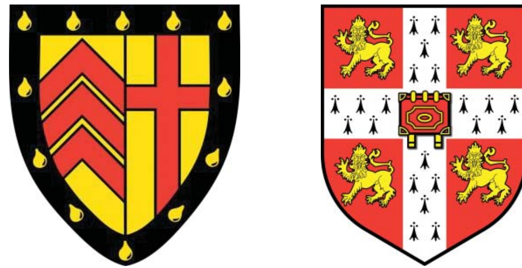
High-throughput reverse genetic screening in *Plasmodium berghei* using barcode sequencing

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

Acknowledgements ... iii
Abstract ... iv
Table of contents ... v
List of Figures ... ix
List of tables ... x
List of Appendices ... x
Abbreviations ... xi
Chapter 1 – Introduction ... 1
1.1 Malaria: A major global parasitic disease ... 2
1.2 The life cycle of *Plasmodium* parasites ... 4
1.3 Next-generation sequencing technologies ... 6
1.3.1 Illumina sequencing ... 7
1.3.1.1 Illumina sequencing overview ... 7
1.3.1.2 Overview of library preparation procedures ... 8
1.3.1.3 Illumina sequencing chemistry ... 10
1.4 Genetics of malaria parasites ... 12
1.5 Recombination in malaria parasites ... 12
1.6 The rodent model of malaria ... 13
1.7 Genetic engineering – new tools for reverse genetics ... 14
1.7.1 The Gateway technology: DNA cloning using site-specific recombination ... 16
1.7.2 Recombineering, a homologous recombination based cloning strategy ... 17
1.7.3 Recombineering in *P. berghei* – the *Plasmo*GEM project ... 19
1.8 High throughput reverse genetic screens ... 22
1.8.1 Signature tagged mutagenesis (STM) ... 22
1.8.2 Epistasis and genetic interactions ... 26
1.9 Protein kinases ... 28
1.9.1 Eukaryotic protein kinases ... 28
1.9.2 Protein kinases in *Plasmodium* parasites ... 29
1.9.3 MAP kinases in *Plasmodium* ... 31
1.9.4 Calcium responsive kinases in *Plasmodium* ... 32
1.9.5 Defining the *Plasmodium* phospho-proteome ... 34

1.10 Project aims 36
Chapter 2 - Materials and Methods 38
2.1 Parasitology 39
2.1.1 Rodents 39
2.1.2 Parasite lines 39
2.1.3 Parasite maintenance 40
2.1.4 Parasite cloning by limiting dilution 40
2.1.5 Selection marker recycling 40
2.1.6 Parasite phenotyping 41
2.1.6.1 Asexual growth curves of *P. berghei* parasites. 41
2.1.6.2 Exflagellation assay 41
2.1.6.3 Ookinete phenotyping 41
2.1.6.3.1 Ookinete culture 41
2.1.6.3.2 Ookinete conversion rate calculation 42
2.1.6.3.3 Ookinete purification with magnetic beads 42
2.1.6.4 Oocyst dissection and counts 42
2.1.6.5 Sporozoite dissection and counts 42
2.2 Generation of targeting vectors 43
2.2.1 Single tube protocol 43
2.2.2 Vectors provided by the *Plasmo*GEM resource 44
2.3 Generation of mutant *P. berghei* parasites 44
2.3.1 Generation of single mutants 44
2.3.2 STM protocols 45
2.3.2.1 Parallel transfection 45
2.3.2.2 Collection and processing of STM time-points. 46
2.4 DNA preparation and genotyping methods 46
2.4.1 White blood cell (WBC) removal 46
2.4.2 Blood lysis 46
2.4.3 DNA extraction methods 47
2.4.3.1 Phenol-chloroform extraction of blood samples 47
2.4.3.2 DNeasy Blood & Tissue Kit 47
2.4.4 PCR genotyping 48
2.4.5 Pulsed-Field Gel Electrophoresis (PFGE) of *P. berghei* chromosomes 48

2.4.6 DNA preparation for Southern blot analysis 49
2.4.7 Southern blotting hybridisation 49
2.5 Illumina sequencing 50
2.5.1 Library preparation 50
2.5.1.1 Adaptor ligation 50
2.5.1.1.1 Whole genome sequencing..................... 50
2.5.1.1.2 STM time-points 50
2.5.1.2 Direct amplification 51
2.5.2 MiSeq run conditions 51
2.5.2.1 Whole genome sequencing 51
2.5.2.2 Barcode sequencing 52
2.6 Data analysis 52
2.6.1 Fitness calculation by barcode counting..................... 52
2.6.2 Genetic interaction coefficients 53
2.7 Western blotting..................... 53
Chapter 3 - Establishment of Signature tagged mutagenesis in *P. berghei* - Setting the scene55
3.1 Introduction..................... 56
3.2 Results..................... 57
3.2.1 Optimisation of operating conditions for transfection..................... 57
3.2.1.1 Choice of electroporator 57
3.2.1.2 Optimal DNA concentration 59
3.2.2 Optimisation of barcode detection using Illumina sequencing 61
3.2.3 Optimisation of Illumina library preparation..................... 64
3.2.4 Optimisation of Illumina MiSeq run conditions..................... 68
3.3 Discussion 70
Chapter 4 - STM analysis of protein kinase genes in *P. berghei* 72
4.1 Introduction..................... 73
4.2 Results..................... 75
4.2.1 Barcode counting in *P. berghei* allows parallel phenotyping of mutants in a single mouse 75
4.2.2 Comparison between barcode counting and a conventional deletion analysis.. 82
4.2.3 Validation of false positives 83
4.2.3.1 *gsk3* KO phenotyping 89
4.3 Discussion..................... 91

Chapter 5 - A genetic interaction screen reveals a new signalling pathway 96
5.1 Introduction 97
5.2 Results 97
5.2.1 Choice of the genetic backgrounds 97
5.2.2 Generation of selection marker free backgrounds 99
5.2.2.1 Phenotypic analysis of the *map* double KO revealed a *map2* KO-like phenotype 100
5.2.3 Revealing epistasis in the *P. berghei* kinome 104
5.2.4 Validation of *cdpk4-pkg* interaction 108
5.3 Discussion 111
Chapter 6 115
References 121
Appendices 139

List of Figures

Fig. 1.1| Malaria endangers half of the world's population. 3
Fig. 1.2| Malaria life cycle. 6
Fig. 1.3| Overview of a library preparation protocol. 9
Fig. 1.4| Illumina sequencing chemistry. 10
Fig. 1.5| Mechanism of Red recombination. 18
Fig. 1.6| First step – Recombineering reaction to generate a KO vector. 20
Fig. 1.7| Second step - Replacement of *zeo-pheS* cassette with *P. berghei* selection marker. 21
Fig. 1.8| Summary of mutagenesis methods used for the generation of pools of mutants. 25
Fig. 1.9| Summary of genetic interactions. 27
Fig. 1.10| Structure of ePKs catalytic domain. 28
Fig. 3.1| The choice of electroporation system and DNA concentration are critical for maximum transfection efficiency. 58
Fig. 3.2| Absence of passenger vectors lacking a selection cassette. 61
Fig. 3.3| The barcodes within the *Plasmo*GEM vectors are compatible with a bar-seq strategy. 63
Fig. 3.4| Increasing the number of PCR cycles does not have a high impact on data quality. 64
Fig. 3.5| Adaptor ligation method overview. 66
Fig. 3.6| Comparison between AL and DA library preparation methods. 67
Fig. 3.7| Miseq run quality analysis. 69
Fig. 3.8| Sequencing reproducibility within and between runs. 70
Fig. 4.1| Proposed experimental design for the *P. berghei* STM-Bar-seq experiments. 75
Fig. 4.2| Distribution of barcode counts for each gene and comparison with the input sample. 76
Fig. 4.3| Parallel transfection of pooled KO vectors generated pools of mutants 78
Fig. 4.4| STM revealed a range of growth phenotypes. 82
Fig. 4.5| Genotyping of the newly obtained mutants *cdpk1* KO, *gsk3* KO, PBANKA_08296 KO, *tkl3* KO, *rio1* KO and *rio2* KO. 85
Fig. 4.6| WGS of RIO kinases - *rio1* locus is disrupted in the *rio1* mutants. 86
Fig. 4.7| RIO kinases KO vector designs. 87
Fig. 4.8| WGS of RIO kinases - *rio2* locus. 88
Fig. 4.9| *gsk3* KO genotyping and phenotypic analysis. 90
Fig. 5.1| The mutation of the gatekeeper amino-acid renders PKG resistant to Compound 1. 98
Fig. 5.2| Genotyping of CDPK KO genetic backgrounds (next two pages) 100
Fig. 5.3| Genotyping and phenotyping of the double KO mutant *map1*$^-$/*map2*$^-$ (dKO). 103
Fig. 5.4| Analysis of the genetic background barcode counts. 105
Fig. 5.5| Interaction coefficients for 258 double and triple mutants (next page). 107
Fig. 5.6| The genetic interaction between *cdpk4* and *pkg* was also detected in independently generated mutants. 109
Fig. 5.7| A role for CDPK4 in merozoites egress – a possible model. 113

List of tables

Table 1.1| Overview of current Illumina Instruments and their applications 8
Table 2.1| Choice of DNA extraction method depending on starting material. 47
Table 2.2| Summary of restriction enzymes used to digest parasite DNA for Southern blot analysis. 49

List of Appendices

Appendix I – List of *Plasmo*GEM IDs for each gene and corresponding annotation. 140
Appendix II – Primers used to genotype cloned mutants. 141
Appendix III – Primers used for genotyping of the STM screen 143
Appendix IV – Primers used for barcode sequencing. 144
Appendix V – Targetability and fitness measurements for ePKs. 145
Appendix VI – Genotyping strategy for *rio1* and *rio2* KO mutants 147
Appendix VII – Genotyping strategy for *gsk3* KO clones 148
Appendix VIII – Genotyping strategy for the KO mutants of the PBANKA_082960 and *tkl3* genes. 149

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