&rendertype=abstract

75. Marinelli LJ, Hatfull GF, Piuri M. Recombineering: A powerful tool for modification of bacteriophage genomes. 2012; 5–14.

76. Wang J, Sarov M, Rientjes J, Fu J, Hollak H, Kranz H, et al. An improved recombineering approach by adding RecA to lambda Red recombination. Mol Biotechnol. 2006;32: 43–53. Available: http://www.ncbi.nlm.nih.gov/pubmed/16382181

77. Hasty P, Rivera-Pérez J, Bradley A. The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol. 1991;11: 5586–91. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=361929&tool=pmcentrez&rendertype=abstract

78. Pierce SE, Davis RW, Nislow C, Giaever G. Genome-wide analysis of barcoded Saccharomyces cerevisiae gene-deletion mutants in pooled cultures. Nat Protoc. 2007;2: 2958–74. doi:10.1038/nprot.2007.427

79. Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature. 2011;474: 337–42. doi:10.1038/nature10163

80. Hayes F. Transposon-based strategies for microbial functional genomics and proteomics. Annu Rev Genet. 2003;37: 3–29. doi:10.1146/annurev.genet.37.110801.142807

81. Smith V, Chou KN, Lashkari D, Botstein D, Brown PO. Functional analysis of the genes of yeast chromosome V by genetic footprinting. Science. 1996;274: 2069–74. Available: http://www.ncbi.nlm.nih.gov/pubmed/8953036

82. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. Simultaneous identification of bacterial virulence genes by negative selection. Science. 1995;269: 400–3. Available: http://www.ncbi.nlm.nih.gov/pubmed/7618105

83. Autret N, Charbit A. Lessons from signature-tagged mutagenesis on the infectious mechanisms of pathogenic bacteria. FEMS Microbiol Rev. 2005;29: 703–717. Available: http://www.sciencedirect.com/science/article/pii/S0168644504000865

84. Darwin AJ, Miller VL. Identification of Yersinia enterocolitica genes affecting survival in an animal host using signature-tagged transposon mutagenesis. Mol Microbiol. 1999;32: 51–62. Available: http://www.ncbi.nlm.nih.gov/pubmed/10216859

85. Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C. Identification of a virulence gene cluster of Mycobacterium tuberculosis by signature-tagged transposon mutagenesis. Mol Microbiol. 1999;34: 257–67. Available: http://www.ncbi.nlm.nih.gov/pubmed/10564470

86. Chiang SL, Mekalanos JJ. Use of signature-tagged transposon mutagenesis to identify Vibrio cholerae genes critical for colonization. Mol Microbiol. 1998;27: 797–805. Available: http://www.ncbi.nlm.nih.gov/pubmed/9515705

87. Mei JM, Nourbakhsh F, Ford CW, Holden DW. Identification of Staphylococcus aureus virulence genes in a murine model of bacteraemia using signature-tagged

mutagenesis. Mol Microbiol. 1997;26: 399–407. Available: http://www.ncbi.nlm.nih.gov/pubmed/9383163

88. Cormack BP, Ghori N, Falkow S. An adhesin of the yeast pathogen Candida glabrata mediating adherence to human epithelial cells. Science. 1999;285: 578–82. Available: http://www.ncbi.nlm.nih.gov/pubmed/10417386

89. Knoll LJ, Furie GL, Boothroyd JC. Adaptation of signature-tagged mutagenesis for Toxoplasma gondii: a negative screening strategy to isolate genes that are essential in restrictive growth conditions. Mol Biochem Parasitol. 2001;116: 11–16. doi:10.1016/S0166-6851(01)00295-X

90. Saenz HL, Dehio C. Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. Curr Opin Microbiol. 2005;8: 612–9. doi:10.1016/j.mib.2005.08.013

91. Mazurkiewicz P, Tang CM, Boone C, Holden DW. Signature-tagged mutagenesis : barcoding mutants for genome-wide screens. Nat Rev Genet. 2006;7: 929–939. doi:10.1038/nrg1984

92. Hunt TA, Kooi C, Sokol PA, Valvano MA. Identification of Burkholderia cenocepacia genes required for bacterial survival in vivo. Infect Immun. 2004;72: 4010–22. doi:10.1128/IAI.72.7.4010-4022.2004

93. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat Genet. 1996;14: 450–6. doi:10.1038/ng1296-450

94. Winzeler EA. Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science (80-). 1999;285: 901–906. doi:10.1126/science.285.5429.901

95. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, et al. Quantitative phenotyping via deep barcode sequencing. Genome Res. 2009;19: 1836–42. doi:10.1101/gr.093955.109

96. Han TX, Xu X-Y, Zhang M-J, Peng X, Du L-L. Global fitness profiling of fission yeast deletion strains by barcode sequencing. Genome Biol. 2010;11: R60. doi:10.1186/gb-2010-11-6-r60

97. Bateson W. FACTS LIMITING THE THEORY OF HEREDITY. Science. 1907;26: 649–60. doi:10.1126/science.26.672.649

98. Guarente L. Synthetic enhancement in gene interaction: a genetic tool come of age. Trends Genet. 1993;9: 362–366. doi:10.1016/0168-9525(93)90042-G

99. Roth FP, Lipshitz HD, Andrews BJ. Q&A: epistasis. J Biol. 2009;8: 35. doi:10.1186/jbiol144

100. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. Nat Rev Genet. 2007;8: 437–49. doi:10.1038/nrg2085

101. Mani R, St Onge RP, Hartman JL, Giaever G, Roth FP. Defining genetic interaction. Proc Natl Acad Sci U S A. 2008;105: 3461–6. doi:10.1073/pnas.0712255105

102. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. Science. 2004;303: 808–13. doi:10.1126/science.1091317

103. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002;418: 387–91. doi:10.1038/nature00935

104. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. 2010;327: 425–31. doi:10.1126/science.1180823

105. Tewari R, Straschil U, Bateman A, Böhme U, Cherevach I, Gong P, et al. The systematic functional analysis of Plasmodium protein kinases identifies essential regulators of mosquito transmission. Cell Host Microbe. 2010;8: 377–87. doi:10.1016/j.chom.2010.09.006

106. Bimbó A, Jia Y, Poh SL, Karuturi RKM, den Elzen N, Peng X, et al. Systematic deletion analysis of fission yeast protein kinases. Eukaryot Cell. 2005;4: 799–813. doi:10.1128/EC.4.4.799-813.2005

107. Tremp AZ, Dessens JT. Malaria IMC1 membrane skeleton proteins operate autonomously and participate in motility independently of cell shape. J Biol Chem. 2011;286: 5383–91. doi:10.1074/jbc.M110.187195

108. Lavazec C, Moreira CK, Mair GR, Waters AP, Janse CJ, Templeton TJ. Analysis of mutant Plasmodium berghei parasites lacking expression of multiple PbCCp genes. Mol Biochem Parasitol. 2009;163: 1–7. doi:10.1016/j.molbiopara.2008.09.002

109. Sharifpoor S, van Dyk D, Costanzo M, Baryshnikova A, Friesen H, Douglas AC, et al. Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs. Genome Res. 2012;22: 791–801. doi:10.1101/gr.129213.111

110. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, et al. Global analysis of protein phosphorylation in yeast. Nature. 2005;438: 679–84. doi:10.1038/nature04187

111. Sharifpoor S, Nguyen Ba AN, Youn J-Y, Young J-Y, van Dyk D, Friesen H, et al. A quantitative literature-curated gold standard for kinase-substrate pairs. Genome Biol. 2011;12: R39. doi:10.1186/gb-2011-12-4-r39

112. Doerig C, Billker O, Haystead T, Sharma P, Tobin AB, Waters NC. Protein kinases of malaria parasites: an update. Trends Parasitol. 2008;24: 570–7. doi:10.1016/j.pt.2008.08.007

113. Lim DC, Cooke BM, Doerig C, Saeij JPJ. Toxoplasma and Plasmodium protein kinases: roles in invasion and host cell remodelling. Int J Parasitol. Australian Society for Parasitology Inc.; 2012;42: 21–32. doi:10.1016/j.ijpara.2011.11.007

114. Sopko R, Andrews BJ. Linking the kinome and phosphorylome--a comprehensive review of approaches to find kinase targets. Mol Biosyst. 2008;4: 920–33. doi:10.1039/b801724g

115. Hanks SK. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. Genome Biol. 2003;4: 111. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=156577&tool=pmcentrez&rendertype=abstract

116. Hanks SK, Hunter T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 1995;9: 576–96. Available: http://www.ncbi.nlm.nih.gov/pubmed/7768349

117. Miranda-Saavedra D, Gabaldón T, Barton GJ, Langsley G, Doerig C. The kinomes of apicomplexan parasites. Microbes Infect. 2012;14: 796–810. doi:10.1016/j.micinf.2012.04.007

118. Ward P, Equinet L, Packer J, Doerig C. Protein kinases of the human malaria parasite Plasmodium falciparum: the kinome of a divergent eukaryote. BMC Genomics. 2004;5: 79. doi:10.1186/1471-2164-5-79

119. Talevich E, Mirza A, Kannan N. Structural and evolutionary divergence of eukaryotic protein kinases in Apicomplexa. BMC Evol Biol. 2011;11: 321. doi:10.1186/1471-2148-11-321

120. Anamika, Srinivasan N, Krupa A. A genomic perspective of protein kinases in Plasmodium falciparum. Proteins. 2005;58: 180–9. doi:10.1002/prot.20278

121. Manning G. Genomic overview of protein kinases. WormBook. 2005; 1–19. doi:10.1895/wormbook.1.60.1

122. Lasonder E, Treeck M, Alam M, Tobin AB. Insights into the Plasmodium falciparum schizont phospho-proteome. Microbes Infect. 2012;14: 811–9. doi:10.1016/j.micinf.2012.04.008

123. Schneider AG, Mercereau-Puijalon O. A new Apicomplexa-specific protein kinase family: multiple members in Plasmodium falciparum, all with an export signature. BMC Genomics. 2005;6: 30. doi:10.1186/1471-2164-6-30

124. LaRonde-LeBlanc N, Wlodawer A. A family portrait of the RIO kinases. J Biol Chem. 2005;280: 37297–300. doi:10.1074/jbc.R500013200

125. Nag S, Prasad KMN, Bhowmick A, Deshmukh R, Trivedi V. PfRIO-2 kinase is a potential therapeutic target of antimalarial protein kinase inhibitors. Curr Drug Discov Technol. 2013;10: 85–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/23082912

126. Lovejoy CA, Cortez D. Common mechanisms of PIKK regulation. DNA Repair (Amst). 2009;8: 1004–8. doi:10.1016/j.dnarep.2009.04.006

127. Pearson G, Robinson F, Beers Gibson T, Xu BE, Karandikar M, Berman K, et al. Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. Endocr Rev. 2001;22: 153–83. doi:10.1210/edrv.22.2.0428

128. Doerig C, Abdi A, Bland N, Eschenlauer S, Dorin-Semblat D, Fennell C, et al. Malaria: targeting parasite and host cell kinomes. Biochim Biophys Acta. Elsevier B.V.; 2010;1804: 604–12. doi:10.1016/j.bbapap.2009.10.009

129. Dorin-Semblat D, Quashie N, Halbert J, Sicard A, Doerig C, Peat E, et al. Functional characterization of both MAP kinases of the human malaria parasite Plasmodium falciparum by reverse genetics. Mol Microbiol. 2007;65: 1170–80. doi:10.1111/j.1365-2958.2007.05859.x

130. Tewari R, Dorin D, Moon R, Doerig C, Billker O. An atypical mitogen-activated protein kinase controls cytokinesis and flagellar motility during male gamete formation in a malaria parasite. Mol Microbiol. 2005;58: 1253–63. doi:10.1111/j.1365-2958.2005.04793.x

131. Harmon AC, Gribskov M, Harper JF. CDPKs - a kinase for every Ca2+ signal? Trends Plant Sci. 2000;5: 154–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/10740296

132. Zhang XS, Choi JH. Molecular evolution of calmodulin-like domain protein kinases (CDPKs) in plants and protists. J Mol Evol. 2001;53: 214–24. doi:10.1007/s002390010211

133. Harper JF, Harmon A. Plants, symbiosis and parasites: a calcium signalling connection. Nat Rev Mol Cell Biol. Nature Publishing Group; 2005;6: 555–66. doi:10.1038/nrm1679

134. Vaid A, Thomas DC, Sharma P. Role of Ca2+/calmodulin-PfPKB signaling pathway in erythrocyte invasion by Plasmodium falciparum. J Biol Chem. 2008;283: 5589–97. doi:10.1074/jbc.M708465200

135. Gantt S, Persson C, Rose K, Birkett AJ, Abagyan R, Nussenzweig V. Antibodies against thrombospondin-related anonymous protein do not inhibit Plasmodium sporozoite infectivity in vivo. Infect Immun. 2000;68: 3667–73. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=97657&tool=pmcentrez&rendertype=abstract

136. Ishino T, Orito Y, Chinzei Y, Yuda M. A calcium-dependent protein kinase regulates Plasmodium ookinete access to the midgut epithelial cell. Mol Microbiol. 2006;59: 1175–84. doi:10.1111/j.1365-2958.2005.05014.x

137. Ono T, Cabrita-Santos L, Leitao R, Bettiol E, Purcell LA, Diaz-Pulido O, et al. Adenylyl cyclase alpha and cAMP signaling mediate Plasmodium sporozoite apical regulated exocytosis and hepatocyte infection. PLoS Pathog. 2008;4: e1000008. doi:10.1371/journal.ppat.1000008

138. Billker O, Dechamps S, Tewari R, Wenig G, Franke-Fayard B, Brinkmann V. Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite. Cell. 2004;117: 503–14. Available: http://www.ncbi.nlm.nih.gov/pubmed/15137943

139. Green JL, Rees-Channer RR, Howell SA, Martin SR, Knuepfer E, Taylor HM, et al. The motor complex of Plasmodium falciparum: phosphorylation by a calcium-dependent protein kinase. J Biol Chem. 2008;283: 30980–9. doi:10.1074/jbc.M803129200

140. Bansal A, Singh S, More KR, Hans D, Nangalia K, Yogavel M, et al. Characterization of Plasmodium falciparum calcium-dependent protein kinase 1 (PfCDPK1) and its role in microneme secretion during erythrocyte invasion. J Biol Chem. 2013;288: 1590–602. doi:10.1074/jbc.M112.411934

141. Lauciello L, Kappes B, Scapozza L, Perozzo R. Expression, purification and biochemical characterization of recombinant Ca-dependent protein kinase 2 of the malaria parasite Plasmodium falciparum. Protein Expr Purif. 2013;90: 170–177. Available: http://www.sciencedirect.com/science/article/pii/S1046592813001125

142. Li JL, Baker DA, Cox LS. Sexual stage-specific expression of a third calcium-dependent protein kinase from Plasmodium falciparum. Biochim Biophys Acta. 2000;1491: 341–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/10760601

143. Kato K, Sudo A, Kobayashi K, Sugi T, Tohya Y, Akashi H. Characterization of Plasmodium falciparum calcium-dependent protein kinase 4. Parasitol Int. 2009;58: 394–400. doi:10.1016/j.parint.2009.08.001

144. Dvorin JD, Martyn DC, Patel SD, Grimley JS, Collins CR, Hopp CS, et al. A plant-like kinase in Plasmodium falciparum regulates parasite egress from erythrocytes. Science. 2010;328: 910–2. doi:10.1126/science.1188191

145. Solyakov L, Halbert J, Alam MM, Semblat J-P, Dorin-Semblat D, Reininger L, et al. Global kinomic and phospho-proteomic analyses of the human malaria parasite Plasmodium falciparum. Nat Commun. Nature Publishing Group; 2011;2: 565. doi:10.1038/ncomms1558

146. Jebiwott S, Govindaswamy K, Mbugua A, Bhanot P. Plasmodium berghei calcium dependent protein kinase 1 is not required for host cell invasion. PLoS One. 2013;8: e79171. doi:10.1371/journal.pone.0079171

147. Coppi A, Tewari R, Bishop JR, Bennett BL, Lawrence R, Esko JD, et al. Heparan Sulfate Proteoglycans Provide a Signal to Plasmodium Sporozoites to Stop Migrating and Productively Invade Host Cells. Cell Host Microbe. 2007;2: 316–327. Available: http://www.sciencedirect.com/science/article/pii/S193131280700248X

148. Möskes C, Burghaus PA, Wernli B, Sauder U, Dürrenberger M, Kappes B. Export of Plasmodium falciparum calcium-dependent protein kinase 1 to the parasitophorous vacuole is dependent on three N-terminal membrane anchor motifs. Mol Microbiol. 2004;54: 676–91. doi:10.1111/j.1365-2958.2004.04313.x

149. Macek B, Gnad F, Soufi B, Kumar C, Olsen J V, Mijakovic I, et al. Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics. 2008;7: 299–307. doi:10.1074/mcp.M700311-MCP200

150. Gruhler A, Olsen J V, Mohammed S, Mortensen P, Faergeman NJ, Mann M, et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics. 2005;4: 310–27. doi:10.1074/mcp.M400219-MCP200

151. Krüger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, et al. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell. 2008;134: 353–64. doi:10.1016/j.cell.2008.05.033

152. Lasonder E, Green JL, Camarda G, Talabani H, Holder AA, Langsley G, et al. The Plasmodium falciparum Schizont Phosphoproteome Reveals Extensive Phosphatidylinositol and cAMP-Protein Kinase A Signaling. J Proteome Res. 2012;11: 5323–5337. doi:10.1021/pr300557m

153. Treeck M, Sanders JL, Elias JE, Boothroyd JC. The Phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii Reveal Unusual Adaptations Within and Beyond the Parasites' Boundaries. Cell Host Microbe. 2011;10: 410–419. doi:10.1016/j.chom.2011.09.004

154. Brochet M, Collins MO, Smith TK, Thompson E, Sebastian S, Volkmann K, et al. Phosphoinositide metabolism links cGMP-dependent protein kinase G to essential Ca^{2+} signals at key decision points in the life cycle of malaria parasites. PLoS Biol. 2014;12: e1001806. doi:10.1371/journal.pbio.1001806

155. Trager W, Gill GS. Enhanced gametocyte formation in young erythrocytes by Plasmodium falciparum in vitro. J Protozool. 1992;39: 429–32. Available: http://www.ncbi.nlm.nih.gov/pubmed/1640389

156. Winger LA, Tirawanchai N, Nicholas J, Carter HE, Smith JE, Sinden RE. Ookinete antigens of Plasmodium berghei. Appearance on the zygote surface of an Mr 21 kD determinant identified by transmission-blocking monoclonal antibodies. Parasite Immunol. 1988;10: 193–207. Available: http://www.ncbi.nlm.nih.gov/pubmed/2453831

157. Saxena AK, Wu Y, Garboczi DN. Plasmodium P25 and P28 Surface Proteins: Potential Transmission-Blocking Vaccines. Eukaryot Cell. 2007;6: 1260–1265. doi:10.1128/EC.00060-07

158. Hamming RW. Error Detecting and Error Correcting Codes. Bell Syst Tech J. 1950;29: 147–160. doi:10.1002/j.1538-7305.1950.tb00463.x

159. Janse CJ, Franke-Fayard B, Mair GR, Ramesar J, Thiel C, Engelmann S, et al. High efficiency transfection of Plasmodium berghei facilitates novel selection procedures. Mol Biochem Parasitol. 2006;145: 60–70. doi:10.1016/j.molbiopara.2005.09.007

160. Janse CJ, Haghparast A, Sperança MA, Ramesar J, Kroeze H, del Portillo HA, et al. Malaria parasites lacking eef1a have a normal S/M phase yet grow more slowly due to a longer G1 phase. Mol Microbiol. 2003;50: 1539–51. Available: http://www.ncbi.nlm.nih.gov/pubmed/14651637

161. Semrov D, Miklavčič D. Numerical modeling for in vivo electroporation. Methods Mol Med. 2000;37: 63–81. doi:10.1385/1-59259-080-2:63

162. Kotnik T, Bobanović F, Miklavc̆ic̆ D. Sensitivity of transmembrane voltage induced by applied electric fields—A theoretical analysis. Bioelectrochemistry Bioenerg. 1997;43: 285–291. doi:10.1016/S0302-4598(97)00023-8

163. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol. 1998;64: 3724–30. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=106531&tool=pmcentrez&rendertype=abstract

164. Tomas AM, Margos G, Dimopoulos G, van Lin LH, de Koning-Ward TF, Sinha R, et al. P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. EMBO J. 2001;20: 3975–83. doi:10.1093/emboj/20.15.3975

165. Dessens JT, Sidén-Kiamos I, Mendoza J, Mahairaki V, Khater E, Vlachou D, et al. SOAP, a novel malaria ookinete protein involved in mosquito midgut invasion and oocyst development. Mol Microbiol. 2003;49: 319–29. Available: http://www.ncbi.nlm.nih.gov/pubmed/12828632

166. Van Dijk MR, van Schaijk BCL, Khan SM, van Dooren MW, Ramesar J, Kaczanowski S, et al. Three members of the 6-cys protein family of Plasmodium play a role in gamete fertility. PLoS Pathog. 2010;6: e1000853. doi:10.1371/journal.ppat.1000853

167. Spaccapelo R, Janse CJ, Caterbi S, Franke-Fayard B, Bonilla JA, Syphard LM, et al. Plasmepsin 4-deficient Plasmodium berghei are virulence attenuated and induce protective immunity against experimental malaria. Am J Pathol. 2010;176: 205–17. doi:10.2353/ajpath.2010.090504

168. Günther S, McMillan PJ, Wallace LJM, Müller S. Plasmodium falciparum possesses organelle-specific alpha-keto acid dehydrogenase complexes and lipoylation pathways. Biochem Soc Trans. 2005;33: 977–80. doi:10.1042/BST20050977

169. Droucheau E, Primot A, Thomas V, Mattei D, Knockaert M, Richardson C, et al. Plasmodium falciparum glycogen synthase kinase-3: molecular model, expression, intracellular localisation and selective inhibitors. Biochim Biophys Acta. 2004;1697: 181–96. doi:10.1016/j.bbapap.2003.11.023

170. Kato N, Sakata T, Breton G, Le Roch KG, Nagle A, Andersen C, et al. Gene expression signatures and small-molecule compounds link a protein kinase to Plasmodium falciparum motility. Nat Chem Biol. 2008;4: 347–56. doi:10.1038/nchembio.87

171. Angermayr M, Roidl A, Bandlow W. Yeast Rio1p is the founding member of a novel subfamily of protein serine kinases involved in the control of cell cycle progression. Mol Microbiol. 2002;44: 309–24. doi:10.1046/j.1365-2958.2002.02881.x

172. Geerlings TH, Faber AW, Bister MD, Vos JC, Raué HA. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20 S Pre-rRNA in Saccharomyces cerevisiae. J Biol Chem. 2003;278: 22537–45. doi:10.1074/jbc.M300759200

173. Widmann B, Wandrey F, Badertscher L, Wyler E, Pfannstiel J, Zemp I, et al. The kinase activity of human Rio1 is required for final steps of cytoplasmic maturation of 40S subunits. Mol Biol Cell. 2012;23: 22–35. doi:10.1091/mbc.E11-07-0639

174. Spahn CM, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G, et al. Structure of the 80S ribosome from Saccharomyces cerevisiae--tRNA-ribosome and subunit-subunit interactions. Cell. 2001;107: 373–86. Available: http://www.ncbi.nlm.nih.gov/pubmed/11701127

175. Schäfer T, Strauss D, Petfalski E, Tollervey D, Hurt E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. EMBO J. 2003;22: 1370–80. doi:10.1093/emboj/cdg121

176. Guler JL, Freeman DL, Ahyong V, Patrapuvich R, White J, Gujjar R, et al. Asexual populations of the human malaria parasite, Plasmodium falciparum, use a two-step genomic strategy to acquire accurate, beneficial DNA amplifications. PLoS Pathog. 2013;9: e1003375. doi:10.1371/journal.ppat.1003375

177. Embi N, Rylatt DB, Cohen P. Glycogen synthase kinase-3 from rabbit skeletal muscle. Separation from cyclic-AMP-dependent protein kinase and phosphorylase kinase. Eur J Biochem. 1980;107: 519–27. Available: http://www.ncbi.nlm.nih.gov/pubmed/6249596

178. Harwood AJ. Regulation of GSK-3: a cellular multiprocessor. Cell. 2001;105: 821–4. Available: http://www.ncbi.nlm.nih.gov/pubmed/11439177

179. Nurul Aiezzah Z, Noor E, Hasidah MS. Suppression of Plasmodium berghei parasitemia by LiCl in an animal infection model. Trop Biomed. 2010;27: 624–31. Available: http://www.ncbi.nlm.nih.gov/pubmed/21399604

180. Trager W, Rudzinska MA, Bradbury PC. The fine structure of Plasmodium falciparum and its host erythrocytes in natural malarial infections in man. Bull World Health Organ. 1966;35: 883–5. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2476278&tool=pmcentrez&rendertype=abstract

181. Sam-Yellowe TY. The role of the Maurer's clefts in protein transport in Plasmodium falciparum. Trends Parasitol. 2009;25: 277–84. doi:10.1016/j.pt.2009.03.009

182. Ingmundson A, Nahar C, Brinkmann V, Lehmann MJ, Matuschewski K. The exported Plasmodium berghei protein IBIS1 delineates membranous structures in infected red blood cells. Mol Microbiol. 2012;83: 1229–43. doi:10.1111/j.1365-2958.2012.08004.x

183. Haase S, Hanssen E, Matthews K, Kalanon M, de Koning-Ward TF. The exported protein PbCP1 localises to cleft-like structures in the rodent malaria parasite Plasmodium berghei. PLoS One. 2013;8: e61482. doi:10.1371/journal.pone.0061482

184. Cromer D, Evans KJ, Schofield L, Davenport MP. Preferential invasion of reticulocytes during late-stage Plasmodium berghei infection accounts for reduced circulating reticulocyte levels. Int J Parasitol. 2006;36: 1389–97. doi:10.1016/j.ijpara.2006.07.009

185. Chang K-H, Tam M, Stevenson MM. Modulation of the course and outcome of blood-stage malaria by erythropoietin-induced reticulocytosis. J Infect Dis. 2004;189: 735–43. doi:10.1086/381458

186. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannou VE. RAG-1-deficient mice have no mature B and T lymphocytes. Cell. Elsevier; 1992;68: 869–877. doi:10.1016/0092-8674(92)90030-G

187. Tullai JW, Graham JR, Cooper GM. A GSK-3-mediated transcriptional network maintains repression of immediate early genes in quiescent cells. Cell Cycle. 2011;10: 3072–7. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3218618&tool=pmcentrez&rendertype=abstract

188. Ojo KK, Pfander C, Mueller NR, Burstroem C, Larson ET, Bryan CM, et al. Transmission of malaria to mosquitoes blocked by bumped kinase inhibitors. J Clin Invest. 2012;122: 2301–5. doi:10.1172/JCI61822

189. Billker O, Lourido S, Sibley LD. Calcium-dependent signaling and kinases in apicomplexan parasites. Cell Host Microbe. 2009;5: 612–22. doi:10.1016/j.chom.2009.05.017

190. Francis SH, Busch JL, Corbin JD, Sibley D. cGMP-dependent protein kinases and cGMP phosphodiesterases in nitric oxide and cGMP action. Pharmacol Rev. 2010;62: 525–63. doi:10.1124/pr.110.002907

191. Shokat K, Velleca M. Novel chemical genetic approaches to the discovery of signal transduction inhibitors. Drug Discov Today. 2002;7: 872–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/12546954

192. Taylor HM, McRobert L, Grainger M, Sicard A, Dluzewski AR, Hopp CS, et al. The malaria parasite cyclic GMP-dependent protein kinase plays a central role in blood-stage schizogony. Eukaryot Cell. 2010;9: 37–45. doi:10.1128/EC.00186-09

193. Diaz CA, Allocco J, Powles MA, Yeung L, Donald RGK, Anderson JW, et al. Characterization of Plasmodium falciparum cGMP-dependent protein kinase (PfPKG):

antiparasitic activity of a PKG inhibitor. Mol Biochem Parasitol. 2006;146: 78–88. doi:10.1016/j.molbiopara.2005.10.020

194. McRobert L, Taylor CJ, Deng W, Fivelman QL, Cummings RM, Polley SD, et al. Gametogenesis in malaria parasites is mediated by the cGMP-dependent protein kinase. Ward GE, editor. PLoS Biol. Public Library of Science; 2008;6: e139. doi:10.1371/journal.pbio.0060139

195. Shimizu H, Nagamori I, Yabuta N, Nojima H. GAK, a regulator of clathrin-mediated membrane traffic, also controls centrosome integrity and chromosome congression. J Cell Sci. 2009;122: 3145–52. doi:10.1242/jcs.052795

196. Dearsly AL, Sinden RE, Self IA. Sexual development in malarial parasites: gametocyte production, fertility and infectivity to the mosquito vector. Parasitology. 1990;100 Pt 3: 359–68. Available: http://www.ncbi.nlm.nih.gov/pubmed/2194152

197. Shimomura T, Ando S, Matsumoto K, Sugimoto K. Functional and physical interaction between Rad24 and Rfc5 in the yeast checkpoint pathways. Mol Cell Biol. 1998;18: 5485–91. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=109133&tool=pmcentrez&rendertype=abstract

198. Fujioka Y, Kimata Y, Nomaguchi K, Watanabe K, Kohno K. Identification of a novel non-structural maintenance of chromosomes (SMC) component of the SMC5-SMC6 complex involved in DNA repair. J Biol Chem. 2002;277: 21585–91. doi:10.1074/jbc.M201523200

199. Ranjan R, Ahmed A, Gourinath S, Sharma P. Dissection of mechanisms involved in the regulation of Plasmodium falciparum calcium-dependent protein kinase 4. J Biol Chem. 2009;284: 15267–76. doi:10.1074/jbc.M900656200

200. Glushakova S, Lizunov V, Blank PS, Melikov K, Humphrey G, Zimmerberg J. Cytoplasmic free Ca2+ is essential for multiple steps in malaria parasite egress from infected erythrocytes. Malar J. 2013;12: 41. doi:10.1186/1475-2875-12-41

201. Collins CR, Hackett F, Strath M, Penzo M, Withers-Martinez C, Baker DA, et al. Malaria parasite cGMP-dependent protein kinase regulates blood stage merozoite secretory organelle discharge and egress. PLoS Pathog. 2013;9: e1003344. doi:10.1371/journal.ppat.1003344

202. Collins CR, Das S, Wong EH, Andenmatten N, Stallmach R, Hackett F, et al. Robust inducible Cre recombinase activity in the human malaria parasite Plasmodium falciparum enables efficient gene deletion within a single asexual erythrocytic growth cycle. Mol Microbiol. 2013;88: 687–701. doi:10.1111/mmi.12206

203. Parsons AB, Brost RL, Ding H, Li Z, Zhang C, Sheikh B, et al. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat Biotechnol. Nature Publishing Group; 2004;22: 62–9. doi:10.1038/nbt919

204. Polley SD, Weedall GD, Thomas AW, Golightly LM, Conway DJ. Orthologous gene sequences of merozoite surface protein 1 (MSP1) from Plasmodium reichenowi and P. gallinaceum confirm an ancient divergence of P. falciparum alleles. Mol Biochem Parasitol. 2005;142: 25–31. doi:10.1016/j.molbiopara.2005.02.012

205. Straimer J, Lee MCS, Lee AH, Zeitler B, Williams AE, Pearl JR, et al. Site-specific genome editing in Plasmodium falciparum using engineered zinc-finger nucleases. Nat Methods. 2012;9: 993–8. doi:10.1038/nmeth.2143

206. Wagner JC, Platt RJ, Goldfless SJ, Zhang F, Niles JC. Efficient CRISPR-Cas9–mediated genome editing in Plasmodium falciparum. Nat Methods. 2014;11: 915–8. doi:10.1038/nmeth.3063

207. Moon RW, Hall J, Rangkuti F, Ho YS, Almond N, Mitchell GH, et al. Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. Proc Natl Acad Sci U S A. 2013;110: 531–6. doi:10.1073/pnas.1216457110

Appendices

Appendix I – List of *Plasmo*GEM IDs for each gene and corresponding annotation.

Gene ID	*Plasmo*GEM ID	Gene name
p230p-tag	PbGEM-226060	p230p-tag
PBANKA_051490	PbGEM-015545	28 kDa ookinete surface protein (P28)
PBANKA_051500	PbGEM-015561	25 kDa ookinete surface antigen precursor (P25)
PBANKA_103780	PbGEM-097822	secreted ookinete adhesive protein (SOAP)
PBANKA_103440	PbGEM-039254	plasmepsin IV (PM4)
PBANKA_110420	PbGEM-122074	3-methyl-2-oxobutanoate dehydrogenase (lipoamide), putative
PBANKA_140160	PbGEM-062476	methyl transferase-like protein, putative
PBANKA_020580	PbGEM-082161	serine/threonine protein kinase, putative (IK2)
PBANKA_030850	PbGEM-009884	protein kinase, putative (TKL1)
PBANKA_031030	PbGEM-072470	protein kinase 7 (PK7)
PBANKA_031140	PbGEM-111762	serine/threonine protein kinase, putative
PBANKA_031420	PbGEM-010677	calcium dependent protein kinase 1 (CDPK1)
PBANKA_040110	PbGEM-084034	serine/threonine protein kinase, putative (SRPK1)
PBANKA_040820	PbGEM-111826	calcium dependent protein kinase 3 (CDPK3)
PBANKA_040940	PbGEM-111754	protein kinase, putative (PKRP)
PBANKA_041040	PbGEM-111698	glycogen synthase kinase 3 (GSK3)
PBANKA_052140	PbGEM-072474	RIO-like serine/threonine kinase, putative
PBANKA_061520	PbGEM-087803	calcium dependent protein kinase 4 (CDPK4)
PBANKA_061670	PbGEM-111690	NIMA related kinase 4 (NEK4)
PBANKA_071730	PbGEM-072538	cdc2-related protein kinase 3 (CRK3)
PBANKA_080560	PbGEM-072522	O-sialoglycoprotein endopeptidase, putative
PBANKA_080800	PbGEM-111786	cdc2-related protein kinase 4 (CRK4)
PBANKA_082710	PbGEM-072502	protein kinase, putative
PBANKA_083560	PbGEM-028140	cAMP-dependent protein kinase catalytic subunit (PKAc)
PBANKA_090110	PbGEM-111746	protein kinase, putative
PBANKA_090380	PbGEM-111794	serine/threonine protein kinase, putative
PBANKA_091210	PbGEM-111738	casein kinase 1 (CK1)
PBANKA_092520	PbGEM-093973	calcium-dependent protein kinase 7 (CDPK7)
PBANKA_092550	PbGEM-111850	calcium dependent protein kinase 6 (CDPK6)
PBANKA_093300	PbGEM-072518	serine/threonine protein kinase, puative
PBANKA_093370	PbGEM-111778	mitogen-activated protein kinase 2 (MAP2)
PBANKA_101330	PbGEM-036210	mitogen-activated protein kinase 1 (MAP1)
PBANKA_101980	PbGEM-111858	serine/threonine protein kinase, putative
PBANKA_112270	PbGEM-111714	protein kinase, putative (TKL4)
PBANKA_112690	PbGEM-099789	protein kinase PK4 (PK4)
PBANKA_122500	PbGEM-111674	serine/threonine protein kinase, FIKK family
PBANKA_130520	PbGEM-053796	serine/threonine protein kinase, putative
PBANKA_130690	PbGEM-104812	serine/threonine protein kinase, putative (SRPK2)
PBANKA_130920	PbGEM-104970	serine/threonine kinase-1, putative
PBANKA_131800	PbGEM-105530	serine/threonine protein kinase, putative
PBANKA_135150	PbGEM-111682	calcium dependent protein kinase 5 (CDPK5)
PBANKA_135260	PbGEM-111802	serine/threonine protein kinase, putative
PBANKA_136210	PbGEM-108848	protein kinase, putative
PBANKA_141450	PbGEM-111842	protein kinase, putative
PBANKA_142160	PbGEM-065291	calcium/calmodulin-dependent protein kinase, putative
PBANKA_144560	PbGEM-111706	protein kinase, putative
PBANKA_146050	PbGEM-072542	serine/threonine protein kinase, putative

Appendix II – Primers used to genotype cloned mutants.

Primer name	Sequence	Target/Comment
arg00059	GAATCACAATTGACCAGG	
arg00060	CAGTAAATTGCTATGATAAATC	
arg00080	GACTTCTGTAGCCATGATAGC	hdhfr::yfcu
arg00081	CGCCACACTACATGGTGAG	hdhfr::yfcu
arg00082	GAGCTCATAGTTATTGCTATTGC	
arg00084	AAAGAATTCTGATGGTTTACAATCACC	RNApol II
arg00085	AAAGCGGCCGCTTTCTTCCTGCATCTCCTC	RNApol II
arg00087	GTGCATTAACAGTTAGAAGAGG	
arg00089	CAGAAACACAATGTTGAAATTC	
arg00094	GTTTCGATTGCACGGACTTTG	
arg00102	GTGTAGAAGTAAATTCATACCC	
arg00214	GAACGGCACTGGTCAACTTG	
arg00215	TCATTCTTCGAAAACGATCT	
arg00216	CGGGGCCCTTATGCATAATC	
arg00218	CTTTGGTGACAGATACTACTG	
arg00255	GGGGGTTTGTGTGGAGGCG	
arg00331	TTGAACATTTGCGCATATATTGG	
arg00332	GTGCCAAATTATTATGGTATACC	
arg00367	CTCCAGCATATACTTGCATAG	
arg00368	CTTCACCAAATGAACCCTTTC	
arg00369	GATTGGGAGATGTTAGTTCTG	
arg00370	GGTTATGAGAAGTTAAACTACG	
arg00371	CCGGTATTTATCATCCAAGAG	
arg00372	CATTTGTTCATGCAATCATTCG	
arg00373	CCCACTACCTAATTTCCGAAC	
arg00374	GCATCATTGACACGAACTCG	
arg00375	GTGAAGAGAGGCTAAGGAGG	
arg00381	TGCACTTCACCAAAAGCGCCA	
arg00382	TCAACGGGAGGTAGCTCCAA	
arg00383	CCACGATCATTTAGAAAACACG	
arg00384	GCTCATATGCATTATACGCTTC	
arg00385	TACACTTTGAGGTTAATGTGC	
arg00386	GGTTATCTATACATTTATTTGTG	
arg00387	CTCTTAAAACTTGGGGGTAGG	
arg00388	GCAACAAGAAAGGAAGCATAAC	
arg00389	CTCCCTTCAAATTTATGCTGAC	
arg00390	GGGAAGTAACCTAATTTGCTG	
arg00396	GGCATGCCGTATTTTCCATG	
arg00397	ATGCATATTCTTTTGTTGCAGC	
arg00398	CACCCATCCAAACATATAAAAG	
arg00399	CTAGTAATAGTCAGTCTGGGG	

Appendix II – (cont) - Primers used to genotype cloned mutants.

Primer name	Sequence	Target/Comment
arg00400	TCCAAAGCGTTATGCCAAGTGT	
arg00401	GCAACTCGTTATATATTTCCG	
arg00402	CTTTTCCAGAACTAACTACTCC	
arg00403	CTTTTCTTGTAACACCCTCAG	
arg00404	TGGAACTACAAAATAGATTTCG	
arg00405	GCTCAAGCAACAGCAGGAC	
arg00406	ATCCAAATATGGTATTTTGAGC	
arg00407	CTGGAGATTCGTTTTGTTCAC	
arg00408	TTGTGTTTGCCGCATGTTGC	
arg00409	GATATTCCTAAAGATCTATCTG	
arg00410	ACGACAATGTGCATGCCTCA	
arg00416	CGTACTTGAATAGCTGTCTAC	
arg00417	CCCCAATATAATGAATATTCTG	
arg00418	GAGCATTCCGCAAAGTATGTC	
arg00419	GGGAACACATCCTTTTAGTTC	
arg00420	AGCCATTACCCGTTGTTTCG	
arg00421	GTGCCAAATTATTATGGTATAC	
arg00422	CCAGAATAATTTTGTTAGAATATAG	
arg00423	GTACAGTTTGTGGTATATATTCC	
arg00424	GGTGATAATAATGCATGCCAAG	
arg00425	GGTTCTATCTGTTTCATGTAC	
arg00426	CCTCAGAAAATGAATGGCAG	
arg00427	GGTGTAATCAAAGTAAAAGTGC	
arg00428	CTCCACCTTCATAAAATTCGG	
arg00429	CATAAAAGATTAAATCGTGCTCG	
arg00430	CAATTATGTGTAGAAAGAAACC	
arg00431	CCCTTTCCTAGTGGTTCTTC	
arg00432	GCCAATGTCCAAATAAATGTATC	
arg00433	CTATGCAAGTATATGCTGGAG	
arg00434	GGGGCAAGAAATGTCTACAC	
arg00435	CGCCACCTTCACATAATTCC	
arg00436	GATTTCCGTCAAATGTATTGGG	
arg00437	CCATCTGTTACTTTCATAGGG	

Appendix III – Primers used for genotyping of the STM screen

Gene ID	Gene specific primer	Sequence	Pairs with
p230p-tag	arg00448	GGAACAATATGGCTGTTCAATG	arg00218
PBANKA_051490	arg00447	GGATTCCGTGAATGATCCCC	arg00216
PBANKA_051500	arg00449	TGTTCCCGTTGTAACAGTGCA	arg00216
PBANKA_103780	arg00446	TTTCCCACTGCGTACCCTTT	arg00218
PBANKA_103440	arg00452	AGACAAACTTTGCCCACAACA	arg00216
PBANKA_110420	arg00451	AAAGCCAGAAACGACATGAA	arg00218
PBANKA_140160	arg00450	CATGGCTATGACCGACAGAG	arg00218
PBANKA_020580	arg00471	CGAAGCGCTTTACCATGTGGGC	arg00218
PBANKA_030850	arg00472	AGTGCATACGCTTCATGACGCT	arg00216
PBANKA_031030	arg00473	AACCGAAGTGCTCTTTGCGA	arg00216
PBANKA_031140	arg00474	TGCTACCTTACGCATTGGACA	arg00218
PBANKA_031420	arg00371	CCGGTATTTATCATCCAAGAG	arg00216
PBANKA_040110	arg00475	TGATGCGGATTTGTGTGTGT	arg00218
PBANKA_040820	arg00386	GGTTATCTATACATTTATTTGTG	arg00218
PBANKA_040940	arg00476	AGCAATGATGTAGGATGTGCA	arg00216
PBANKA_041040	arg00470	AGCGAGTTCCCGTGCACTCA	arg00218
PBANKA_052140	arg00477	TCCAAAGCGTTATGCCAAGTGT	arg00218
PBANKA_061520	arg00255	GGGGGTTTGTGTGGAGGCG	arg00216
PBANKA_061670	arg00234	GCACACTCACCTGAAATGTCG	arg00216
PBANKA_071730	arg00478	TGGTTCAATTGTTGAGCAAAGTCCT	arg00218
PBANKA_080560	arg00479	TCGGAAAAACCTTGAAAGCGCT	arg00218
PBANKA_080800	arg00480	TCGCTAGTTATATATGCTCACGCT	arg00216
PBANKA_082710	arg00481	ACATTACCAGCAGTTGCCCA	arg00216
PBANKA_083560	arg00482	TCAAGTGAAACGGAATAGAAGCGA	arg00218
PBANKA_090110	arg00483	TCAGAAAGGTATACGTCAACGGT	arg00216
PBANKA_090380	arg00484	AGCTTGTATGTCGATTCGAGA	arg00218
PBANKA_091210	arg00485	ACGATGTGTGCAGCAGGTCT	arg00216
PBANKA_092520	arg00486	TGTCTCCCTAAAAGGCATGTGCA	arg00218
PBANKA_092550	arg00381	TGCACTTCACCAAAAGCGCCA	arg00218
PBANKA_093300	arg00487	AGCAGTGCACACAAAAGAAGA	arg00218
PBANKA_093370	arg00256	ACCATGAGTGCATGCATAGGA	arg00216
PBANKA_101330	arg00453	CGCGTGGAAAACTGTGGGC	arg00216
PBANKA_101980	arg00488	TGCCCGGAATGCACATATGTTGC	arg00216
PBANKA_112270	arg00489	TGGGGAGTACCTTGCCCATGCA	arg00216
PBANKA_112690	arg00490	AGTATTGCCCCATCCATTGCT	arg00218
PBANKA_122500	arg00491	TGTCTGACTCTCCATGGTGTCCCA	arg00218
PBANKA_130520	arg00456	GAGTACCTGTTGGTCACGC	arg00216
PBANKA_130690	arg00225	TGCCCTTTTGATGCCAAGACG	arg00216
PBANKA_130920	arg00492	TCACGCATCGGGGATTTGTCA	arg00216
PBANKA_131800	arg00493	ACGGAGCACAATGTATGCCATGGA	arg00218
PBANKA_135150	arg00494	TCGACGGTACTGTCTGACTGGTCA	arg00216
PBANKA_135260	arg00259	CTGGCGCACGGCAAAACCC	arg00218
PBANKA_136210	arg00262	ACGACAATGTGCATGCCTCA	arg00218
PBANKA_141450	arg00272	CCACAAAGCAATTCCGGTGC	arg00218
PBANKA_142160	arg00454	TCTAAATCGCGGCTTTCACA	arg00218
PBANKA_144560	arg00495	TGCTCAAGCAACAGCAGGACA	arg00216
PBANKA_146050	arg00173	CCTGGAATTGTTTCCCCACAC	arg00218

Appendix IV – Primers used for barcode sequencing.

Primer name	Sequence
arg00444	TCGGCATTCCTGCTGAACCGCTCTTCCGATCTGTAATTCGTGCGCGTCAG
arg00445	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTCAATTTCGATGGGTAC
PE1.0	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T 1)
iPCRindex1	CAAGCAGAAGACGGCATACGAGATTGCTAATCACTGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex2	CAAGCAGAAGACGGCATACGAGATTAGGGGGATTCGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex3	CAAGCAGAAGACGGCATACGAGATAGTTTCCCAGGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex4	CAAGCAGAAGACGGCATACGAGATCCTGGGAGGTAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex5	CAAGCAGAAGACGGCATACGAGATATACCACAAATGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex6	CAAGCAGAAGACGGCATACGAGATGATCTCTCGGGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex7	CAAGCAGAAGACGGCATACGAGATACCCTATACTCGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex8	CAAGCAGAAGACGGCATACGAGATCTCAATTAAGAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex9	CAAGCAGAAGACGGCATACGAGATCGACAGAACGTGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex10	CAAGCAGAAGACGGCATACGAGATTCGCCATTATGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex11	CAAGCAGAAGACGGCATACGAGATATGTTCCGGCCGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex12	CAAGCAGAAGACGGCATACGAGATTCTTGAAGTGAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex13	CAAGCAGAAGACGGCATACGAGATGAAGGCCAGCTGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex14	CAAGCAGAAGACGGCATACGAGATCCAATGTGCAGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex15	CAAGCAGAAGACGGCATACGAGATATCGAAGGACCGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex16	CAAGCAGAAGACGGCATACGAGATTCGGGTGCGAAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex17	CAAGCAGAAGACGGCATACGAGATGTAATTTACGGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex18	CAAGCAGAAGACGGCATACGAGATATATCGACTACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex19	CAAGCAGAAGACGGCATACGAGATTGATTCTTACAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex20	CAAGCAGAAGACGGCATACGAGATACGGCGGGCCTGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex21	CAAGCAGAAGACGGCATACGAGATCTTGCGTGGAGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex22	CAAGCAGAAGACGGCATACGAGATTAATCAAAGACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex23	CAAGCAGAAGACGGCATACGAGATGGCGGGCTCTAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex24	CAAGCAGAAGACGGCATACGAGATCCTCCATTTCTGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex25	CAAGCAGAAGACGGCATACGAGATAACCAGCGCTGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex26	CAAGCAGAAGACGGCATACGAGATTATTCGTCAACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex27	CAAGCAGAAGACGGCATACGAGATGCGCTGATGCAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex28	CAAGCAGAAGACGGCATACGAGATCTCATATGGCTGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex29	CAAGCAGAAGACGGCATACGAGATACAGGGGCAGGGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex30	CAAGCAGAAGACGGCATACGAGATGGTTTTATACCGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex31	CAAGCAGAAGACGGCATACGAGATGCATGACTTTAGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T
iPCRindex32	CAAGCAGAAGACGGCATACGAGATTTCTGAGTTCTGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T

Appendix V – Targetability and fitness measurements for ePKs.

				Day 5 p. t.			Day 6 p. t.			Day 7 p. t.			Day 8 p. t.				
P. berghei **gene ID**	**Gene name**	**Tewari *et al.* 2010**	**This study**	**Fitness**	**SD**	**p**	**Fitness**	**SD**	**p**	**Fitness**	**SD**	**p**	**Fitness**	**SD**	**p**	**Average fitness, days 5-8**	**Assessment**
PBANKA_135150	*cdpk5*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_092520	*cdpk7*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_031140		Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_083560	*pka*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_090380		Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_091210	*ck1*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_071730	*crk-3*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_130920	*lammer/(CLK1)*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_093300	*prk4*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_080800	*crk-4*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_090110		Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_093860	*ck2*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_101090	*tkl5*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_144300	*nek-1*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_135090	*pk6*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_080560		Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_122500	*fikk*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_100820	*pkg*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_112690	*pk4*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential
PBANKA_052140	*rio2*	Possibly essential	Possibly essential	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	**NA**	NA	NA	NA	Possibly essential†□
PBANKA_040110	*srpk*	KO confirmed	Targetable	**0.59**	*0.18*	2.5E-02	**0.65**	*0.24*	1.3E-01	**0.80**	*0.08*	7.8E-02	**0.83**	*0.15*	2.1E-01	0.72	Confirmed KO
PBANKA_040820	*cdpk3*	KO confirmed	Targetable	**1.00**	*0.09*	1.0E+00	**0.97**	*0.05*	2.2E-01	**1.01**	*0.07*	7.6E-01	**1.04**	*0.06*	2.9E-01	1.01	Confirmed KO
PBANKA_040940	*pkrp*	KO confirmed	Targetable	**0.98**	*0.21*	9.2E-01	**1.03**	*0.09*	7.4E-01	**0.98**	*0.09*	9.0E-01	**1.02**	*0.05*	8.5E-01	1.00	Confirmed KO
PBANKA_061520	*cdpk4*	KO confirmed	Targetable	**0.94**	*0.13*	5.8E-01	**0.89**	*0.10*	4.1E-02	**0.98**	*0.07*	7.4E-01	**0.81**	*0.08*	1.4E-02	0.90	Confirmed KO
PBANKA_061670	*nek-4*	KO confirmed	Targetable	**0.87**	*0.19*	2.0E-01	**1.00**	*0.16*	9.7E-01	**1.02**	*0.11*	7.4E-01	**1.02**	*0.16*	8.5E-01	0.98	Confirmed KO
PBANKA_082710		KO confirmed	Targetable	**0.38**	*0.22*	2.5E-03	**0.65**	*0.06*	1.7E-04	**0.91**	*0.08*	2.9E-02	**1.03**	*0.38*	8.4E-01	0.74	Confirmed KO
PBANKA_092550	*cdpk6*	KO confirmed	Targetable	**0.86**	*0.12*	8.9E-02	**0.90**	*0.05*	8.7E-03	**0.79**	*0.10*	8.1E-03	**0.97**	*0.12*	6.8E-01	0.88	Confirmed KO
PBANKA_093370	*map-2*	KO confirmed	Targetable	**0.99**	*0.11*	9.5E-01	**0.95**	*0.06*	1.8E-01	**1.05**	*0.06*	2.1E-01	**1.03**	*0.05*	4.2E-01	1.00	Confirmed KO
PBANKA_101330	*map-1*	KO confirmed	Targetable	**1.04**	*0.11*	5.8E-01	**0.92**	*0.04*	4.1E-03	**1.06**	*0.04*	4.1E-02	**1.04**	*0.04*	8.5E-02	1.02	Confirmed KO
PBANKA_101980	*cdlk*	KO confirmed	Targetable	**0.20**	*0.02*	5.0E-06	**0.53**	*0.13*	1.7E-02	**0.62**	*0.42*	4.2E-01	**1.13**	*0.19*	8.0E-01	0.62	Confirmed KO
PBANKA_112270	*tkl4*	KO confirmed	Targetable	**0.46**	*0.33*	7.8E-02	**0.28**	*0.11*	2.9E-03	**0.55**	*0.16*	3.0E-02	**0.69**	*0.11*	6.2E-02	0.49	Confirmed KO
PBANKA_130690	*srpk2*	KO confirmed	Targetable	**0.83**	*0.32*	5.3E-01	**1.00**	*0.09*	9.7E-01	**1.01**	*0.14*	9.5E-01	**1.05**	*0.09*	3.9E-01	0.97	Confirmed KO
PBANKA_131800	*kin*	KO confirmed	Targetable	**1.08**	*0.29*	6.9E-01	**0.99**	*0.10*	1.9E-01	**0.97**	*0.07*	2.0E-01	**1.00**	*0.11*	8.4E-01	1.01	Confirmed KO
PBANKA_135260		KO confirmed	Targetable	**0.88**	*0.27*	5.8E-01	**0.98**	*0.07*	5.6E-01	**1.00**	*0.09*	9.5E-01	**1.03**	*0.04*	3.9E-01	0.97	Confirmed KO
PBANKA_146050	*gak*	KO confirmed	Targetable	**0.83**	*0.12*	8.9E-02	**0.74**	*0.09*	2.6E-03	**0.83**	*0.05*	1.4E-04	**1.17**	*0.11*	8.5E-02	0.89	Confirmed KO
PBANKA_030850	*tkl1*	Possibly essential	Targetable	**0.86**	*0.28*	5.3E-01	**0.96**	*0.21*	8.4E-01	**1.14**	*0.18*	1.7E-01	**1.04**	*0.06*	3.9E-01	1.00	New KO
PBANKA_031420	*cdpk1*	Possibly essential	Targetable §	**0.95**	*0.14*	5.5E-01	**0.89**	*0.05*	1.4E-03	**1.10**	*0.05*	2.7E-03	**1.00**	*0.06*	9.5E-01	0.98	New KO Ω
PBANKA_041040	*gsk-3*	Possibly essential	Targetable	**0.99**	*0.03*	8.2E-01	**1.00**	*0.04*	9.3E-01	**1.05**	*0.05*	5.6E-01	**1.01**	*0.03*	8.5E-01	1.01	New KO Ω
PBANKA_082960		Possibly essential	Targetable	**1.03**	*0.04*	7.5E-01	**1.02**	*0.08*	7.5E-01	**1.01**	*0.06*	7.5E-01	**1.05**	*0.02*	2.5E-02	1.03	New KO ◊
PBANKA_130520		Possibly essential	Targetable	**1.02**	*0.12*	9.5E-01	**0.98**	*0.13*	8.4E-01	**1.01**	*0.10*	8.3E-01	**0.99**	*0.08*	9.5E-01	1.00	New KO *
PBANKA_136210	*tkl3*	Possibly essential	Targetable	**1.05**	*0.22*	7.7E-01	**0.99**	*0.09*	9.6E-01	**1.06**	*0.14*	4.6E-01	**0.92**	*0.09*	2.0E-01	1.01	New KO *◊
PBANKA_141450		Possibly essential	Targetable	**0.91**	*0.33*	7.3E-01	**1.00**	*0.16*	9.7E-01	**0.94**	*0.07*	1.7E-01	**0.99**	*0.10*	8.5E-01	0.96	New KO *
PBANKA_142160		Possibly essential	Targetable	**1.01**	*0.13*	1.0E+00	**0.93**	*0.07*	3.4E-02	**1.10**	*0.16*	2.4E-01	**1.01**	*0.12*	9.5E-01	1.01	New KO *
PBANKA_144560	*rio1*	Possibly essential	Targetable	**0.23**	*0.00*	2.0E-03	**0.67**	*0.04*	1.1E-02	**0.91**	*0.08*	2.7E-01	**0.94**	*0.01*	5.9E-02	0.69	New KO ◊†
PBANKA_020580	*eik2; uis1*	KO confirmed	No integration	**NA**	*NA*	NA	**NA**	*NA*	NA	**NA**	*NA*	NA	**NA**	*NA*	NA	NA	False negative
PBANKA_031030	*pk7*	KO confirmed	No integration	**NA**	*NA*	NA	**NA**	*NA*	NA	**NA**	*NA*	NA	**NA**	*NA*	NA	NA	False negative

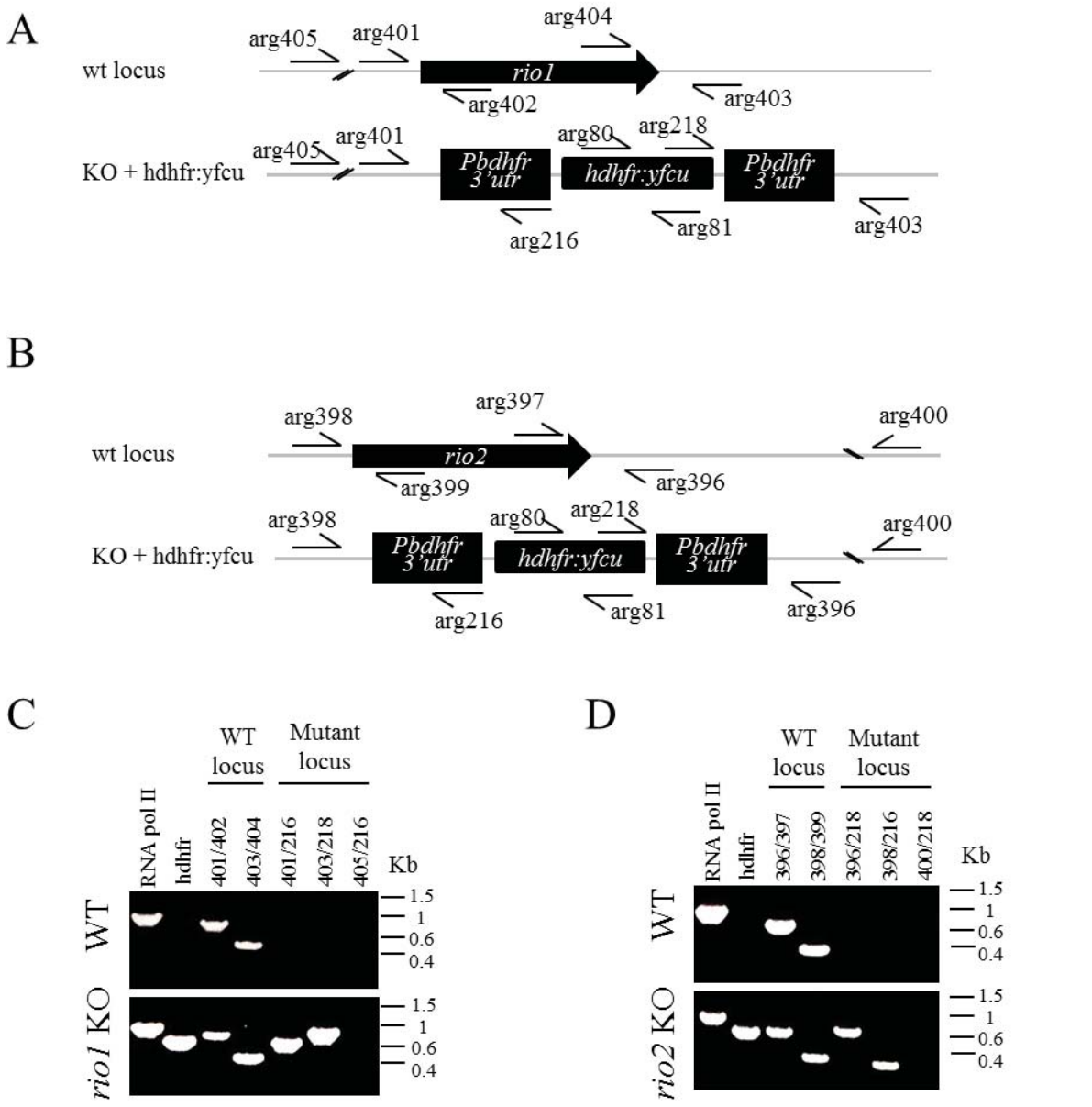
Appendix V – (cont) Targetability and fitness measurements for ePKs.

				Day 5 p. t.			Day 6 p. t.			Day 7 p. t.			Day 8 p. t.				
***P. berghei* gene ID**	**Gene name**	**Tewari *et al.* 2010**	**This study**	**Fitness**	**SD**	**p**	**Fitness**	**SD**	**p**	**Fitness**	**SD**	**p**	**Fitness**	**SD**	**p**	**Average fitness, days 5-8**	**Assessment**
PBANKA_051490	*p28*		Targetable	**0.92**	*0.15*	5.3E-01	**0.98**	*0.10*	9.2E-01	**0.98**	*0.05*	7.4E-01	**1.01**	*0.06*	8.5E-01	0.97	Normal growth reference
PBANKA_051500	*p25*		Targetable	**1.01**	*0.12*	9.8E-01	**1.03**	*0.02*	1.0E-01	**1.04**	*0.07*	2.4E-01	**0.98**	*0.04*	3.9E-01	1.02	Normal growth reference
PBANKA_103780	*soap*		Targetable	**1.00**	*0.08*	1.0E+00	**0.95**	*0.07*	1.1E-01	**1.06**	*0.05*	1.1E-01	**1.03**	*0.05*	3.8E-01	1.01	Normal growth reference
PBANKA_030600	*p230p*		Taggable	**1.07**	*0.12*	5.8E-01	**1.04**	*0.07*	4.7E-01	**0.91**	*0.03*	8.1E-03	**0.98**	*0.09*	8.5E-01	1.00	Normal growth reference
PBANKA_140160			Targetable	**0.51**	*0.12*	1.1E-04	**0.46**	*0.06*	1.3E-05	**0.58**	*0.07*	2.1E-05	**0.87**	*0.07*	1.4E-02	0.60	Attenuated reference
PBANKA_110420	*bckdh e1b*		Targetable	**0.74**	*0.12*	6.9E-03	**0.58**	*0.07*	1.3E-05	**0.67**	*0.05*	2.8E-06	**0.61**	*0.06*	6.6E-05	0.65	Attenuated reference
PBANKA_103440	*pm4*		Targetable	**0.79**	*0.20*	8.9E-02	**0.73**	*0.05*	1.3E-05	**0.71**	*0.12*	4.3E-03	**0.70**	*0.08*	2.2E-03	0.73	Attenuated reference

* PCR genotyping evidence for targetability but no independent clone generated.
§ This study and Jebiwott *et al*., 2013
† Confirmed by WGS of uncloned population
□ Confirmed by selection for target duplication
Ω Confirmed by genotyped clone
◊ Confirmed by PFGE analysis of uncloned population

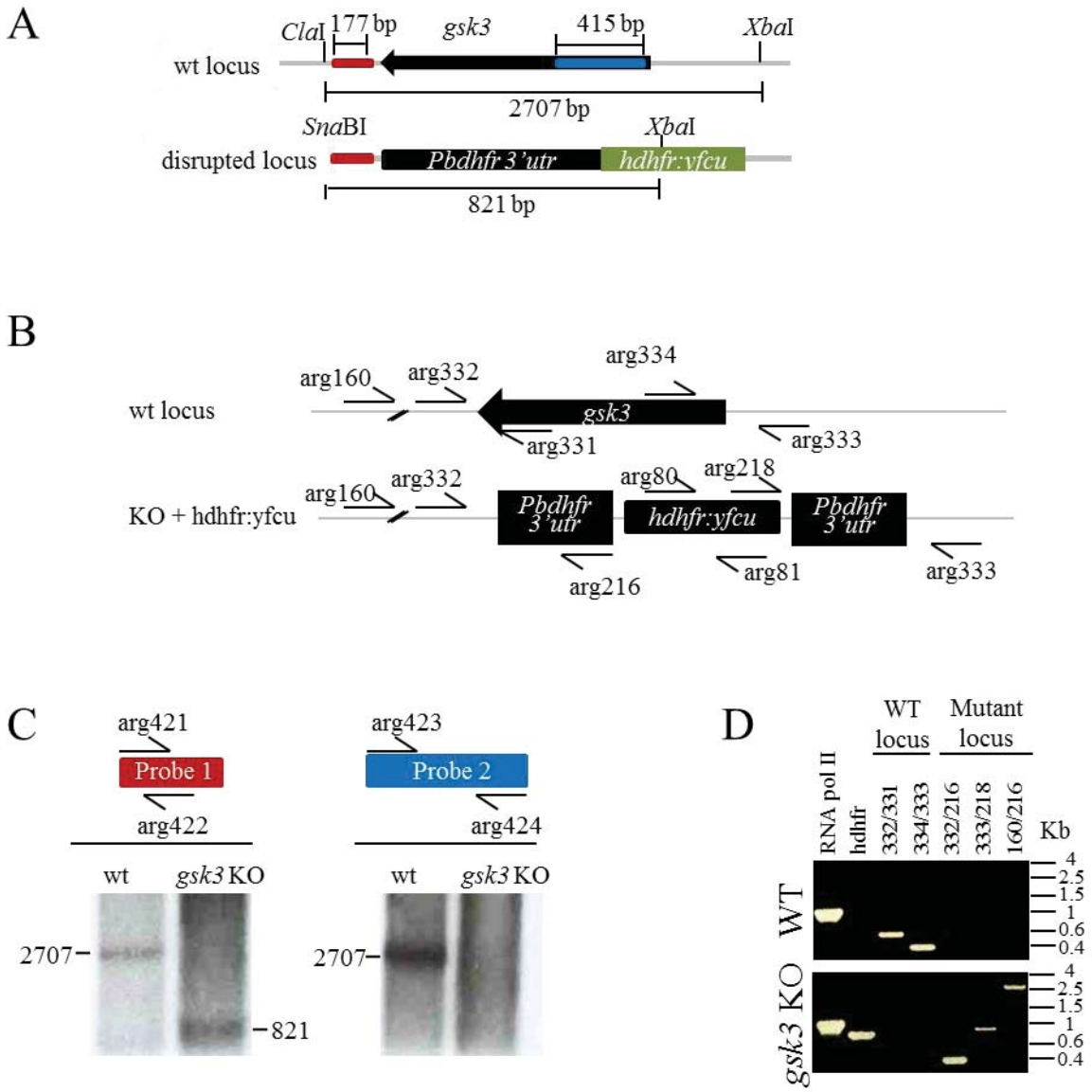
Targetability and fitness measurements for 46 eukaryotic protein kinase genes as determined by barcode sequencing, compared to data from a previous study by Tewari et al.[105], which used conventional gene targeting. p values are adjusted for multiple testing.

Appendix VI – Genotyping strategy for *rio1* and *rio2* KO mutants



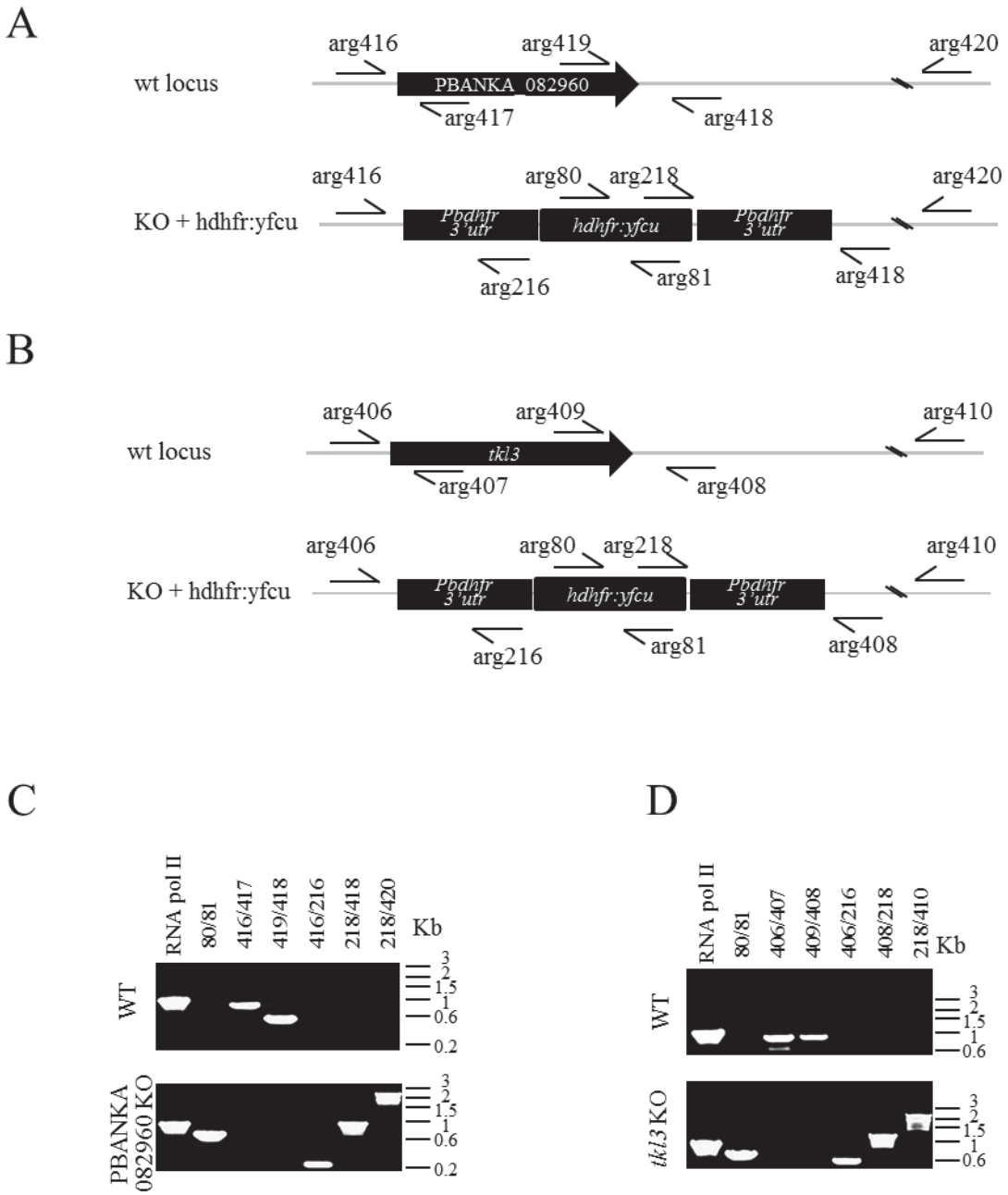
Genotyping of *rio* KO mutants.
(A,B) PCR genotyping strategy for *rio1* and *rio2* mutants, respectively; (C,D) PCR genotyping results showing the presence of both the KO vector and the WT gene in the target locus.

Appendix VII – Genotyping strategy for *gsk3* KO clones



Genotyping of *gsk3* KO clones.
(A) Southern blot strategy; (B) PCR genotyping strategy; (C) Southern blot results showing the correct diagnostic digestion patterns; (D) PCR genotyping results confirming *gsk3* gene deletion.

Appendix VIII – Genotyping strategy for the KO mutants of the PBANKA_082960 and *tkl3* genes.



Genotyping of the PBANKA_082960 and *tkl3* KO mutants.
(A,B) PCR genotyping strategy for the PBANKA_082960 and *tkl3* mutants, respectively; (C,D) PCR genotyping results showing the presence of the KO vector in the target locus and absence of WT contamination. This genotyping was performed in cloned mutants