Investigations into the within-host genomic diversity and phenotypic variation of *Plasmodium falciparum*

Lee Murray

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of London

2016

Department of Pathogen Molecular Biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the BBSRC

Research supervisor: Professor David Conway

Abstract

Clinical isolates of *Plasmodium falciparum* frequently consist of multiple genotypes, particularly within areas of high infection endemicity. Understanding these infections is important, as they have been linked to traits including drug resistance and virulence. I compared interpretations on within-host diversity from the allele-frequency based F_{WS} index on whole-genome sequence data within a set of clinical isolates with microsatellite genotyping. Whilst there is significant correlation between the two methodologies, over half of isolates consist predominantly of a single genotype by $F_{WS,}$ yet possess multiple genotypes through the microsatellite typing, thus providing correlative and complementary information. Furthermore, genomic diversity within an infection can be produced by meiotic recombination between heterozygous haploid gametes, classical superinfection, or arise through *de novo* mutation. I applied F_{WS} as a genome-wide scan of diversity within four West African populations. Through two approaches, utilising gene and bootstrapped SNP scores of F_{WS}, I attempted to identify loci with consistently high levels of within-host homozygosity relative to their within-population heterozygosity, that were posited to be under within-host directional selection. These approaches did not consistently identify outlier loci across the four populations, but I discuss alternative approaches to identify targets of within-host selection.

Experimental work performed within rodent malaria species has reported that growth and gametocytogenesis are two plastic phenotypes affected by the presence of a competitor genotype. Utilising qPCR assays, comparison of the expression levels of the gametocyte commitment markers AP2-G and GDV-1, in addition to the sex-specific markers Pfs25 and Pfs230p, between two West African populations of differing transmission intensity, were all not significantly different. In addition, I designed a multi-cycle low starting density growth assay, initially using four laboratory clones, finding mean parasite multiplication rate (PMR) of these clones to range from 7.59 to 10.53 over the assay. When applied to fourteen recently adapted *ex vivo* clinical isolates, mean PMR was much lower at 2.90, with a range of 1.97 to 4.10. In combination with allele-specific qPCR, the growth assay was applied to assess competitive growth of the four laboratory clones, with no evidence of a significant induced plastic growth response of any clone. Together, these approaches will enable competitive growth studies of clinical isolates. The baseline quantitative data provided in this thesis supports the investigation of plastic phenotypes of *P. falciparum*.

Acknowledgements

I need to acknowledge several people for the support that they have put into this thesis over the past three and a bit years. Firstly, I should thank my supervisor Dave for all of his insight and encouragement over this period- his passion for malariology has certainly rubbed off on me and I am grateful for him leading such an interdisciplinary research group. In addition, I would like to thank his collaborators and their staff for collection and processing of clinical isolates for the sequence analyses and phenotyping performed in this thesis including; Dr Alfred Amambua-Ngwa, Dr Gordon Awandare, Dr Kovana Loua, Dr Victor Mobegi, Dr Ambroise Ahouidi and Dr Mahmadou Diakite, in addition to Prof Dominic Kwiatkowski and colleagues at the Wellcome Trust Sanger Institute for producing the genomic sequences used. Also, thanks to Prof Robert Sinden and Dr Sara Marques at Imperial College London for provision of the pEFGFP plasmid.

Moreover, I incredibly appreciate the day-to-day training provided in the respective dark arts of genomics, molecular biology and parasitology given to me by Dr Craig Duffy, Dr Sarah Tarr and Ms Lindsay Stewart. They have gone beyond their job remit to train and encourage me, which I will never be able to repay fully. I truly hope that they all go on to achieve everything that they desire in science, as they all thoroughly deserve it. I should definitely thank Craig (and Sammy too) for introducing me to the artisan Bloomsbury coffee scene and to Sarah for enthusiastically supporting my appreciation of it when she arrived. Lindsay in particular has had to endure me at times when I've been most frustrated, so I'm very grateful to her for the ad hoc counselling sessions, and doing her best to get me to see the positives in those moments, even if it didn't always appear to help immediately.

Thank you also to Dr Samuel Assefa, Harvey Aspeling-Jones, Paul Divis and Dr Ofelia Diaz for occasional help and interesting conversations through my thesis. Also, Dr Sam Alsford has provided good advice to me as part of my studies in his role as Department Research Degrees Coordinator and as part of my upgrade panel. Moreover, thanks to Ms Eloise Thompson for running the malaria culture lab, to Prof Jurg Bahler for discussions in the early stages and to Dr Laura Drought and Dr Avnish Patel for dragging me away from my computer occasionally and to the BBSRC for funding my thesis.

Finally, I should thank my family and Alice for their support over the course of my thesis- in lieu of having not produced a Nature paper (spoilers!), AI is undoubtedly the best finding of my PhD.

Declaration of work

I, Lee Murray, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Within Chapter 3, read alignment and SNP calling was performed by Dr Craig Duffy, whilst the microsatellite genotyping used for comparison was performed by Dr Victor Mobegi. I performed all downstream SNP analyses.

Within Chapters 3, 4, 5 and 6 all clinical isolates were collected and cryopreserved by collaborators located within West Africa as part of the GENINVADE project and and genome sequences used have previously been published for separate analyses; Dr Alfred Amambua-Ngwa, Dr Marcel Loua, Dr Gordon Awandare, Dr Ambroise Ahouidi and Dr Mahmadou Diakite led local collections. Whole-genome sequencing was performed by the Wellcome Trust Sanger Institute and was published for other studies.

Within Chapter 6, *ex vivo* culture adaptation was performed for the most part by Ms. Lindsay Stewart, with some assistance from myself.

Within Chapter 7, the mKate2 gBlocks gene fragment was designed by Dr Sarah Tarr.

.

Contents

Abstract.....2

Acknowledgements.....3

Declaration of work.....4

Contents.....5

List of figures and tables.....11

Relevant publications.....16

1 Introduction.....17

1.1 Clinical burden of *Plasmodium falciparum*.....17

1.2 The life cycle of *P. falciparum*.....19

1.2.1 Pre-erythrocytic stages.....19

1.2.2 Erythrocytic asexual stages.....19

1.2.3 Sexual stages.....22

1.3 *P. falciparum* genomics.....23

1.3.1 The genome of *P. falciparum*.....23

1.3.2 Population genomic scans of selection in *P. falciparum*.....26

1.4. Sexual recombination within *P. falciparum*.....28

1.5 Investigating the within-host diversity of *P. falciparum* infections.....30

1.5.1 PCR-based approaches.....30

1.5.2 Sequencing based approaches.....32

1.6 Importance of understanding multiple genotype infections of *Plasmodium*.....34

1.6.1 Drug resistance and virulence evolution.....35

1.6.2 Gametocyte investment plasticity.....37

1.7 Phenotypic and transcriptional variation of *P. falciparum*.....40

1.8 Linking phenotypic variation to genetic determinants.....42

1.9 Aims of this thesis summary.....45

2 Materials and Methods.....46

2.1 Bioinformatic whole genome sequence analysis of within-host diversity.....46

2.1.1 Sample collection and whole-genome sequencing.....46

2.1.2 SNP and coverage filtering procedures.....47

2.1.3 Genome-wide calculation of the within isolate fixation index F_{WS}.....48

2.1.4 Population comparison of isolate F_{WS} scores.....49

2.2 Clinical isolate collection for *ex vivo* studies.....49

2.3 *P. falciparum* cell culture techniques.....50

2.3.1 Parasites studied.....50

2.3.2 Thawing and cryopreservation of *P. falciparum* parasite.....50

2.3.3 Erythrocyte preparation.....51

2.3.4 General parasite culture conditions.....51

2.3.5 Synchronisation of parasite stages.....52

2.4 Genomic DNA extraction and preparation of genomic standards.....52

2.5 Parasite genotyping.....53

2.5.1 Sequencing and genotyping of MSP-1 block 2 for laboratory clones.....53

2.5.2 Microsatellite genotyping of clinical isolates.....55

2.5.3 Primer selection and design for allele-specific qPCR assays.....55

2.6 Standardised conditions for quantitation of growth rates of *P. falciparum*.....56

2.6.1 Standardised culture conditions.....56

2.6.2 Monoculture laboratory clone growth assay.....56

2.6.3 *Ex vivo* clinical isolate growth assays.....57

2.6.4 Competitive growth assays of *P. falciparum*.....57

2.7 qPCR for quantitation of laboratory clones, *ex vivo* isolates and competing clones... 58

2.7.1 qPCR for analysis of laboratory clone and *ex vivo* isolate PMR.....58

2.7.2 PMR analysis of qPCR data for laboratory clones and *ex vivo* isolates.....59

2.7.3 qPCR for analysis of pairwise competition growth rates of laboratory clones.....60

2.8 Generation of fluorescent *P. falciparum* parasites for *in vitro* competition studies... 61

2.8.1 Fluorescent protein selection.....61

2.8.2 *E.coli* competent cell production.....61

2.8.3 Restriction digests and ligation reactions.....62

2.8.4 DNA sequencing.....63

2.8.5 Integration into the *Pf*47 locus.....63

2.8.6 Integration PCR.....64

2.8.7 Transfection of *P. falciparum* parasites.....66

2.8.8 Limiting dilution cloning of integrated parasite transfections.....67

2.8.9 Live parasite imaging.....67

2.8.10 Comparison of quantitation of *in vitro* competition using a combination of..... allele-specific qPCR and fluorescent imaging.....68

3. Genome-wide assessment of *P. falciparum* within-host diversity within a highly endemic West African population.....69

3.1 Introduction.....69

3.2 Materials and Methods.....72

3.3 Results.....74

3.3.1 Isolate F_{WS} indices within the Republic of Guinea.....74

3.3.2 Identifying SNPs for assay of within-host diversity with a small number of loci...75

3.3.3 Concordance of F_{ws} with other proposed measures of within-host diversity.....78

3.4 Discussion.....83

4. Assessing evidence of non-neutral allele-frequency based polymorphism at the..... within-host level across the *P. falciparum* genome within West Africa.....86

4.1 Introduction.....86

4.2 Materials and Methods.....88

4.3 Results.....91

4.3.1 Population comparison of isolate F_{WS} indices across West Africa.....91

4.3.2 A gene-based scan of F_{WS} across the genome.....97

4.3.3 Bootstrapping of F_{WS} SNP scores across the genome.....102

4.4 Discussion.....105

5 Quantifying variation of gametocyte commitment in clinical samples of *P.*..... *falciparum* from two populations in different transmission environments.....108

5.1 Introduction.....108

5.2 Materials and Methods.....112

5.3 Results.....114

5.3.1 Quantification of mRNA concentrations and quality filtering.....114

5.3.2 Varied isolate expression and population comparison of gametocytogenesis..... transcripts and sex ratios.....118

5.4 Discussion.....123

6 Quantifying variation in parasite multiplication rate during the ex vivo culture..... adaptation process of clinical isolates of P. falciparum.....126

6.1 Introduction.....126

6.2 Materials and Methods.....128

6.3 Results.....130

6.3.1 Optimisation of a low starting density assay for PMR quantitation.....130

6.3.2 Quantifying growth rates of laboratory clones.....132

6.3.3 Quantifying growth rates of ex vivo clinical isolates.....137

6.3.4 Repeated profiling of PMR over the ex vivo culture adaptation process.....142

6.4 Discussion.....148

7 Development of tools and assays to investigate potential competitive interactions between *P. falciparum* genotypes in culture.....152

7.1 Introduction.....152

7.2 Materials and Methods.....154

7.3 Results.....157

7.3.1 Validation of allele-specific qPCR assays for *in vitro* competition.....157

7.3.2 Quantifying *in vitro* competitive growth rates of *P. falciparum*.....157

7.3.3 Generation of differentially fluorescent *P. falciparum* parasites.....160

7.4 Discussion.....168

8 Discussion.....170

8.1 Whole genome sequenced based investigation of within-host diversity.....170

8.2 Identification of molecular targets of within-host selection.....173

8.3 Disparity between the availability of genomic and phenotypic information.....176

8.4 Identifying genomic determinants of phenotypic and transcriptional variation from recently adapted *P. falciparum* clinical isolates.....178

9 References.....181

10 List of appendices.....207

List of figures and tables

Figure 1.1 Distributions of global transmission intensity and infection endemicity within the African continent.....18

Figure 1.2 The life cycle of *Plasmodium falciparum*.....20

Figure 1.3 Schematic diagram of variation detected when looking at loci under A) balancing selection, B) neutrality and C) directional selection within a haploid population.....25

Figure 1.4 Outcomes of sexual recombination within the mosquito vector for *P. falciparum*.....29

Figure 1.5 High levels of relatedness between six genotypically distinct parasites isolated by limiting dilution cloning from a clinical isolate.....33

Figure 1.6. Predictions of male gametocyte proportions within *Plasmodium* infections with differing clonal compositions.....39

Figure 2.1 Schematic of the pEFGFP plasmid.....65

Figure 3.1. Schematic diagram representing the presence of multiple genotypes within natural infections of *P. falciparum*.....70

Figure 3.2 Quality control filtering pipeline of SNPs and isolates for calculation of the F_{ws} metric.....73

Figure 3.3 Genome-wide computation of Guinean isolate F_{WS} indices.....76

Figure 3.4 Frequency of within-host polymorphisms of the Guinean population.....77

Figure 3.5 Correlation coefficients of isolate F_{WS} indices derived from randomly sampled small numbers of SNPs compared with indices using all genome-wide SNPs analysed in the local population.....79

Figure 3.6 Concordance of Guinea isolate F_{WS} scores with alternative measures of within-host diversity.....81

Figure 4.1 Schematic diagram of the within-host heterozygosity of three isolates and the corresponding within-population heterozygosity for calculation of the F_{WS} metric at four example nucleotide positions.....89

Figure 4.2. Population minor allele frequency distributions for all four West African population SNP sets. 93

Figure 4.3 Cumulative density plot of isolate F_{WS} indices from four West African populations.....95

Figure 4.4 Genome-wide within-host heterozygosity of three example Guinean isolates.....96

Figure 4.5 Distribution of F_{WS} gene scores within the four West African populations..99

Figure 4.6 Genome-wide plots of gene F_{WS} scores within four West African populations..... 101

Figure 4.7 Concordance between gene-based calculations of F_{WS} across four West African populations.....103

Figure 4.8 Relationship between SNP population minor allele frequency and estimation of a mean SNP F_{WS} score through bootstrapping.....104

Figure 4.9 Concordance between SNP-based calculations of F_{WS} across four West African populations.....105

Figure 5.1 Normalised expression profiles for a subset of A) 22 Ghanaian isolates and B) 10 Senegalese isolates.....119

Figure 5.2 Population male gametocyte transcript and sex ratio comparison of Ghanaian and Senegalese population subsets.....121

Figure 5.3 Pairwise correlations of normalised expression levels between Pfs25, AP2-G and GDV-1.....122

Figure 6.1 Effect of starting density on quantified PMR estimated over either 6 or 8 days *in vitro* from two biological replicates using laboratory clone Dd2.....133

Figure 6.2 Quantifying parasite multiplication rates for four laboratory clones of *P. falciparum*.....135

Figure 6.3 Concordance between parasite numbers as determined by giemsa-based microscopy and qPCR based quantitation.....138

Figure 6.4 Quantifying parasite multiplication rates for a selection of four clinical isolates.....140

Figure 6.5 Parasite multiplication rates (PMR) per 48 hours for 14 West African clinical isolates quantified over 6 days using a general linear model approach.....141

Figure 6.6 Profiles of 6 *ex vivo* clinical isolate which had their PMR quantified more than once during the culture adaptation process. 146

Figure 6.7. Normalised repeated profiling of clinical isolates during the *ex vivo* culture adaptation process.....147

Figure 7.1 Fluorescent melt curve analysis confirms allele-specificity of qPCR assays used for *in vitro* competition studies..... 158

Figure 7.2 Allele-specific quantitation of controlled mixtures of genomic DNA of four laboratory clones..... 159

Figure 7.3 Quantitation of PMR for four laboratory clones across three intraerythrocytic cycles for all six possible pairwise *in vitro* competitive scenarios...161

Figure 7.4 Confirmation of fluorescent protein expression in *P. falciparum*.....163

Figure 7.5 PCR-based integration genotyping for the fluorescent parasites at the Pf47 locus.....164

Figure 7.6 Examples of fluorescent overlay live parasite images for correlation with genome copy numbers by allele-specific qPCR.....166

Figure 7.7 Concordance of measurement of controlled mixtures of parasites expressing different fluorescent proteins through live imaging microscopy and allele-specific qPCR. 167

Table 3.1 Estimations of multiplicity of infection as calculated by estMOI software.....82

Table 4.1 Summary statistics of quality filtered SNP sets and isolate F_{WS} indices from four West African populations.....92

Table 5.1 Measurements of gametocyte sex ratio by microscopy within four *P. falciparum* populations.....109

Table 5.2 Primers used for gametocytogenesis transcript profiling of clinical isolates 113

Table 5.3 Mean copy number transcripts for all 5 genes profiled within the 32 Ghanaian isolates that had a housekeeping transcript concentration of >100 copies per µL.....115

Table 5.4 Mean copy number transcripts for all 5 genes profiled within the 36 Senegalese isolates that had a housekeeping transcript concentration of > 100 copies per µL.....116

Table 5.5 Summary of the numbers and proportions of clinical isolates within each population for which each of the transcripts profiled were quantifiable above certain thresholds.....117

Table 6.1 Sensitivity of collecting 200 µL from a 3% haematocrit culture using serially diluted ring-stage 3D7A-GFP parasites.....131

Table 6.2 Assessing *in vitro* DNA degradation of *P. falciparum*.....131

Table 6.3 Parasite multiplication rates (PMR) per 48 hours estimated for the Dd2 clone performed in biological duplicate, with upper and lower 95% confidence intervals for assays measured over 6 and 8 days for four different starting parasite densities.....134

Table 6.4 Parasite multiplication rates (PMR) per 48 hours for the four laboratory clones (with upper and lower 95% confidence intervals) quantified over the duration of a 6 day assay.....136

Table 6.5 Parasite multiplication rates (PMR) per 48 hours for nineteen clinical isolates (with upper and lower 95% confidence intervals) quantified over the duration of a 6 day assay using a general linear model approach.....139

Table 6.6 Microsatellite genotype profiles for four loci for 14 clinical isolates that had their PMR quantified.....143

Table 6.7 Repeated profiling of *ex vivo* PMR for six clinical isolates through the culture adaptation process.....145

Table 7.1. Allele-specific qPCR assay primer information for assessment of *in vitro* competitive growth.....155

Table 7.2 Parasite multiplication rates (PMR) calculated for the four laboratory clones in pairwise competitive growth experiments with upper and lower confidence intervals taken from growth across three consecutive 48 hour periods.....162

Relevant publications

Murray, L.*, Mobegi, V. A.*, Duffy, C. W., Assefa, S. A., Kwiatkowski, D. P., Laman, E., Loua, K. M., Conway, D. J. (2016) Microsatellite genotyping and genome-wide single nucleotide polymorphism-based indices of *Plasmodium falciparum* diversity within clinical infections. *Malar. J.*, 15, 275.

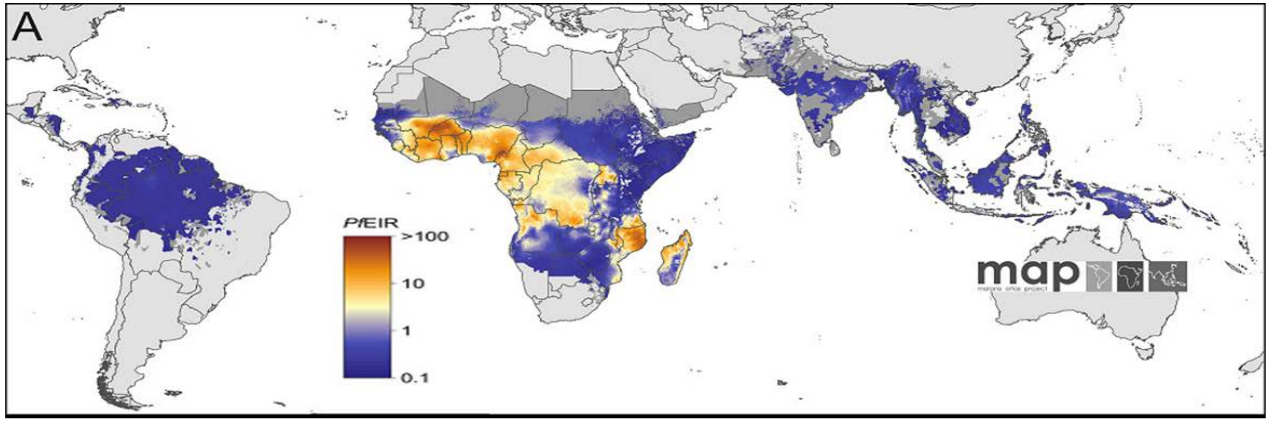
Mobegi, V. A., Duffy, C. W., Amambua-Ngwa, A., Loua, K. M., Laman, E., Nwakanma, D. C., MacInnis, B., Aspeling-Jones, H., Murray, L., Clark, T. G., Kwiatkowski, D. P., Conway, D. J. (2014) Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol. Biol. Evol.*, 31, 1490-1499.

* indicates joint-first authorship

Chapter 1. Introduction

1.1 Clinical burden of *Plasmodium falciparum*

P. falciparum is a protozoan parasite within the phylum Apicomplexa and is the major causative agent of human malaria cases globally. Within the rest of the genus *Plasmodium*, five other species are also known to cause human infections; *P. vivax*, *P.malariae*, *P. ovale wallikeri, P. ovale curtisi* and *P. knowlesi*. Greater than 2 billion people are estimated to be at risk of infection from *P. falciparum* (Gething *et al.* 2011), with endemic regions stretching across Africa, the Middle East, Central and South-East Asia and Latin America (Figure 1.1). Recent estimates by the World Health Organization have put the annual total number of cases of the disease caused by *P. falciparum* at approximately 200 million globally, with *P. vivax* being the second species of *Plasmodium* most greatly responsible for cases with just under 14 million annually (World Health Organization 2015). Despite public health interventions within sub-Saharan Africa having decreased the incidence of *P. falciparum* by 40% within the last 15 years (Bhatt *et al.* 2015), this region continues to be the major focus of cases, with greater than 85% of malaria cases occurring here, of which 70% of fatalities are seen within children under the age of five (World Health Organization 2015). In addition, the presence of widespread malaria cases within a country has been estimated to limit economic growth by 1.3% annually when other socioeconomic factors including human capital are taken into account (Gallup & Sachs 2001). Across sub-Saharan West Africa, there is a range of endemicity due to the presence of a north-south transmission gradient that is driven by the season and volume of rainfall across the region (Hay *et al.* 2009), but is also associated with the intensity of local control programmes.



B)

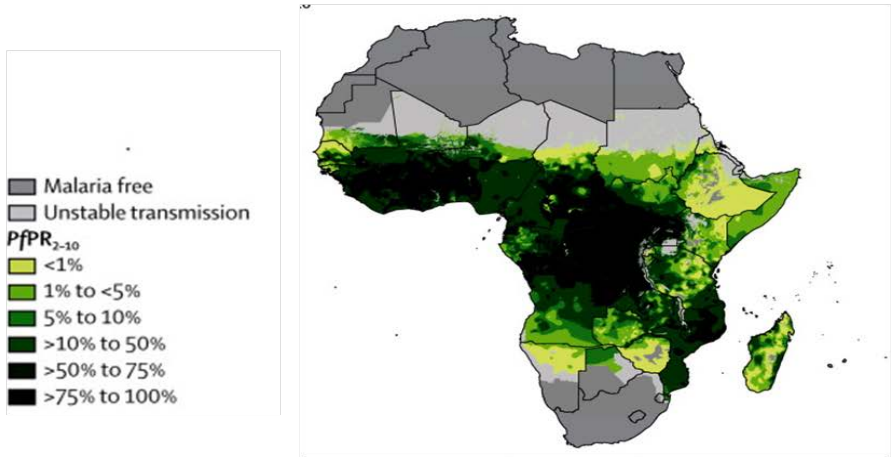


Figure 1.1 Distributions of global transmission intensity and infection endemicity within the African continent. Panel A) highlights differences in the entomological inoculation rate (*Pf*EIR*)* across different global regions, while panel B) is a 1km^2 spatial resolution map where *P. falciparum* parasite rate (PfPR) was calculated from the prevalence of infection within children between the ages of 2 and 10 in 2010 (adapted from Gething *et al.* 2011 and Noor *et al.* 2014 respectively).

1.2The life cycle of *P. falciparum*

1.2.1 Pre-erythrocytic stages

P. falciparum possesses an indirect life cycle consisting of a human host and female *Anopheles* species mosquito vector (Figure 1.2). Human infection is initiated when the mosquito vector takes a blood meal from the host, thus biting and injecting sporozoites into the human dermis (Ponnudurai *et al.* 1991), prior to reaching the bloodstream. Although approximately 10000 sporozoites have been estimated to be present within the salivary glands of the infected mosquito, a median of <50 sporozoites have been found to be injected into the human host (Rosenberg *et al.* 1990). The sporozoites make their way through the vascular system towards the liver where they invade hepatocytes, form a parasitophorous vacuole and develop into hepatocytic schizonts, with thousands of daughter merozoite cells produced for each of the initial sporozoites that successfully invaded (Prudencio *et al.* 2006). This represents the end of pre- erythrocytic development, with the daughter merozoite cells being fully developed after approximately 7 days (Vaughan *et al.* 2012; Soulard *et al.* 2015), which rupture and invade into erythrocytes.

1.2.2 Erythrocytic asexual stages

The intraerythrocytic asexual cycle of *P. falciparum* is approximately 48 hours in duration, although this can vary considerably between 43 and 51 hours for different cultured lines *in vitro* (Reilly Ayala *et al.* 2010). A recently invaded merozoite will develop into a ring stage, which feeds on the host erythrocyte through a cytostome, which is an infolding of the parasitophorous vacuolar membrane and the parasite plasma membrane. Erythrocyte cytoplasm is taken up through the cytostome and haemoglobin is trafficked to the parasite digestive

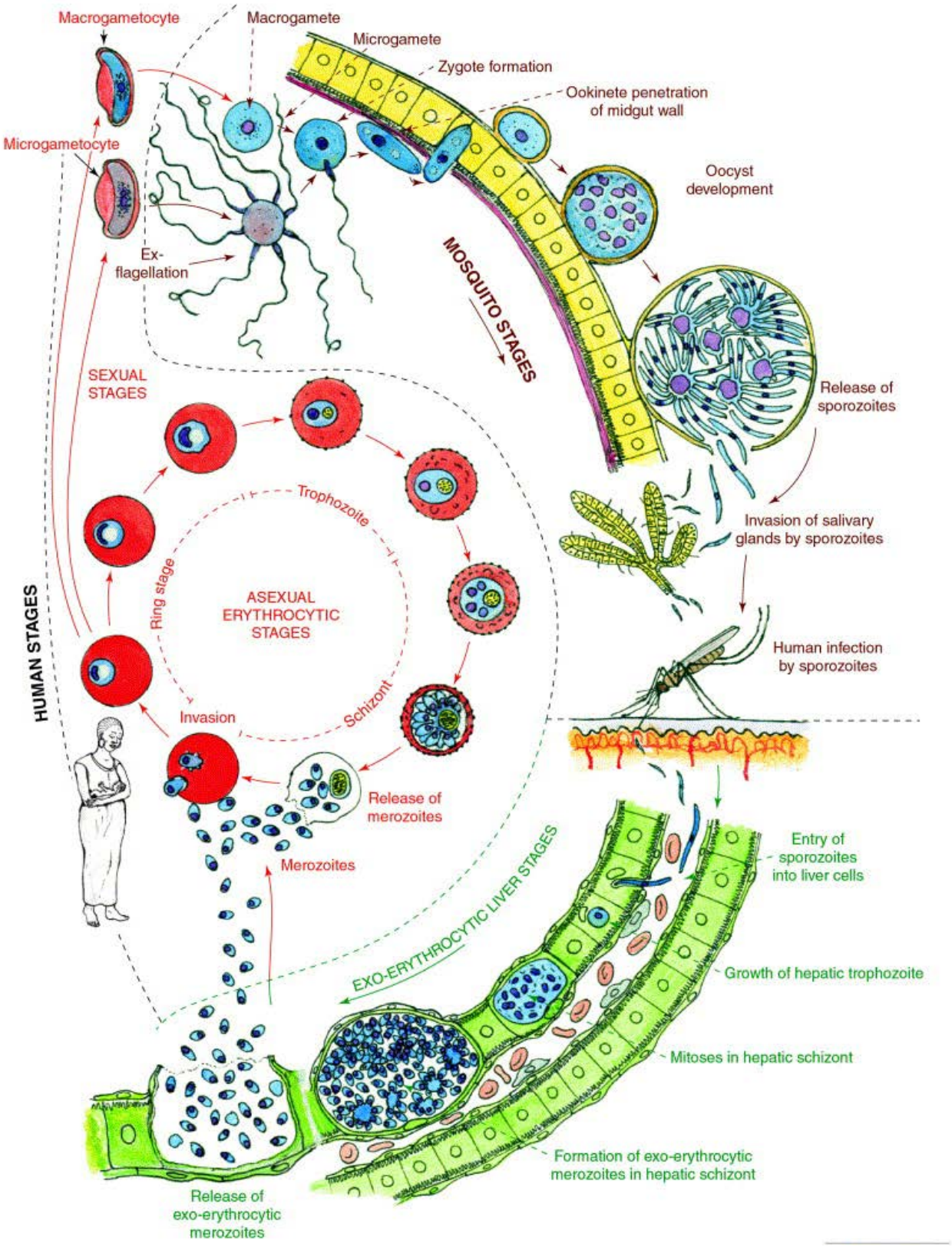


Figure 1.2 The life cycle of *Plasmodium falciparum*. Included within the figure are the hepatocytic, intraerythrocytic and mosquito stages (Bannister & Mitchell 2003).

vacuole inside double-membrane vesicles (Milani *et al.* 2015). The ring stage is commonly seen by microscopy within clinical infections and subsequently develops into the trophozoite stage, during which period the majority of the surrounding host cell cytoplasm is ingested (Francis *et al.* 1997). Following this, at approximately the 32 hour timepoint for a *P. falciparum* parasite with a 48 hour intraerythrocytic cycle, the trophozoite begins to undergo nuclear division, developing into erythrocytic schizonts consisting of between approximately 16 and 32 daughter merozoite cells. Trophozoite and schizont stages are not seen microscopically within the peripheral blood system in clinical infections due to parasite sequestration within organs (Miller 1969). Synchronous egress of large numbers of merozoites from schizonts is responsible for the periodic fevers consistent with the length of the replication cycle (Kwiatkowski & Nowak 1991), due to the release of pyrogenic parasite toxins that are recognised by innate host immune cells, which induce a thermoregulatory response through the production of inflammatory cytokines (Oakley *et al.* 2011).

The asexual erythrocyte stages are responsible for the clinical manifestations of the disease (Figure 1.2), and may reach a level of 1×10^{12} parasites or more within an individual who has a case of severe malaria (Dondorp *et al.* 2005). A considerable proportion of adults within highly endemic areas can harbour asymptomatic infections (Okell *et al.* 2009), whilst approximately 1% of symptomatic infections result in severe malaria (Miller *et al.* 2013). The clinical syndromes seen within severe cases of malaria can include anaemia, respiratory difficulties, metabolic acidosis and cerebral malaria (Mackintosh *et al.* 2004). The trophozoite and schizont stages of the intraerythrocytic cycle play a significant role within the clinical pathogenesis of the disease, as it is during this period that the protein *P. falciparum* erythrocyte membrane protein-1

(PfEMP1) is exported to the surface of the erythrocyte and subsequently induces sequestration of these stages to endothelial cells, preventing immune clearance and potentially causing organ damage (Miller *et al.* 2013).

1.2.3 Sexual stages

Within the intraerythrocytic stages, a small proportion of parasites become committed to development into the sexual life cycle stages, known as gametocytes, which facilitate transmission from the human host into the mosquito vector. The process of gametocyte development takes between 8 to 12 days (Sinden 2009) and has been reported through histological examination to be principally undertaken within the bone marrow of the host (Joice *et al.* 2014). All merozoites present within an erythrocytic schizont are either committed to undergo gametocytogenesis (Bruce *et al.* 1990) or to remain asexually replicating. Those that are committed to sexual development, following invasion of a new erythrocyte, slowly develop through the five stages of gametocytes that have been morphologically characterised, resulting in distinct female and male stage V gametocytes (Talman *et al.* 2004), which are released from the bone marrow into the peripheral bloodstream.

For *P. falciparum* to complete its life cycle, the male and female gametocytes are ingested in a mosquito blood meal, differentiating into their sex-specific gamete forms, before the exflagellated male microgamete penetrates the female macrogamete to produce a diploid zygotic stage. It is during this stage that meiotic recombination can occur (Sinden & Hartley 1985) and if the macrogamete and microgamete are of different genotypes, then this will result in the generation of novel allelic combinations across the genome (Walliker *et*

al. 1987) (section 1.4). The diploid zygote develops within 24 hours into a motile ookinete form, which navigates itself into a position on the midgut wall and differentiates into an oocyst. Subsequently, the oocyst undergoes around ten rounds of nuclear division (Gerald *et al.* 2011), forming thousands of haploid sporozoites over approximately two weeks (Angrisano *et al.* 2012; Baton & Ranford-Cartwright 2005).

1.3 *P. falciparum* genomics

1.3.1 The genome of *P. falciparum*

The haploid nuclear genome of *P. falciparum* is just over 23 million nucleotides in length and is organised into 14 linear chromosomes. Chromosome sizes range from ~0.65Mb up to ~3.29Mb (Gardner *et al.* 2002) and approximately 5400 genes have been described. In addition to the nuclear genome, *P. falciparum* also possesses two organellar genomes, a ~35kb apicoplast genome (Wilson *et al.* 1996) and a ~6kb mitochondrial genome (Feagin 1994). For the majority of its life cycle, the genome of *P. falciparum* is haploid, becoming diploid briefly only within the mosquito stages following the fusion of male and female gametes uptaken during a blood meal (Figure 1.2). The rate of meiotic crossover within *P. falciparum* has been shown to range from approximately 10 to 15kb per centimorgan (Su *et* al. 1999; Jiang *et al.* 2011). Genomic variation can also be generated in the human intraerythrocytic stages of the disease through mitotic recombination and is particularly concentrated in immune evasion gene families located within sub-telomeric chromosomal regions (Bopp *et al.* 2013). Pairwise single nucleotide polymorphism (SNP)

diversity has been estimated at 1.16 x 10^{-3} from sequencing of laboratory clones and clinical isolates (Volkman *et al.* 2007).

A highly AT-rich nuclear genome, at greater than 80% of A and T, gives difficulty in sequencing some parts, particularly intergenic regions. Whilst the original reference genome sequence for *P. falciparum* clone 3D7 was produced through a whole chromosome shotgun sequencing strategy (Gardner *et al.* 2002), the development of next-generation sequencing technologies provided a comparatively cheap method to sequence entire genomes in a single run, of which Illumina quickly became a market leader using fluorescent reversible nucleotides (Metzker 2010; Bentley *et al.* 2008). Due to the AT-bias, standard library preparation methods for Illumina sequencing, which utilise polymerase chain reaction (PCR) amplification, were shown to introduce artefacts through preferential amplification of genome regions with balanced nucleotide content (Kozarewa *et al.* 2009). However, with sequencing of clinical *P. falciparum* isolates often required to be performed with low quantities of initial parasite genomic material, novel methods have begun to reduce the challenges of sequencing these biased genomes (Oyola *et al.* 2012). One advantage of applying a next-generation sequencing technology such as Illumina to clinical infections of *P. falciparum* is to be able to go 'deep' beyond the consensus sequence to assay genomic variation of *P. falciparum* within an individual (Manske *et al.* 2012). In addition, this technological advance has begun to allow for large scale population level sequencing analyses aimed at identifying genome regions under selective processes (section 1.3.2 and Figure 1.3) due to pressures such as drug resistance and host immune evasion (Park *et al.* 2012; Amambua-Ngwa *et al.* 2012), with global consortia efforts having sequenced

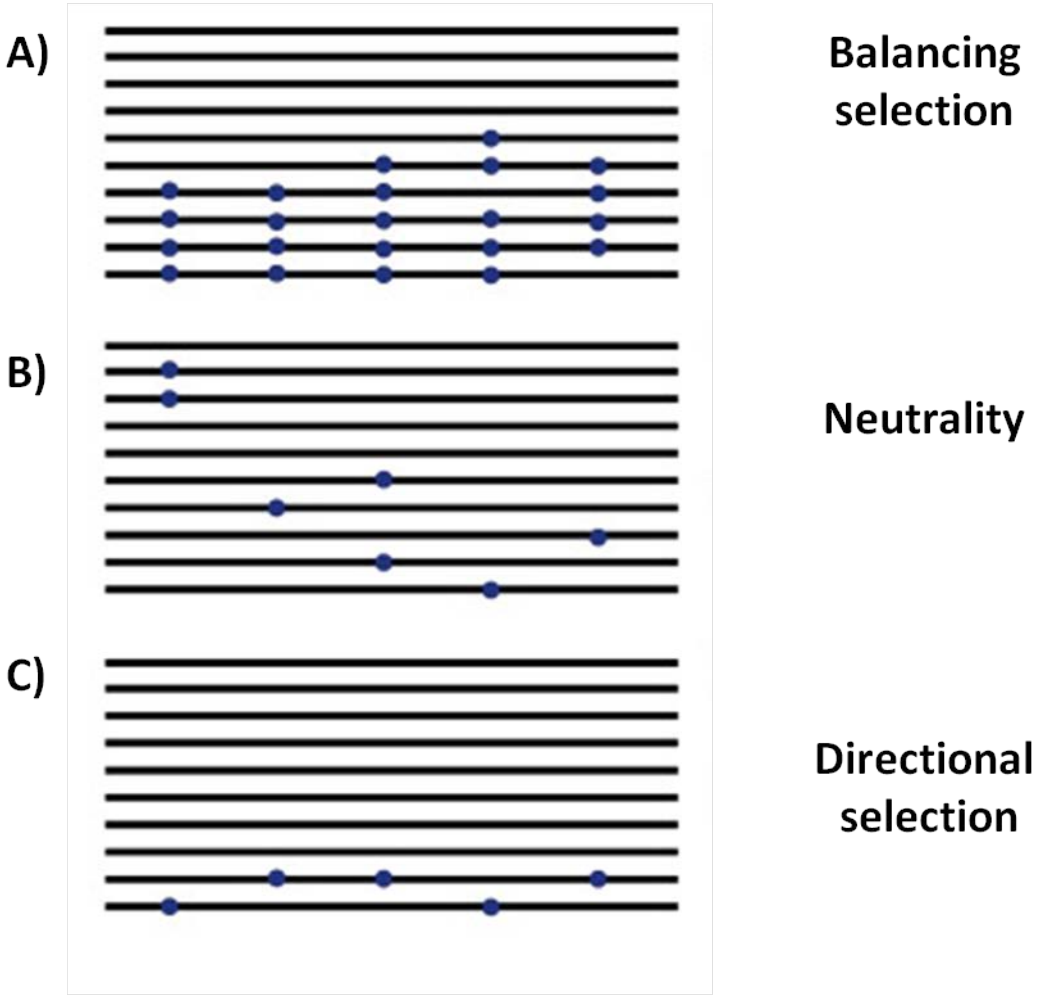


Figure 1.3 Schematic diagram of the type of allele frequency distributions at loci under A) balancing selection, B) neutrality and C) directional selection within a haploid population. Each black line represents an individual haploid *P. falciparum* genome within the intraerythrocytic stages of infection, with ten individual genomes in each case, whilst blue dots represent the location of alleles differing from the majority type (adapted from Weedall & Conway 2010). Whilst under neutrality, the accumulation and maintenance of polymorphism is effectively random, being only influenced by genetic drift or mutation and is not under selection. Balancing selection will keep population alleles at relatively intermediary frequencies, whilst directional selection will skew the frequency of one allele over another.

over 2500 genomes of *P. falciparum* (MalariaGEN 2016). Gene expression throughout the intraerythrocytic cycle has been profiled using both microarray based and RNA-sequencing approaches (Le Roch *et al.* 2003; Bozdech *et al.* 2003; Otto *et al.* 2010). Moreover, population level transcriptomic studies have elucidated the upregulation of transcripts involved in artemisinin resistance from natural populations (Mok *et al.* 2015) and probed correlations between host and parasite transcriptional responses (Yamagishi *et al.* 2014).

1.3.2 Population genomic scans of selection in *P. falciparum*

Allele frequency based tests have historically been applied to assess evidence of selection operating on the nucleotide sequence of *P. falciparum* on a per locus-of-interest basis. These studies utilised metrics such as Tajima's D (Tajima 1989), which assesses the number of segregating sites and the average number of pairwise nucleotide differences within a given sequence, to investigate evidence for potential balancing selection signatures in proposed vaccine candidate genes (Polley & Conway 2001; Conway *et al.* 2000; Baum *et al.* 2003), calculated measures of inter-population divergence at particular loci through the use of F_{ST} (Drakeley *et al.* 1996; Conway *et al.* 2001) or for evidence of heterozygous deficiency within the diploid oocyst stage of infection from dissected mosquitoes (Anthony *et al.* 2007; Mzilahowa *et al.* 2007; Hill *et al.* 1995). With the increased availability of *P. falciparum* DNA sequence data, application of such tests has progressed from analyses of individual loci, to comparisons of multiple candidate loci (Tetteh *et al.* 2009; Ochola *et al.* 2010).

Subsequently, the advent of next-generation sequencing technologies (section 1.3.1) has paved the way for the use of whole genome approaches to identify evidence of natural selection on a variety of model and natural organisms

across a range of taxa (Haasl & Payseur 2015). High density maps of polymorphic genomic markers have allowed for genome-wide analyses of variation and selection upon the *P. falciparum* within geographically distinct parasites, in addition to comparative genomic studies across the *Plasmodium* genus that have uncovered signatures of adaptive evolution within the *P. falciparum* genome (Otto *et al.* 2014; Sundararaman *et al.* 2016).

Supported by sequencing projects (Manske *et al.* 2012; Miotto *et al.* 2013; MalariaGEN 2016), genome scan based methods have been adopted for the detection of signatures of selection operating on whole *P. falciparum* genomes sequenced from clinical isolates. These analyses have applied haplotype-based methods to identify genome regions that show evidence of directional selection (Cheeseman *et al.* 2012; Park *et al.* 2012; Takala-Harrison *et al.* 2013; Miotto *et al.* 2013; Nwakanma *et al.* 2014; Ocholla *et al.* 2014), which is often associated with drug resistance evolution. In addition, a whole genome scan approach has been utilised to identify candidate loci under balancing selection (Amambua-Ngwa *et al.* 2012, Mobegi *et al.* 2014). Some of the latter signatures are predicted to be under host immune selective forces (Weedall & Conway 2010). Furthermore, F_{ST} based genome scans have also been applied to look for loci that are extremely geographically divergent between *P. falciparum* populations and thus may be under local selective pressures. This approach has previously been used to identify loci that have been implicated within a sub-population specific artemisinin slow clearance phenotype and gametocytogenesis (Cheeseman *et al.* 2012; Miotto *et al.* 2013; Mobegi *et al.* 2014; Duffy *et al.* 2015).

1.4. Sexual recombination within *P. falciparum*

P. falciparum is known to have a high recombination rate, consistent with cross-fertilisation within experimental genetic crosses between laboratory clones (Walliker *et al.* 1987; Ranford-Cartwright *et al.* 1993) and measurement of reduced levels of linkage disequilibrium over increasing physical genetic distance along a chromosome (Conway *et al.* 1999; Manske *et al.* 2012). Within areas of higher transmission intensity there is very little linkage disequilibrium (Anderson *et al.* 2000) and a high level of outbreeding as shown by oocyst stages containing products of meiosis at polymorphic loci (Ranford-Cartwright *et al.* 1991; Paul *et al.* 1995; Mzilahowa *et al.* 2007; Anthony *et al.* 2007).

Recent comparative analyses of the frequency of novel SNP and indel variants found in progeny from three distinct genetic crosses (Walliker *et al.* 1987; Wellems *et al.* 1990; Hayton *et al.* 2008) have shown that whilst the generation of indels appears to occur approximately at twice the level of SNPs across the whole genome, within coding regions, SNPs account for around two-thirds of the novel variants produced (Miles *et al.* 2016). Whilst experimental estimations of the meiotic recombination rate derived from crosses appears to be much higher than that of other Apicomplexans (Katzer *et al.* 2011; Tanriverdi *et al.* 2007; Ranford-Cartwright & Mwangi 2012), the effective rate of recombination that is actually seen within natural populations is significantly lower than this. This is due to the possibility of selfing occurring at meiosis between two genetically identical haploid gametes (Figure 1.4). The probability of selfing happening is lower within natural populations that exist within areas of higher transmission intensity, where a high proportion of infections consist of more than one individual genotype. Thus within many African parasite populations, rates of effective recombination are higher than within populations where

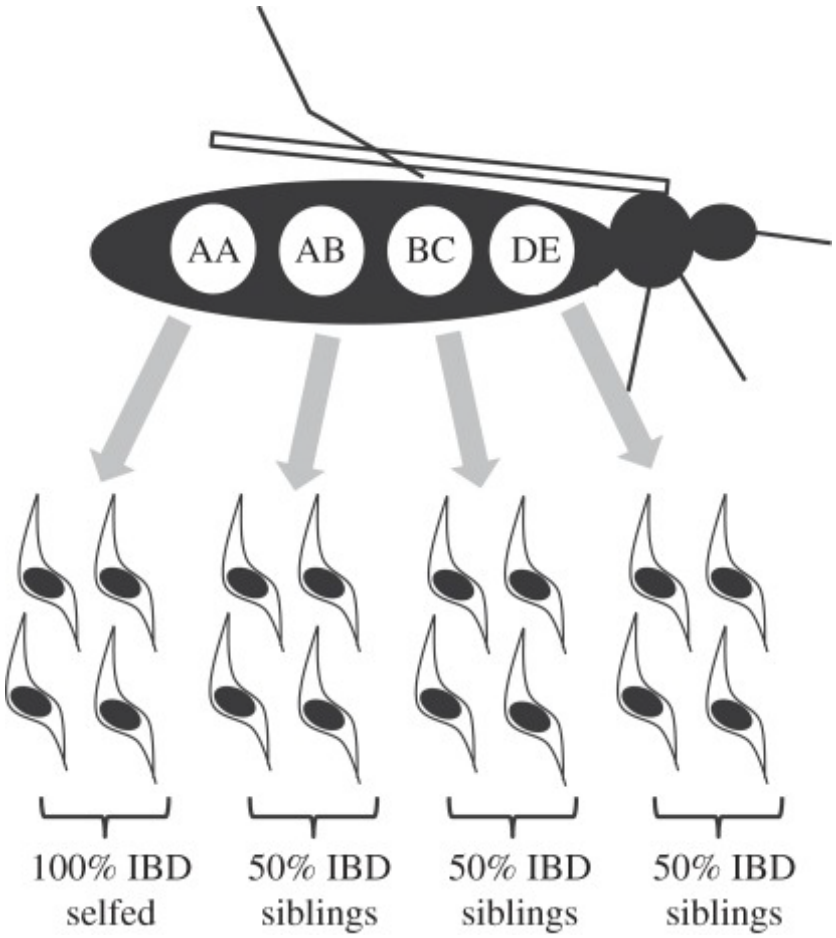


Figure 1.4 Outcomes of sexual recombination within the mosquito vector for *P. falciparum*. Within this example, 5 haploid gametocytes of a different genotypic background (A-E) have been uptaken within a blood meal by the mosquito and have produced random diploid zygotic stages. Following meiotic recombination, the pairwise levels of identity-by-descent (IBD) that exist between the haploid, sporozoite progeny are displayed, with one example of selfing and thus pairwise IBD of 100% having occurred, whilst all other progeny are produced by meiosis of different genotypes (adapted from Nkhoma *et al.* 2012).

infection endemicity is lower (Conway *et al.* 1999; Anderson *et al.* 2000; Anderson 2004).

1.5 Investigating the within-host diversity of *P. falciparum* infections

1.5.1 PCR-based approaches

Whilst pioneering studies utilised isoenzyme electrophoresis (Carter & McGregor 1973) and differential monoclonal antibody staining using indirect immunofluorescence (Conway *et al.* 1991), since the early 1990s and the advent of widespread use of the polymerase chain reaction (PCR), studies into *P. falciparum* within-host diversity have focused on the use of PCR-based genotyping to distinguish between distinct parasite clones. Such assays offer low cost, sensitive methods able to discriminate *P. falciparum* genotypes present at certain loci. Commonly analysed targets including the single-copy genes encoding merozoite surface protein-1 (MSP1), merozoite surface protein-2 (MSP2) and glutamate rich protein (GLURP) (Snounou *et al.* 1993; Farnert *et al.* 2001). Focusing on the measurement of multiplicity of infection (MOI) at such loci has continued to be used widely to provide information on overall infection complexity (e.g. Paul *et al.* 1995; Owusu-Agyei *et al.*2009; Stresman *et al.* 2014). These approaches however are limited by the number of loci that are used to discriminate between genotypes and varying conclusions on the multiplicity of infection have been seen even in assays undertaken by different laboratories (Farnert *et al.* 2001).

To further detail the within-host diversity and population structure of *P. falciparum* infections, selectively neutral polymorphic microsatellite markers have been utilised to provide a measure of the number of alleles detectable at

multiple loci across the genome (Anderson *et al.* 1999). Profiling of the genetic diversity present within multiple populations globally, revealed population structures that ranged from being highly clonal and in strong linkage disequilibrium within areas of low transmission towards more interbreeding and almost panmictic populations within areas with higher rates of transmission (Anderson *et al.* 2000). Subsequently, application of these methods has characterised the genetic structure of *P. falciparum* within many different locations (e.g. Machado *et al.* 2004; Schultz *et al.* 2010; Pumpaibool *et al.* 2009; Iwagami *et al.* 2009), which have advanced local understanding of epidemiology. A recent systematic characterisation of population structure within the West African region using ten locus microsatellite typing within the region has found more genotypically mixed infections within more southern populations, where rainfall is generally higher and infection endemicity is greater (Hay *et al.* 2009; Mobegi *et al.* 2012).

In addition, SNP-based barcode methods have also been designed using SNPs with high population minor allele frequency across global populations at >=24 positions across the genome, with the possibility to be performed on a real-time PCR machine that allows for simpler application of multi-locus genotyping (Daniels *et al.* 2008; Sisya *et al.* 2015). A combination of PCR-based approaches, *ex vivo* manipulation and propagation of clinical isolates offers a powerful system for the dissection of the within-host diversity of *P. falciparum* infections. Barcode based genotyping methods have been used alongside such low-throughput physical separation of genetically distinct *P. falciparum* clones by limiting dilution (Nkhoma *et al.* 2012) to highlight that *P. falciparum* parasites can show a spectrum of relatedness, confirming the

findings of an earlier study which had analysed clones from only two isolates (Druihle *et al.* 1998).

The genetic spectrum of relatedness seen within *P. falciparum* infections is driven by their recent history of sexual recombination, which can occur between distinct or clonal genotypes (section 1.4). Within an infection, pairwise comparisons of nucleotide variation between isolated clones can reveal chromosomal haplotypes which show evidence of having been inherited from a shared ancestor, known as identity-by-descent (Browning & Browning 2012). A pair of parasites whose recent ancestors have undergone sequential rounds of meiosis with another distinct genotype, will possess relatively short and infrequent identical chromosomal regions, whilst two parasites that have experienced recent rounds of recombination with a closely related genotype at meiosis, will have longer shared regions. Genome-wide profiling of clones isolated from infections has revealed that average levels of relatedness, as determined through calculation of pairwise allele-sharing, is higher between parasites found within than between infections and attributable to the inoculation of genetically related sporozoites within the mosquito vector (section 1.2.1). In addition, genotypically distinct haplotypes within an infection can show evidence of extreme levels of relatedness (Figure 1.5), with a low proportion of pairwise allelic differences or contain unrelated parasites that are alternatively the result of classical superinfection (Nkhoma *et al.* 2012; Nair *et al.* 2014).

1.5.2 Sequencing based approaches

The first deep-sequencing study to investigate within-host diversity at a small selection of loci applied massively parallel pyrosequencing to quantify the number of unique allelic variants found within certain commonly used

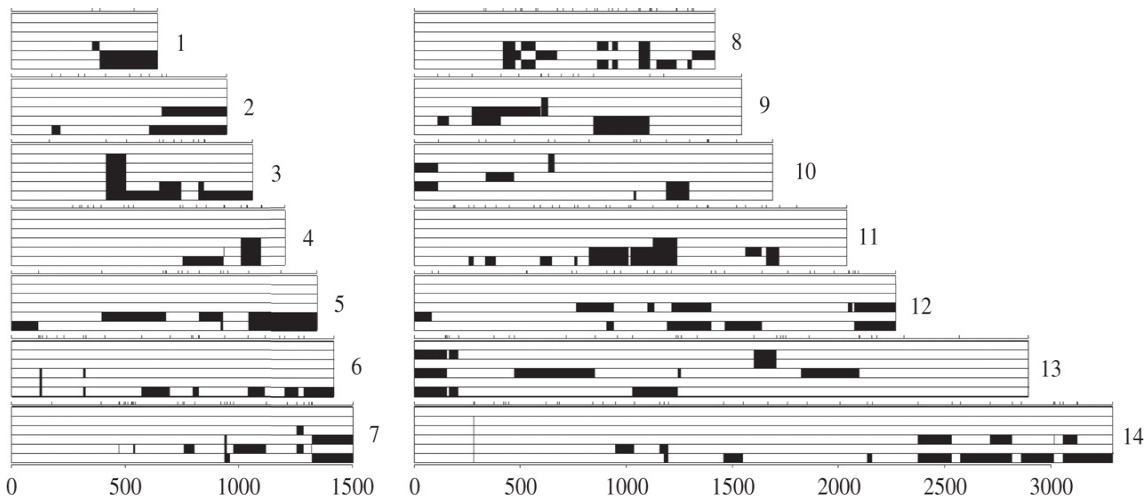


Figure 1.5 High levels of relatedness between six genotypically distinct parasites isolated by limiting dilution cloning from a clinical isolate. The x-axis marks the chromosomal position of each SNP, which are found at the ticks at the top of each of the 14 chromosomes of *P. falciparum*. The six different haplotypes are represented by the six different blocks, which are clear at chromosomal regions that are shared with the haplotype seen within the first block and black at loci which possess the alternate SNP. Within this Malawian isolate, the top two haplotypes are distinct at only 9 SNP positions out of 316, which can be seen on chromosomes 3, 10 and 13 (adapted from Nkhoma *et al.* 2012).

genotyping loci, MSP1-block 2 and MSP2, uncovering consistently greater numbers of variants using the deep sequence approach over PCR-based genotyping for a small number of Cambodian and Malawian isolates (Juliano *et al.* 2010). This approach has been extended further to characterise the haplotype diversity of the circumsporozoite protein, which is the target of the RTS/S vaccine candidate (Bailey *et al.* 2012) and to probe longitudinal within-infection subpopulation specific responses during artemisinin treatment (Mideo *et al.* 2016).

Due to the complex levels of relatedness of different clones within infections uncovered by PCR-based methods, an alternative deep-sequencing based approach was developed to provide quantitative information on this. The F_{WS} metric (Manske *et al.* 2012; Auburn *et al.* 2012) was proposed to provide a measure of heterozygosity using deep sequencing data from the haploid erythrocytic stages of the parasite life cycle within an individual infection relative to the rest of the population. Calculation of this metric across multiple populations globally has shown that a combination of local transmission intensity and genetic isolation have an effect on the metric (Manske *et al.* 2012). Development of single-cell genome sequencing methods have also been applied to characterise the genomic diversity of individual *P. falciparum* infected erythrocytes from *in vivo* infections and identify genomic regions that show evidence of being identical-by-descent between different parasites (Nair *et al.* 2014).

1.6 Importance of understanding multiple genotype infections of *Plasmodium*

1.6.1 Drug resistance and virulence evolution

The presence of co-infecting genotypes has been theoretically linked to important parasite traits including drug resistance and virulence evolution. The historical and genetic origins of *P. falciparum* resistance to artemisinin, chloroquine and sulfadoxine-pyrimethamine have consistently been found within Western Cambodia (White 2010). Distinct subpopulations within this region have been characterised as having high levels of haplotype homozygosity and if a genotype arises through meiotic recombination that possesses both drug resistance alleles and an appropriate level of biological fitness, the inbreeding seen within these subpopulations will aid the emergence of drug resistance within the area (Dye & Williams 1997; Miotto *et al.* 2013). Theoretical modelling across different transmission intensities has predicted a reduction in the capacity of drug resistant parasites to successfully increase in frequency within a population with high levels of multiple genotype infections (Klein *et al.* 2012). The repeated appearance of drug resistance within the same region raises the question of whether drug treatment policy should differ depending on local population structure.

Recent experimental work on *P. chabaudi* has highlighted some of the difficulty in deciding whether aggressive drug treatment would be the most appropriate course of action to limit the evolution of drug resistance (Kouyous *et al.* 2014). Once resistance has appeared within a population, a more concentrated drug dose within a multiple genotype infection will give a higher degree of competitive release to the resistant parasite due to increased suppression of its drug sensitive competitors and lead to its greater propagation within the population (Huijben *et al.* 2013; Pollitt *et al.* 2014). The interaction between local population structure, inbreeding and competition between clones within an

infection clearly plays an important role in drug resistance emergence (Hastings 2006), but requires further investigation within natural populations of *P. falciparum*.

Several studies have also attempted to associate clinical disease severity with the genotypic mixedness of an infection, although conclusions have been predominantly negative and are likely to suffer from the difficulties of single time point sampling of infections as a survey of genotypic diversity (Conway 2007; Farnert *et al.* 1997). In addition, studies of within-host genotype competition using different rodent malaria species have yielded different conclusions on the intrinsic competitive ability of particular genotypes and the subsequent transmission ability. Studies on *P. chabaudi* have reported that competitive ability strongly correlates with clone virulence and transmission success to the mosquito vector, thus the process of within-host competition is posited to select for virulence within *Plasmodium* parasites (de Roode *et al.* 2005; Bell *et al.* 2006). Conversely, studies with *P. yoelli* report the opposite of this, with less virulent clones within single infections outcompeting more virulent clones within mixed infections and subsequently the less virulent clone also going on to have greater transmission ability (Abkallo *et al.* 2015), with a resulting hypothesis that the more virulent clone was subject to immune-mediated clone-specific selection. Thus, it is unclear as to which, if any, of these findings might also hold true for *P. falciparum* and underlines the need to translate for studies on human infectious *Plasmodium* species where possible. If within-host competitive processes did play a role in driving virulence evolution of *P. falciparum*, it may be possible that parasite virulence is higher in areas of higher infection endemicity or that the fitness costs associated with drug resistance may cause parasites with such alleles to be competitively suppressed (Conway 2007).

1.6.2 Gametocyte investment plasticity

A few studies have investigated differences in gametocyte commitment amongst co-infecting genotypes when compared to single clone infections, particularly within rodent malaria species. Restrained investment in gametocytes has been reported in multiple genotype infections of *P. chabaudi*, as parasites seek to maximise their competitive ability within the host (Pollitt *et al.* 2011). Whilst this model system allows for initiation of known starting genotypes within a host system, a limitation of this work is the limited number of clones that are used, which is not truly reflective of the diversity or relatedness spectrum seen within natural populations of *P. falciparum*. In addition, plasticity of clone gametocyte sex ratio has been studied as a phenotype which is predicted to be affected by the clonal composition of a *Plasmodium* infection. Hamilton's local mate competition theory (Hamilton 1967) has underpinned such work by proposing that in the presence of multiple genotypes, a higher proportion of sexual stage investment is to be allocated to male gametocytes. If a single *Plasmodium* clone does infect a mosquito, then gametocyte sex ratio would be predicted to be heavily female biased as one male gametocyte can fertilise more than one female gametocyte, due to up to 8 male gametes being produced per male gametocyte (Read *et al.* 1992). Production of male gametocytes above the minimum required to fertilise all female gametocytes would be a wasteful investment of resources (Leggett *et al.* 2014). Hence, the optimum *Plasmodium* sex ratio within an infection is predicted to be dependent on the male fecundity, in terms of the number of viable gametes produced per male gametocyte and the likelihood of selfing at meiosis (Schall 2009).

In agreement with this, controlled infections of again a limited number of genotypes within *P. chabaudi* and *P. mexicanum* have reported alterations to

the sex ratio of clones in response to the presence of other genotypes within the infection (Reece *et al.* 2008; Neal & Schall 2014; Figure 1.6). Changes in sex ratio within *P. falciparum* can have a significant impact upon parasite transmission success at certain gametocyte densities (Mitri *et al.* 2009), demonstrating their importance. However, analyses on the sex ratio of *P. mexicanum* in natural populations, with robust sample sizes and using microsatellite determination of within-host diversity, has seen no significant variations in sex ratio to correlate to within-host diversity of the infection (Neal & Schall 2014).

Furthermore, sex ratio changes have not solely been linked to within-host diversity, with host anaemia also shown to have a significant effect on the sex ratio of *P. gallinaceum* and *P. vinckei* (Paul *et al.* 2000). In addition, experiments using *P. relictum* have recorded significantly increased levels of transmission to their mosquito vector when subjected to biting (Cornet *et al.* 2014), whilst changes in the genotypic composition of asymptomatic infections of *P. falciparum* longitudinally have also been recorded (Farnert *et al.* 1997), reinforcing the difficulty in truly being able to relate the sex ratio of gametocytes to the within-host diversity present. In order to transmit successfully, *P. falciparum* parasites are required to be able to adapt and respond to a changing environment consisting of heterogeneous host, parasite and environmental factors. Understanding parasite plasticity and potential transcriptional variation within gametocyte stage commitment could therefore have important consequences.

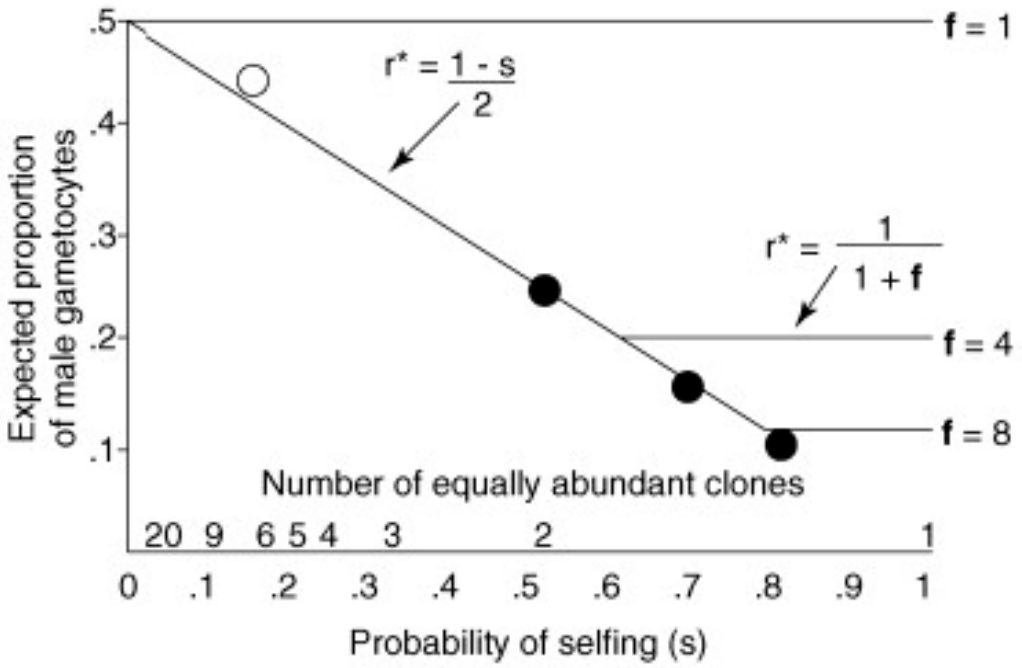


Figure 1.6. Predictions of male gametocyte proportions within *Plasmodium* infections with differing clonal compositions. Within this figure, r*=proportion of male gametocytes, s=predicted selfing rate and f=mean number of viable gametes produced by one male gametocyte (adapted from Schall 2009). In addition, the open circle represents the gametocyte sex ratio of approximately 0.45 that was calculated from an experiment with *P. chabaudi* when six clones were in competition. Closed black circles represent the calculated gametocyte sex ratio of two different clones of *P. chabaud*, both with an estimated f=8, which had their relative proportions and gametocyte sex ratios calculated using reverse transcription PCR assays (Reece *et al.* 2008). Assuming f=5 as suggested for *P. falciparum*, the lower limit of r* for *P. falciparum* would thus equal 0.17, which is in line with its estimation of 0.18 within a highly inbred population of Papua New Guinea (Read *et al.* 1992).

1.7 Phenotypic and transcriptional variation of *P. falciparum*

A small number of laboratory clones of *P. falciparum* are regularly used for *in vitro* investigations, with different parasites available from a range of original populations, with 3D7A, a clone of an airport malaria isolate from the Netherlands (Ponnudurai *et al.* 1981), being the most commonly cultured *P. falciparum* genotype. Laboratory cloned parasites which have been adapted to *ex vivo* culture for potentially long periods of time, offer a genetically tractable system with relatively robust growth rates (Reilly *et al.* 2007) for *in vitro* propagation and manipulation, including targetedtransfection strategies. Profiling of laboratory clones has revealed remarkable levels of transcriptional variation that are found within even laboratory clones of *P. falciparum*, with heterogeneity existing in the expression of genes involved in a variety of functions including metabolic processes, host cell remodelling and host-parasite interactions such as antigenic variation and erythrocyte invasion (Rovira-Graells *et al.* 2012).

Several *ex vivo* phenotyping studies of *P. falciparum* have been performed within the first asexual replication cycle, when parasites are in the red blood cell they were parasitizing within the host. Studies of erythrocyte invasion have characterised substantial phenotypic and transcriptional diversity within local endemic populations (Baum *et al.* 2003; Gomez-Escobar *et al.* 2010; Bowyer *et al.* 2015). In addition, *ex vivo* first-cycle parasite multiplication rate (PMR) has also been characterised (Chotivanich *et al.* 2000; Deans *et al.* 2006), reporting a range of initial replication rates. Thus, whilst *ex vivo* adaptation and phenotyping of clinical isolates of *P. falciparum* is clearly achievable, there remain several challenges including the low *ex vivo* PMR reported, particularly

within certain populations (Deans *et al.* 2006) and the logistics of sampling, cryopreservation and storage.

In addition, to the asexual erythrocytic stages, it has been well established *ex vivo* that different clones of *P. falciparum* vary in their baseline rates of commitment and their sex ratios. A study of clones derived from clinical isolates by limiting dilution cloning determined that rates of gametocytogenesis varied significantly between different parasite clones that had come from the same parent isolate (Graves *et al.* 1984). Across a collection of 22 isolates from cases of malaria imported within the Netherlands, the ratio of male to female gametocytes ranged from 1:3 to 1:8 (Ponnudurai *et al.* 1982), whilst the sex ratio of one clone was shown to be relatively stable over 15 weeks of subculturing (Burkot *et al.* 1984). Continuous asexual culture of *P. falciparum* parasites does however appear to lead to a reduction in the viability of gametocytes for transmission (Delves *et al.* 2016). Long-term laboratory adapted clones have been found to no longer be able to undergo gametocytogenesis, due to the deletion of a particular genome region (Alano *et al.* 1995), complicating the investigation of gametocyte commitment.

Asexual parasite replication, gametocyte sex ratio and overall gametocyte commitment are thus clinically and epidemiologically important parasite phenotypes that show evidence of variation from *ex vivo* analyses of isolates and have also been suggested to be inducible plastic phenotypes that can respond to the presence of a competitor genotype (Reece *et al.* 2008; Pollitt *et al.* 2011), although inferences on replication rate have come from the assessment of the end point density of genotypes in competition (Bell *et al.* 2006; Wacker *et al.* 2012).

1.8 Linking phenotypic variation to genetic determinants

Identification of genes that are accountable for phenotypic variation can be determined using either reverse or forward genetics, with both approaches having been utilised within the study of *Plasmodium* parasites. Reverse genetics attempts to uncover gene function by knocking out, knocking down or performing allelic replacement of a specific target locus, although this can be laborious within *Plasmodium* due to the relatively slow rates of growth and low *in vitro* parasite densities of any potential transfectants. In addition, reverse genetic approaches have been applied to confirm the essential role of genes involved within the early stages of commitment to gametocytogenesis. The complementation of gametocyte development 1 gene (GDV-1) to a laboratory clone that had lost the ability to produce gametocytes, restored this phenotype, whilst overexpression of this gene in a laboratory clone that could produce gametocytes enhanced sexual commitment (Eksi *et al.* 2012). Furthermore, the essential role of the transcription factor AP2-G, a member of the Apicomplexan AP2 protein family, was confirmed through production of a knockout mutant without AP2-G function through double homologous recombination and inducible knockdown of gametocytogenesis from targeting PfAP2-G (Kafsack *et al.* 2014).

Classically, reverse genetics has also relied on *a posterior* knowledge of a particular molecular pathway to investigate gene function, although recent approaches have begun to systematically probe this at least for genes expressed within the asexual erythrocytic stages. A genome-wide set of >2000 gene-specific plasmid vectors have been developed for interrogation of gene function in *P. berghei* (Schwach *et al.* 2014; Rita Gomes *et al.* 2015) and high-

throughput approaches using transposon mutants have begun to be applied to *P. falciparum* (Bronner *et al.* 2016). Whilst approximately 5400 genes are present within the *P. falciparum* genome, less than half have a known function and thus systematic investigation of gene function is much desired. Molecular tools for reverse genetic investigation of *P. falciparum* are also improving with the use of systems such as efficient CRISPR-Cas9 genome editing (Ghorbal *et al.* 2014).

Forward genetics-based approaches to gene function discovery have been applied using genetic crosses of genetically distinct clones of *P. falciparum* (Walliker *et al.* 1987), requiring the synchronised production of gametocytes and transmission to mosquito colonies. Quantitative trait loci (QTL) analyses of these have classically played a major role in uncovering the chloroquine resistance transport locus, providing resistance to chloroquine, which was further confirmed by reverse genetic manipulation (Wellems *et al.* 1991; Fidock *et al.* 2000; Culleton & Abkallo 2015). In addition to drug resistance, a genetic cross based approach utilising QTL has uncovered loci associated with traits including cell cycle duration (Reilly Ayala *et al.* 2010), erythrocyte invasion (Hayton *et al.* 2008) and transcriptional regulation (Gonzales *et al.* 2008). In addition, linkage group selection, which relies on phenotype-specific selection of uncloned cross progeny and subsequent comparison of genome-wide marker frequencies to the unselected progeny has also been used to identify loci associated with drug resistance, strain-specific targets of immunity and growth rates (Culleton *et al.* 2005; Pattaradilokrat *et al.* 2009; Cheesman *et al.* 2010).

An alternative, more recent forward genetics approach for *P. falciparum* has used a combination of clinical phenotyping of artemisinin clearance rates within

a South-East Asian population and whole-genome scans of selection to identify loci associated with the artemisinin slow clearance phenotype. These studies first used identical parasite genotypes found within different individuals in an analogous manner to human 'twin' studies to estimate genetic heritability of the artemisinin slow clearance trait (Anderson *et al.* 2010) and subsequently performed F_{ST} and haplotype-based genome scans of selection to identify associated loci (Cheeseman *et al.* 2012). This approach however makes the assumption that environmental and genetic factors of ‘twin’ genotypes could be considered as independent, but due to the increased probability of spatial or temporal clustering of these identical genotypes, such an assumption is unlikely to be true (Hastings *et al.* 2016). Nevertheless, these studies highlight the possibility of using underlying phenotypic variation of clinical isolates, phenotype-specific screens and whole-genome sequencing for forward genetics.

1.9 Aims of this thesis summary

This thesis investigates the within-host diversity of *P. falciparum* infections through a combination of whole-genome sequence analyses, *in vitro* competition studies and characterisation of the phenotypic variation seen within two reportedly induced plastic phenotypes through examining the following questions:

- How do conclusions from allele-frequency based genome-wide estimates of within-host diversity concord with interpretations from traditional microsatellite genotyping?
- Utilising deep-sequence data on within-host and within-population allele frequencies from four distinct West African populations, is it possible to identify genomic loci that are consistently under directional within-host selection?
- Within two West African populations of contrasting infection endemicity, are there differences in the expression levels of two key early gametocytogenesis markers and the gametocyte sex ratio?
- What baseline variation in parasite multiplication rate is present within four commonly cultured laboratory clones and a selection of West African isolates undergoing *ex vivo* culture adaptation?
- Does the parasite multiplication rate of *P. falciparum* laboratory clones in monoculture change when placed into competition with an additional genotype?

Chapter 2. Materials and Methods

2.1 Bioinformatic whole genome sequence analyses of within-host diversity

2.1.1 Sample collection and whole genome sequencing:

For the whole genome sequence analyses of Guinean isolates collected within the N'Zérékoré region described within section 3, Illumina paired-end short-read genome sequencing has previously been described and made available at the European Nucleotide Archive (Mobegi *et al.* 2014). Briefly, sequencing was performed on an Illumina HiSeq machine at the Wellcome Trust Sanger Institute and mapping of reads was done using SMALT (http://www.sanger.ac.uk.resources/software/smalt/), against the *P. falciparum* 3D7 reference sequence (version 3). Subsequent SNP calling was executed using the SAMtools software (http://www.samtools.sourceforge.next), with SNPs being removed if they were positioned within var, rifin or stevor gene families or in any sub-telomeric chromosomal regions. Finally, SNPs were also removed if they were found within loci identified as possessing highly repetitive sequences as defined by Tandem Repeat Finder (Benson 1999).

Whole genome sequence data of *P. falciparum* clinical isolates used for the independent population comparison of an F_{WS} based selection scan within section 4 were taken from four previously published sequence read data sets available at the European Nucleotide Archive collected within the N'Zérékoré region of the Republic of Guinea (Mobegi *et al.* 2014), the Greater Banjul area of The Gambia (Amambua-Ngwa *et al.* 2012) and two distinct locations within Ghana, Kintampo and Navrongo, (Duffy *et al.* 2015), which are located within a holoendemic and a hyperendemic area respectively. Briefly, sequencing was

performed on an Illumina HiSeq machine at the Wellcome Trust Sanger Institute as part of the MalariaGEN *P. falciparum* community project (Manske *et al.* 2012; MalariaGEN 2016), read alignments were done using the Burrow- Wheels Aligner program (http:\\bio-bwa.sourceforge.net) and subsequent SNP calling was executed using SAMtools software (http:\\www.samtools.sourceforge.net). To further minimise potential alignment errors, they underwent a further alignment step using SNP-o-matic (Manske & Kwiatkowski 2009). A variant call format (VCF) file for each of these four populations was downloaded from the MalariaGEN website (http://www.malariagent.net/welcome), containing metadata on all potential SNPs for each isolate and was used for all downstream analyses.

2.1.2 SNP and coverage filtering procedures

For calculation of the F_{WS} metric, allele frequencies of biallelic SNPs are needed at the within-host level, which requires deep sequencing data beyond the consensus sequence of an isolate (Manske *et al.* 2012). For this reason, additional filtering of SNPs and isolates was performed on the consensus-level sequences used for previous analysis of this population (Mobegi *et al.* 2014), to ensure the validity of the SNPs included. Rare alleles, defined as not being found at greater than 10 reads within at least one isolate and not being at a greater than 1% minor allele frequency within the population were removed along with SNPs which were located within non-protein coding regions. Population total coverage for each SNP was calculated through the summation of coverage levels across all isolates, with SNPs discarded if their population total coverage was not found within an approximate Gaussian distribution of 15-85% around the median population total coverage. Furthermore, SNPs within isolates which were not above a coverage threshold of 5 reads for an individual

allele or reach a total coverage threshold of 20 reads across both potential nucleotides for each biallelic SNP were classed as missing data and ignored within the analysis. These coverage filter designations also led to the removal of all samples that did not have read count information remaining for at least 80% of all SNPs present within the catalogue. Similarly, SNPs were also removed which were missing data for greater than 10% of the population.

2.1.3 Genome-wide calculation of the within isolate fixation index F_{WS}

Using the filtered SNP and isolate coverage datasets, allele frequencies were calculated at each SNP position for each isolate individually, through the formula p= 1-q, whereby p and q represent the proportions of read counts for the major and minor alleles. All SNPs were then assigned to ten minor allele frequency (MAF) intervals, representing the proportional frequency of the minor allele at each SNP across each population, with the ten equally sized intervals ranging from 0-5% up to 45-50%. From this, levels of within-host (H_W) and within-population (H_S) heterozygosity at each SNP were calculated as H_w=2*p_w*q_w and H_s=2*p_s*q_s. The mean H_W and H_S of each MAF interval were then computed from the corresponding heterozygosity scores of all SNPs within that particular MAF interval. The resulting isolate specific plot of H_W against H_S was produced and a linear regression model was used to determine a value for the gradient H_w/H_s, with F_{WS} = 1- (H_W/H_S). Isolates with an F_{WS} score of >=0.95 were regarded as predominantly single genotype infections in agreement with published definitions (Auburn *et al.* 2012), with all other isolates being characterised as being genotypically diverse infections. All F_{WS} analyses were performed using custom scripts in R, with an example script found within Appendix 1.

2.1.4 Population comparison of isolate F_{WS} scores:

For each of the four West African populations used for the genome-wide scan of F_{WS}, the VCF file format was manipulated using custom Bash scripts, to produce a dataframe for each population containing all necessary information for downstream processing and filtering in R, including candidate SNP annotation information and the biallelic sequencing read counts of all candidate SNPs. Population minor allele-frequency distribution analyses were performed using R-based scripts. SNP and coverage filtering was performed as described previously in section 2.1.2, leaving four separate SNP sets for each of the West African populations. Subsequently, isolate wide F_{WS} scores were calculated using linear regression based estimation within population minor allele frequency intervals of the relationship H_W/H_S (section 2.1.3).

2.2 Clinical isolate collection for *ex vivo studies*

For all *ex vivo* clinical isolate gametocyte transcript expression profiling and *ex vivo* growth (section 5 and 6 respectively) studies, blood samples of up to 5 mL were collected and cryopreserved using glycerolyte (section 2.3.2) from patients who had gone to local area health centres and provided written informed consent to have their blood collected. These isolates were collected by teams within each region led by Dr Alfred Amambua-Ngwa in Guinea, Dr Gordon Awandare in Ghana, Dr Ambroise Ahouidi in Senegal and Dr Mahmadou Diakite in Mali.

For *ex vivo* gametocyte transcript expression profiling (section 5), all Senegalese isolates were collected from Pikine within late 2013 and Ghanaian isolates were collected from Navrongo and Kintampo between August and October 2013 and May and August 2013 respectively. For *ex vivo* clinical

isolate growth studies (section 6.3.3), Guinean isolates were collected from near Faranah in November 2012, Malian isolates were collected Nioro de Sahel in 2013 and Senegalese isolates were collected from Pikine within late 2014. Once in glycerolyte, these isolates were frozen (section 2.3.2) and transported on dry ice to LSHTM, where they were stored in LN_2.

2.3 *P. falciparum* cell culture techniques

2.3.1 Parasites studied

For all monoculture and competitive growth studies (sections 6.3.2 and 7.3.2 respectively), four different laboratory parasites of were cultured; 3D7A-GFP, Dd2, D10 and HB3. These four clones offered a range of geographical origins and were also being used for other purposes in Chapter 4. Clone 3D7A was derived from isolate NF54 which originated from a case of airport malaria within the Netherlands (Ponnudurai *et al.* 1981; Walliker *et al.* 1987) with 3D7A-GFP subsequently being a clone of 3D7A that had been transfected to constitutively express green fluorescent protein without any reported effect on parasite growth or gametocytogenesis phenotype (Talman *et al.* 2010). Dd2 is a clone of the isolate W2 (Guinet *et al.* 1996), which originated from South East Asia (Campbell *et al.* 1982), whilst HB3 was cloned from the Honduran I/CDC isolate (Bhasin & Trager 1984) and D10 was cloned from the FCQ27 isolate that originated from Papua New Guinea (Chen *et al.* 1980). These four clones were also used for transfection experiments (section 7.3.3), with the exception that a wild-type 3D7A clone was used instead of 3D7A-GFP.

2.3.2 Thawing and cryopreservation of *P. falciparum* parasites

Parasites were thawed from long-term storage in liquid nitrogen by initially placing in a 37°C waterbath for 2 minutes. Subsequently, the volume of blood

(V) within the cryovial was determined and 0.1x the volume of NaCl was added to this and mixed gently by shaking, before being left for 2 minutes. Following this, a 10x volume of V of 1.6% NaCl was also added and mixed gently again by shaking. This solution was then centrifuged for 5 minutes at 500 g and the supernatant was removed, washed with completed medium and the remaining blood pellet was added to 10 mL of complete medium in a T25 flask, at a haematocrit of approximately 3%.

When cryopreservation was required, parasites were grown to have a ring-stage parasitemia of at least 2% and the culture was spun down at 500 g for 3 minutes, before the supernatant culture medium was aspirated. The volume of the blood pellet was estimated and 0.33 x the volume of glycerolyte solution was added dropwise, shaken gently and left at room temperature for 5 minutes. Subsequently, an additional 1.33 x the blood pellet volume of glycerolyte solution was added, the tube was gently shaken and the cell suspension was then transferred to a cryovial tube and put into a liquid nitrogen freezer.

2.3.3 Erythrocyte preparation:

For general parasite culture, A+ erythrocytes (National Blood Transfusion Service) were washed thrice with incomplete medium (RPMI-1640, Sigma Aldrich) through centrifugation at 2500 g for 5 minutes, prior to being diluted to a haematocrit of 50% using complete medium and kept at 4°C. These cells were kept for up to 10 days, or for a shorter period if they appeared to show any evidence of haemolysis or inability to support parasite growth.

2.3.4 General parasite culture conditions

Laboratory clone parasites were generally cultured asynchronously in complete medium consisting of RPMI 1640 medium (Sigma-Aldrich) supplemented with

25 mM HEPES, 2 µg/mL sodium bicarbonate, 25 µg/mL gentamycin, 0.2% glucose and 0.5% albumax, at a haematocrit of between 3 and 5% in a 10mL total culture volume, with a gaseous environment of 95% Nitrogen and 5% carbon dioxide and a temperature of 37°C (Trager and Jensen 1976). Media was changed in accordance with flask parasitemia, which was monitored by creating a blood smear on a microscopy slide that was subsequently giemsa stained and parasitemia determined through light microscopy. Clinical isolates undergoing *ex vivo* adaptation were maintained under the same conditions, with the exception of a gaseous environment of 1% O_2, 3% CO_2, 96% N_2 within an orbital shaking incubator operating at 50 rpm and a haematocrit of no more than 3%.

2.3.5 Synchronisation of parasite stages

Ring-stage laboratory clone parasites were synchronised using a 5% sorbitol solution (Lamrbos & Vanderberg 1979) to remove trophozoite and schizont stages. *P. falciparum* cell cultures were centrifuged at 500 g for 3 minutes, before being resuspended within 10 x the blood pellet volume of 5% sorbitol for 5 minutes and incubated at room temperature. These solutions were subsequently washed three times with complete culture medium and resuspended once again at between 3% and 5% haematocrit.

2.4 Genomic DNA extraction and preparation of genomic standards

Genomic DNA was prepared to produce DNA standards from each of the four cultured laboratory clones. For this, parasites were grown asynchronously under general culture conditions (section 2.3.4) to a parasitemia of approximately 10%, frozen at -20°C and were subsequently extracted using a QIAamp DNA Blood Midi Kit (Qiagen). Once extracted these standards were

quantified against a calibrator PCR product generated from two primers that were designed outside the target region of a pan-genotype Pfs25 assay (Schneider *et al.* 2015, Table 7.1): Pfs25 outer forward 5'-GGAACTTTGCCTATATTACATGAGCA-3' and Pfs25 outer reverse 5'-CGAAAGTTACCGTGGATACTGTATG-3' using a 2x 5 µL SYBR Select Master Mix (Life Technologies), with 300 nM primers, 2.9 µL of nuclease-free water and 1.5 µL of 3D7A-GFP DNA. The generated PCR product was quantified using a Qubit 2.0 Fluorometric Quantitation device and the dsDNA HS assay kit (Qubit). Subsequently, this PCR product was diluted tenfold in 5 ng/µL salmon sperm DNA and used as an additional calibrator standard to quantify the genomic DNA standards of the four clones against. The concentration of the Pfs25 product dilution series ranged from 0.01 ng/µL down to 1 x 10^{-7} ng/µL. For this, a Kapa Probe Fast kit (Kapa Biosystems) was used, with a 10 µL total reaction volume consisting of 2x Kapa Probe Fast Mastermix, 250 nM primers of the highly conserved pan-genotype Pfs25 assay (Table 7.1), 250 nM of the probe 6FAM-5'-CCGTTTCATACGCTTGTAA-3'-MGB (Schneider *et al.* 2015), 5.5 µL of H_2O and 3 µL of DNA per reaction. All quantitative PCR (qPCR) reactions were carried out using a Rotor-Gene 3000 qPCR machine (Corbett). Tenfold genomic DNA standards were subsequently created for each laboratory clone, with 5 ng/µL salmon sperm DNA used as diluent and all genomic DNA standards were kept at -20°C.

2.5 Parasite genotyping

2.5.1 Sequencing and genotyping of MSP-1 block 2 for laboratory clones

For the four sets of laboratory clone DNA, their identity was initially confirmed by capillary sequencing of the highly polymorphic MSP-1 block 2 region. PCR products were generated by nested PCR (Snounou *et al.* 1999) using the pan-

genotype M1-OF/M1-OR primer set for the first nest, with thermocycler conditions of 95°C for 10 minutes, followed by 25 cycles of 94°C for 1 minute, 58°C for 2 minutes and 72°C for 2 minutes, with a final extension of 72°C for 5 minutes, using a Kapa blood PCR kit (Kapa Biosystems). Products from the first nest were diluted 1 in 100 and 1 µL was added to the second nest, where the three allelic type specific PCR sets were used (K1, MAD20 and RO33), with thermocycler conditions of 95°C for 10 minutes, followed by 30 cycles of 94°C for 1 minute, 61°C for 2 minutes and 72°C for 2 minutes, with a final extension of 72°C for 5 minutes. PCR products were visualised on a 3% agarose gel, providing initial confirmation of the clone identity.

For additional confirmation, the MSP-1 block 2 region underwent capillary sequencing using all of the allelic-type specific primers using a BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), on PCR product from nest 1, which had been purified using a QIAquick PCR Purification kit (Qiagen). Thermocycling conditions for the cycle sequencing were 25 cycles of 96°C for 30 seconds, 50°C for 15 seconds and 60°C for 4 minutes. The sequencing product then underwent ethanol precipitation prior to being resuspended in highly de-ionised formamide. Sequencing products were assessed using an ABI 3730 DNA Analyzer (Applied Biosystems) and the results were visualised using SeqMan alignment software (DNAStar). Laboratory clone parasites being cultured were intermittently re-checked for their identity by MSP-1 block 2 nested PCR typing as described or through allele-specific qPCR (Table 7.1).

2.5.2 Microsatellite genotyping of *ex vivo* clinical isolates

For *ex vivo* phenotyping of clinical isolates parasite multiplication rate (section 6.3.3), DNA from the one of the extracted DNA samples taken at Day 2 underwent microsatellite genotyping at four polymorphic loci (TA1, Polyα, PfPK2 and TAA109), which were selected due to their previously published high levels of expected heterozygosity within the Republic of Guinea (Mobegi *et al.* 2012). Genotyping of each marker was undertaken using the hemi-nested protocol reported previously (Anderson *et al.* 1999; Mobegi *et al.* 2012). Both nests were performed in separate tubes for each loci, in a total volume of 11 µL, consisting of 1x ThermoPol II reaction buffer (New England Biolabs), 0.2 U Taq DNA Polymerase, 0.25 µL of 10 µM forward primer, 0.25 µL of 10 µM reverse primer, 0.1 µL of 100mM $MgSO_4$, and 1 µL of template, 6.55 µL of H_2O and 0.8 µL of a 10 mM dNTP mix. PCR products were run separately using Genescan and subsequently analysed on a Genetic Analyzer 3730 capillary electrophoresis machine (Applied Biosystems). Allelic electrophoresis sizes were determined using GENEMAPPER version 4.0 software, with the predominant allele identified as the peak with the greatest height, whilst minor alleles were also noted if their peak height was at least 25% the size of the predominant allele recorded.

2.5.3 Primer selection and design for allele-specific qPCR assays:

In silico visualisation of the DNA sequences of the four aforementioned laboratory clones was performed in Jalview 2.8.1 (Waterhouse *et al.* 2009) for merozoite surface protein 1 (MSP1, PF3D7_0930300), merozoite surface protein 6 (MSP6, PF3D7_1035500), merozoite surface protein 7 (MSP7, PF3D7_1335100), duffy binding-like merozoite surface protein (DBLMSP, PF3D7_1035700) and erythrocyte binding antigen-175 (EBA-175,

PF3D7_0731500). A literature review found two suitable *P. falciparum* allele-specific quantitative PCR assays for these clones which both are targeted to the MSP1 gene (Wacker *et al.* 2012; Chen *et al.* 2014). This indicated that the two published assays should be able to distinguish between 5 of the 6 different pairwise scenarios desired, with an additional allele-specific assay needed to distinguish the Dd2 and D10 clones. Primer design was aided by use of the Primer 3 (version 4.0, Untergasser *et al.* 2012) software. In addition, the qPCR assay targeting the highly conserved Pfs25 gene locus was used to provide pan-genotype quantification of parasite DNA (Schneider *et al.* 2015).

2.6 Standardised conditions for quantitation of growth rates of *P. falciparum*

2.6.1 Standardised culture conditions

For all monoculture, *ex vivo* and competitive growth assays performed (sections 6.3 and 7.3.2), biological replication was implemented, with each replicate defined as having a different erythrocyte donor, whose blood group was unknown. These donors were collected from the blood service at LSHTM one day prior to being put into a growth assay and coagulation was prevented through the addition of 50 units of heparin to each 50 mL falcon tube of blood. Subsequently, erythrocytes were prepared as previously described (see section 2.3.3).

2.6.2 Monoculture laboratory clone growth assays

For all monoculture laboratory clone growth assays (section 6.3.2), the four clones were cultured simultaneously, with 6 biological replicates being performed. Culture conditions for all laboratory clone growth assays were as follows: 10mL cultures consisting of complete medium with 0.5% albumax, 5%

starting haematocrit of erythrocytes and a gas composition of 95% N_2 and 5% CO_2 with atmospheric O_2 at a temperature of 37°C. For setup on Day 0 of each assay, the asynchronous parasitemia of the parent flask was calculated from counting 1000 red blood cells of a giemsa smear using a light microscope. Replicates were asynchronously initiated at a parasitemia of 0.02% in fresh complete medium. Complete medium changes were performed on Day 2, 4, 5, 6 and 7. Every 48 hours (Day 0, 2, 4, 6 and 8) a 300 µL sample of resuspended culture medium was taken, from which 100 µL was pelleted and used to make a giemsa-stained slide, with the remaining 200 µL frozen at -20°C.

2.6.3 *Ex vivo* clinical isolate growth assays

For quantitation of *ex vivo* clinical isolate growth rates a total of 61 isolates from three different locations across West Africa (section 2.2) underwent adaptation to *ex vivo* culture. To minimise any effect upon the parasites of the freezing, transport and thawing of red blood cells from multiple locations, isolates underwent several replication cycles of culture adaptation. This involved the thaw and maintenance of clinical isolates under the conditions stated in section 2.3.4. Assay conditions were set as described (section 2.6.1), with the exception of a 3% haematocrit and a gaseous composition of 1% O_2, 3% CO_2 and 96% N_2, within an orbital shaking incubator operating at 50 rpm. These growth assays were performed for 6 days, with media being changed on Day 2, 4 and 5 and a 200 µL sample of total culture was taken every 48 hours and frozen at -20°C. In addition, to provide a control that would allow for normalisation across the triplicates of erythrocyte donors used, all performed clinical isolate growth assays were run alongside a 3D7A-GFP control.

2.6.4 Competitive growth assays of *P. falciparum*

Pairwise competitive growth assays were set up between the four aforementioned laboratory clones. Use of these parasites provided a total of 6 co-culture scenarios; 3D7A vs Dd2, 3D7A vs HB3, 3D7A vs D10, Dd2 vs HB3, Dd2 vs D10 and HB3 vs D10. These were set up identically to the monoculture laboratory clone growth assays, except that for each competitive scenario, such as Dd2 vs D10, there were three monoculture flasks of Dd2 and D10 each begun at a parasitemia of 0.02% and three competitive flasks where 0.01% of each parasite was added to the starting culture in order to try to ensure that the starting density of the control and experimental groups remained equivalent. Allele-specific qPCR assays were used to monitor growth rates of individual clones experiencing competition (section 7.2).

2.7 qPCR for PMR quantitation of laboratory clones, *ex vivo* isolates and competing clones

2.7.1 qPCR for analysis of laboratory clone and *ex vivo* isolate PMR

To ensure further standardisation of the assay, all biological replicates where the same blood batch donor was used were grown simultaneously in the same batch of complete medium and all subsequent DNA extractions for each biological replicate across the clone combinations were also performed in unison for all monoculture, *ex vivo* and competitive growth experiments.

All collected blood samples were extracted using the centrifuge-based protocol of a Qiagen DNA Blood Mini kit (Qiagen), with all extracted DNA eluted in 40 µL of elution buffer and stored at -20°C. For absolute quantitation of the growth rate of each laboratory clone, the corresponding genomic DNA standard was used, for example Dd2 was quantified against the Dd2 series. For absolute quantitation of the growth rate of all clinical isolates profiled, this was compared

against the 3D7A-GFP dilution series. All genomic standard controls were run in technical triplicate for each run, with a concentration range of 1 million copies per µL down to 10 copies per µL. qPCR reactions for all DNA extractions for each timepoint of each biological replicate were also performed in triplicate and run alongside a no template control sample. Primers targeting the conserved Pfs25 locus were used, with the forward primer of 5'-CCATGTGGAGATTTTTCCAAATGTA-3' and the reverse primer of 5'-CATTTACCGTTACCACAAGTTACATTC-3' (Schneider *et al.* 2015). These were used in a total reaction volume of 10 µL, consisting of 5 µL SYBR select master mix (Applied Biosystems), 0.3 µL of 10 µM forward primer, 0.3 µL of 10 µM reverse primer and 2.9 µL of H_2O. All qPCR reactions were run on a Rotor-Gene 3000 qPCR machine, with thermo cycler conditions of 50°C for 2 minutes, 95°C for 2 minutes, 40 cycles of 95°C for 15 seconds and 60°C for 1 minute with acquisition being performed in the FAM channel with a gain setting of 7. Following each machine run, a melt curve was performed, which comprised ramping from 60°C to 99°C in 1°C increments, with a 30 second pause on the first temperature step and 5 seconds pauses on each step afterwards.

2.7.2 PMR analysis of qPCR data for laboratory clones and *ex vivo* isolates

For each growth assay timepoint, the geometric mean of the estimated parasite genome copy number per µL of each technical triplicate was calculated by comparison against the corresponding genomic DNA series using the Rotor-Gene 6 software, with a consistent fluorescence threshold of 0.05 implemented. This cut off consistently gave quantification within the log-linear phase of the amplification process. These values were collated for all isolates run and using the R statistical environment, parasite genome copy densities were log

transformed and expressed as the number of genome copies per µL of total culture for each isolate. Subsequently, to quantify parasite growth rate, a general linear model was fitted within R across all biological replicates and 95% confidence intervals of the determined parasite multiplication rate, defined as the fold change in growth per 48 hours across the entirety of the assay, were calculated using the confint function available within R.

For normalisation of the clinical isolate growth rates across the sets of different biological replications, all independent rates of 3D7A-GFP growth were averaged. The difference between the calculated growth rate of the clinical isolate and its particular 3D7A-GFP control estimation was calculated and used to normalise against the averaged rate of 3D7A-GFP growth across all replications. These normalised values were subsequently used to compare the repeated growth rate analysis of clinical isolates through the culture adaptation process.

2.7.3 qPCR for analysis of pairwise competition growth rates of laboratory clones

For all monoculture growth assays, growth rate analysis was performed using the pan-genotype Pfs25 assay, whilst for all pairwise competition assays, the appropriate allele-specific assay (Table 7.1) was used to provide allele-specific quantitation. All qPCR reactions were performed in technical triplicate and absolutely quantified against the aforementioned genomic DNA standards. The absolute copy number of each clone within a competition assay, was quantified by the appropriate allele-specific assay to monitor growth rate through general linear modelling and determination of 95% confidence intervals within R. All extracted timepoints were performed in technical triplicate and run alongside no template control samples for qPCR.

2.8 Generation of fluorescent *P. falciparum* parasites for *in vitro* competition studies

2.8.1 Fluorescent protein selection

Whilst GFP has been used previously to provide a constiutively expressed fluorescent parasite tool for *P.falciparum* (Wilson *et al.* 2010; Talman *et al.* 2010), for potential competition studies *in vitro*, a two-colour system would be needed. Due to excitation wavelength and emission detection filters available using the in-house BD FACSAria Fusion cell sorter two red fluorescent proteins were selected; mCherry (Shaner *et al.* 2004) and mKate2 (Shcherbo *et al.* 2009), following visualisation of their emission and excitation spectra at nic.ucsf.edu/FPvisualization/. Whilst mCherry has previously regularly transfected into *P. falciparum* (Mancio-Silva *et al.* 2013), mKate2 has not previously been used within *P. falciparum*, but was chosen due to its reported increased brightness and improved spectral suitability for the in-house FACSAria Fusion flow cytometer. The plasmid containing the mKate2 sequence was synthesised as a gBlocks gene fragment from Integrated DNA Technologies using nucleotide sequence derived from a publically available amino acid sequence (Shcherbo *et al.* 2009) and the mCherry protein was PCR amplified from a plasmid kindly donated by Dr Hans Dessens.

2.8.2 *E.coli* competent cell production

For use in non-ligation transformations, chemically competent stellar *E.coli* cells were made by growing a colony of cells overnight within a 400 mL conical flask of LB-broth supplemented with 100 µg/mL ampicillin, until an OD600 measurement from a spectrophotometer reached a reading of 0.4. These cultures were then centrifuged in 50 mL tubes at 4000 rpm for 15 minutes at 4°C, These were then resuspended in 200 mL of 100 mM $MgCl_2$ and placed on

ice for 10 minutes, prior to repeated centrifugation, aspiration and resuspension in 200 mL of $CaCl_2$, prior to being left on ice for 1 hour. Following this, the cells were once again centrifuged and aspirated, before cell pellets were resuspended in 4mL of 100 mM $CaCl_2$, 15% glycerol solution, aliquoted in 200 µL and stored at -80°C.

2.8.3 Restriction digests and ligation reactions:

Restriction enzymes for digestion were obtained from either New England Biolabs or ThermoFisher. All digestions were performed overnight at 37°C. For all digestion products and PCR amplified inserts, the product underwent agarose gel electrophoresis, before target bands were visualised under ultraviolet light and cut out with a scalpel, prior to being purified using a Nucleospin Gel and PCR clean-up kit using manufacturer's instructions and stored subsequently at -20°C. Ligation reactions between the digested plasmid vector and PCR amplified insert were performed using an In-Fusion HD Cloning Kit (Clontech) in reaction volumes of in 5 µL total reaction volumes, consisting of vector to insert ratio of 2:1 and were carried out in accordance with manufacturer's instructions.

Ligation transformations were performed using the competent cells provided within the In-Fusion HD Cloning kit (Clontech). To screen potential transformants, colony PCR was performed, whereby between 5 and 10 colonies were picked from an agar plate, placed in 30 µL of H_2O and lysed at 95°C for 5 minutes. These then underwent PCR using a Clontech 2x PCR mix, 0.3 µL of each 10 µM primer and 4.4 µL of H_2O, prior to being visualised using agarose gel electrophoresis. Correctly sized colonies were then picked for overnight culture in LB-broth and 100 µg/mL ampicillin, prior to a QIAprep Spin Miniprep

extraction (Qiagen) and DNA sequencing to confirm successful cloning of the insert into the vector.

2.8.4 DNA sequencing

Capillary sequencing of ligation colonies was performed using a BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), on plasmid DNA which had been isolated using a QIAprep Spin Miniprep kit (Qiagen). Thermocycling conditions for the cycle sequencing were 25 cycles of 96°C for 30 seconds, 50°C for 15 seconds and 60°C for 4 minutes. The sequencing product then underwent ethanol precipitation prior to being resuspended in highly de-ionised formamide (ThermoFisher). Sequencing products were assessed using an ABI 3730 DNA Analyzer (Applied Biosystems) using either an in-house faculty service or outsourced to Source Bioscience and the results were visualised using FinchTV software (Geospiza).

2.8.5 Integration into the Pf47 locus

The parental plasmid, pEFGFP, used for subsequent modification was obtained from Professor Bob Sinden (Talman *et al.* 2010, Figure 2.1) and targets integration through single homologous recombination at the Pf47 locus (PF3D7_1346800). Visualisation of the homology region locus within publically available sequence data from PlasmoDB revealed a high degree of nucleotide sequence conservation at this locus. For this approach, the pEFGFP plasmid was initially modified to contain an additional restriction site to facilitate replacement of the GFP coding sequencing with that of the mCherry sequence. pEFGFP underwent restriction digestion using BamHI and KpnI (New England BioLabs), with a total reaction volume of 35 µL. In addition, primers LM20FA, LM20FB and LM20R were used to generate a 1.3 kb PCR product from plasmid

pEFGFP, which was done to In-Fusion clone the restriction site AVRII in place of the GFP sequence producing plasmid pEFGFPINT, which was confirmed by colony PCR using primer A0153 and ST516. Subsequently pEFGFPINT underwent restriction digestion using KpnI and AvrII in a total reaction volume of 30 µL. The mCherry coding sequence was PCR amplified using primers LM23mCherryF and LM23mCherryR , whilst the GFP sequence was PCR amplified using primers LM24GFPF and LM24GFPR. These products underwent agarose gel electrophoresis, gel extraction under UV light, before being In-Fusion cloned in the digested pEFGFPINT vector. Colony PCR was used to screen colonies for insertion using primers ST516 and A0153 and candidate plasmids underwent DNA sequencing, leading to the production of pEFmCherryLM. For replacement of the GFP sequence with the mKate2 sequence, the pEFGFP plasmid was digested using KpnI and CfrI in a 35 µL reaction volume, and the mKate2 insert was In-Fusion cloned into the vector following PCR amplification using primers LM22mKateF and LM22mKateR, prior to colony PCR screening using primers ST516 and A0153 and sequencing by A0153 to produce pEFmKate2LM. All primers used within the fluorescent parasite generation work can be found in Appendix 2.

2.8.6 Integration PCR

Integration at the Pf47 locus was monitored using primers AT1 and AT2 (Talman *et al.* 2010) to detect the presence of a wild type Pf47 locus, whilst primers AT1 and ST509 were used to detect the presence of plasmid integration at the Pf47 locus. Cycling conditions for the wild type PCR were 95°C for 2minutes, followed by 25 cycles of 95°C for 15 seconds, 56°C for 30 seconds and 68°C for 70 seconds, with a final elongation step of 68°C for 5

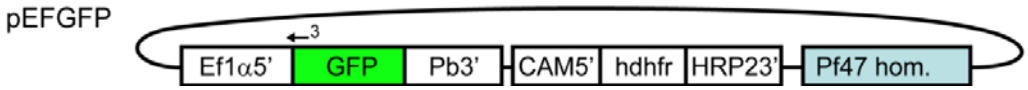


Figure 2.1 Schematic of the pEFGFP plasmid. This plasmid was modified to express mCherry or mKate2 instead of GFP (adapted from Talman *et al.* 2010). The human dihydrofolate reductase region provides resistance to the antifolate compound WR99210 (de Koning-Ward *et al.* 2000).

minutes. Cycling conditions for the integration PCR were 95°C for 2minutes, followed by 28 cycles of 95°C for 15 seconds, 57°C for 30 seconds and 68°C for 70 seconds, with a final elongation step of 68°C for 5 minutes. All PCR products were subsequently run on a 1% agarose gel, alongside genomic wild-type and 3D7A-GFP integration controls.

2.8.7 Transfection of *P. falciparum* parasites

For plasmid preparation, at least 50 µg of each plasmid were grown up in competent *E. coli* cells and then precipitated overnight at -20°C using 0.1 x V of NaOAc and 2.5 x V of absolute EtOH. The following morning plasmids were microcentrifuged at 14000 rpm, the subsequent pellets were washed in 70% EtOH, prior to being microcentrifuged again at 14000 rpm for 5 minutes. Following aspiration and air drying, these pellets were resuspended under sterile conditions using 30 µL of 0.2 µM filtered H_2O and subsequently 30 µL of cytomix, consisting of 120 mM KCL, 0.15mM CaCl2, 2 mM EGTA, 5 mM $MgCl_2$, 10 mM K_2HPO_4 pH 7.6 and 25 mM HEPES.

For all transfections, a spontaneous uptake transfer (Deitsch *et al.* 2001) was performed on parasite cultures with ~5% levels of ring stage parasites. For each transfection, a 1 mL aliquot of fresh 50% haematocrit erythrocytes was washed in 10 mL of cytomix, aspirated and 400 µL of the pelleted erythrocytes were transferred to a 2 mm electroporation cuvette and the 60 µL aliquot of plasmid resuspended in cytomix was added to the cuvette and mixed gently. The cuvette was placed on ice for 30 minutes, prior to being electroporated using a Bio-Rad device (Bio-Rad) with settings of 0.31 kV and 960 uFd capacitance. Electroporated erythrocytes were then transferred to pre-prepared falcon tubes containing incomplete RPMI and subsequently underwent two rounds of

resuspension and aspiration. The final pellet was resuspended in 10mL of complete media and parasitemia was diluted down to ~0.5%. Culture media was changed daily for the next 7 days, with either 2.5 nM or 5 nM of WR99210 being added to the culture on Day 2. The health of the red blood cells was monitored by microscopy once a week alongside a change of WR-supplemented media and when deemed necessary, a 1 in 10 dilution of erythrocytes was performed.

2.8.8 Limiting dilution cloning of integrated parasite transfections

In order to clone out individual parasites possessing an integrated fluorescent plasmid copy, the parasitemia of the parent culture was calculated from a 1000 red blood cell count of a Giemsa stained microscope slide. Following this, the haematocrit of the parent culture and the fresh erythrocyte donor was calculated using a C-chip haemocytometer (Neubauer Improved), using the manufacturer's instructions. Parasite cultures were subsequently diluted down to a concentration of 1 parasite per mL and 250 µL was loaded into each well of a 96-well plate at a haematocrit of 1%. This low haematocrit allowed for the identification of parasite positive wells using a plaque-based approach (Thomas *et al.* 2016) under the 4x lens of a light microscope. For each transfection plate, 6 parasite positive wells were taken into a 6 well plate and bulked up to a parasitemia of ~1%, prior to 200 µL of 50% haematocrit culture being frozen at -20°C and subsequently extracted using a QIAamp DNA Blood Midi Kit (Qiagen).

2.8.9 Live parasite imaging

All parasite images were taken using live parasites, prepared by placing a droplet of approximately 30% haematocrit culture under a coverslip on a glass microscope slide. These were subsequently imaged under the 40x lens of an

EVOS FL Cell Imaging System (EVOS). When required for counting of the proportion of parasites that were fluorescent, DNA was stained prior to imaging using Hoeschst stain 33342 (ThermoFisher) at a concentration of 1 µg/mL to 50% haematocrit cell culture for 45 minutes at 37°C.

2.8.10 Comparison of quantitation of *in vitro* competition using a combination of allele-specific qPCR and fluorescent imaging

Following PCR screening of Pf47 integration and limiting dilution cloning, measurements of *in vitro* mixedness were determined using both the allele-specific qPCR assays and microscopic fluorescence, ring-stage parasites were sorbitol synchronised and allowed to develop into early stage trophozoite stages. Parasitemia of the parent flask was calculated from a 1000 red blood cell count to confirm the absence of multi-nucleate schizont stages and five pairwise artificial competitive scenarios were produced consisting of 20:1, 10:1, 1:1, 1:10 and 1:20 mixtures of the two genotypes. 200 µL aliquots of these mixtures were taken for qPCR and processed as described (section 2.7.1), whilst microscopic counts were performed on the two parasites with different fluorophores using the alternate EVOS Texas Red and GFP Light cubes, for distinction between red and green fluorescent parasites respectively. Images of these mixtures were taken (section 2.7.9) and proportions of each genotype were calculated by counting 200 parasites.

3. Genome-wide assessment of *P. falciparum* within-host diversity within a highly endemic West African population

3.1 Introduction

Clinical infections of *P. falciparum* regularly contain a mixture of different haploid *P. falciparum* genotypes, particularly in areas of high endemicity where superinfection can frequently occur (Anderson et al., 2000). As previously described (section 1.5), analyses of the within-host diversity of *P. falciparum* have commonly used assays to discriminate alleles at a small number of highly polymorphic antigen genes loci (Farnert et al., 2001). However, as these loci are subject to immune selection from the host, genotyping of multiple putatively neutral polymorphic microsatellite marker loci is preferable for most general analyses (Anderson et al., 2000). These studies have clearly demonstrated a wide variation in prevalence of multiple genotype infections and an inverse correlation with multilocus linkage disequilbrium in different geographical populations of *P. falciparum* (Anderson et al., 2000; Machado et al., 2004; Anthony et al., 2005; Schultz et al., 2010).

Further dissection of the within-host diversity of *P. falciparum* infections has recently been performed using genome-wide SNP data. Studies on different clones from individual infections have revealed that a high degree of haplotype relatedness is exhibited by some distinct clones within infections, in comparison to those sampled from separate infections (Nkhoma et al., 2012; Nair et al., 2014, section 1.5.1). This illustrates a challenge in characterising or quantifying the within-host diversity of *P. falciparum*, as natural infections may consist of a single clone, multiple genetically unrelated clones (Figure 3.1), or a mixture of closely related non-identical genotypes.

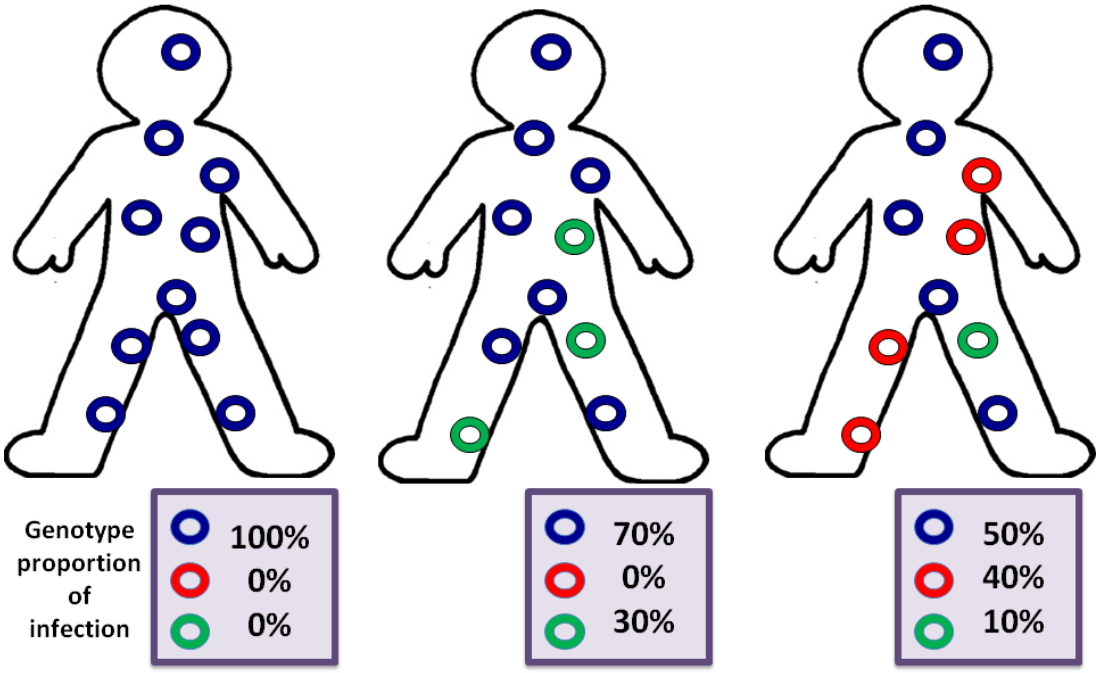


Figure 3.1. Schematic diagram representing the presence of multiple genotypes within natural infections of *P. falciparum*. Three infections are shown consisting of a single genotype, two genotypes and three genotypes respectively, in different relative proportions. This scenario is further complicated by different degrees of relatedness exhibited by clones within infections (Nkhoma *et al.* 2012, section 1.5.1).

Moreover, these combinations within a host may be present at different relative proportions within an infection, further complicating their analysis and can change in proportion over time (Farnert *et al.* 1997).

While genome-wide SNP data of clones from an infection can provide resolution on closely related but distinct clones following intensive *ex vivo* culturing, whole-genome deep sequence data of uncloned samples can also be used to study parasite diversity. Due to *P. falciparum* being haploid during the intraerythrocytic cycle, the relative proportions of SNP alleles in whole-genome sequence data from an uncloned infection can be estimated from the proportions of reads mapping to a reference sequence, and this has been used for computation of a within-isolate fixation index (Auburn et al., 2012; Manske et al., 2012). Taking this approach, the F_{WS} index has been developed as a quantitative measure which compares the extent of average genome-wide within-host diversity (H_W) to the overall parasite population (or local subpopulation, H_S) diversity. This F_{WS} index has an upper limit of 1.0, indicative of an extremely clonal isolate where sequence diversity is absent or at least very limited, whilst an isolate score of zero is indicative of the level of within-host heterozygosity being equal to the level of within-population heterozygosity. Thus, the index is influenced by the relative proportions as well as the degree of relatedness of genotypes that are present within an infection.

This chapter focuses on the bioinformatic analysis of clinical *P. falciparum* isolate sequence data from the N'Zérékoré region of the Republic of Guinea, which has previously been characterised as a highly endemic region through ten locus microsatellite marker typing (Mobegi *et al.* 2012), where the vast majority of infections are found to consist of more than a single genotype. The

within-host diversity of *P. falciparum* infections has been assessed using the computation of the allele-frequency based F_{WS} index from whole genome sequence data. Further analyses have explored the feasibility of estimating genome-wide F_{WS} from small numbers of SNP loci when whole-genome sequence data are not obtainable. In addition, measurement of within-host diversity through the F_{WS} index has been compared against three other proposed methods; the well established method of multilocus microsatellite typing (Anderson *et al.* 1999) and the more recently proposed methodologies of estMOI (Assefa *et al.* 2014) and a complexity of infection likelihood (COIL) calculator (Galinsky *et al.* 2015). These two latter methodologies, rather than characterising within-host diversity *per se*, provide a clone-counting measure of an infection through analysis of whole genome sequence alignment files and biallelic SNP barcodes respectively.

3.2 Materials and Methods

From an initial population of 120 Guinean whole genome sequences that were produced from isolates collected within the N'Zérékoré region of the Republic of Guinea (section 2.1.1), a total of 95 Guinean isolates and 50082 SNPs passed through several quality control filtering procedures (section 2.1.2, Figure 3.1) and isolate F_{WS} indices were calculated (section 2.1.3). Subsequently, for estimation of isolate F_{WS} from limited numbers of SNPs, F_{WS} indices were subsequently derived from sampling a small number (between 1 and 20) of randomly selected SNPs and comparing the calculated isolate scores with the genome-wide indices previously determined. A Pearson's correlation coefficient with the genome-wide SNP estimate of F_{WS} was determined across all isolates. For each limited SNP selection of *n* SNPs, 100 iterations were performed to randomly allocated subset groups of *n* SNPs and the mean Pearson's

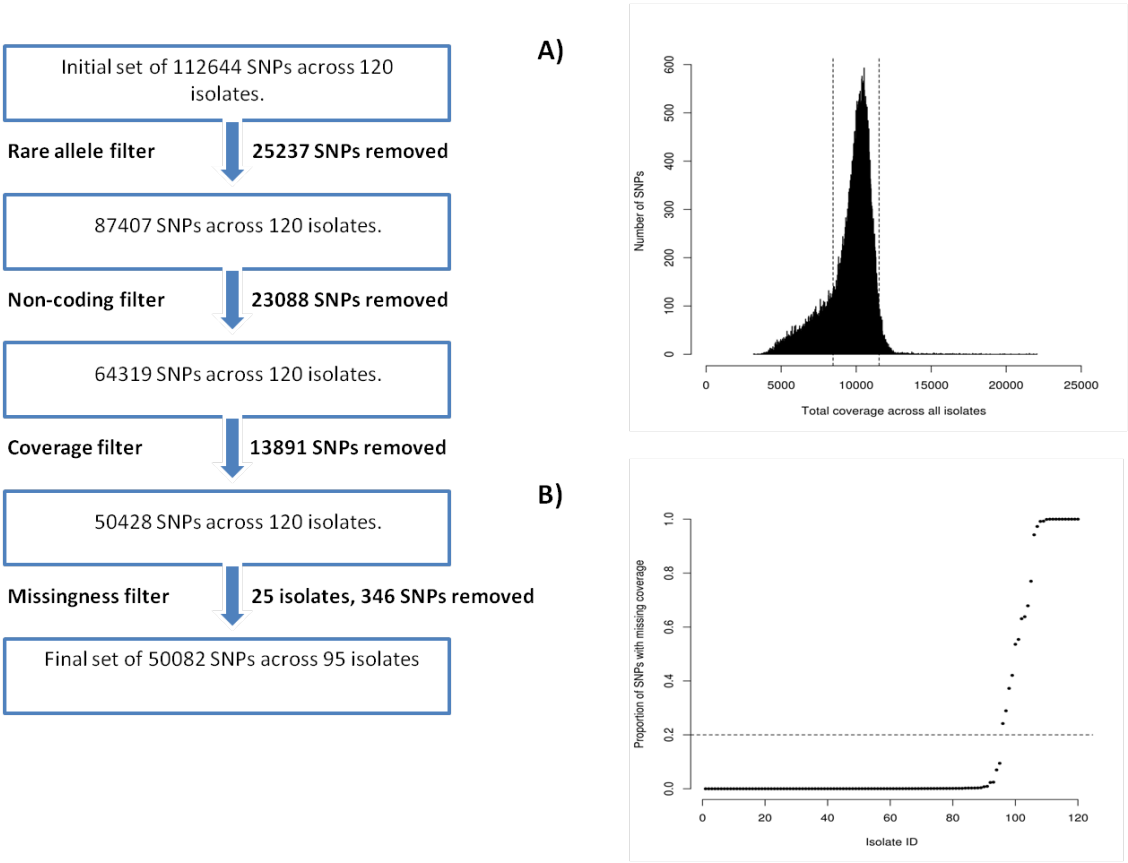


Figure 3.2 Quality control filtering pipeline of SNPs and isolates for calculation of the F_{WS} metric. Example plots represent A) the population coverage filter implemented, which removed 13891 SNPs lying outside of the dashed lines which indicate an approximate Gaussian distribution and B) the isolate missingness filter, whereby 25 isolates were removed which were missing coverage data for greater than 20% of the filtered population SNP set.

coefficient across these iterations was calculated. This analysis was repeated and comparisons performed between sets of SNPs with different ranges of population minor allele frequencies.

Isolate F_{WS} scores were compared with three other proposed measures of within-host diversity; PCR based genotyping of ten polymorphic microsatellite loci (Anderson *et al.* 1999) performed by Dr Victor Mobegi (Victor Mobegi, LSHTM PhD thesis 2014), a deep sequence based estimator of infection multiplicity (estMOI, Assefa *et al.* 2014) and a bi-allelic SNP barcode estimator of infection multiplicity (COIL, Galinsky *et al.* 2015). For the running of the software estMOI, a custom perl based script was written which iterated over all of the Guinean isolates sequence alignment files in BAM format. For the implementation of COIL, a set of bi-allelic SNP genotype barcodes were generated from the Guinean whole genome sequence data for 96 randomly selected SNPs (Appendix 3) that had passed all filtering procedures which were calculated to have a population minor allele frequency of greater than 20%. These barcodes were subsequently uploaded to the COIL programme available at http://www.broadinstitute.org/infect/malaria/coil/.

3.3 Results

3.3.1 Isolate F_{WS} indices within the Republic of Guinea

Genome-wide SNP data for each of the isolates were used to generate within-host F_{WS} fixation indices, for comparison to the microsatellite profiling of the same isolates. Using the Illumina short-read sequence reads to estimate relative frequencies of SNP alleles within each isolate, the within-isolate diversity (H_W) was derived, and the gradient of H_W/H_S (where H_S is the diversity in the entire local population sample) was calculated within each minor allele

frequency bin to derive the F_{WS} index for each isolate (Figure 3.3A). Over all 95 clinical isolates, the mean genome-wide F_{WS} index was 0.79, with a range of 0.16 to 1.00 (Appendix 4). Of these, 50 Isolates (52.6%) had an F_{WS} index approaching 1 (values of >0.95) indicating that they each predominantly consist of a single haploid genome sequence (Figure 3.3B), with any additional genotypes being rare, or closely related to the predominant genotype. Conversely, 45 of the clinical isolates (47.4%) had an F_{WS} index of <0.95 and were thus clearly genotypically mixed infections. No significant correlation was seen between average genome sequence read mapping depth and F_{WS} score (Pearson's r = 0.02), indicating that there was no bias in estimating within-host diversity due to sequence coverage (Figure 3.3C).

When viewing within-host heterozygosity of all SNPs for each isolate individually (Figure 3.4), all 95 isolates had individual SNPs that were biallelic, with the number of heterozygous SNPs within each host ranging from a minimum of 49 SNPs (0.1% of all 50,082 within-population SNPs) within isolate PA0149 to a maximum of 6762 SNPs (13.5% of all within-population SNPs), with a mean number of 1995 polymorphic SNPs found within the Guinea population, which is to be expected due to the vast majority of population SNPs being rare within the population and thus unlikely to be found to be heterozygous within an individual isolate.

3.3.2 Identifying SNPs for assay of within-host diversity with a small number of loci

Calculation of F_{WS} indices from genome-wide data may not always be possible for clinical samples, due to the absence of genome-wide SNP data. In order to

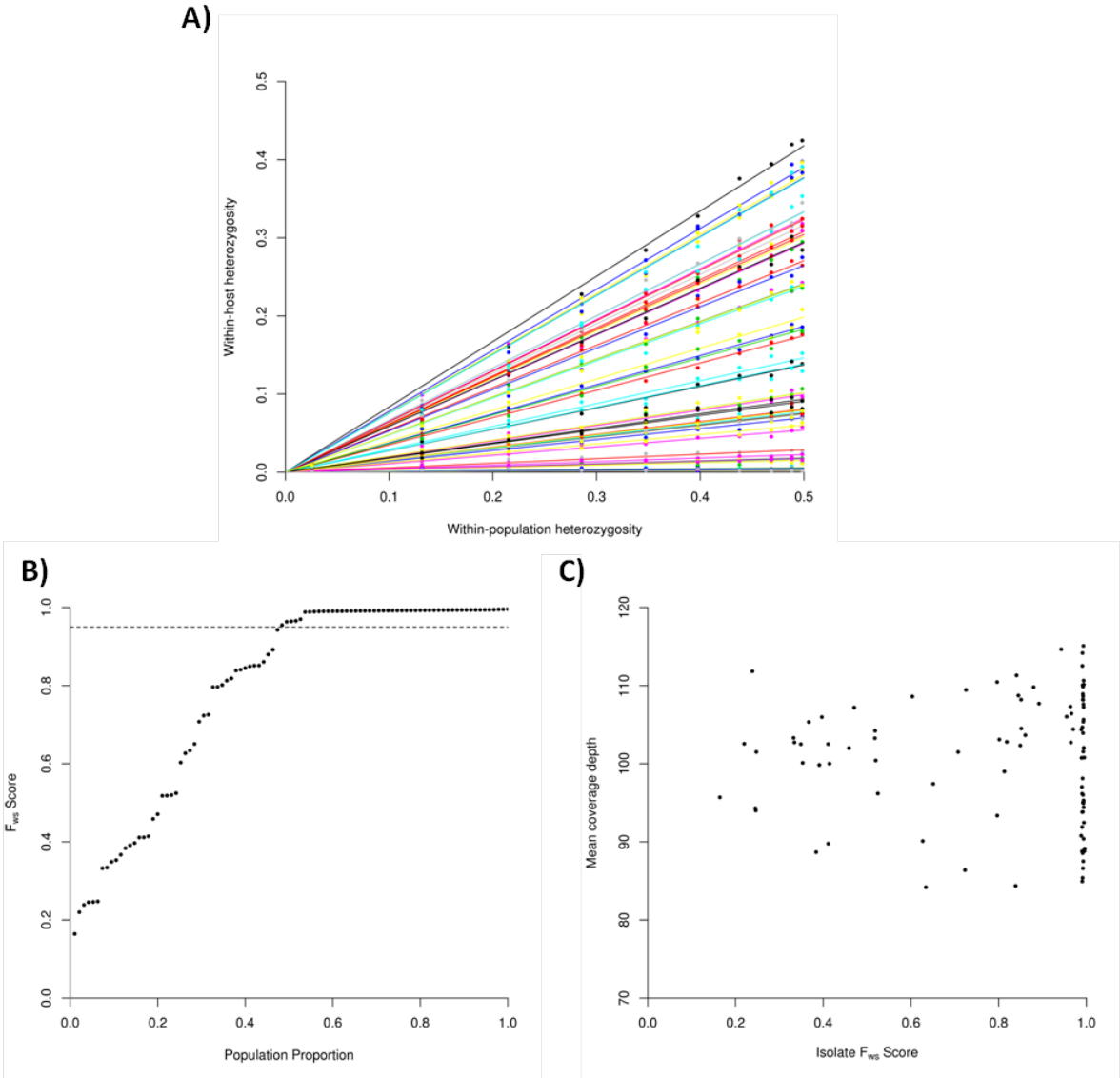


Figure 3.3 Genome-wide computation of Guinean isolate F_{WS} indices. A) The linear relationship between within-host and within-population heterozygosity for all 95 isolates used to calculate F_{WS}, where dots representing the mean values for each corresponding population minor allele frequency bin. The steeper the gradient of the slope H_w/H_s, represents an isolate with a higher level of within-host diversity. B) Cumulative density plot of isolate F_{WS} scores, with the horizontal line on the y-axis at 0.95, above which isolates were defined as predominantly single-genotype infections. C) No correlation is seen between mean coverage depth and isolate F_{WS}, providing support for the quality filtering procedures implemented (Pearson's r = 0.02).

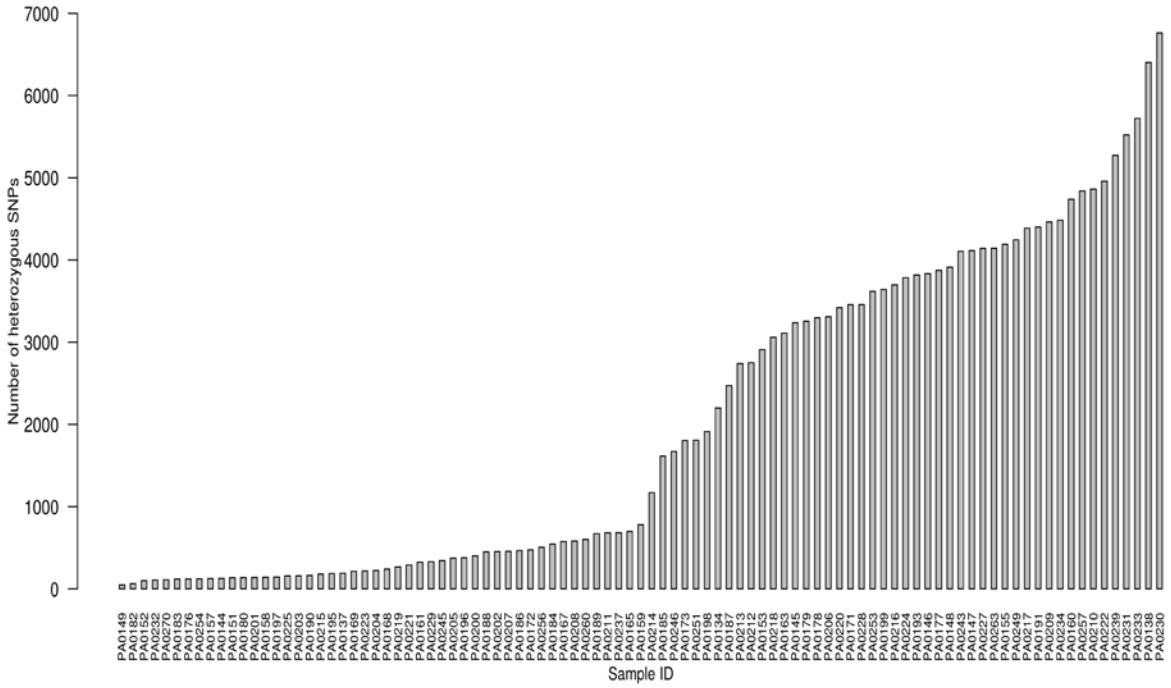


Figure 3.4 Frequency of within-host polymorphisms of the Guinean population. Whilst 43.1% of isolates (n=41) had an F_{WS} score of >=0.99, 100% (n=95) of the Guinean population possessed individual biallelic SNPs. As expected there was a significantly negative correlation between the number of within-host heterozygous SNPs and the isolate F_{WS} score (Pearson's r = -0.91, p-value <0.001)

investigate the potential use of small numbers of SNPs, data from up to 20 randomly chosen SNPs were used to estimate F_{WS} indices and correlations with values derived from genome-wide SNP data (Figure 3.5) were determined. By selecting from all SNPs, regardless of allele frequency, correlations with genome-wide estimations were modest (for 20 SNPs Pearson's r<0.71). However, by selecting SNPS with minor allele frequencies greater than a specified threshold (either <5%, 10%, 20%, 30% or 40%), much higher correlations were seen. Correlation for each selection began to plateau above ~10 SNPs (Figure 3.5). Randomly choosing ten or more SNPs with minor allele frequencies of greater than 10% gave correlations of r > 0.9 with the genome-wide *F*ws indices.

3.3.3 Concordance of F_{WS} with other proposed measures of within-host diversity

Microsatellite typing allows an assessment of the proportion of loci that have multiple alleles within an infection, as well as a minimum estimate of the number of different parasite genotypes, as defined by the highest number of alleles detected at any locus. As expected, there was a highly significant negative correlation between the F_{WS} fixation index derived from the SNP analysis and the number of microsatellite loci with multiple alleles detected within each infection (Pearson's r = -0.88, P < 0.001) (Figure 3.6A). There was also a highly significant negative correlation between the isolate F_{WS} index and the number of different genotypes within each infection detected by microsatellite typing (Pearson's r= -0.80, P< 0.001). However, many of the infections that appeared to contain a single dominant genotype (on the basis of an F_{WS} index of >0.95) were seen to contain multiple microsatellite genotypes, and some of these had multiple alleles at several of the loci genotyped (Appendix 4).

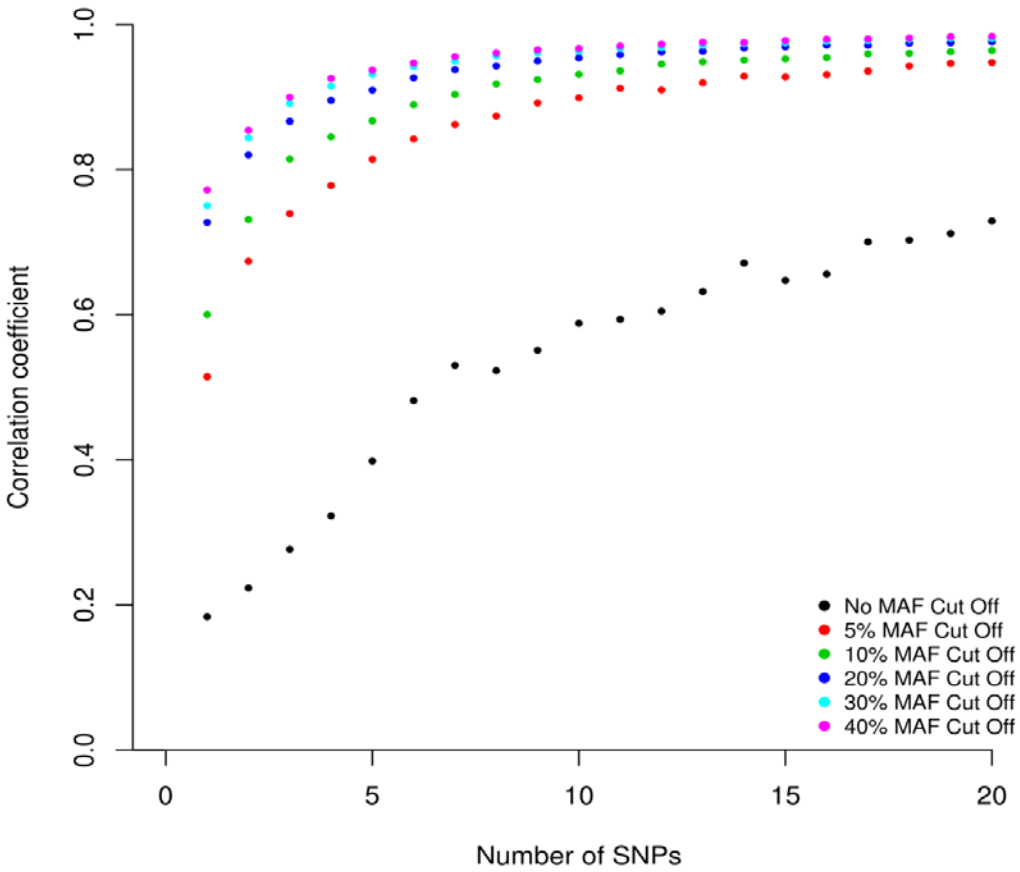


Figure 3.5 Correlation coefficients of isolate F_{WS} indices derived from randomly sampled small numbers of SNPs compared with indices using all genome-wide SNPs analysed in the local population. The effect of using SNPs with allele frequencies above different levels is shown. Within the overall SNP set sampled from, 7684 SNPs have a MAF of >5%, 5004 SNPs have a MAF of >10%, 2854 SNPs have MAF of >20%; 1719 SNPs have MAF of >30% and 833 SNPs have MAF >40%. Replications of 100 random samples of different numbers of SNPs (between 1 and 20) were performed and the mean correlation coefficient for each is plotted.

Moreover, there was strong concordance between the isolate F_{ws} indices and the corresponding outputs derived from COIL (Pearson's r= -0.93, p-value<0.001). COIL can determine the presence of up to a maximum of five genotypes from a 96 SNP barcode as implemented here and defines its estimation of infection complexity as the number of individual genotypes present. The mean number of clones per infection estimated by COIL was 2.01, with 51.6% (n=49) of the isolates estimated to contain a single genotype and 10.5% (n=10) of the isolates were estimated to contain five different genotypes.

Alternatively, the estMOI software is designed to use deep sequence data to infer minimum numbers of distinct genotypes found within an infection by calculating the number of haplotypes derived across SNPs from sequencing read pairs (Assefa *et al.* 2014). However, despite iterations using different input conditions for the minimum haplotype frequency to be considered, the output estimated number of clones derived from the estMOI software was not concordant with the other measures implemented. Across both iterations used, not a single isolate was estimated to have greater than or equal to four clones present within an infection. In addition, the range of estimated clone numbers produced by the software was extremely limited, with greater than 80% of all isolates apparently estimated to have two clones present. Furthermore, when a minimum haplotype frequency of two was used, only one isolate was estimated to consist of a single clone, whilst when a minimum haplotype frequency of three was used, no isolates were found to have >=3 estimated clones. When these results are considered alongside the estimates of within-host diversity provided by the microsatellite typing, the F_{WS} index and COIL, the outputs of estMOI did not concur.

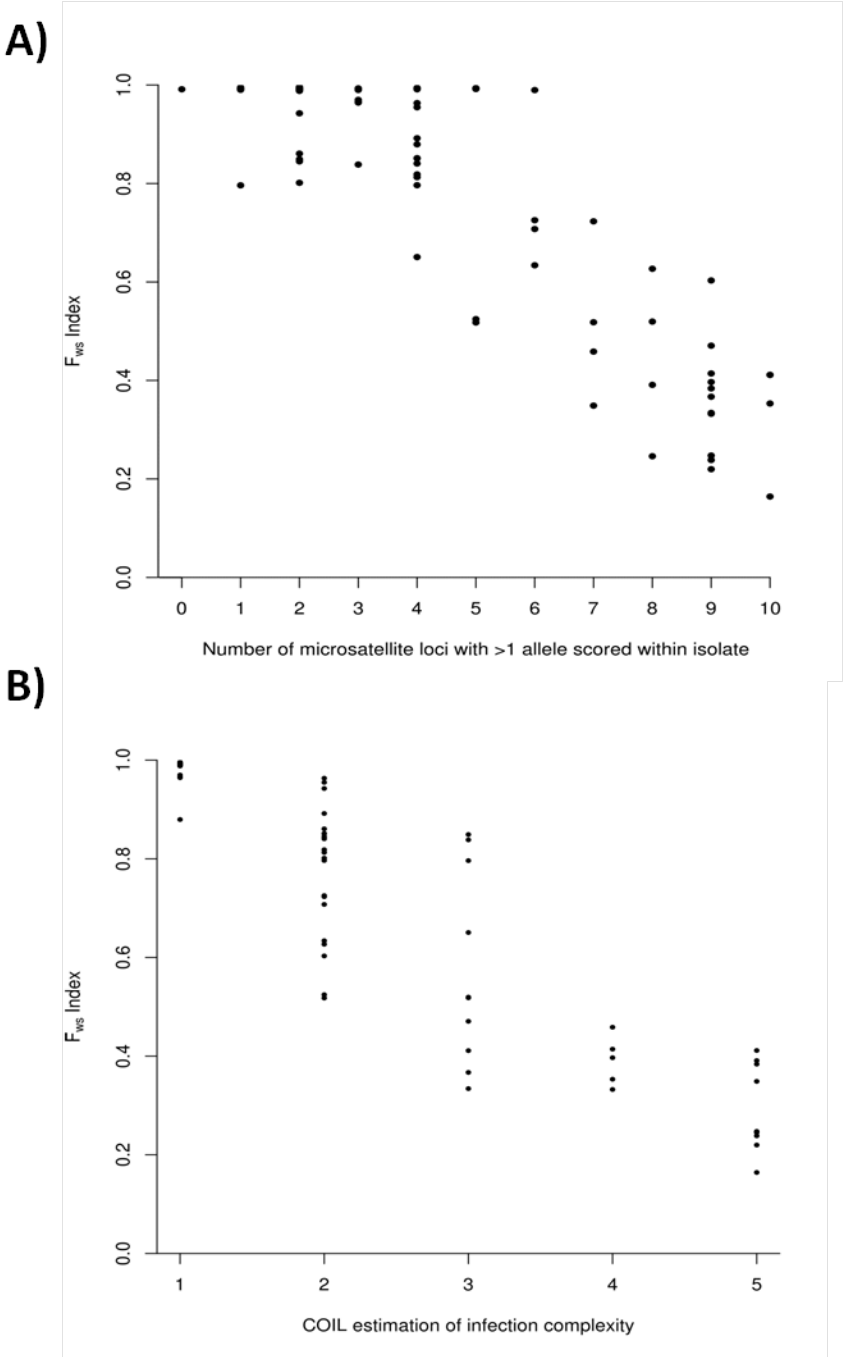


Figure 3.6 Concordance of Guinea isolate F_{WS} scores with alternative measures of within-host diversity. A) Concordance between microsatellite genotyping and F_{WS} indices derived from genome-wide SNP data for *P. falciparum* clinical isolates from Guinea (N = 93 with complete 10-locus genotype profiles)**.** A highly negative relationship (Pearson's coefficient = -0.88, P < 0.001) was seen between isolate F_{WS} index and the number of microsatellite loci with >1 allele detected within the isolate. B) Concordance between COIL-based estimation of multiplicity and F_{WS} indices.

	Minimum haplotype frequency implemented	
Estimated number of clones	**1**	**2**
1	1	19
2	77	76
3	17	0

Table 3.1 Estimations of multiplicity of infection as calculated by estMOI software. Two separate minimum haplotype frequencies were implemented, but neither iteration showed concordance with the microsatellite profiling or F_{WS} scores derived.

3.4 Discussion

Within this chapter, genome-wide surveying of SNP-based allele frequencies has been used to characterise *P. falciparum* diversity within clinical infections from the Republic of Guinea. The results are indicative of a high degree of transmission intensity in the Guinean population studied, and are consistent with previous microsatellite data from other samples taken locally (Mobegi *et al.* 2012). The genome-wide SNP data here indicate more than half of all infections to each be composed predominantly of single genotypes (Figure 3.3), which is in agreement with previously published calculations of isolate F_{WS} indices from West African populations (Auburn *et al.* 2012).

Random sampling of the genome-wide SNP data shows that the within-isolate F_{ws} fixation indices may be estimated from modest numbers of SNPs, and correlated with the indices derived from genome-wide data. Therefore, to estimate genotypic mixedness of isolates without whole genome sequencing, it may be feasible to quantitate alleles of between 10 and 20 SNPs with other genotyping tools (Daniels *et al.* 2008), particularly focusing on SNPs with high overall minor allele frequencies. Due to the MalariaGEN catalogue of SNPs (MalariaGEN 2016), it is now possible to select such genotyping targets. Additionally, it is preferable that any such SNPs should be neutral across multiple populations, so that estimates are not biased by selection acting on the parasite.

Furthermore, the F_{WS} index has been compared against three other proposed alternative measures of within-host diversity within this study. Microsatellite typing has been well established to allow for the sensitive detection of distinct parasite genotypes present at low proportions within an infection, although

cloning or single-cell analysis of isolates would be needed to estimate the degree of relatedness among the different parasite. As expected within the Guinean population (Mobegi *et al.* 2012), the microsatellite genotyping was able to detect multiple clones in all but one of the 50 isolates which were found to consist predominantly of a single genotype by F_{WS}, in agreement with previously reported PCR-based methods (Auburn *et al.* 2012). These comparisons highlight that estimates of genotypic complexity of malaria parasite infections using very different methods give correlated and complementary information, apart from those obtained by the estMOI software. With a wide range of methods now available, studies can choose genotyping and analytical approaches to suit personalised goals, recognising relative advantages of each in relation to the costs and available resources.

The terminology used to describe the within-host diversity of *P. falciparum* infections is quite varied and can include multiplicity of infection (MOI), infection diversity or infection complexity. The term 'single-clone' infection is regularly used to describe infections which have, in most instances, been determined as monoclonal from a small number of polymorphic loci using PCR based typing methods. Understanding processes affecting different parasite genotypes within an infection could offer insight into mechanisms that are clinically relevant. It is possible that some clones within an infection have low proportions due to intrinsic growth rate differences, competitive suppression by other *P. falciparum* genotypes, or host selective pressures. Analyses of parasites in culture could examine whether particular *P. falciparum* clones outcompete others in terms of asexual replication, or whether relative growth rates are dependent on particular co-cultured combinations. Further dissection of the patterns of parasite genotypic diversity in clinical isolates, and possible

interactions between genotypes, is a priority for further dissection of the within-host diversity of *P. falciparum* parasites.

Chapter 4. Assessing evidence of non-neutral allele-frequency based polymorphism at the within-host level across the *P. falciparum* genome within West Africa

4.1 Introduction

F_{ST} based genome scans between *P. falciparum* populations have been applied to investigate loci associated with a sub-population specific artemisinin slow clearance phenotype (Cheeseman *et al.* 2012) and to more generally scan for local selection within particular endemic areas (Mobegi *et al.* 2014; Duffy *et al.* 2015). In an analogous manner to F_{ST} based genome scans, within this chapter the F_{WS} index (Manske *et al.* 2012) is applied using a genome-wide scan to identify loci which exhibit deviant patterns of within-host heterozygosity (H_W) relative to their local level of within-population heterozygosity (H_S).

For some other pathogenic organisms, studies to identify loci under within-host selection pressures are classically performed using time course experiments within a host (Wright *et al.* 2011; Golubchik *et al.* 2013; Illingworth *et al.* 2014, Didelot *et al.* 2016), but longitudinal sampling of human *P. falciparum* infections is only possible in asymptomatic individuals (Farnert *et al.* 1997). The genomic variation within a *P. falciparum* infection is determined by two main processes. The initial genomic diversity is representative of the haploid sporozoite progeny that are produced following the process of meiosis within the diploid zygotic stages of the parasites life cycle (Sinden & Hartley 1985). If this occurs between heterozygous haploid gametes, then meiotic recombination will take place, resulting in the production of numerous recombinant haploid progeny that have inherited different genomic regions from the heterozygous parents (section 1.4). Four studies producing experimental crosses in *P. falciparum* have been reported to date, three having involved inoculation of progeny into a

chimpanzee host (Walliker *et al.* 1987; Wellems *et al.* 1990; Hayton *et al.* 2008; Ranford-Cartwright & Mwangi 2012) and one with the use of human-liver chimeric immunocompromised mice (Vaughan *et al.* 2015). These crosses have each reported the generation of tens of distinct recombinant progeny, highlighting that in natural parasite populations where endemicity is high and the population is outbreeding (Conway *et al.* 1999; Manske *et al.* 2012), there is potential for high levels of genomic diversity within the initial inoculation of an infection.

In addition to this genomic heterogeneity that can already be present within infections, *de novo* mutations will also occur. Estimation of the mutation rate of *P. falciparum* during the intraerythrocytic cycle on the 3D7 laboratory clone has been put at 1.7×10^{-9} per base pair per generation. When adjusted for the estimated proportion of deleterious mutations produced, this means that after only 25 asexual rounds of replication, the daughter parasite would be expected to have accumulated a non-synonymous mutation from the original asexual parent (Bopp *et al.* 2013). When this is extended further to account for the mean parasite load found in an acutely infected person of 7.0×10^{11} (Dondorp *et al.* 2005) and the possibility of infections persisting for several years (Sama *et al.* 2004; Ashley & White 2014), there is significant potential for the generation of novel allelic variants during the intraerythrocytic cycle within the human host. Finally, particularly within regions of high endemicity, superinfection of unrelated parasite genotypes occurs and can also contribute substantially to the genomic diversity found within a host (Nkhoma *et al.* 2012).

Within this chapter, I test the hypothesis that some of the genomic diversity present within an infection may be subject to directional selective forces at the

within-host level and assess whether particular loci under these processes can be detected using an F_{WS} based scan approach. Directional selection acting on genes will result in either purifying or positive selection that will lead to the fixation of alleles that improve the parasites ability to grow within the host. Thus, using an F_{WS} genome based scan I set out to identify loci with consistently high levels of within-host homozygosity relative to their within-population heterozygosity (Figure 4.1) analysing data collected from four different populations within West Africa. Underlying potential causes of directional within-host selective forces that would result in increased within-host homozygosity at particular loci (Hastings 1996) could include drug selection of parasites possessing resistance mutations (Robinson *et al.* 2011), specific host receptor polymorphisms that prevent establishment or growth of certain recombinant progeny or alleles, or allele-specific immune selection (Conway *et al.* 2000; Osier *et al.* 2010).

4.2 Materials and Methods

Publicly available whole-genome sequence data from population samples of *P. falciparum* clinical isolates of four West African locations (section 2.1.1) was utilised, which consisted of an initial set of 131 isolates from the Republic of Guinea, 79 isolates from The Gambia, 68 isolates from Central Ghana and 48 isolates from Northern Ghana. All isolates underwent quality control filtering (section 2.1.2) and isolate F_{WS} indices were calculated accordingly (section 2.1.4). The within-population heterozygosity calculations and data filtering procedures were performed within each of the four populations independently. Two separate strategies were then used to apply a scan of F_{WS} across the genome, with all analyses being performed using custom scripts within R.

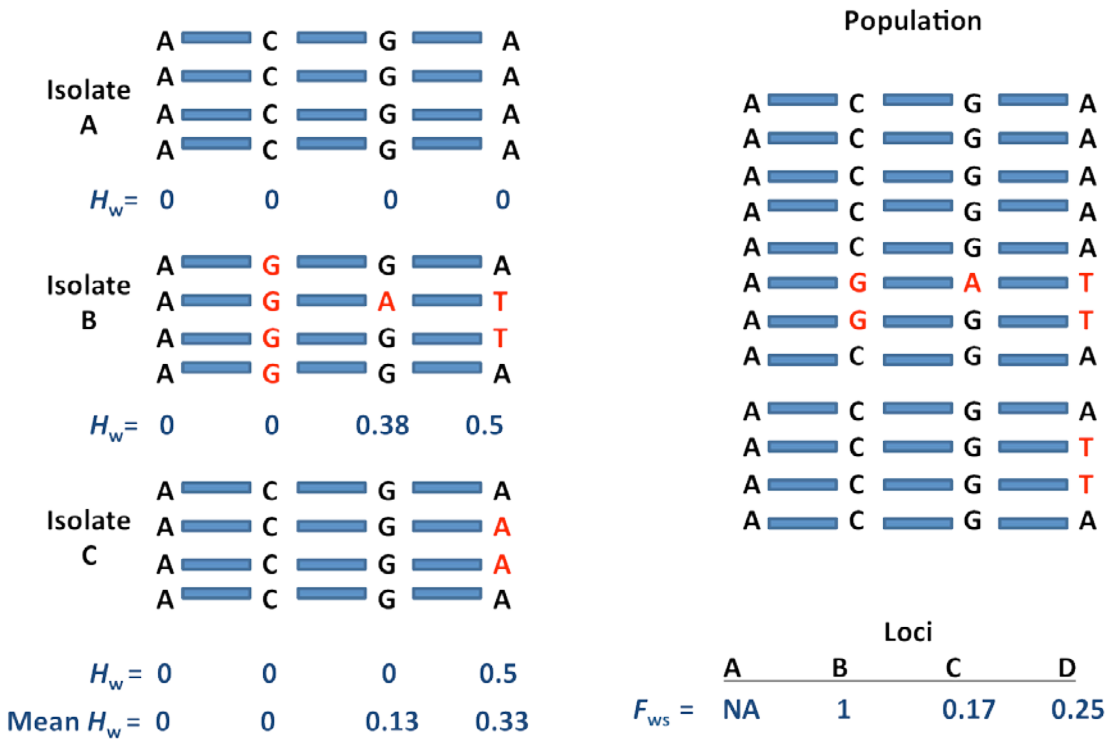


Figure 4.1 Schematic diagram of the within-host heterozygosity of three infections and the corresponding within-population heterozygosity for calculation of the F_{WS} metric at four example nucleotide positions. At locus A, no polymorphism is seen within the population. At locus B, all three isolates are homozygous, but across the whole population there is heterozygosity. At loci C and D, different levels of within-host and within-population heterozygosity are derived. Thus, potential F_{WS} values across the population at a particular locus has a potential range from 0 to 1, where 0 indicates that the average within-host heterozygosity is equal to the within-population heterozygosity, whilst a score of 1 indicates that the locus is homozygous at the within-host level. Within this figure, each isolate only has four sequences sampled per locus per isolate, whereas the deep sequence data filter used in this study had a minimum coverage of 20 reads.

Firstly, a gene-based genome-wide scan of F_{WS} was employed, where mean H_W and H_S were calculated across all SNPs within each gene. Thus, genic F_{WS} values were effectively calculated as a 'ratio of averages' (Bhatia *et al.* 2013; Weir & Cockerham 1984), in which the numerator H_W and H_S are averaged independently. Following this, a minimum cut off of 10 SNPs was required to produce a gene F_{WS} score. Genes exhibiting putative F_{WS} values with increased levels of within-host homozygosity were defined as those where the value was found to be >2 standard deviations away from the mean score across all genes within that population. The calculation of gene scores also facilitated the use of GO-term enrichment analysis to examine the potential presence of significantly enriched gene functions across the different populations using Fisher's exact test and the weighted algorithm within the TopGO package in R (Alexa *et al.* 2006).

The second methodology which was implemented was a SNP-based approach designed to attempt to identify individual non-synonymous mutations that could offer a within-host selective advantage. For this purpose, 95% confidence intervals were generated through 100 bootstrapped replications of mean non-synonymous F_{WS} SNP scores, where H_W and H_S were calculated individually for each SNP. Initially, the Guinean population dataset was used for a pilot analysis to determine the relationship between population minor allele frequency and estimations of an F_{WS} SNP score, prior to setting a population minor allele frequency cut off of 5% across all populations. Following this, only SNPs encoding non-synonymous mutations present within all four West African populations were retained, leaving a total of 3906 SNPs. 95% confidence intervals of bootstrapped mean SNP-based F_{WS} scores were then calculated across each population individually and SNPs were considered significantly

more homozygous at the within-host level than expected from the overall within-population data when calculated to have an F_{WS} value with a 95% confidence lower limit boundary higher than the calculated mean across the population. This cut off for individual loci followed that previously used for analyses of potential non-neutral oocyst inbreeding coefficients (Anthony *et al.* 2007).

4.3 Results

4.3.1 Population comparison of isolate F_{WS} indices across West Africa

From an original sample set of 326 isolates across the four locally sampled populations, 90.2% of isolates (n=294) remained following filtering procedures, with respective SNP set sizes of 49740, 60148, 69291 and 86403 within The Gambia, Northern Ghana, Central Ghana and the Republic of Guinea respectively (Table 4.1). Within each of the four populations, the minor allele frequency distribution recorded was highly skewed towards low frequency variants, with 79.2% of Gambian SNPs and 90.8% of Guinean SNPs found to possess a population minor allele frequency of 5% or less (Figure 4.2), whilst proportions of the two sites in Ghana were intermediate between these. This high frequency of relatively rare minor allele frequency variants, including a higher proportion of Gambian SNPs with a minor allele frequency of greater than 5% compared to the other sites is in agreement with previous population genome analyses of these datasets (Amambua-Ngwa *et al.* 2012; Mobegi *et al.* 2014; Duffy *et al.* 2015).

When isolate F_{WS} scores were compared pairwise between the populations (Table 4.1), only the Gambian population was found to have a significantly different population mean isolate score from all of the other three populations

	Republic of Guinea (N'Zérékoré)	The Gambia (Greater Banjul)	Northern Ghana (Navrongo)	Central Ghana (Kintampo)
Number of isolates	111	73	47	63
Number of SNPs	86403	49740	60148	69291
Mean F_{WS} score	0.81	0.93	0.77	0.81
Isolate F_{WS} range	0.19-1.00	0.33-1.00	0.16-1.00	0.35-1.00
Proportion of scores >0.95	54.1% (n=60)	76.7% (n=56)	43.8% (n=21)	50.8% (n=32)
Number of SNPs with >5% MAF	8035	10366	8302	8961
Number of genes with >10 SNPs	2492	1655	1912	2172

Table 4.1 Summary statistics of quality filtered SNP sets and isolate F_{WS} indices from four West African populations. F_{WS} scores were calculated across all quality filtered SNPs within each population individually, thus for the Guinean population this was performed across 86403 SNPs.

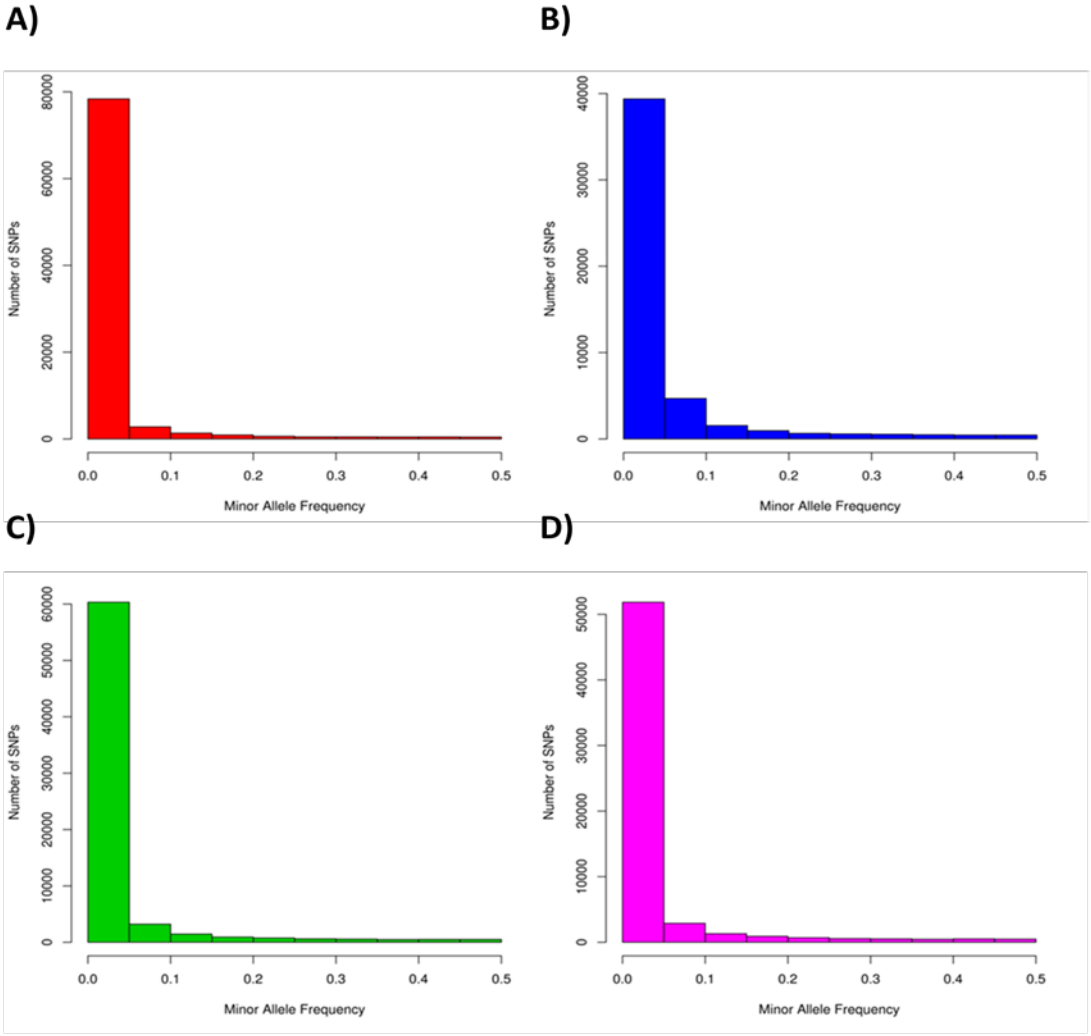


Figure 4.2. Population minor allele frequency distributions for all four West African population SNP sets. All SNPs within each panel (A= Republic of Guinea, B= The Gambia, C= Central Ghana and D= Northern Ghana) have been placed into equally sized minor allele frequency intervals, with steps and bin sizes of 0.05.

(Wilcoxon rank sum test with Bonferroni correction, Gambia vs Guinea: p-value=0.008, Gambia vs Northern Ghana: p-value=0.009, Gambia vs Central Ghana: p-value= 0.006, Guinea vs Northern Ghana: p-value=1, Guinea vs Central Ghana: p-value= 1, Northern Ghana vs Central Ghana: p-value= 1). Isolate F_{WS} scores averaged across all SNPs in an isolate ranged from 0.16 to 1.00. In addition, as previously noted for the Guinean population (Murray *et al.* 2016, Figure 3.3) , a considerable proportion of isolates in all four populations were found to have sequence data that reflected predominantly single-genotype infections, with between 43.8% of isolates from Navrongo and 76.7% of Gambian isolates having a F_{WS} value of >0.95 (Figure 4.3, Appendix 5).

Visualisation of within-host heterozygosity (H_W) across the genome for three isolates of varying F_{WS} scores illustrates the markedly different levels of heterozygosity at the within-host level (Figure 4.4). The isolate within panel A (Figure 4.4), with an F_{WS} value of 0.97 may be defined as consisting predominantly of a single-genotype infection, but the modality of the distribution of its H_W values were found to be at least bimodal (Hartigan's dip test, p-value=0.003). Unsurprisingly, this was also the case for isolate B with a F_{WS} value of 0.33 (Hartigan's dip test, p-value< 2.2e-16), but the H_W distribution of isolate C with an F_{WS} value of 1.00, was found to have a unimodal distribution (Hartigan's dip test, p-value=1). This suggests that even within apparently highly clonal isolates across the genome, a small number of loci have apparent within-host heterozygosity, although it is possible that this may reflect artefacts due to sequencing or mapping errors at a minority of SNPs.

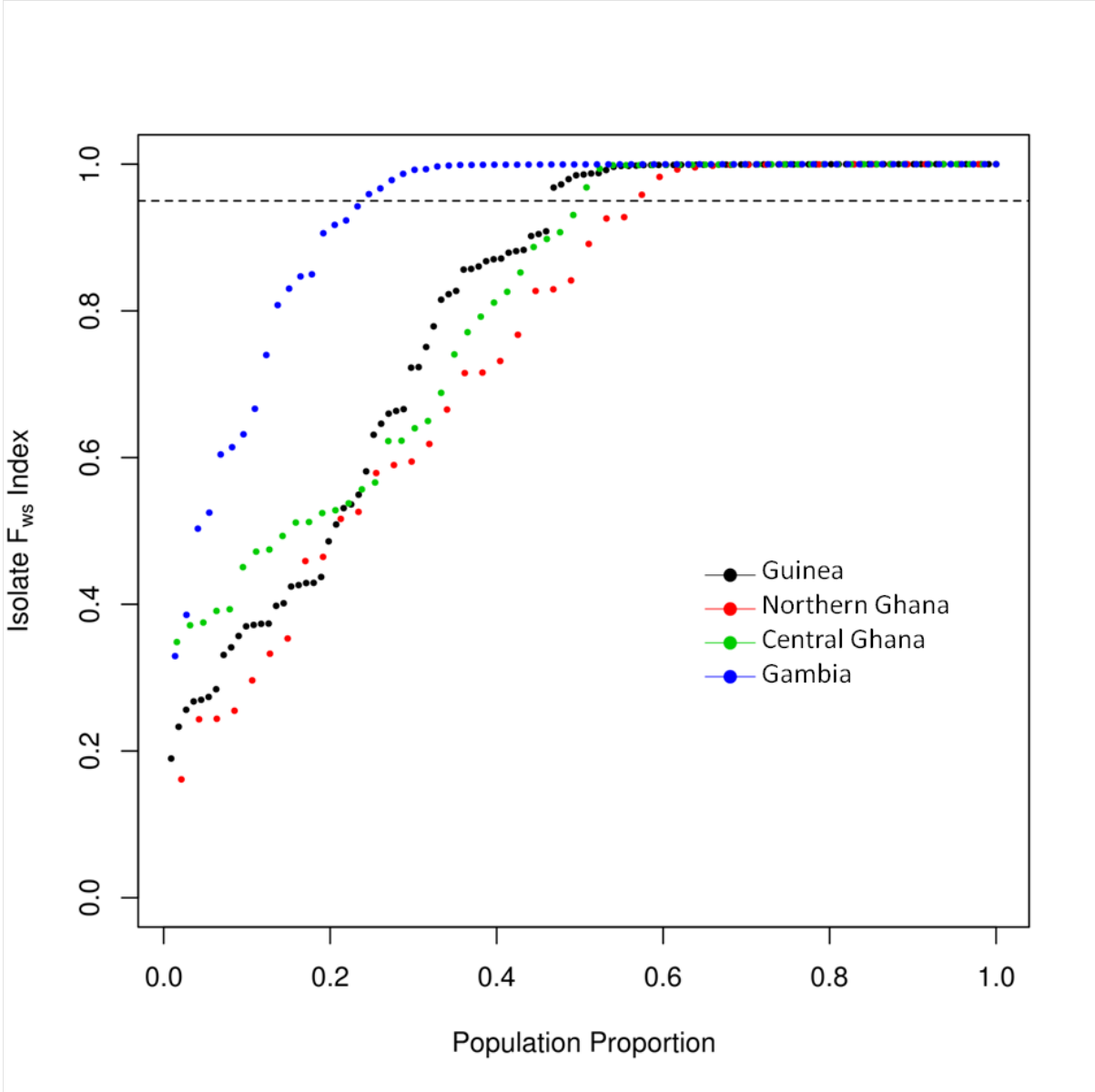


Figure 4.3 Cumulative density plot of isolate F_{WS} indices from four West African populations. The horizontal dashed line on the y-axis represents an F_{WS} score of 0.95, above which isolates were defined as consisting predominantly of a single genotype. Scores were calculated through linear regression of the relationship H_W/H_S across minor allele frequency intervals (section 2.1.3, Figure 3.3A).

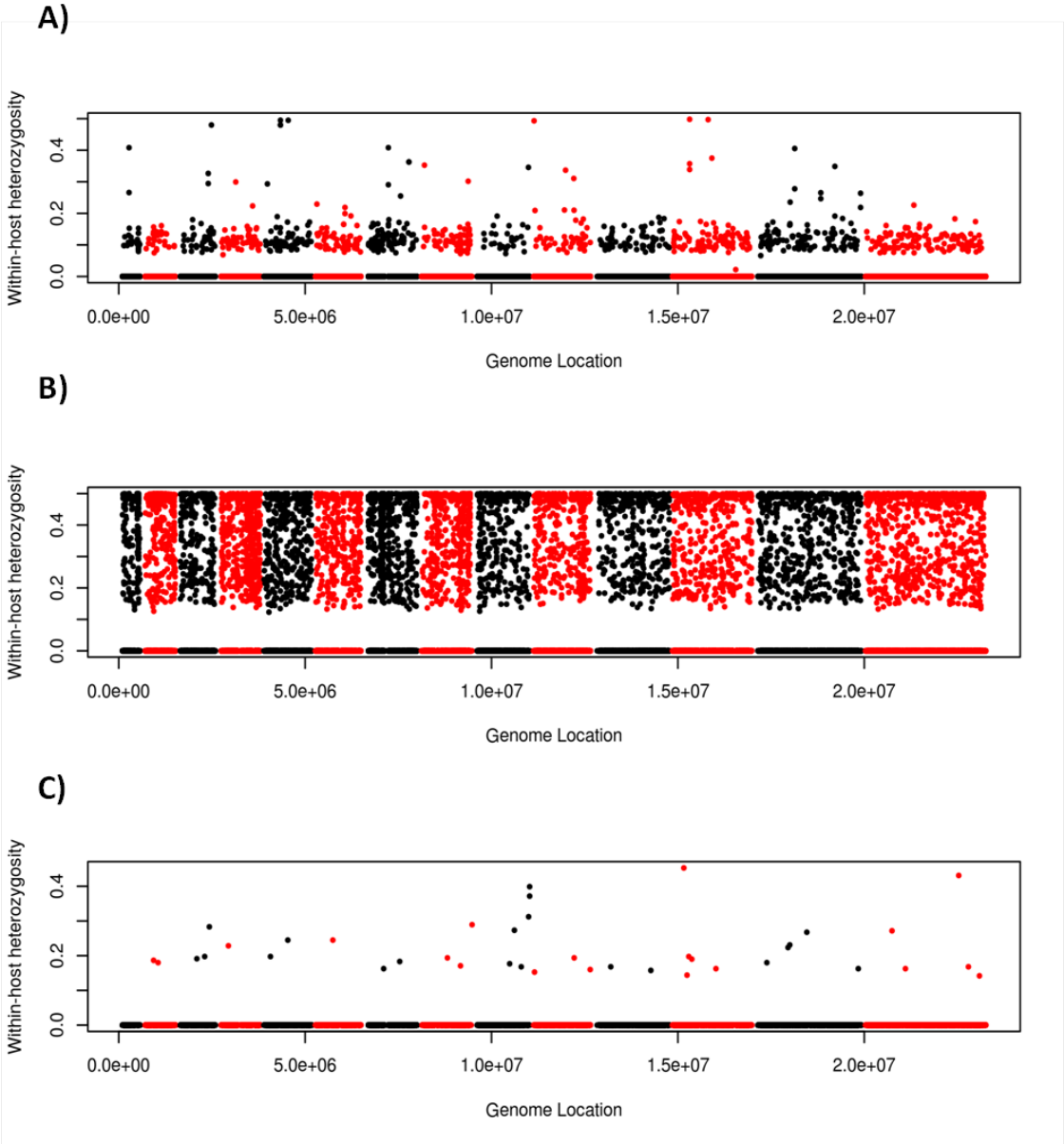


Figure 4.4 Genome-wide within-host heterozygosity of three example Guinean isolates. Three isolates are shown with varying isolate F_{WS} scores of A) 0.97, B) 0.33 and C) 1.00 from top to bottom respectively, where all dots represent individual SNPs and alternating colours represent the boundaries of the fourteen chromosomes of *P. falciparum*.

4.3.2 A gene-based scan of F_{WS} across the genome

For identification of genes with outlying F_{WS} scores relative to the rest of the population, genes with scores that were found to be greater than two standard deviations above the mean population gene score of each population were identified (Figure 4.5 and Figure 4.6; Guinea mean F_{WS} gene score= 0.828, SD= 0.0345, outliers > 0.897; Gambia mean F_{WS} gene score= 0.919, SD= For identification of genes with outlying F_{WS} scores relative to the rest of the population, genes with scores that were found to be greater than two standard 0.0274, outliers > 0.974, Central Ghana mean F_{WS} gene score= 0.804, SD= 0.0468, outliers > 0.898; Northern Ghana mean F_{WS} gene score= 0.766, SD= 0.0521, outliers >0.870). This procedure left a total of 78, 10, 35 and 37 outlier genes across the Guinean, Gambian, Central and Northern Ghanaian populations respectively.

Across the four populations this identified a total of 154 genes that were found within at least one of these outlier subsets, but only three genes were present within two of the different population outlier subsets (PF3D7_0906600 in Central Ghana and Guinea; PF3D7_1013900 in Central Ghana and Guinea; PF3D7_1369000 in Central Ghana and Northern Ghana). These three genes putatively encode a zinc finger protein, an initiation factor subunit protein and a glycosylphosphatidynlionositol anchor attachment 1 protein respectively. This was inconsistent with an initial hypothesis that certain genes would be consistently under within-host selection. Some genes did not have sufficient numbers of SNPs in all populations to be included in all analyses, but PF3D7_1369000 had enough SNPs within all four populations to be present within the final analyses without always being present within the outlier subsets.

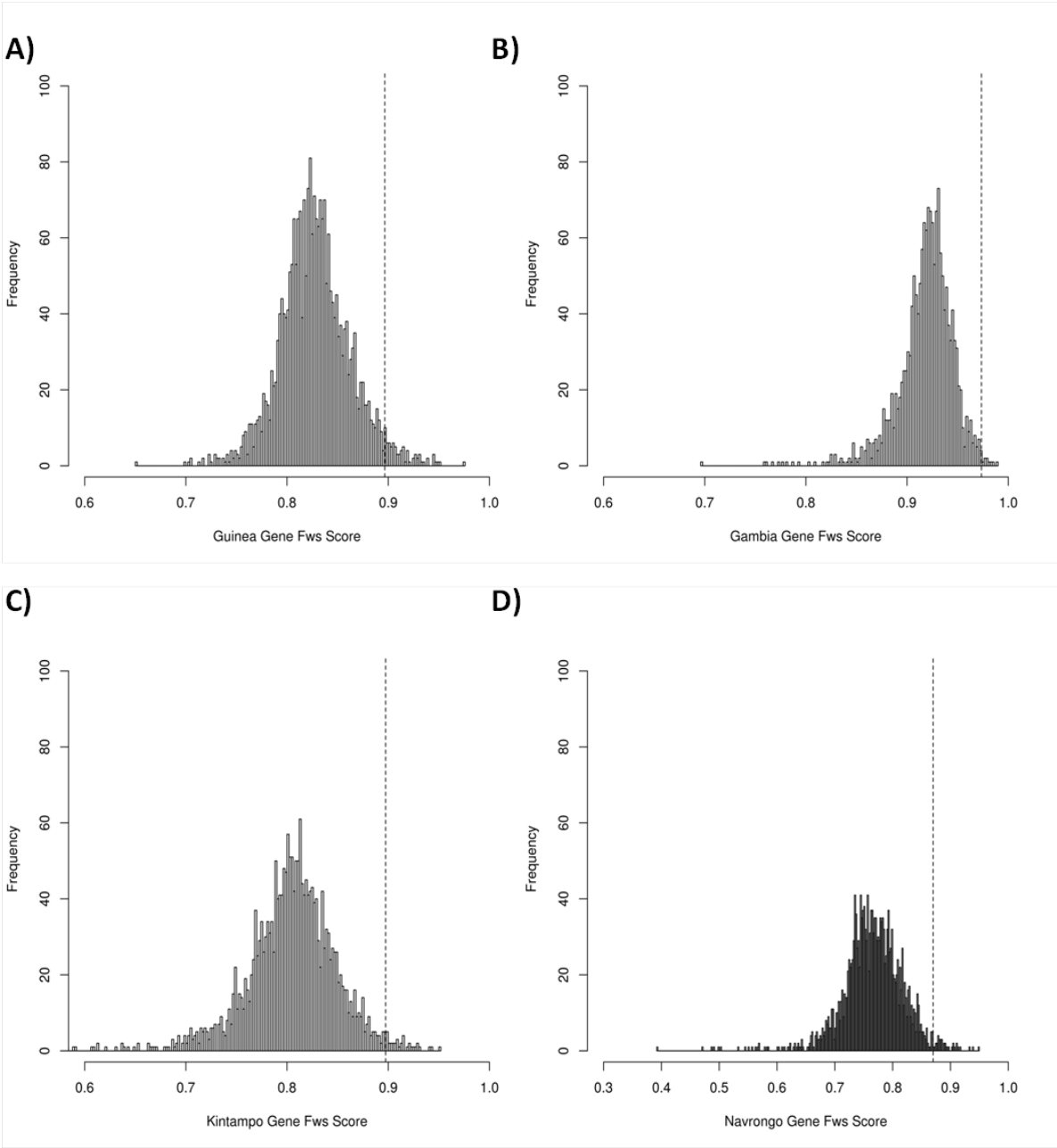


Figure 4.5 Distribution of F_{WS} gene scores within the four West African populations. Within each panel, the dashed vertical line represents the cut off score implemented for the identification of outlier genes of mean plus 2 SD (A) Guinea outliers = >0.897; B) Gambia outliers = >0.974, C) Central Ghana outliers = >0.898; D) Northern Ghana outliers = >0.870).

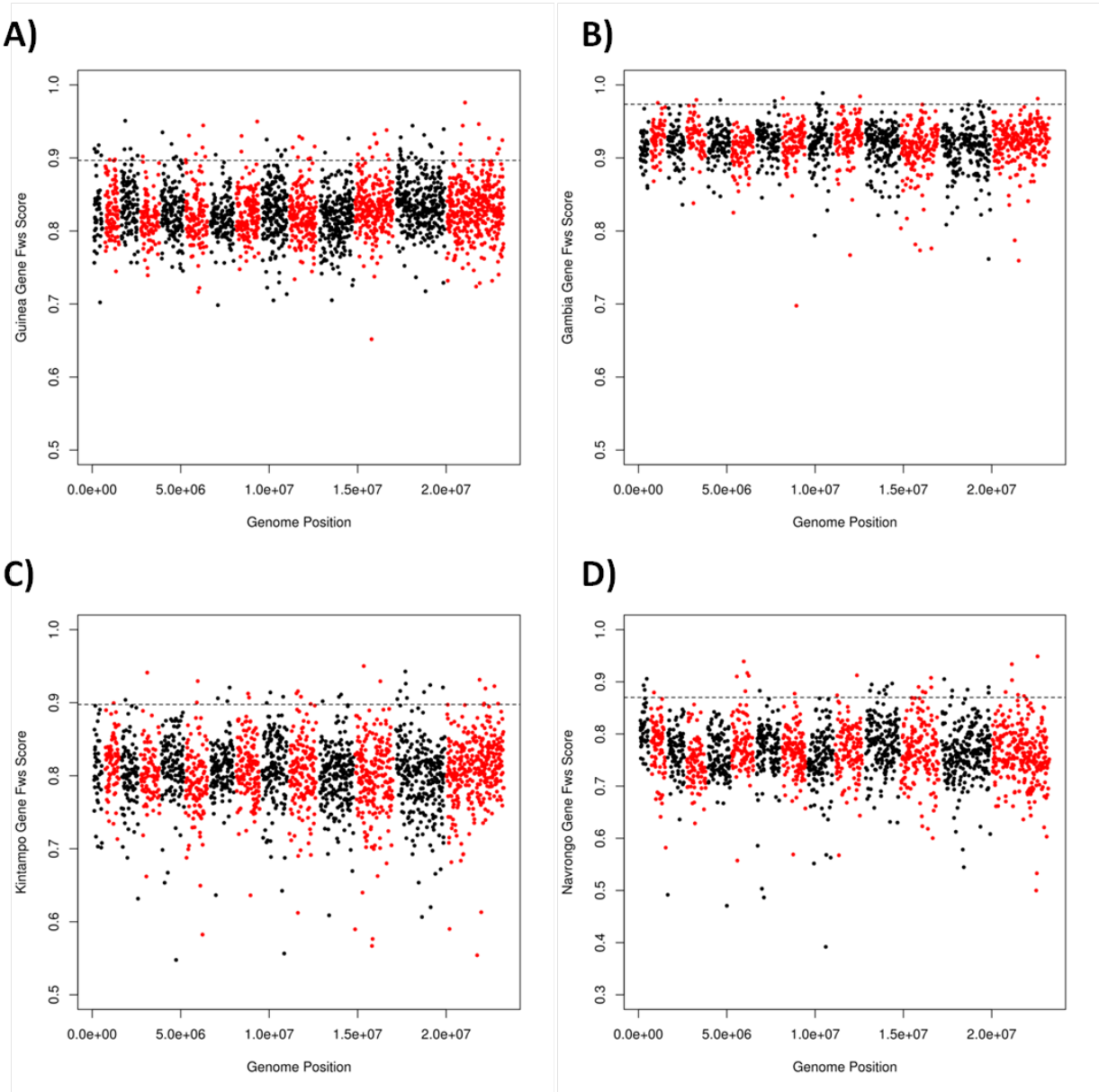


Figure 4.6 Genome-wide plots of gene F_{WS} scores within four West African populations. Within each panel, the dashed horizontal line represents the cut off score implemented for the identification of outlier genes of mean plus 2 SD (A) Guinea outliers = >0.897; B) Gambia outliers = >0.974, C) Central Ghana outliers = >0.898; D) Northern Ghana outliers = >0.870). Each dot represents a gene with >=10 SNPs in that population.

In addition, whilst PF3D7_0906600 and PF3D7_1013900 did not possess enough SNPs to be present within the final gene score analysis for all populations, they did both have 5 or more SNPs within all four populations. Subsequent gene-based calculation of F_{WS} for these two genes within the corresponding populations in which they had not been performed did not produce genic F_{WS} scores that would have been found within the outlying group of scores in any other population (Gambia gene F_{WS} of PF3D7_1013900 = 0.924, Northern Ghana gene F_{WS} of PF3D7_1013900 = 0.770; Gambia gene F_{WS} of PF3D7_0906600 = 0.828, Northern Ghana gene F_{WS} of PF3D7_0906600 = 0.734, Appendices 6 to 9).

Furthermore, pairwise correlation plots between the four West African populations for the remaining set of 1492 genes which had greater than 10 SNPs within all populations showed no evidence of correlation (Figure 4.7, Pearson's r ranged between -0.05 and 0.1 for all six pairwise comparisons), which had been previously been shown to be strongly concordant for other allele-frequency spectra based scans of selection within these populations (Mobegi *et al.*2014; Duffy *et al.* 2015).

To determine whether there were any putative functional properties that were consistently found within the outlier F_{WS} gene scores across the four West African populations, gene ontology (GO) term analysis was applied. This found no GO terms to be consistently enriched across the four West African populations, with only rRNA processing (GO:0006364) found to be significantly enriched (p-value <0.005) within only the Navrongo population (Appendices 10 to 13).

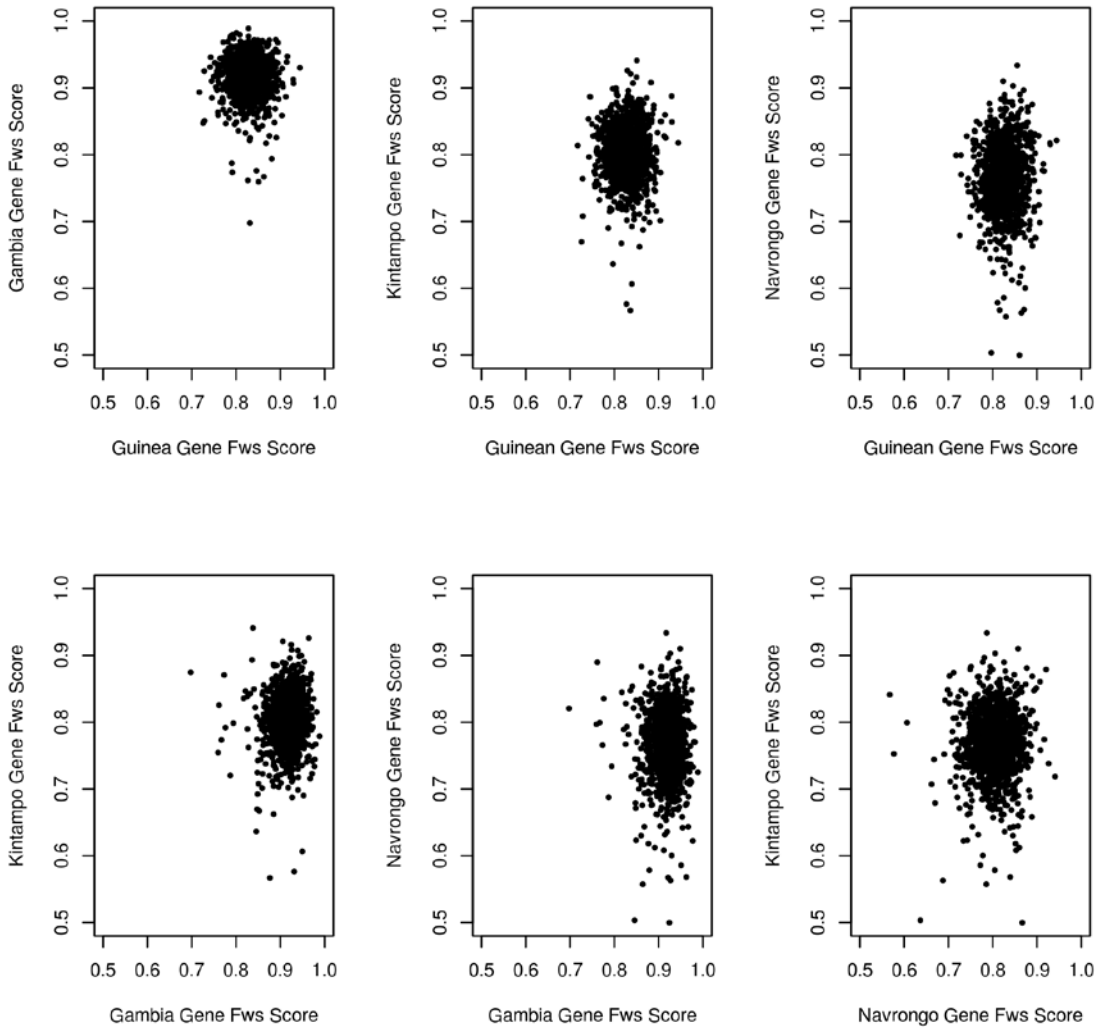


Figure 4.7 Concordance between gene-based calculations of F_{WS} across four West African populations. Pairwise correlations of the gene-based F_{ws} scores for four West African populations, using all 1492 genes for which >=10 SNPs were present within each population.

4.3.3 Bootstrapping of F_{WS} SNP scores across the genome

For the calculation of SNP-based F_{WS} scores, initially the Guinean population set was used to determine the effect of population minor allele frequency on the size of the confidence intervals produced through 100 bootstrap replications. Examination of the bootstrapping of SNPs within the interval consisting of SNPs with a population minor allele frequency of 5% or less indicated the intrinsic variation when estimating a SNP-based F_{WS} within this group (Figure 4.8). For instance, bootstrapping had no effect for SNPs that were either only found to be homozygous at the within-host, but heterozygous within the population (F_{WS} = 1), nor for SNPs that were only found to be present as a within-host minor allele within one isolate and not found within the rest of the population (F_{WS} = <1), as there was no within-host variation in heterozygosity to be able to resample for bootstrapping to produce an alternative F_{WS} score.

Thus, SNPs from this interval were removed prior to bootstrapping across all populations, which produced genome-wide mean SNP F_{WS} scores of 0.802 within Guinea, 0.925 within The Gambia, 0.802 within Kintampo and 0.748 within Navrongo. As with the identification of outliers using the gene-based F_{WS} scan, the lower 95% confidence limits of no bootstrapped SNP F_{WS} scores was found to be greater than the respective genome-wide mean population F_{WS} score consistently across all four populations. Furthermore, pairwise correlation plots between all of the four West African populations for the bootstrapped mean F_{WS} score calculated for each of the remaining set of 3906 SNPs showed no evidence of correlation (Figure 4.9, Pearson's ranged between -0.05 and 0.05 for all six pairwise comparisons).

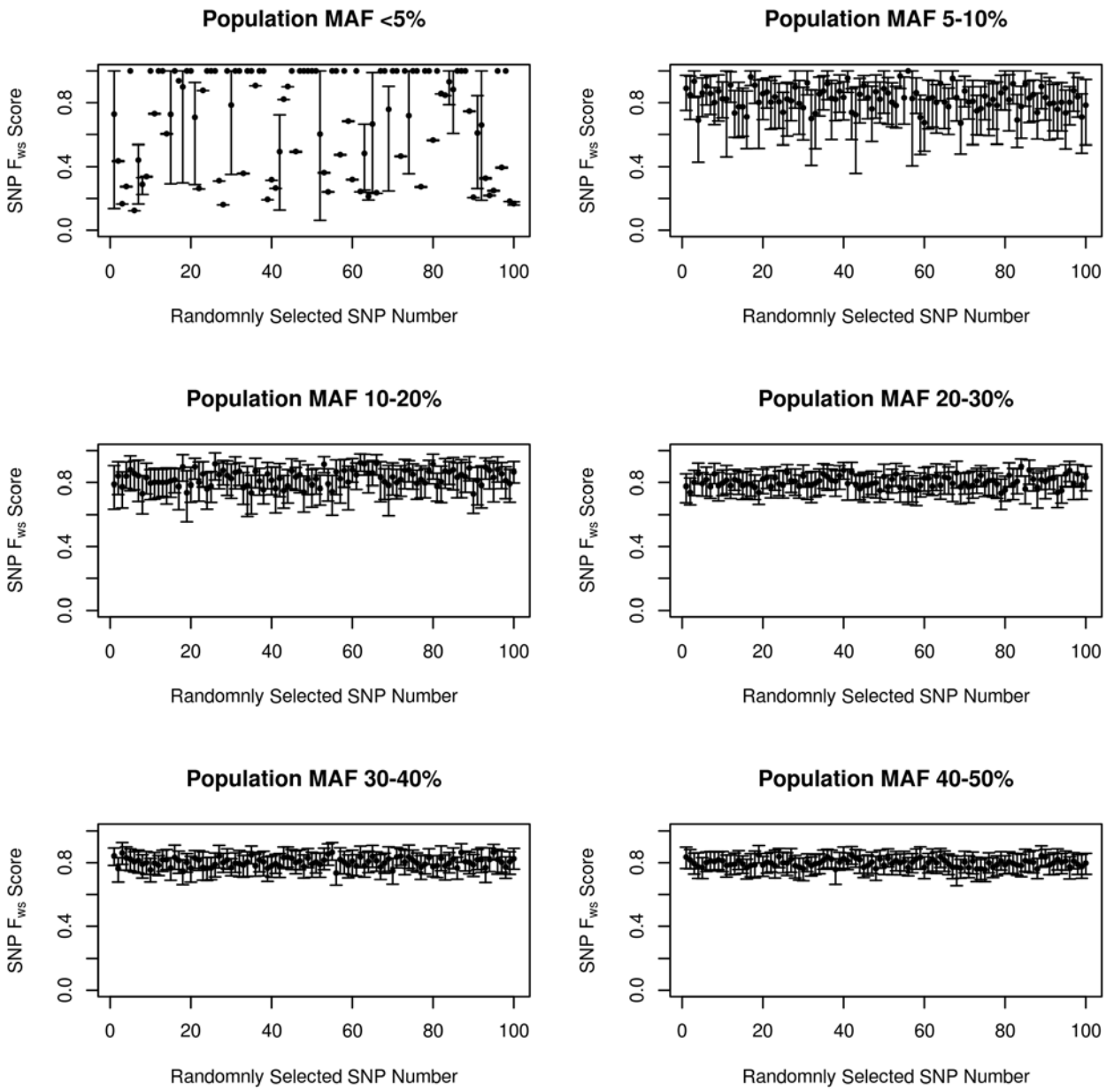


Figure 4.8 Relationship between SNP population minor allele frequency and estimation of a mean SNP F_{WS} score through bootstrapping. Bootstrapped mean F_{WS} scores and 95% confidence intervals of 100 randomly selected SNPs from 6 population minor allele frequency (MAF) intervals from the Guinean population.

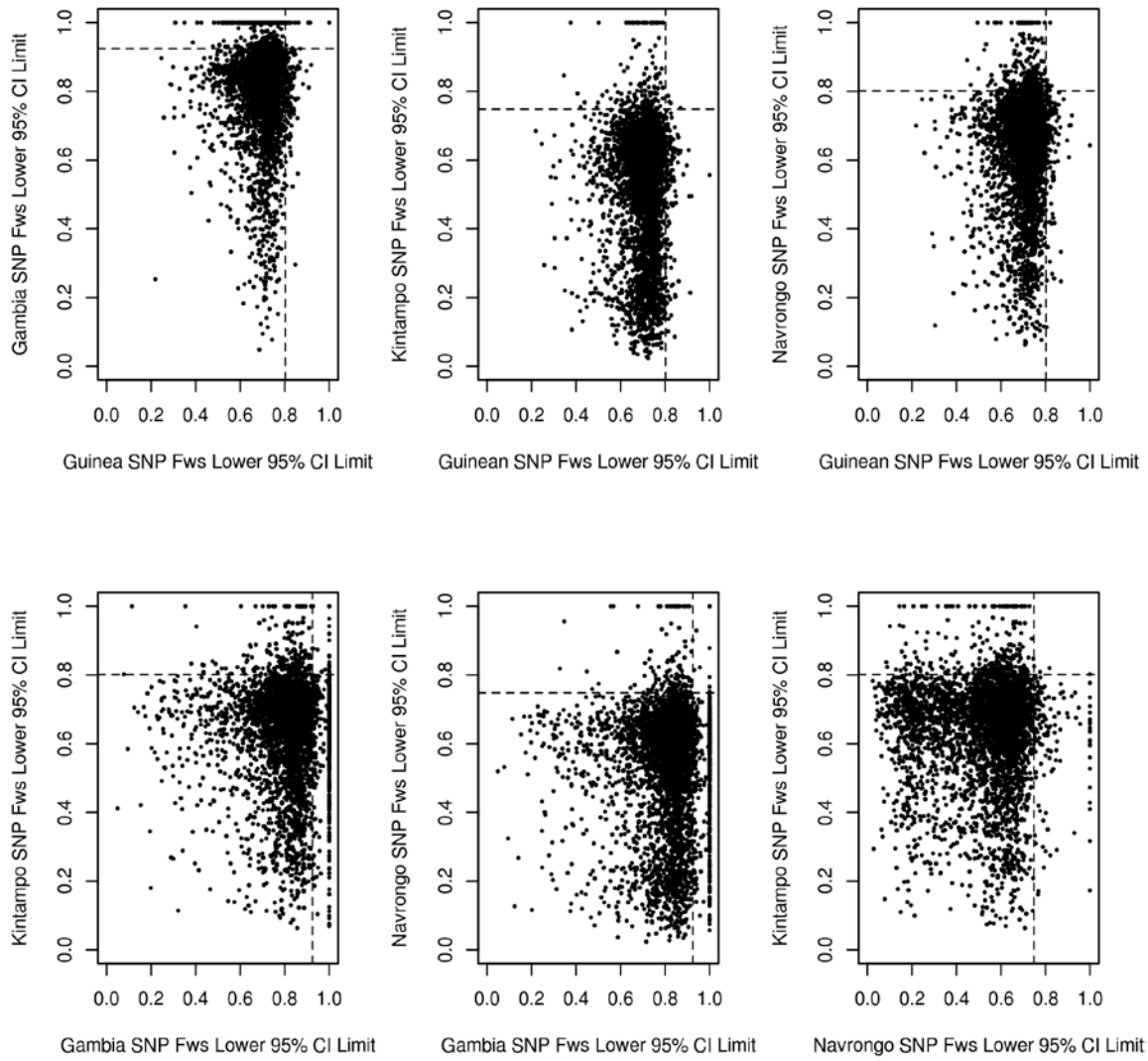


Figure 4.9 Concordance between SNP-based calculations of F_{WS} across four West African populations. Pairwise correlations of the lower 95% confidence interval of 3906 non-synonymous SNPs present within all four West African populations. Dashed lines represent the cut offs within each population of the mean genome-wide F_{WS} score for each population. The top right quadrant within each panel is the region in which outlier SNPs were identified within the two populations compared.

4.4 Discussion

Within this chapter, I present the first attempt to identify loci from deep sequence data that are putatively under within-host selective forces in natural populations of *P. falciparum*. Distributions of isolate scores of F_{WS} are in general agreement with previous microsatellite based reports of within-host host genotypic diversity within the region and generally consistent with higher levels of genomic mixedness being found within areas of higher infection endemicity within West Africa (Mobegi *et al.* 2012). Despite the reported differences in transmission intensity within the Ghanaian regions of Kintampo and Navrongo, which are holoendemic and seasonally endemic respectively (Asante *et al.* 2011; Kasasa *et al.* 2013; Duffy *et al.* 2015), isolate scores of F_{WS} did not differ significantly within these two populations nor the Guinean population, with only the Gambian population having a significantly different mean. Furthermore, the considerable proportion of isolates found to consist predominantly of a single-genotype is also consistent with previous calculations of F_{WS} within areas of high infection endemicity (Manske *et al.* 2012; Auburn *et al.* 2012).

Population replication of a gene-based and bootstrapped SNP-based approach to implement an F_{WS} based scan across the genome did not identify any loci that were consistently found within the outlier strata across all four West African populations as originally hypothesised and only weak correlation in gene and SNP F_{WS} scores was seen between the different populations. Genome-wide approaches using fixation indices and other allele-frequency based metrics have previously identified targets of selection within multiple populations (Manske *et al.* 2012; Mobegi *et al.* 2014). Whilst this approach was thus ultimately unsuccessful, within this section I discuss alternative potential

avenues to identify potential molecular targets of within-host selection in *Plasmodium* parasites.

The approach presented within this chapter shares some conceptual similarity with linkage group selection analyses of genetic cross progeny, which can test for loci associated with a particular phenotype through drug, growth or strain-specific immune selection of uncloned progeny (Culleton & Abkallo 2015) using genome-wide polymorphic markers. These analyses attempt to identify loci amongst the progeny pool following phenotype-specific selection pressure that effectively show higher levels of homozygosity when compared to the original unselected progeny pool. There are however notable differences, as the F_{WS} scan approach within a natural population does not have information on the progeny that have been produced within the previous round of meiotic recombination within the mosquito (section 1.4).

Moreover, there are other alternatives to further examine potential evidence of within-host selection of *P. falciparum* from clinical isolates using sequencing based analyses. Whilst within this study, an allele-frequency based approach has not consistently identified any outlier loci across multiple populations, but a similar approach could be used to look for differences in transcriptome-wide expression levels of genes when particular allelic types are present. Transcriptional variation within laboratory clones of *P. falciparum* has been previously reported (Rovira-Graells *et al.* 2012) and this 'bet-hedging' heterogeneity is likely to play an important role in ensuring successful parasite replication, antigenic variation and transmission within the variable host environment (Waters 2016). Thus, any within-host selection operating may be noticeably more detectable in terms of differential gene expression. Deep

sequence SNP calling on transcriptome data has previously been applied to clinical isolates of *P. falciparum* (Yamagishi *et al.* 2014), although this was performed directly on peripheral blood material and investigation of the variable expression of merozoite invasion ligands would be of interest, due to the well documented variation in expression at this life cycle stage (Mitchell *et al.* 1986; Hadley *et al.*1987; Bei & Duraisingh 2012; Amambua-Ngwa *et al.* 2012; Gomez-Escobar *et al.* 2010). At present however, no population level RNA-Seq datasets are available at the *ex vivo* late schizont stage for *P. falciparum* and difficulties do exist in the differences that exist in the length of the cell cycle between different clones (Reilly *et al.* 2007; Lemieux *et al.* 2009).

Furthermore, the development and application of long-read sequencing, at >10 kb in length, to population studies of *P. falciparum*, will allow for complete chromosome sequencing and *de* novo assembly of the highly polymorphic sub-telomeric regions of *P. falciparum*, which are thus difficult to map against the 3D7 reference genome. Sub-telomeric regions have previously been shown to have increased rates of mitotic recombination in comparison to the core genome (Bopp *et al.* 2013; Claessens *et al.* 2014) and thus effects of within-host selection on genes within this region could be more pronounced.

Chapter 5. Quantifying variation in gametocyte commitment in clinical samples of *P. falciparum* from two populations in different transmission environments

5.1 Introduction

Controlled experimental infections of multiple-genotypes within model *Plasmodium* species have assessed the effect of competition upon gametocyte sex ratio and commitment phenotypes (Pollitt *et al.* 2011; Reece *et al.* 2008; Neal & Schall 2014). However, analyses of natural populations have not seen such significant effects (Neal & Schall 2014, section 1.6.2). A handful of individual population studies have microscopically quantified gametocyte sex ratios for *P. falciparum* (Table 5.1). These have reported small differences in sex ratio commitment, although only the Papua New Guinea population contains mostly single clone infections (Paul *et al.* 1995), while all three other populations within West Africa possessing higher proportions of multiple-genotype infections (Konaté *et al.* 1999; Sowunmi *et al.* 2009; Ntoumi *et al.* 2000). Even with a direct measure of selfing probability, it should be noted that an existing issue would be the changes in the clonal composition over the course of an infection, as any sexual specific gametocyte commitment decision could only be reflective of a brief window during the infection.

Theoretically, the lower limit of male gametocytes present within an individual, known as r*, is dependent on the number of viable gametes produced per gametocyte (fecundity, f) and is equal to 1/(1+f). Within *P. falciparum* infections, male gametocytes usually produced 4-6 gametes out of a theoretical maximum of 8, so with an assumption of f=5, the lower limit of r* for *P. falciparum* would equal 1/(1+5) and thus 0.17. This theoretical expectation of r* is similar to the value estimated for the Papua New Guinea population sample of of 0.18 (Read

Country	Mean r*	95% CI	Number of samples	Reference
Papua New Guinea	0.18	0.15-0.22	28	Read *et al.* (1992)
Senegal	0.26	0.24-0.27	284	Robert *et al.* (2003)
Nigeria	0.22	0.15-0.28	~100	Sowunmi *et al.* (2009)
Cameroon	0.22	0.20-0.23	54	Robert *et al.* (1996)

Table 5.1 Measurements of gametocyte sex ratio by microscopy within four *P. falciparum* populations. Mean r* represents the proportion of male gametocytes reported, as calculated through r*=Number of male gametocytes/(Number of male gametocytes + Number of female gametocytes). Whilst the Papua New Guinea population has a high proportion of single-clone infections (Paul *et al.* 1995), the other three populations within West Africa possess higher proportions of multiple-genotype infections (Konaté *et al.* 1999; Sowunmi *et al.* 2009; Ntoumi *et al.* 2000).

et al. 1992), where malaria parasites are relatively highly inbred. Any population level analysis would be most appropriate by having a measurement of predicted selfing rate (s) within the host across the genome, which has not been tested for. Above its lowest limit, r* will rise within *P. falciparum* based on the increasing genotypic mixedness of infection as r* = (1-*F*), where *F* is selfing probability. With inference of selfing probability within West African populations possible using the F_{WS} index (Manske *et al.* 2012), the expected sex ratio of gametocytes in response to having a particular F_{WS} score can be predicted. From r*, it is also possible to estimate $F_{WS,}$ with an r* = 0.17, deemed as the theoretical limit, equalling an F_{WS} of 0.66 through F_{WS} = 1-(2r*). From this, it can be predicted that infections with an F_{WS} of <= 0.65 would begin to produce high values of r* and thus greater proportions of male gametocytes within an infection.

Within the West African populations that have had their population mean F_{WS} derived within this thesis (section 4.3.1), the population means ranged from 0.77 within the endemic Northern Ghanaian population to 0.93 within the relatively less endemic Gambian population. Within all of these populations however, a significant proportion of isolates were calculated as highly clonal infections, where F_{WS} >=0.95. Thus, the skewed distributions of isolate F_{WS} within these populations were prohibitive to testing an association between inferred probability of selfing within individual isolates and the proportion of male gametocytes. Furthermore, whilst F_{WS} offers an inference of the probability of selfing, this can only experimentally be determined by oocyst based genotyping from large numbers of mosquitoes (Mzilahowa *et al.* 2007).

Alternatively, this chapter examines whether local parasite populations sampled from contrasting areas of transmission intensity have undergone microevolutionary processes, leading to a local population adaptation of optimal sex ratio that sees parasites in local populations continuously producing overall optimal sex ratios, rather than adjusting to the level of diversity in a current infection. Genomic loci within these localities involved in gametocytogenesis processes have been identified as exhibiting evidence of local population selective processes (Mobegi *et al.* 2014; Duffy *et al.* 2015) and thus selection could be operating at the population level due to transmission differences caused by basal variation in gametocyte sex ratio commitment.

Gametocyte prevalence within endemic regions of *P. falciparum* has been widely determined using a combination of microscopy and molecular methods for the detection of stage V gametocytes (Molineaux & Gramiccia 1980; Stepniewska *et al*. 2008; Bousema *et al.* 2006). Within this chapter, *in vivo* commitment to gametocyte production has been characterised by profiling the expression of genes that have been shown to quantify gametocyte sex ratios using sensitive molecular detection assays (Schneider *et al.* 2015). In addition, to further investigate the possibility of different basal rates of gametocytogenesis, the expression of genes shown to be upregulated during the early stages of gametocytogenesis and essential for gametocyte formation within 3D7A are profiled; GDV-1 and AP2-G (Eksi *et al.* 2012; Kafsack *et al.* 2014). These investigations have been performed upon samples from two West African populations which have different levels of local transmission within Ghana and Senegal (Mobegi *et al.* 2012; Mobegi *et al.* 2014). Although the Senegalese population within Table 5.1 had a high proportion of multiple-genotype infections (Konaté *et al.* 1999), the Senegalese samples used within

this chapter were collected from a region near Dakar where parasite transmission had substantially reduced between 2006 and 2011 due to increased control efforts (Daniels *et al.* 2013). The Ghanaian samples used consisted of 12 samples from Kintampo in Central Ghana and 21 samples from Northern Ghana, for which whole-genome sequence analyses had characterised similar levels of infection diversity to other highly endemic West African populations and little genetic differentiation between the two populations (Duffy *et al.* 2015).

5.2 Materials and Methods

Clinical isolates that were used within the gametocyte RT-PCR (section 2.2) had at least 100 µL of packed red blood cells thawed (section 2.3.2) and placed straight into TRIzol reagent (Invitrogen), before being stored at -80°C immediately. RNA extraction was performed using an RNeasy Micro kit (Qiagen) and eluted in 20 µL of elution buffer, prior to storage at -80°C. Subsequently, mRNA was reverse transcribed with oligo(dT) using TaqMan reverse transcription reagents (Life Technologies) to produce cDNA. To monitor any potential carryover of genomic DNA for the reverse transcription process, 16 samples had no-RTs control set up and assessed by qPCR of the housekeeping transcript used. Following cDNA production, all samples were diluted 1 in 2 to give 40 µL of cDNA material for qPCR.

Five different transcripts were quantified from each clinical sample; a commonly used housekeeping transcript serine tRNA ligase (PF3D7_0717700), the female-specific gametocyte transcript Pfs25 (PF3D7_1031000), the male-specific gametocyte transcript Pfs230p (PF3D7_028900), and the early markers of gametocyte commitment transcripts GDV-1 (PF3D7_0935400) and AP2-G

Primer Identity	Primer Sequence (5' to 3')	Target Gene	Origin of primers
Pfserylt-F	AAGTAGCAGGTCATCGTGGTT	Serine tRNA ligase	Eksi *et al.* 2012
Pfserylt-R	GTTCGGCACATTCTTCCATAA	Serine tRNA ligase	Eksi *et al.* 2012
Pfs25-F	CCATGTGGAGATTTTTCCAAATGTA	Pfs25	Schneider *et al.* 2015
Pfs25-R	CATTTACCGTTACCACAAGTTACATTC	Pfs25	Schneider *et al.* 2015
Pfs230p-F	CCCAACTAATCGAAGGGATGAA	Pfs230p	Schneider *et al.* 2015
Pfs230p-R	AGTACGTTTAGGAGCATTTTTTGGTAA	Pfs230p	Schneider *et al.* 2015
GDV1-F	TAGGCGTCGAAATAGTGCTAGTAGAAA	GDV-1	Eksi *et al.* 2012
GDV1-R	GTCCTCACAACCAGCATCATTAGTA	GDV-1	Eksi *et al.* 2012
AP2G -F	AACAACGTTCATTCAATAAATAAGG	AP2-G	Rovira-Graells *et al.* 2012
AP2G-R	ATGTTAATGTTCCCAAACAACCG	AP2-G	Rovira-Graells *et al.* 2012

Table 5.2 Primers used for gametocytogenesis transcript profiling of clinical isolates.

(PF3D7_1222600), with all primers listed in table 5.2. All primers were checked using sequence data available at www.plasmodb.org to ensure that no SNPs were present within their target annealing sites. For all qPCRs, reaction and thermocycler conditions were the same as previously used for qPCR experiments for PMR quantitation (section 2.7.1). All transcripts were quantified in duplicate against 3D7A genomic DNA standard curves and the four gametocyte commitment transcripts were normalised against the seryl tRNA ligase housekeeping transcript using standard curve based absolute quantitation.

5.3 Results

5.3.1 Quantification of mRNA concentrations and quality filtering

From the original set of 81 West African clinical isolates that underwent RNA extraction (48 from Senegal and 33 from Ghana), 33.3% of Senegalese isolates (n=16) and 84.8% of Ghanaian isolates (n=28) had a concentration of parasite mRNA of the housekeeping serine-tRNA ligase transcript of >1000 copies per µL, which was implemented as the cut off threshold for subsequent profiling of the four gametocytogenesis linked transcripts. From this set of 28 Ghanaian isolates, mRNA concentration was quantified above a threshold of 10 copies per µL within 27 isolates (93.8%) for AP2-G, 26 isolates (81.3%) for GDV-1, 14 isolates (87.5%) for Pfs25 and 8 isolates (50%) for Pfs230p (Table 5.5). Within the set of 16 Senegalese isolates, mRNA concentration was quantified above the same threshold within 15 isolates (69.4%) for AP2-G, 13 isolates (63.9%) for GDV-1, 25 isolates (69.4%) for Pfs25 and 11 isolates (30.6%) for Pfs230p. The total parasite material within each of the clinical samples, as measured by quantification of the housekeeping transcript was significantly greater for the

Isolate ID	Population	Housekeeping (copies/µL)	Pfs25 (copies/µL)	Pfs230p (copies/µL)	AP2-G (copies/µL)	GDV-1 (copies/µL)
INV136	Ghana	29707	366	270	410	514
INV137	Ghana	14553	356	96	202	232
INV138	Ghana	3578	0	0	0	0
INV139	Ghana	7644	100	94	290	78
INV140	Ghana	33528	248	192	420	305
INV141	Ghana	23502	460	158	378	356
INV155	Ghana	15811	66	52	46	33
INV156	Ghana	1832	42	32	104	87
INV157	Ghana	1867	24	15	44	15
INV158	Ghana	4959	81	49	184	132
INV159	Ghana	32729	204	262	802	114
INV160	Ghana	6148	20	30	298	40
N180	Ghana	277	205	0	23251	263
N166	Ghana	7950	174	45	23	64
N247	Ghana	1174	23	0	121	0
N100	Ghana	6485	103	14	18	35
N235	Ghana	2339	22	0	21	188
N019	Ghana	18738	265	135	32	61
N209	Ghana	218	17	0	0	0
N224	Ghana	1932	226	0	2573	94
N036	Ghana	12056	191	35	30	86
N223	Ghana	1422	33	0	40	55
N113	Ghana	57197	944	58	80	56
N176	Ghana	4984	146	10	114	35
N086	Ghana	16315	312	43	904	71
N022	Ghana	1233	53	0	1414	10
N032	Ghana	40729	15466	248	20312	831
N175	Ghana	14681	592	132	17019	34
N005	Ghana	748	386	0	544	0
N292	Ghana	90467	295	48	12	86
N098	Ghana	10299	319	50	121	183
N237	Ghana	809	10	0	0	59

Table 5.3 Mean copy number transcripts for all 5 genes profiled within the 32 Ghanaian isolates that had a housekeeping transcript concentration of >100 copies per µL. All concentrations are given in copies per µL, from duplicate replicates against a 3D7A genomic DNA series which had a lowest concentration limit of 10 copies per µL, with values of 0 representing transcripts which were found to be at a concentration of below 10 copies per µL.

Isolate ID	Population	Housekeeping (copies/µL)	Pfs25 (copies/µL)	Pfs230p (copies/µL)	AP2-G (copies/µL)	GDV-1 (copies/µL)
PK85	Senegal	491	12	0	90	43
PK66	Senegal	14052	12	0	80	15
PK46	Senegal	66654	452	16	508	73
PK61	Senegal	9777	39	20	299	105
PK74	Senegal	3214	6421	0	134	135
PK58	Senegal	959	0	0	0	0
PK56	Senegal	238	24	0	0	0
PK76	Senegal	711	297	12	63	49
PK62	Senegal	940	0	0	0	0
PK84	Senegal	587	14	0	0	38
PK75	Senegal	622	776	20	3902	162
PK83	Senegal	822	53	0	187	26
PK68	Senegal	442	26	0	720	0
PK45	Senegal	720	0	0	0	10
PK64	Senegal	160	45	0	0	26
PK39	Senegal	1249	0	0	10	0
PK77	Senegal	2708	147	13	805	31
PK25	Senegal	6428	344	49	82	67
PK52	Senegal	22550	166	66	731	498
PK44	Senegal	3405	76	0	243	0
PK41	Senegal	352	0	0	0	0
PK36	Senegal	2056	195	0	204	927
PK40	Senegal	214	0	0	22	0
PK50	Senegal	768	19	0	18	13
PK80	Senegal	107	21	0	0	0
PK55	Senegal	979	16	0	0	11
PK71	Senegal	3169	109	0	374	20
PK67	Senegal	173	0	0	13	0
PK18	Senegal	4628	39	19	165	34
PK07	Senegal	4182	70	16	16	1556
PK24	Senegal	122	0	0	596	0
PK23	Senegal	220	0	0	0	0
PK16	Senegal	3078	0	0	0	0
PK81	Senegal	48334	225	138	233	61
PK65	Senegal	2816	86	0	391	50
PK59	Senegal	885	0	12	39	15

Table 5.4 Mean copy number transcripts for all 5 genes profiled within the 36 Senegalese isolates that had a housekeeping transcript concentration of >100 copies per µL. All concentrations are given in copies per µL, from duplicate replicates against a 3D7A genomic DNA series which had a lowest concentration limit of 10 copies per µL, with values of 0 representing transcripts which were found to be at a concentration below 10 copies per µL.

Population	Senegal	Ghana
RNA extractions performed	48	33
Number of samples with housekeeping transcript concentration >1000 copies/µL	16 (33.3%)	28 (84.8%)
Number of samples with AP2-G transcript concentration >10 copies/µL	15 (93.8%)	27 (96.4%)
Number of samples with GDV-1 transcript >10 concentration copies/µL	13 (81.3%)	26 (92.9%)
Number of samples with Pfs25 transcript >10 concentration copies/µL	14 (87.5%)	27 (96.4%)
Number of samples with Pfs230p transcript >10 concentration copies/µL	8 (50%)	22 (78.6%)

Table 5.5 Summary of the numbers and proportions of clinical isolates within each population for which each of the transcripts profiled were quantifiable above certain thresholds. For the housekeeping transcript, percentages are given above based on the total number of RNA extractions performed. For the four other gametocyte specific transcripts, percentages are given based on the number of isolates with a concentration of the housekeeping transcript of >1000 copies/ µL.

Ghanaian samples than the Senegalese population (Wilcoxon rank-sum test, p-value < 1 x 10^{-5}). This may explain why there were fewer Senegalese isolates which reached the detection threshold of >=10 copies per µL for each of the other gene transcripts. Individual mean copy number concentrations for each of the 36 Senegalese and 32 Ghanaian isolates with > 100 copies per µL of the housekeeping transcript are reported within Table 5.3 and Table 5.4.For each transcript, any isolates with a calculated concentration of <10 copies per µL, as determined by the lower range of the genomic DNA standard controls, were regarded as having a concentration of 0 copies per µL for all downstream analyses. Previous estimations of the proportion of isolates containing gametocytes has been put at about 90% using a molecular QT-NASBA approach which targeted the Pfs25 gene, although this was undertaken within a highly endemic Kenyan population (Bousema *et al.* 2006). This is broadly in agreement with the 96.4% of Ghanaian isolates and 87.5% of Senegalese isolates which had a Pfs25 concentration of >10 copies per µL when >1000 copies of the housekeeping transcript were quantified.

5.3.2 Varied isolate expression and population comparison of gametocytogenesis transcripts and sex ratios

To provide quantitative information on the relative levels of each of the four gametocyte transcripts at the isolate level, all four gametocytogenesis implicated transcripts were normalised as the proportion of the housekeeping gene for all isolates which had a calculated housekeeping copy concentration of > 1000 copies per µL (Table 5.3 and Table 5.4). From this selection of 28 Ghanaian and 16 Senegalese isolates, a subset of 22 Ghanaian and 10 Senegalese isolate expression profiles for the four genes can be found within Figure 5.1, which were selected to support visual clarity of the varied distribution

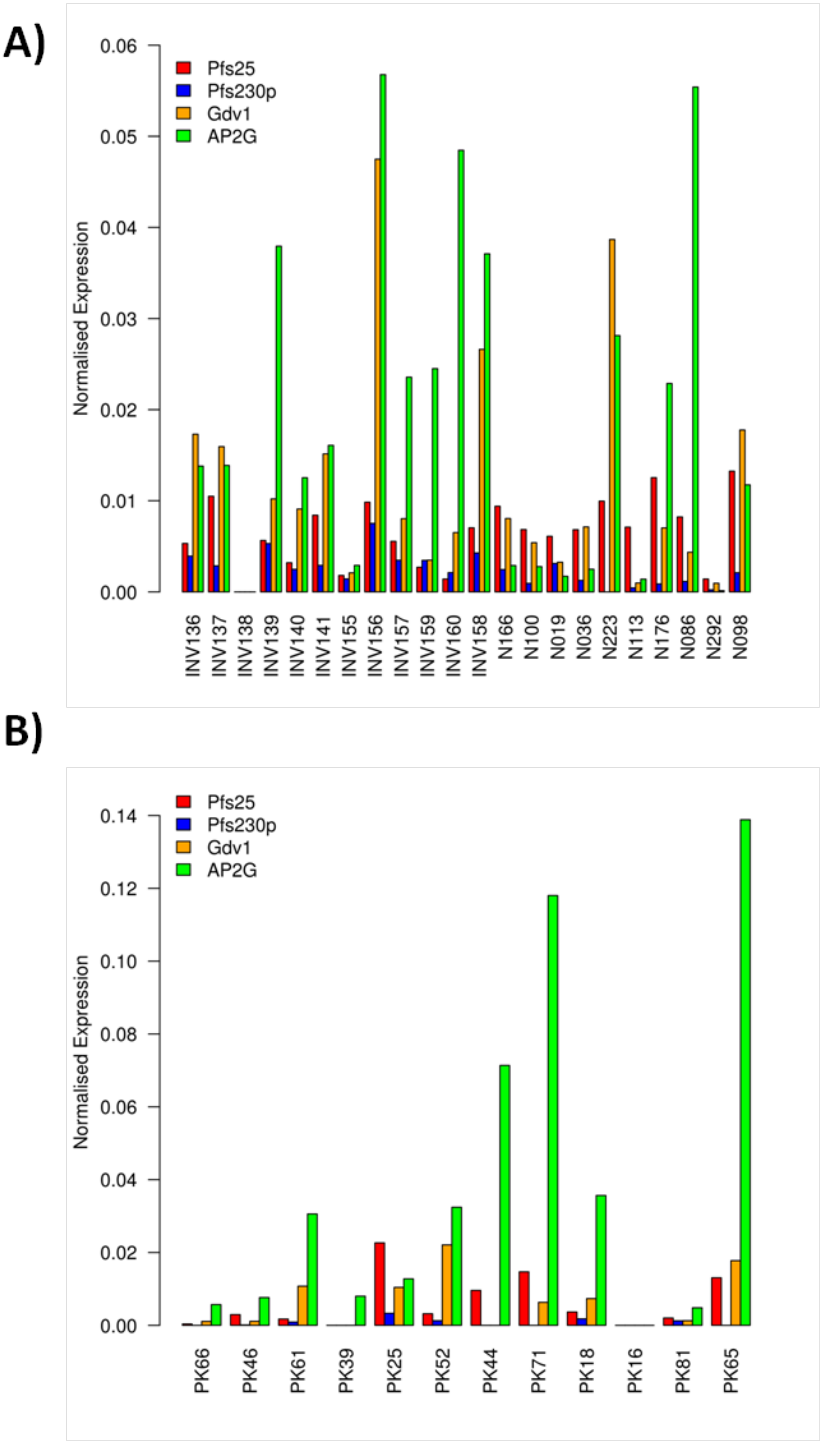


Figure 5.1 Normalised expression profiles for a subset of A) 22 Ghanaian isolates and B) 10 Senegalese isolates. Individual isolate IDs are found on the x-axis, whilst the representative colour scheme for each of the four transcripts can be found within the legend. For each of the four transcripts, the normalised proportion of that transcript against the prescribed housekeeping transcript is plotted on a log scale. These overall 32 isolates were selected on the basis they did not possess any outlying levels of a particular transcript, to aid visual clarity. Isolate INV138 in panel A) shows no evidence of expression of any of the four gametocytogenesis transcripts despite robust levels of expression of the housekeeping gene (Table 5.3).

of normalised expression profiles for the four gametocytogenesis transcripts. For instance, Senegalese isolate PK74 was not shown on the basis of showing markedly higher expression of the Pfs25 transcript relative to its housekeeping transcript expression. Within the plotted isolates, considerable variation is visible in the normalised expression levels recorded. For example, within the Senegalese samples, isolate PK25 showed higher expression of thefemale gametocyte specific Pfs25 gene than all other Senegalese samples, whilst for PK16 no expression of any of the four gametocytogenesis linked transcripts was detected. Meanwhile, within the Ghanaian samples, expression of the male gametocyte specific transcript Pfs230p is found to be highest within sample INV156.

For pairwise population comparisons of expression of all four gametocytogenesis linked transcripts, all 28 Ghanaian and 16 Senegalese isolates were used, with any individual transcripts that were calculated to have a concentration of <10 copies per µL were regarded as being 0. From these comparisons, the normalised expression levels of Pfs25, AP2-G and GDV-1 were all not significantly different (Mann Whitney U test, p-values > 0.5).The normalised expression level of Pfs230p within the Senegalese samples was lower than within the Ghanaian samples and was marginally significant when corrected for multiple comparison testing (Mann Whitney U test, Bonferroni corrected p-value = 0.048). For a comparison of sex ratios between the two populations, a further data filtering measure was implemented, whereby isolates were only included if they satisfied the criteria of having both a housekeeping transcript concentration of >1000 copies per µL and either an additive copy number of >100 copies for the two sex-specific markers Pfs25 and Pfs230p or the Pfs230p transcript was calculated to be at least one-tenth the concentration

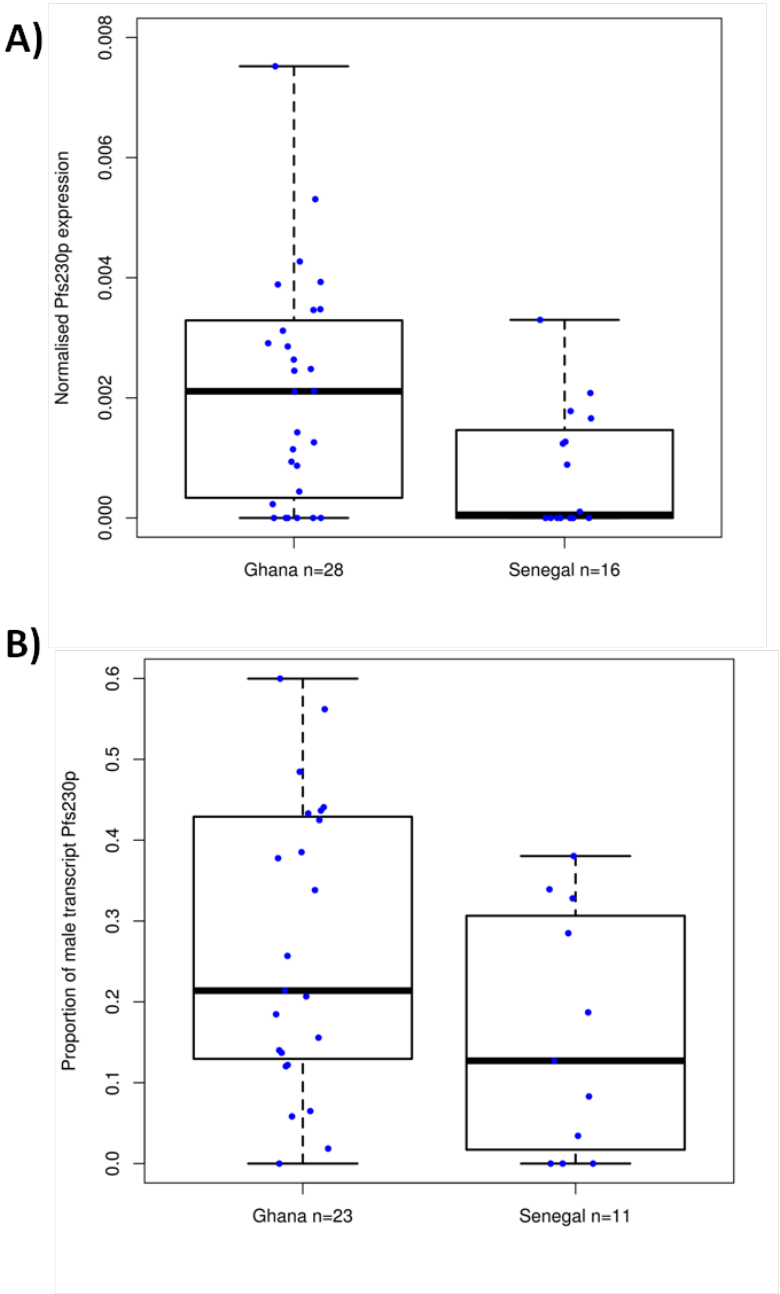


Figure 5.2 Population male gametocyte transcript and sex ratio comparison of Ghanaian and Senegalese population subsets. Panel A) shows a slightly significant lower normalised expression of Pfs230p within 16 Senegalese samples compared to 28 Ghanaian samples (Mann Whitney U , Bonferroni corrected p-value= 0.048). Panel B) is a population sex ratio comparison between 23 Ghanaian isolates and 11 Senegalese isolates, where the normalised male transcript proportion was calculated as normalised Pfs230p/(normalised Pfs230p + normalised Pfs25). Although the mean was male proportion was lower within Senegal, no significant difference was seen (Mann Whitney U, p-value = 0.08).

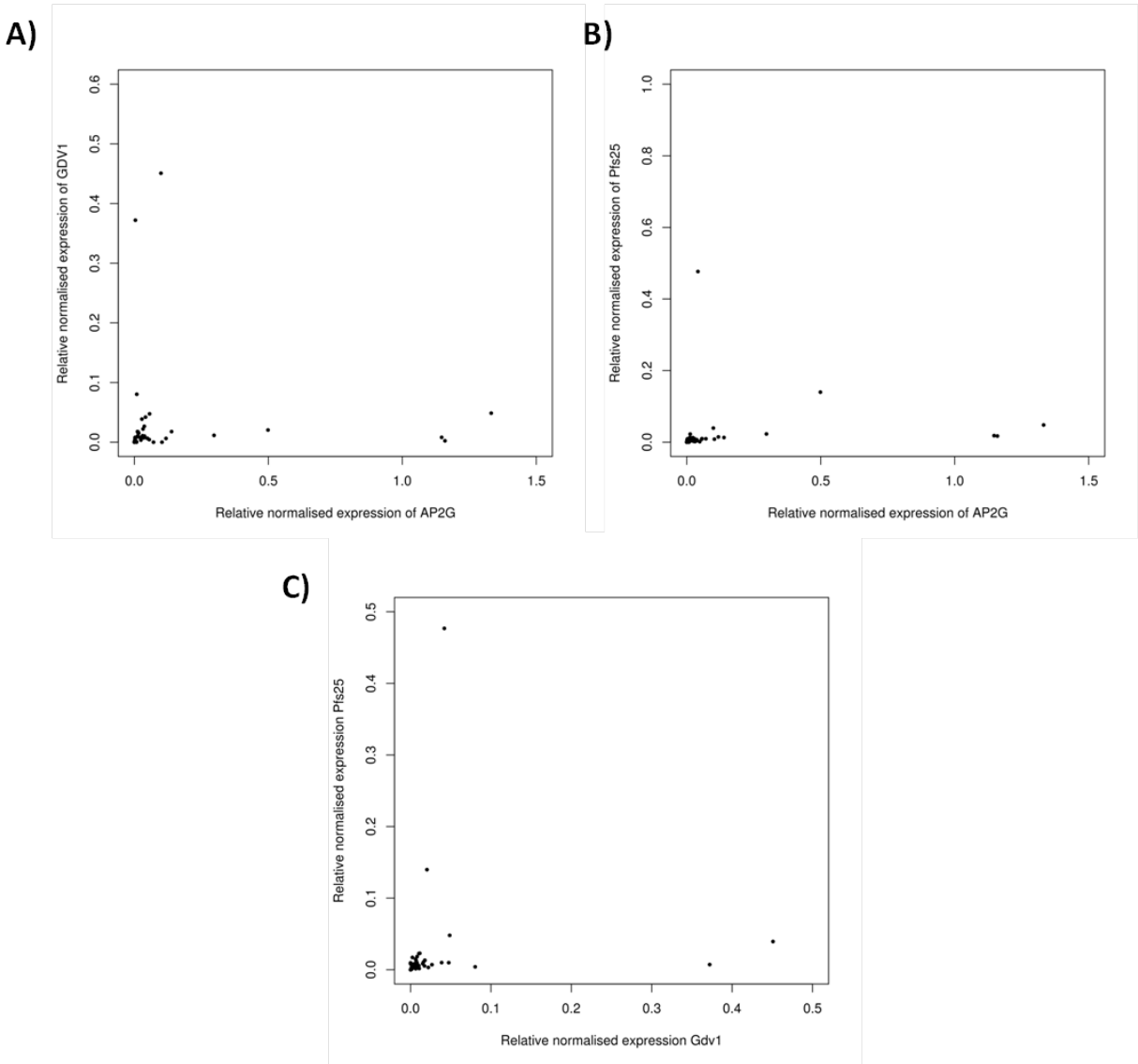


Figure 5.3 Pairwise correlations of normalised expression levels between Pfs25, AP2-G and GDV-1. No evidence of correlation was seen between A) AP2-G and GDV-1 (Pearson's r = -0.02), B) Pfs25 and AP2-G (Pearson's r = 0.09) or C) Pfs25 and GDV-1 (Pearson's r = 0.07) for 44 clinical isolate samples pooled across the Senegalese and Ghanaian populations.

of the Pfs25 transcript. This led to the inclusion of 23 Ghanaian isolates and 11 Senegalese isolates, with the removal of INV138, N247, N235, N223 and N022 from the Ghanaian population and PK66, PK77, PK44, PK16 and PK65 from the Senegalese population. Whilst the Senegalese population had a lower mean normalised sex transcript ratio of 0.16 compared to the Ghanaian population of 0.268, this difference was not significant (Mann Whitney U test, p-value = 0.08). In addition, pairwise correlations of the normalised expression levels of the Pfs25, AP2-G and GDV-1 genes were performed by pooling together the Ghanaian, but did not reveal any significant pairwise correlation (Figure 5.3).

5.4 Discussion

The analyses presented within this chapter represent the first attempt to quantify the expression levels of gametocyte sex specific transcripts in a comparison across two populations of differing transmission intensity and to quantify expression of the essential gametocytogenesis genes GDV-1 and AP2-G from clinical isolates from two West African populations. In general, characterisation of the expression levels of specific gametocyte or early gametocytogenesis transcripts has not been well studied *in vivo*. Eksi *et al.* (2012) did quantify gametocyte expression profiles of a number of early stage gametocytogenesis implicated transcripts, including GDV-1, from a collection of twenty clinical isolates that were sampled upon the border of Thailand and Myanmar, finding a spectrum of isolate gametocyte expression profiles. However, the recently described essential early regulator of gametocytogenesis, AP2-G (Kafsack *et al.* 2014), has not had its expression profile characterised previously within clinical infections *in vivo* and no

molecular-based comparison of sex ratios has previously been performed for *P. falciparum*.

Quantitative data on the distribution of profiles reported within this chapter provide important findings for designing further studies that may investigate sex ratio distributions or transcriptional variation within early or late-stage gametocyte markers *in vivo*, with great interest in these life cycles stages as potential transmission-blocking vaccine candidates (Kapulu *et al.* 2015; Nilsson *et al.* 2015). Whilst the relatively low levels of gametocytemia present within clinical isolates make such studies difficult, a potential approach for investigating stage V gametocyte-specific antigens would be to use a combination of monoclonal antibodies for antigens such as Pfs25 and Pfs230p that are available from the MR4 reagent depository (Quakyi *et al.* 1987; Barr *et al.* 1991) to specifically bind to and sort gametocytes for limited-cell based RT-PCR or whole transcriptome amplification.

Although the analysis of sex ratio within this chapter did not find a significantly lower proportion of male gametocytes within the Senegalese population as posited by evolutionary sex ratio theory, this population-level comparison was hampered by the relatively low expression levels of the Pfs230p transcript and the lower general parasite densities that were seen within the Senegalese population, which led to relatively small sample sizes. Alternatively, recent studies have characterised male and female gametocyte specific proteomes (Lasonder *et al.* 2016; Tao *et al.* 2014) and thus an RT-PCR assay targeting a more highly expressed male-specific protein may be a useful tool for such studies. In addition, a combination of an increased volume of sampling or the progression to analyses of limited cell amplification assays could overcome these issues.

In addition, the population level comparisons of the relative normalised levels of AP2-G, GDV-1 and Pfs25 expression reveal no significant differences, despite the varied levels of transmission intensity within this region. Expression of gametocyte specific transcripts is likely to be mediated by subtle differences in the seasonality of the local area, with experimental analyses of *P. relictum* having revealed that in the presence of its natural vector *Culex pipens*, infections have higher levels of parasitemia which lead to increased mosquito transmission levels (Cornet *et al.* 2014). The relationship between host-specific, parasite-specific, vector-specific and environmental factors that influence gametocytogenesis in *P. falciparum* remain challenging to dissect, but merits further investigation.

Chapter 6. Quantifying variation in parasite multiplication rate during the *ex vivo* culture adaptation process of clinical isolates of *P. falciparum*.

6.1 Introduction

Due to the aforementioned success of *P. falciparum* whole genome sequencing (section 1.3) ventures such as the Pf3K project (Manske *et al.* 2012; Miotto *et al.* 2013; MalariaGEN 2016), over 2,500 whole-genome sequences are publically available for analysing clinical isolates from a range of populations globally (https://www.malariagen.net). However, there is comparably much less quantitative data available on many medically and epidemiologically relevant *P. falciparum* phenotypes derived from *ex vivo* clinical isolates.

Whilst experiments on a rodent malaria model using *P. chabaudi* in mice have reported associations between parasite multiplication rate (PMR) and the virulence of infections (Mackinnon & Read 1999; Mackinnon & Read 2004), there have been conflicting reports from studies on the presence of an association between *P. falciparum* PMR within the first *ex vivo* cycle and the virulence of the infection from studies of isolates from South-East Asia (Chotivanich *et al.* 2000) and Sub-Saharan Africa (Deans *et al.* 2006). Whilst these studies provide quantitative estimations of PMR at the earliest possible period *ex vivo*, they were however only taken using a single biological replicate per isolate, from one replication cycle. In addition, measurements of *ex vivo* PMR and parasite invasion can be confounded by the generally low growth rates and any heterogeneous effects of cryopreservation and storage on measurements of PMR within that first cycle *ex vivo* (Nsobya *et al.* 2008; Bowyer *et al.* 2015). The most detailed analyses of *in vivo* growth rate across multiple intraerythrocytic cycles for three different ‘strains’ of *P. falciparum*,

using historical data from *P. falciparum* treatment of neurosyphillis cases reported differences in PMR across the strains and a 90% prediction interval for PMR across more than 200 cases of between 5.5 and 12.3 fold per 48 hour period (Simpson *et al.* 2002). Thus, this chapter focuses on quantifying PMR in a standardised manner from a panel of clinical isolates as they undergo culture adaptation *ex vivo*.

Whilst parasite growth rates have been assessed within laboratory clones of *P. falciparum* previously (Reilly *et al.* 2007) and from cross progeny to identify genomic loci associated with cell cycle duration (Reilly Ayala *et al.* 2010), laboratory clones may not be representative of natural parasite populations. For example, it has been shown that laboratory clones have lost various chromosomal regions during the culture adaptation process (Cheeseman *et al.* 2009). In addition to detailing baseline information on an important phenotype of *P. falciparum*, determining the spectrum of PMR seen in clinical isolates *ex vivo* would be an important step in being able to identify other potential genomic or transcriptomic determinants of growth and to being able to perform controlled *ex vivo* competitive growth studies of *P. falciparum* to test for any evidence of multiple-genotype interactions. Conclusions from within-host competition studies of different model *Plasmodium* species (section 1.6.1) have given contrasting conclusions upon potential competitive suppression of certain genotypes (de Roode *et al.* 2005; Abkallo *et al.* 2015) and ideally would be performed within *ex vivo* clones or clonal isolates of *P. falciparum*. Within this chapter, I thus focus on quantifying the baseline variation *ex vivo* PMR and test the hypothesis that this is likely to increase over time during culture adaptation.

6.2 Materials and Methods

Laboratory clone parasites (section 2.3.1) were thawed (section 2.3.2) and cultured (section 2.3.4). The limit of detection of the assay when a 3% haematocrit, 200µL sample was taken was assessed by using a serially diluted sorbitol synchronised ring stage culture of the clone 3D7A. Furthermore, the degradation of 3D7A-GFP genomic DNA was assessed *in vitro* by heating the volume of parasite material going into the assay at 95°C for 10 minutes, prior to sampling at 48 hour intervals over 6 days, alongside an unheated control set at 0.02% parasitemia in standardised growth assay conditions (2.6.1 and 2.6.2). Furthermore, initial starting density of the assay was determined by setting up laboratory clone Dd2 in two different erythrocyte donors at four different starting parasitemias (0.01%, 0.0316%, 0.1% and 0.316%) to provide a measurement of PMR at a range of starting densities.

For all laboratory clone assays, DNA extractions and qPCR reactions were carried out (2.7.1) and PMR was analysed (section 2.7.2). As a quality filtering step, outlying replicate time point density measurements for a laboratory clone were removed if they were >=0.5 log genome copies away from all other replicates for that clone at that timepoint and for an entire erythrocyte donor replication for an isolate where >=2 timepoints met the individual timepoint criteria. These steps only affected one Day 4 timepoint for clone Dd2 and one biological replicate for D10; thus PMR was estimated across 6 biological replications of HB3 and 3D7A-GFP, 5 biological replications for D10 and 6 biological replications for Dd2, with the exception of a single omitted Day 4 timepoint. In addition, giemsa stained microscopy counting of parasitemia from 1000 red blood cells of 90 growth assay slides taken from Day 2 to Day 8 across the four laboratory clone was performed, with a singular Dd2 Day 4 time

point omitted for qPCR analysis and 3 Day 2 time points for which no parasites were observed from a Day 2 count (2 from 3D7A-GFP, 1 from D10 and 1 from HB3), along with a 3D7A-GFP Day 4 count for which a poor quality smear was prepared and subsequently not counted. From the microscopy counts, parasite numbers were adjusted to account for the asynchronous presence of multi-nucleate schizonts, with each schizont considered to have 8 genome copies.

For *ex vivo* clinical isolate growth assays, isolates were thawed (section 2.3.2) and allowed to adapt to *ex vivo* conditions (section 2.3.4) for a minimum of two weeks, which was performed by Ms Lindsay Stewart. Day 0 of each initial standardised growth assay was determined by the availability of the triplicate biological donors, and was undertaken between Day 16 and 37 of the *ex vivo* culture adaptation process for each isolate. On Day 0 of the standardised growth assay, isolates were included which matched the threshold of having an asexual parasitemia of 0.2% based on a 1000 red blood cell count. This adopted cut off was used to ensure that a minimum of 90% of the erythrocytes that were going into each flask were from the erythrocyte donor. This led to the inclusion of a total of 19 clinical isolates (consisting of 6 Guinean isolates, 9 Senegalese isolates and 4 Malian isolates) for which growth assays were set up accordingly (section 2.6.3) alongside the 3D7A-GFP laboratory clone control and qPCR was performed (section 2.7).

Culture adaptation of these 19 isolates was continued and for those for which parasites were still microscopically present a minimum of 4 weeks later (n=6), their growth rate phenotype was requantified. A maximum of three measurements of *ex vivo* PMR were thus recorded during the culture adaptation process. DNA extractions and qPCR were carried out (section 2.7) and

subsequently underwent PMR analysis within R (2.7.2). For one of the 6 isolates which continued to successfully adapt to culture *ex vivo*, PMR was calculated at only two timepoints during *ex vivo* culture adaptation, whilst the other 5 isolates had their PMR quantified three times. Individual outlier replicates were quality filtered as described for the laboratory clones, which led solely to the removal of a Day 4 and Day 6 timepoint for INV248 and of an entire biological replicate for four isolates (INV228, INV232, INV035 and INV243) where the laboratory clone control also did not grow within that erythrocyte donor. Across the initial *ex vivo* profiling and the repeated profiling for six isolates, a total of 8 sets of triplicate erythrocyte donors were used for the assay. Microsatellite genotyping of four loci was performed for the initial *ex vivo* estimation of PMR as described, except for isolate INV035 which was performed from the second *ex vivo* growth assay (section 2.5.2).

6.3 Results

6.3.1 Optimisation of a low starting density assay for PMR quantitation

The limit of detection of the assay when a 200µL sample was taken from a 3% haematocrit culture as performed using a sorbitol synchronised ring stage culture of the clone 3D7A-GFP was found to be at least 0.0066% (Table 6.1). In addition, the test of DNA degradation over the length of the assay found that whilst parasite DNA was still detectable within the culture which had been heated at 95°C prior to inoculation on Day 6, the quantified number of genome copies present had decreased approximately sevenfold over this period, indicative of at least some DNA degradation occurring and no parasite growth (Table 6.2). For all initial starting densities, linear growth rates were calculated

Parasitemia	Calculated genome copy concentration per µL total culture
6.6%	20840
0.66%	2467
0.066%	196
0.0066%	20

Table 6.1 Sensitivity of collecting 200 µL from a 3% haematocrit culture using serially diluted ring-stage 3D7A-GFP parasites. Calculated genome copy concentrations are shown per µL total culture, with a parasitemia of at least 0.0066% determined to be detectable through this sampling.

Day	Control genome copies per µL total culture	95°C heated genome copies µL total culture
0	211	225
2	2758	79
4	15598	34
6	48977	33

Table 6.2 Assessing *in vitro* DNA degradation of *P. falciparum*. qPCR based genome copy number quantitation of a 200µL sample from a control 0.02% 3D7A-GFP culture which was collected 48 hourly over a 6 day period and to an experimental culture where 0.02% of parasites were heated at 95°C for 10 minutes prior to inoculation, in order to allow for assessment of *in vitro* DNA degradation. Within the 95°C heated culture by Day 6, genome copy number had reduced by close to sevenfold and by this point was > 1000 fold less than the genome copy number of the control sample.

across both 6 and 8 days to determine which format would provide an optimal assay length for PMR quantitation (Figure 6.1). The 6 day assay estimate consistently gave a slightly higher estimate of PMR across all parasite starting densities and was thus selected for use in later experiments (Table 6.3). Although not a primary aim of this investigation, it should be noted how the rates of growth were lower for assays at higher starting parasitemia, as these reached density limiting conditions by the end of the assay. This suggests that *in vitro* PMR could be highly sensitive to parasite density, with the low starting parasitemia adopted here designed to minimise any such effect. To offer a compromise between the least density affected rate of growth over 6 days, minimising any potential effect of variation from initiating low parasite density cultures and keeping the vast majority of erythrocytes at Day 0 as fresh as possible, the starting parasitemia was set at 0.02% for the subsequent experiments.

6.3.2 Quantifying growth rates of laboratory clones

Six biological replications, with an individual erythrocyte donor being placed into each flask, were set up for quantifying PMR for each of the four laboratory clones used within the study; 3D7A-GFP, Dd2, D10 and HB3 and were monitored over an 8 day period to test whether the slightly lower PMR estimated for the 8 day assay in comparison to the 6 day assay for Dd2 was also observed for the three other laboratory clones. As expected, a lower estimate of PMR was seen for the 8 day assays for all four clones (Figure 6.2), so the 6 day assay was confirmed as appropriate. Although mean PMR as determined through general linear modelling of the four laboratory clones ranged from 7.59 in HB3 to 10.53 in Dd2 per 48 hour period over the 6 day assay (Table 6.4), the calculated confidence intervals for all four laboratory

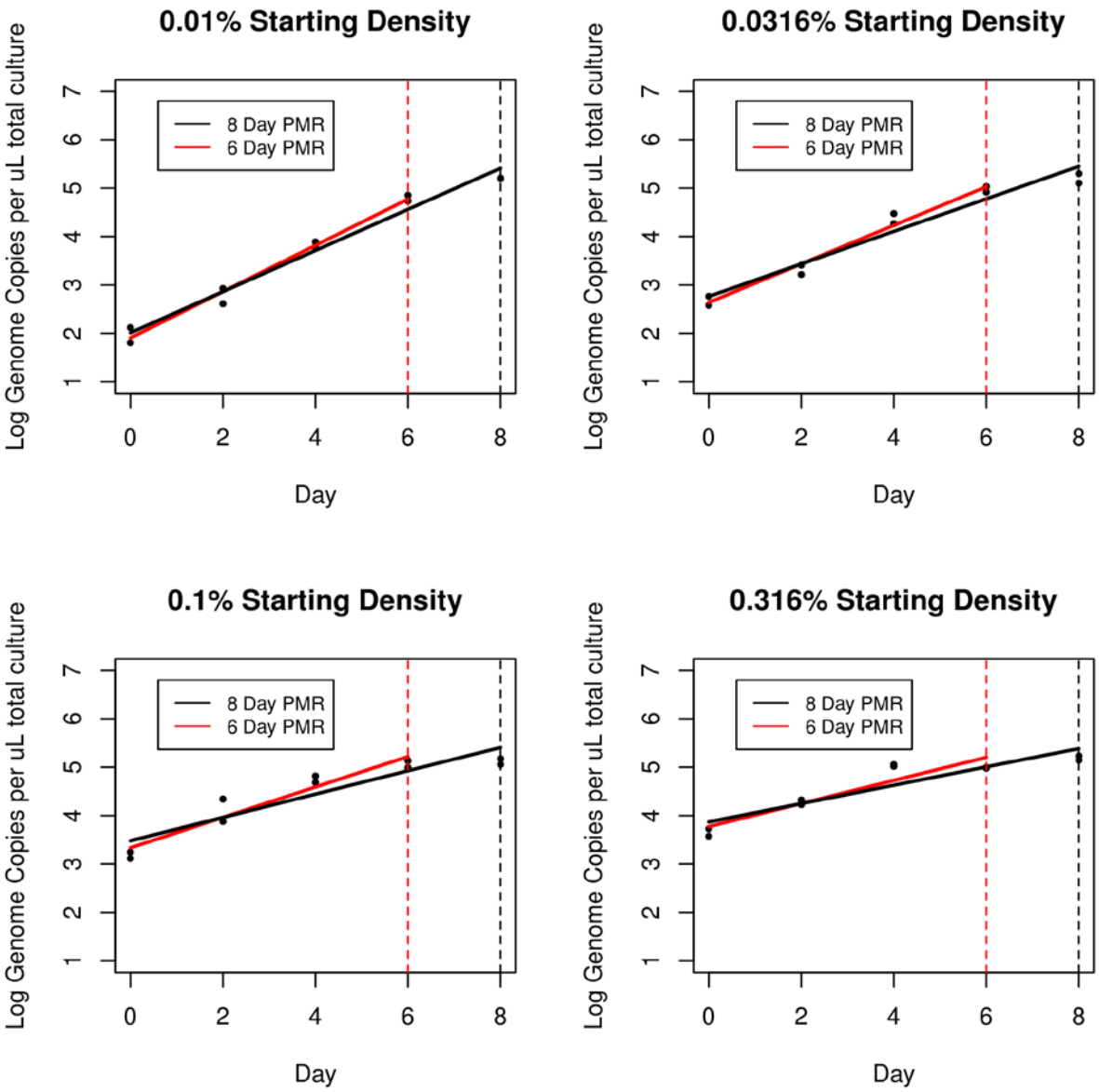


Figure 6.1 Effect of starting density on quantified PMR estimated over either 6 or 8 days *in vitro* from two biological replicates using laboratory clone Dd2. General linear models were fitted across the biological duplication for 6 days (as represented by the red line and dashed red x-axis cut-off) and 8 days (as represented by the black line and dashed x-axis cut-off). At all starting densities, PMR is estimated to be slightly higher when measured over 6 days rather than 8 days (Table 6.3).

Initial Density	Days	PMR	Upper 95% CI	Lower 95% CI
0.01 %	6	9.02	11.81	6.90
0.01 %	8	7.08	9.10	5.51
0.0316 %	6	6.26	8.22	4.77
0.0316 %	8	4.70	6.27	3.52
0.1 %	6	4.25	6.44	2.80
0.1 %	8	3.03	4.38	2.10
0.316 %	6	2.39	4.56	1.97
0.316 %	8	2.39	3.25	1.76

Table 6.3 Parasite multiplication rates (PMR) per 48 hours estimated for the Dd2 clone performed in biological duplicate, with upper and lower 95% confidence intervals for assays measured over 6 and 8 days for four different starting parasite densities.

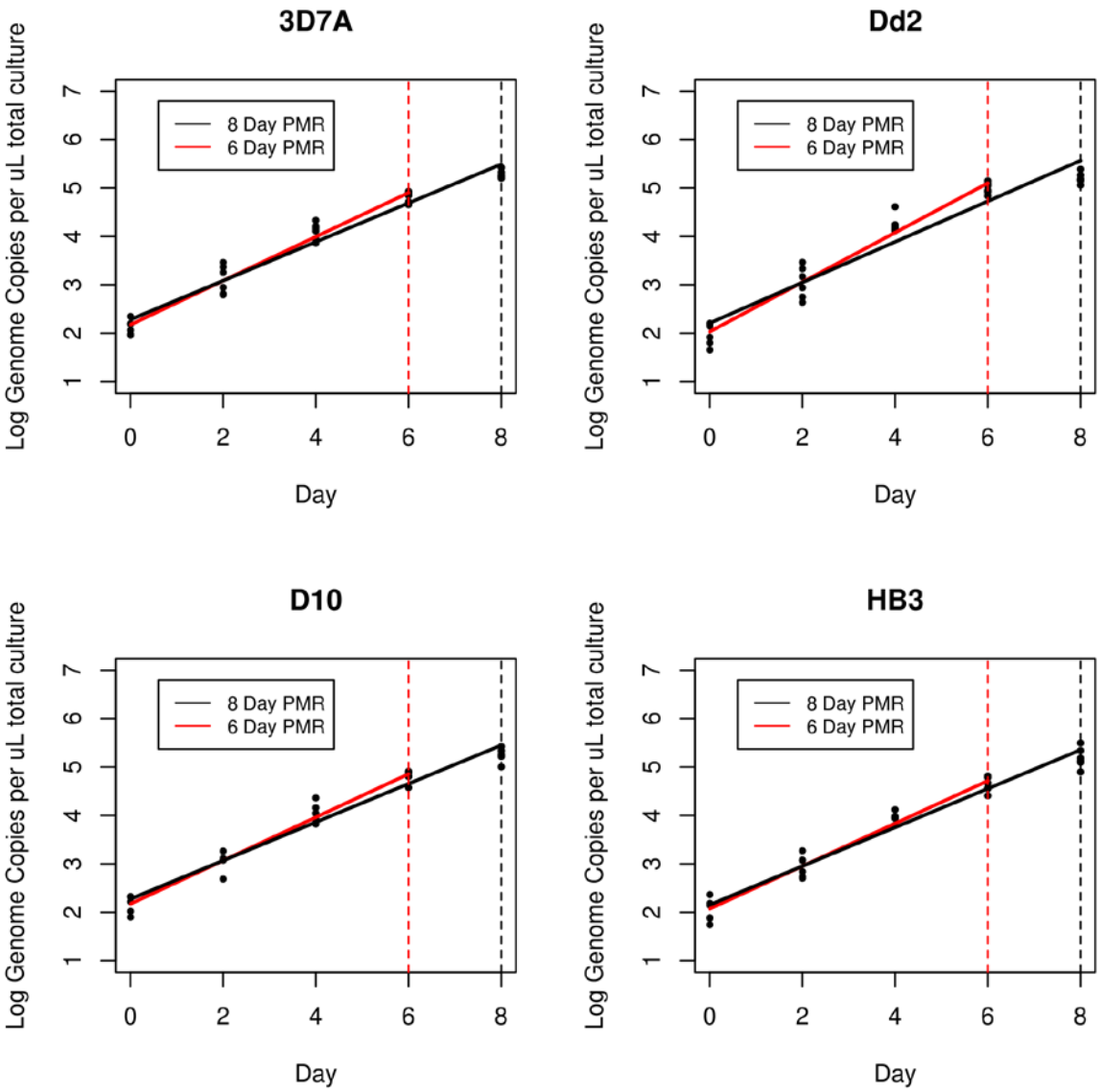


Figure 6.2 Quantifying parasite multiplication rates for four laboratory clones of *P. falciparum*. Whilst 8 day assays were performed, as also seen within Figure 6.1, the Day 8 measurement of genome copy density appeared to be affected by a possible density-dependent reduction in PMR in this period. Thus PMR was calculated over 6 days as indicated through fitting of a general linear model across the biological replications, as represented by the red line and dashed red x-axis cut off.

Clone	PMR	Upper 95% CI	Lower 95% CI	Coefficient of determination (R^2)
3D7A-GFP	8.09	9.64	6.79	0.97
Dd2	10.53	13.07	8.48	0.96
D10	7.83	9.41	6.52	0.97
HB3	7.59	9.00	6.40	0.97

Table 6.4 Parasite multiplication rates (PMR) per 48 hours for the four laboratory clones (with upper and lower 95% confidence intervals) quantified over the duration of a 6 day assay.

clones overlapped (3D7A-GFP 95% CI [6.79 – 9.64], Dd2 95% CI [8.48 – 13.07], D10 95% CI [6.52-9.41], HB3 95% CI [6.40 – 9.00]).

In addition, correlation of the microscopy count data and qPCR estimation was highly significant (Pearson's r = 0.93, p-value < 1 x 10^{-10}), whilst the mean difference and 95% confidence limits of agreement between the two methods, when the microscopy measurement was subtracted from the qPCR based measurement, was 0.216 log units (-0.454 log units to 0.885 log units) (Figure 6.3), which was similar to previously published comparisons of microscopy and qPCR (Nwakanma *et al.* 2009). The slightly higher measurement seen could potentially be explained by the sensitivity of the qPCR assay, which could include the quantitation of genome copies which have been generated during asexual replication, but have not successfully invaded or developed within a red blood cell and have subsequently not yet been degraded within the culture, or alternatively the presence of mature trophozoites which have begun to undergo mitosis, but not yet progressed to the schizont stage.

6.3.3 Quantifying growth rates of *ex vivo* clinical isolates

Of the 19 clinical isolates that were put into the 6 day estimation of PMR following at least two weeks of *ex vivo* culture adaptation (Table 6.5), two isolates were characterised as having not grown successfully over the course of the assay, as determined by possessing a PMR of <1 (INV221 and INV243). Coefficient of determination values for mean PMR across the remaining 17 isolates were generally high, with R^2 values for fourteen isolates > 0.75. This was implemented as a threshold above which the general linear model fit was considered as acceptable, which three other isolates (INV228, INV248 and INV264) did not meet. This threshold was well within the R^2 range of previous logistic *in vivo* (Bejon *et al.* 2005) and linear *in vitro* (Petersen *et al.* 2015) model

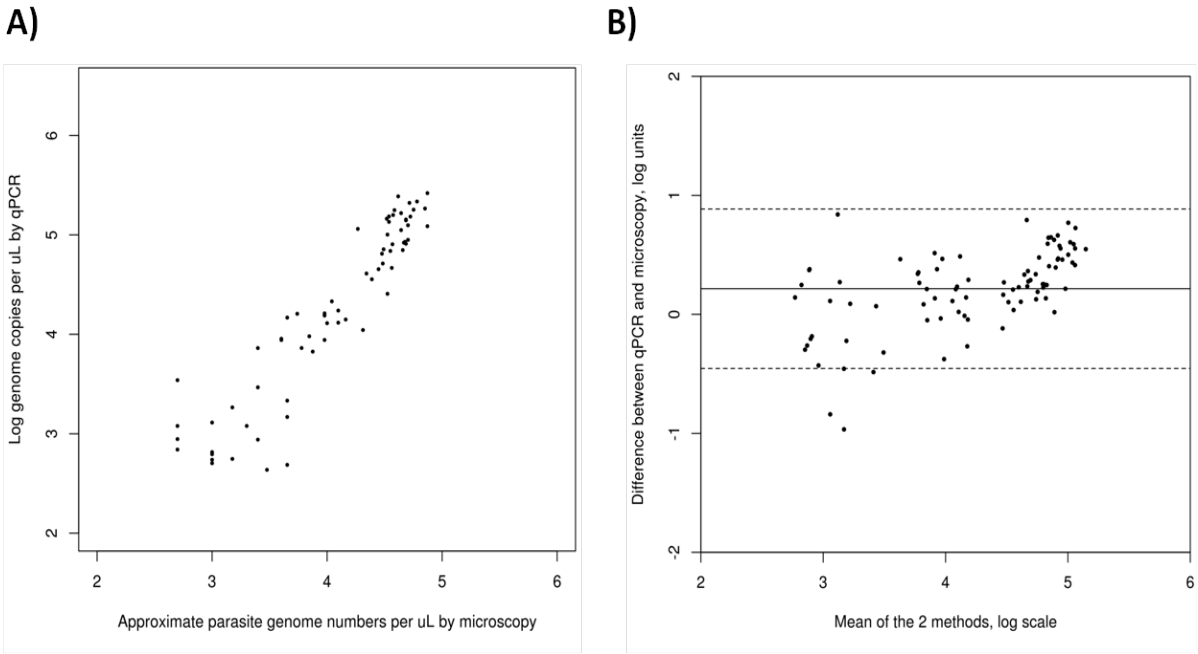


Figure 6.3 Concordance between parasite numbers as determined by giemsa-based microscopy and qPCR based quantitation. A) Correlation plot of the two measures is highly significant from comparison of 90 measurements for four laboratory clones across the 8 day assay format (Pearson's r = 0.93, p-value < 1 x 10^{-10}). B) Bland-Altman plot where dashed lines represent 95% confidence limits of agreement between the two methods (using a mean ± 1.96 x standard deviation approach), where the solid line represents the mean difference of the microscopy approach from the qPCR approach.

Isolate ID	Days *ex vivo*	PMR	95% Upper CI	95% Lower CI	R^2 value
3D7A	NA	8.79	9.80	7.88	0.95
INV216	32	1.97	2.56	1.51	0.78
INV218	32	3.15	4.22	2.35	0.88
INV220	32	2.45	3.03	1.79	0.90
INV221	32	0.65	0.85	0.50	0.56
INV226	25	2.12	2.59	1.73	0.87
INV231	18	3.79	5.38	2.67	0.87
INV228	25	1.12	1.30	0.98	0.41
INV232	25	2.52	3.16	2.01	0.94
INV243	18	0.72	0.98	0.53	0.53
INV246	16	4.09	5.55	3.01	0.92
INV247	16	2.48	2.96	2.08	0.93
INV249	16	3.68	4.94	2.74	0.91
INV250	16	3.87	4.24	3.53	0.99
INV251	16	2.23	2.60	1.91	0.93
INV248	31	2.01	5.08	0.79	0.36
INV264	15	1.26	1.42	1.11	0.67
INV211	35	2.21	2.75	1.78	0.88
INV215	37	3.56	4.95	2.57	0.89
INV035	15	2.47	3.36	1.82	0.90

Table 6.5 Parasite multiplication rates (PMR) per 48 hours for nineteen clinical isolates (with upper and lower 95% confidence intervals) quantified over the duration of a 6 day assay. Isolates were put into the assay between 15 and 37 days during *ex vivo* adaptation, depending on erythrocyte donor availability.

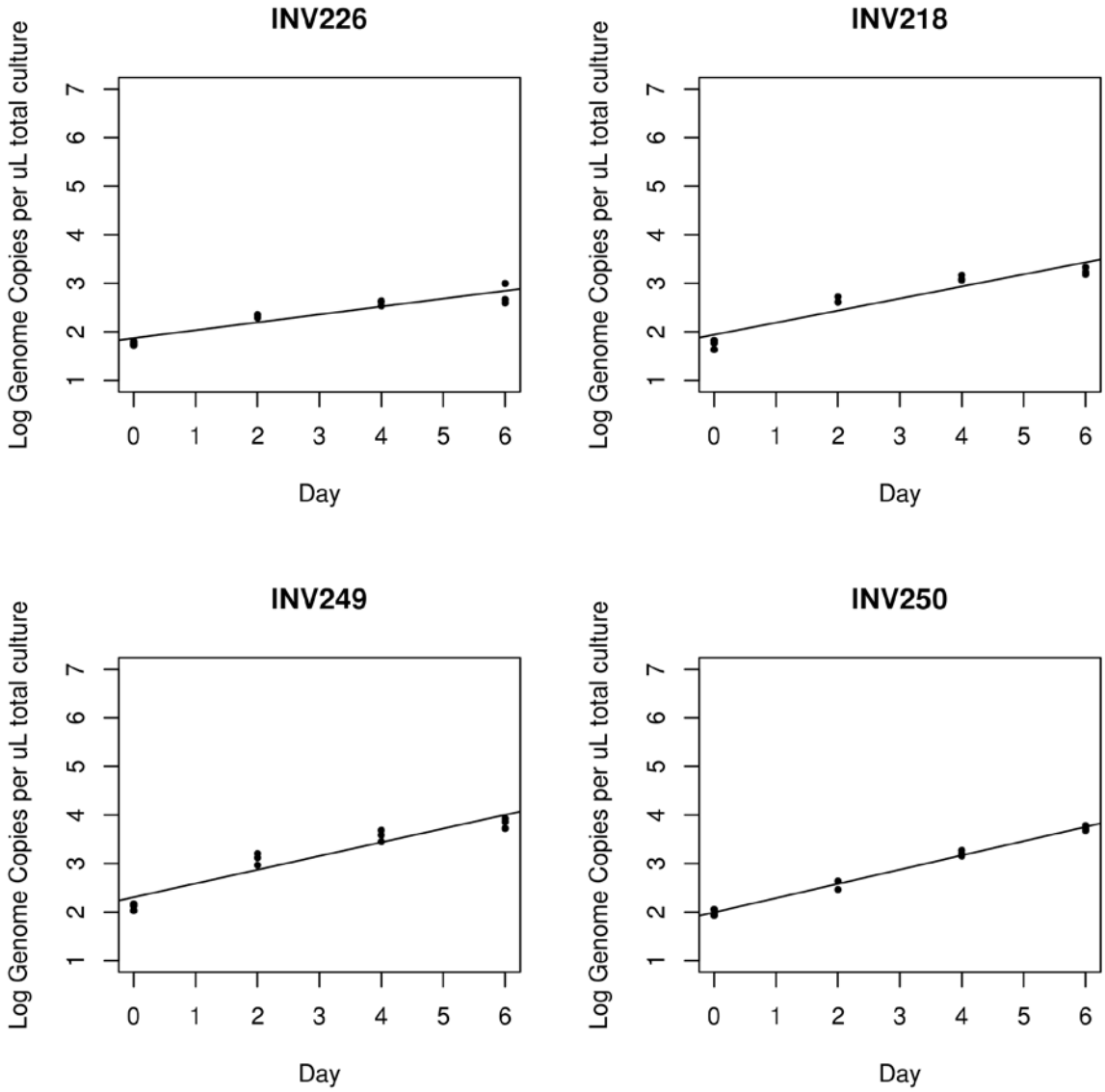


Figure 6.4 Quantifying parasite multiplication rates for a selection of four clinical isolates. PMR was quantified across the 6 day assay format previously described, to provide a measurement through multiple asexual replication cycles.

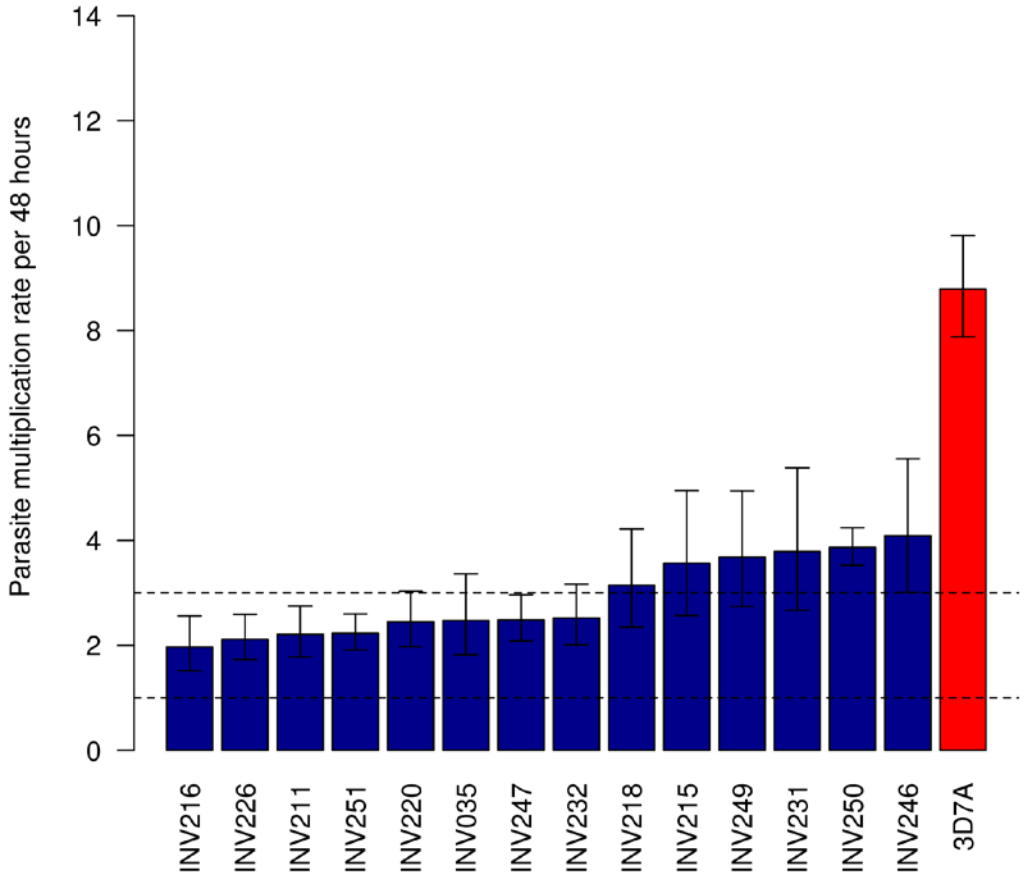


Figure 6.5 Parasite multiplication rates (PMR) per 48 hours for 14 West African clinical isolates quantified over 6 days using a general linear model approach. All isolate PMRs were quantified within biological triplicate, except for INV232 and INV035, which were performed in biological duplicate. The 14 isolates shown within this figure were those from table 6.6 with a PMR calculated to be >=1 fold per cycle, with dashed horizontal lines representing a PMR of 1 and 3. Mean PMR as calculated across the 14 clinical isolates was 2.90.

agreements for calculation of PMR for *P. falciparum* infections. The replication rates amongst these 14 remaining isolates ranged from 1.97 for isolate INV216to 4.10 for isolate INV246, with a mean PMR of 2.90 (Figures 6.4 and 6.5). In addition, four locus microsatellite genotype profiles that were generated for these 14 clinical isolates defined to have grown during the assay from the extracted DNA material used for qPCR analysis confirmed that all 14 isolates were genotypically distinct and had not been cross contaminated during culture adaptation (Table 6.6).

6.3.4 Repeated profiling of PMR over the *ex vivo* culture adaptation process

For 6 of the initial 19 isolates that were put into the growth assay, parasites which were still visible by giemsa stained slide microscopy a minimum of four weeks after they had originally had their PMR profiled were put into another assay and PMR was determined across 6 days as previously seen (INV035, INV221, INV231, INV246, INV249 and INV250), with INV035 being profiled at two timepoint and all others isolates being profiled at three timepoints. These assays were subjected to the previously described filtering procedures, which led to the removal of one biological replicate within the final temporal PMR profile generated for each of the six isolates, due to this replicate not appearing to support growth within the 3D7A-GFP laboratory clone.

Visualisation of these longitudinal PMR estimations during *ex vivo* adaptation, revealed that INV221 and INV035 had increase in growth rate during this period (Figure 6.6), however the other four isolates had maintained a relatively stable PMR throughout the adaptation process, although confidence intervals generated for the final temporal point were large due to having only been calculated across a biological duplicate. This analysis methodology however

Sample ID	TA1	PolyA	PfPK2	TAA109
INV216	169	156	163	NA
INV218	166	159	172	172
INV220	160	153	160	160
INV226	166	156	160	160
INV231	166	156	169	160
INV232	160	156	172	160
INV246	166	**168,** 162	**166**, 169	175
INV247	166	117	160	160
INV249	163	159	166	160
INV250	**196**, 178	159	**163**, 156	175
INV251	**160,** 166	**153**, 156	**169**, 160	160
INV211	166	156	172	**160**, 196
INV215	160	156	172	160
INV035	163	168	169	172

Table 6.6 Microsatellite genotype profiles for four loci for 14 clinical isolates that had their PMR quantified. The numbers under the four loci refer to the allelic trinucleotide bin size that was recorded using Genescan. Where two separate numbers are present for the same isolate at the same loci, multiple alleles were scored, with the major allele present in bold typing. For isolate INV216, the TAA109 locus failed to produce an allele call, but its three locus profile confirmed it as distinct from the other isolates.

did not account for any potential batch heterogeneity of culture media or erythrocyte donor on influencing isolate PMR. Due to this, a normalised PMR estimate for each isolate was determined using the PMR data obtained from the laboratory clone control that was run alongside every biological triplicate performed. For this, all independent rates of 3D7A-GFP growth as calculated were averaged (Table 6.6). The difference between the growth rate of the clinical isolate and its particular 3D7A-GFP control estimation was calculated and used to subsequently normalise against the averaged rate of 3D7A-GFP growth across all relevant estimations of 3D7A-GFP growth. Hence for isolate INV231, this was done using 3D7CG1, 3D7AG5 and 3D7CG7 (Table 6.6), whilst for isolates INV246, this was done using 3D7CG4, 3D7CG6 and 3D7CG7. Following this normalisation process (Figure 6.7), across all repeated isolates there was a significant increase in normalised PMR after a minimum of 4 or 8 additional weeks undergoing *ex vivo* culture adaptation (Wilcoxon rank sum test; Timepoint 1 vs Timepoint 2 p- value <0.005, Timepoint 1 vs Timepoint 3 p-value <0.003), although no difference was seen between the second and third timepoints recorded (Wilcoxon rank sum test p-value =0.79). The mean PMR values across all repeated isolates for the three normalised timepoint data sets were respectively 2.72, 4.57 and 4.86.

Assay ID	Isolate ID	Days *ex vivo*	*Ex vivo* repeat	PMR	Upper 95% CI	Lower 95% CI
CG1	3D7A-GFP			8.83	11.43	6.83
	INV221	32	1	0.65	0.85	0.50
	INV231	18	1	3.79	5.38	2.67
CG4	3D7A-GFP			10.86	14.70	8.01
	INV246	16	1	4.09	5.55	3.01
	INV249	16	1	2.48	4.94	2.74
	INV250	16	1	3.87	4.24	3.53
CG5	3D7A-GFP			8.25	10.67	6.38
	INV221	81	2	4.00	4.97	3.21
	INV231	67	2	5.47	6.64	4.51
CG6	3D7A-GFP			8.00	9.53	6.72
	INV246	45	2	4.04	4.54	3.60
	INV249	45	2	3.22	3.77	2.76
	INV250	45	2	4.96	6.06	4.06
CG7	3D7A-GFP			7.53	9.57	5.93
	INV035	15	1	2.47	3.36	1.82
	INV221	114	3	4.49	5.54	3.64
	INV231	100	3	4.56	6.82	3.06
	INV246	76	3	3.04	8.08	1.14
	INV249	76	3	3.49	5.26	2.32
	INV250	76	3	3.90	5.32	2.87
CG8	3D7A-GFP			9.16	11.88	7.06
	INV035	114	2	7.80	10.09	6.03

Table 6.7 Repeated profiling of *ex vivo* PMR for six clinical isolates through the culture adaptation process. The Isolate ID shows the isolate and the Assay ID the biological triplicate (e.g. CG7) in which their profile was taken. For all rows shown, biological triplicates were used for calculation of PMR, except for all CG7 isolates for which biological duplicate was used.

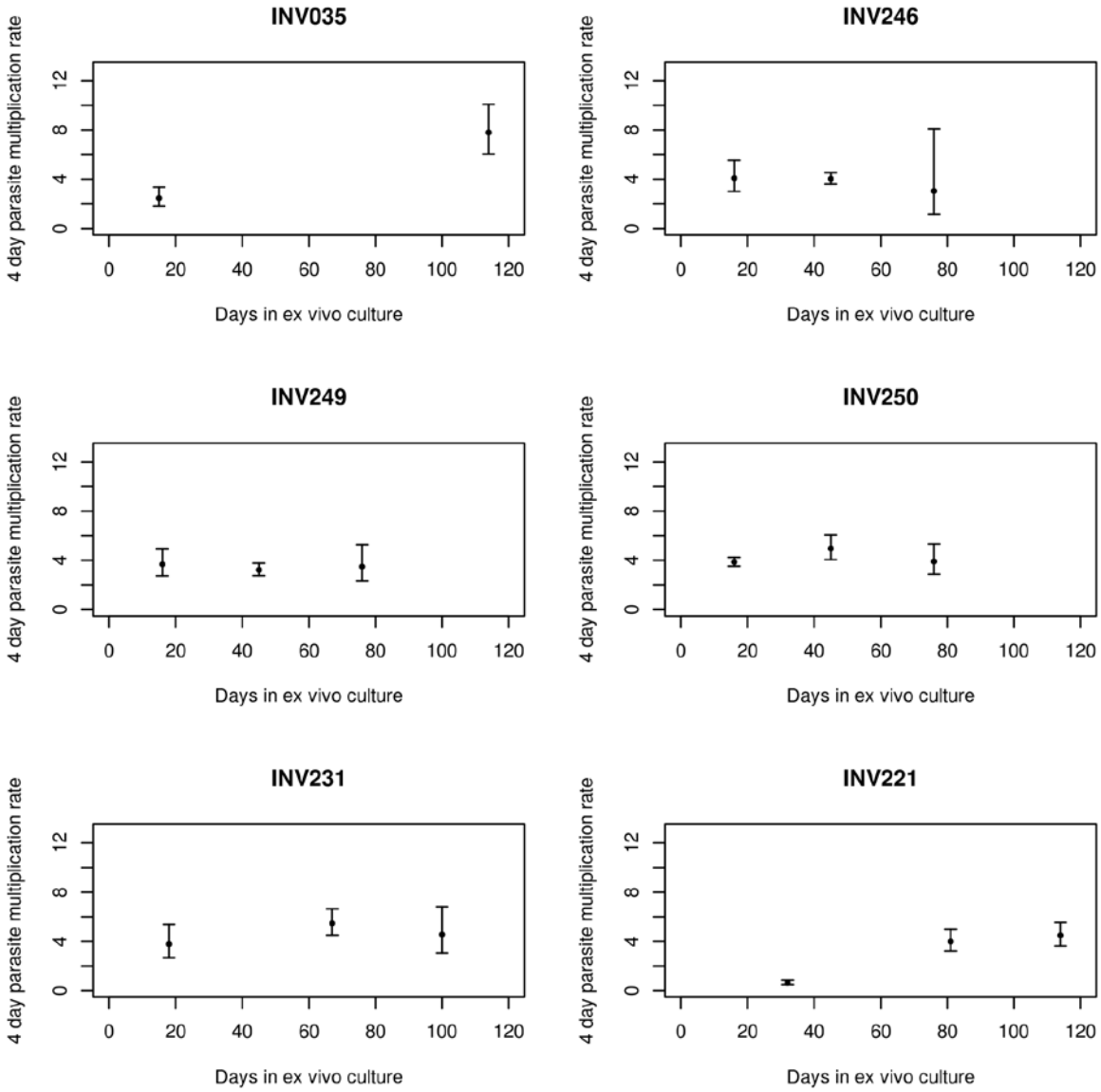


Figure 6.6 Profiles of 6 *ex vivo* clinical isolate which had their PMR quantified more than once during the culture adaptation process. Growth rates were requantified for 5 clinical isolates three times and twice for one isolate, with a minimum of 4 weeks between the sequential time points taken. The x-axis refers to how long each isolate had been in culture on the day each growth assay was initiated.

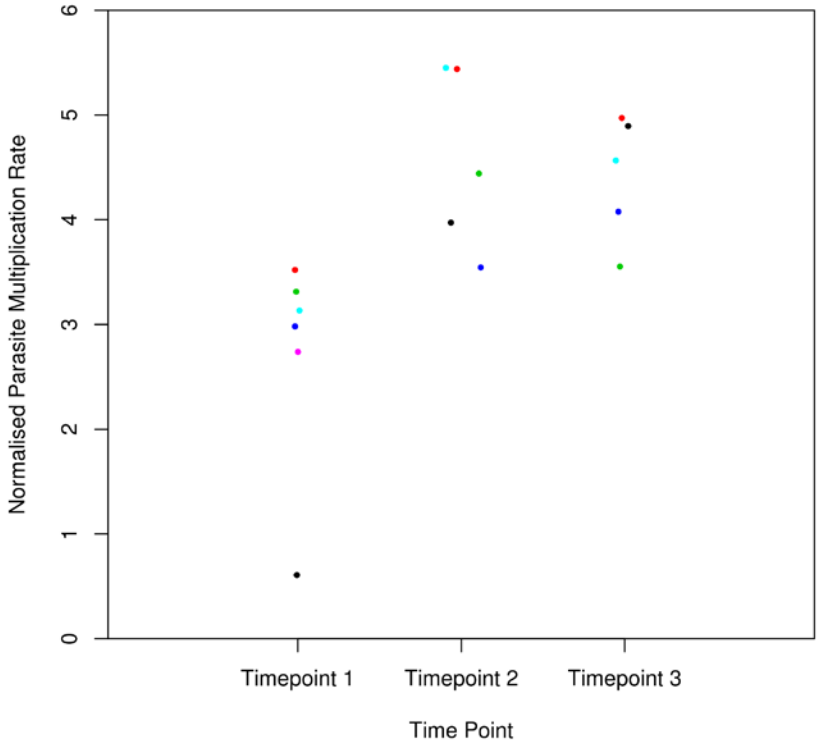


Figure 6.7 Normalised repeated profiling of clinical isolates during the *ex vivo* culture adaptation process. Normalised PMR for the six clinical isolates within Figure 6.6 against the 3D7A laboratory clone control present within set of biological replicates. The minimum period between time points was 4 weeks, with highly significant differences seen between timepoint 1 and timepoints 2 and 3. Each colour across the three different timepoint represents the same isolate.

6.4 Discussion

This chapter focuses on quantifying the variation of an important phenotype of clinical isolates of *P. falciparum* that has been previously relatively neglected; *ex vivo* PMR during the culture adaptation process. *Ex vivo* culture adaptation of clinical isolates of *P. falciparum* remains challenging even over short time periods (Nsobya *et al.* 2008) as underlined by the initially low PMR reported amongst clinical isolates within this study. Only 31.1% (n=19 out of 61) of thawed clinical isolates reached the first timepoint for *ex vivo* growth phenotyping, which was taken between 16 and 37 days in *ex vivo* culture based on the availability of triplicate blood donors at the same point. Of these nineteen isolates, 31.5% of isolates (n=6) successfully continued to adapt to *ex vivo* culture and had at least one additional point at which their *ex vivo* growth rate was quantified, with 5 of them having two additional timepoints profiled. These proportions of successful *ex vivo* culture adaptation are in agreement with a previous study which measured copy number evolution during *ex vivo* culture adaptation (Dr Ian Cheeseman, LSHTM PhD thesis 2010). Whilst extrinsic factors such as isolate storage conditions can affect the cultivability of clinical isolates of *P. falciparum*, the potential contribution of other possible intrinsic phenotypic, genomic or transcriptional properties of parasites collected from natural populations has not been investigated widely, although a recent study of clinical Indian isolates suggests that *ex vivo* 'cultivability' of specific isolates is consistent from different cryopreservation aliquots within separate erythrocyte batches (White Jr *et al.* 2016).

Within this chapter, standardised quantitation of *ex vivo* PMR with biological replication over multiple intraerythrocytic cycles was performed for a set of 19 West African isolates. Previous reports of the multiplication rates of *P.*

falciparum clinical isolates have been recorded at the first *ex vivo* cycle from South East Asian and African populations (Chotivanich *et al.* 2000; Deans *et al.* 2006), reporting vastly different mean PMRs of 8.3 and 2.3 respectively. The PMR seen within this chapter is thus much closer to the low multiplication rates previously reported for the Malian and Kenyan populations performed by Deans *et al.* (2006). These previous studies were however performed without biological replication of erythrocyte donors, although they do provide an *ex vivo* measure of PMR after the shortest possible period of *ex vivo* adaptation. Whilst *in vivo* PMR of *P. falciparum* has previously been calculated from blood smear microscopy (Simpson *et al.* 2002) to show approximately 8-fold rates of growth, similar to those exhibited by the four laboratory clones profiled here, the calculated mean PMR of the *ex vivo* isolates was much lower than this at 2.90 fold.

Moreover, for the isolates that continued to successfully adapt to *ex vivo* culture, PMR significantly increased when normalised against the PMR of the 3D7A-GFP control profiled. Transcript expression changes of the erythrocyte membrane protein 1 (PfEMP1) during *ex vivo* culture adaptation have been previously reported (Peters *et al.* 2007; Zhang *et al.* 2011), as the parasite grows in a vastly different nutritional and physiological environment (Le Roux *et al.* 2009). Epigenetic mediated regulation of clonally variable transcriptional variation has been reported within a range of laboratory adapted parasite clones (Rovira-Graells *et al.* 2012), but has yet to be examined within *ex vivo* clinical isolate culture adaptation. Such a study would be complicated however by the difficulties of lower parasite multiplication rates and the need for tight synchronisation of parasites, in addition to discerning the potential contributions of separate genotypes within the same infection.

The phenotyping performed within this chapter broadly quantifies parasite growth rate over a set 48-hour period to enable comparison between different isolates. There are several other cellular parasite features that could influence parasite growth rate such as merozoite numbers per schizont, intraerythrocytic duration and red blood cell invasion efficiency. For instance, the laboratory clone Dd2 has a mean intraerythrocytic cycle length of 44 hours and mean merozoite number per schizont of 18, whereas HB3 takes 50 hours per intraerythrocytic cycle and possesses a mean of 16 merozoites per schizont respectively, whilst invasion efficiency of Dd2 is approximately 35% of merozoites compared to 25% for HB3 (Reilly *et al.* 2007). Any potential spectrum of these intrinsic, cellular parasite features has not been characterised for *ex vivo* isolates and thus not previously correlated with *ex vivo* PMR. Although more challenging with *ex vivo* isolates due to the lower rates of PMR seen, a potential method to examine at least the variation of merozoite number and cell cycle duration would be to use a combination of early ring-stage sorbitol synchronisation, careful monitoring of parasite development for each isolate through the intraerythrocytic cycle and use of the E64 protease inhibitor (Hill *et al.* 2014) to block merozoite egress from the schizont and provide a simultaneous count of merozoite numbers from a proportionally enriched schizont-stage culture and through regular giemsa-stained smearing of the development cycle for each isolate, determine a distribution of cell cycle duration for each isolate, where the read out of the experiment would be the proportion of parasites that possess multi-nucleate schizonts at that timepoint post-invasion.

Quantitative information on the variation seen within natural populations *P. falciparum* is important baseline parasitological information and is the first

necessary step of any clinical isolate forward genetic based screens (Anderson *et al.* 2010; Cheeseman *et al.* 2012) to identify potential genomic or transcriptional markers that underlie phenotypic variation within traits of *P. falciparum* from natural populations such as *ex vivo* PMR or gametocyte commitment (section 1.7, section 5 and section 8.4).

7. Development of tools and assays to investigate potential competitive interactions between *P. falciparum* genotypes in culture

7.1 Introduction

Multiple-genotype infections of *P. falciparum* are widespread, particularly within regions of high infection endemicity such as West Africa (Mobegi *et al.* 2012; Manske *et al.* 2012). Furthermore, these multiple-genotype infections are further complicated by the presence of a spectrum of relatedness between different genotypes (section 1.5.1). Several studies within rodent and lizard *Plasmodium* species have (section 1.6.1 and 1.6.2) investigated potential genotype interactions. A paucity of studies have however extended this research to an *in vitro* system of *P. falciparum*. Two studies have examined the potential effect of competition on *in vitro* growth rates for two clones each using qPCR based approaches, but came to different conclusions. One study reported a genotype specific competitive suppression effect, which was determined by comparing differences in the end point densities of competition assays between Dd2 and HB3 with monoculture control cultures (Wacker *et al.* 2012). The other published study however emphasises an issue in assessing such potential interactions, as it is difficult to separate the direct response of a parasite due to the presence of a competitor with potential indirect responses caused by possible density dependence and reported that alterations to the growth rate of two laboratory clones, Dd2 and K1, was due to the additive density of all clones present within the culture, rather than genotype specific suppression (Zervos *et al.* 2012).

If, as suggested for *P. chabaudi* and *P. mexicanum*, the parasite is able to induce a phenotypic response due to the genotypic composition of the infection (Reece *et al.* 2008; Neal & Schall 2014), it raises the prospect of a variety of

other transcriptional responses due to the presence of different genotypes that have not been investigated previously in *P. falciparum* due to the difficulty of physical clone separation. Whilst potential changes in gametocyte commitment could be investigated, such a genotype specific fluorescence system would allow for the investigation of any transcriptional changes associated with genotype competition, potentially across the parasite life cycle. Mechanisms that are required for any non-self recognition system mediating transcriptional responses would require the strong balancing selective forces seen operating on *P. falciparum* populations to maintain the diversity required for genotype distinction (Rousset & Roze 2007). Synchronisation of phenotypic responses amongst clone mates would also likely require efficient intercellular communication pathways, which could possibly exist in microvesicle trafficking systems between infected red blood cells that have been found to promote sexual commitment (Dyer & Day 2003; Mantel *et al.* 2013; Regev-Rudzki *et al.* 2013). To aid such studies, the most versatile fluorescent parasite tools for *in vitro* competition would have constitutive expression of a fluorescent protein across the entire parasite life cycle, as previously reported for a transfected clone of 3D7A (Talman *et al.* 2010). In addition, an integrated copy of the plasmid within the parasite genome would be desired due to cultures containing episomally fluorescent transfected parasites potentially consisting of a significant proportion of non-fluorescent parasites (Babbitt *et al.* 2012).

Within this chapter, I use allele-specific qPCR assays designed to assess competition in an alternative manner to other *Plasmodium* studies, by focusing on growth rate of each individual genotype within a mixed *in vitro* culture. If competitive suppression is a genotype-specific effect and not due to potential density-dependent effects, it can be hypothesised that any potential

suppressive effect would be visible by observing a reduced PMR of parasites under controlled *in vitro* competition assays across multiple cycles.

I also hypothesise that the production of different laboratory clones expressing differentially coloured fluorescent proteins (Gibson *et al.* 2008; Regev-Rudzki *et al.* 2013;) would be a novel tool to further investigate potential genotype interactions of *P. falciparum* in an *in vitro* setting. Such an approach would allow for study of direct parasite-parasite interactions from potentially confounding factors that are present within mouse malaria models such as immune-mediated competition (Raberg *et al.* 2006; Santhanam *et al.* 2014; Abkallo *et al.* 2015). Thus, I developed a small panel of fluorescent parasite genotypes expressing different fluorescent proteins, with the aim of integration at the Pf47 locus (Talman *et al.* 2010), for potential *in vitro* investigation of genotype competition in *P. falciparum*.

7.2 Materials and Methods:

For *in vitro P. falciparum* competitive growth experiments, laboratory clones 3D7A-GFP, Dd2, D10 and HB3 (section 2.3.1) were thawed (section 2.3.2) and cultured (section 2.3.4). The pan-genotype and allele-specific and qPCR assays that were used were either found following a literature review or designed with the aid of Primer 3 (section 2.5.3). Due to the use of SYBR green for all qPCR assays, all qPCR assays run underwent melt-curve analysis using Rotor-Gene 6 software to ensure specificity of the generated amplicons (Table 7.1). For the melt-curve, post-amplification products were subjected to an additional ramping melt step up to 99 degrees centigrade. Only single peaks were generated by

Primer Identity	Primer Sequence (5' to 3')	Amplicon Size (bp)	Target Gene	Primer usage (target clone in bold)	Origin of primers
Pfs25-F	CCATGTGGAGATTTTTCCAAATGTA	143	Pfs25	Pan-genotype	Schneider *et al*. 2015
Pfs25-R	CATTTACCGTTACCACAAGTTACATTC				
HB3 MSP1-F	GAAATTGCCAAAACTATTAAATTTAATATTGATAG	133	MSP-1	**HB3** + Dd2, **HB3** + 3D7, **HB3** + D10	Wacker *et al*. 2012
HB3 MSP1-R	GGTTCAGTTGATTCCTTTGTTTCAAC				
Dd2 MSP1-F	AATTGCCAAAACTATTAAATTTAACATTGATAG	131	MSP-1	HB3 + **Dd2,** HB3 + **3D7**, HB3 + **D10**	Wacker *et al.* 2012
Dd2 MSP1-R	TGAACAGATTTCGTAGGATCTTGTGA		MSP-1		Wacker *et al.* 2012
K1 MSP1-F	AAATGAAGAAGAAATTACTACAAAAGGTGC	235	MSP-1	**3D7** + Dd2, **3D7** + D10	Chen *et al.* 2014
K1 MSP1-R	GCTTGCATCAGCTGGAGGGCTTGCACCAGA		MSP-1		Chen et al/ 2014
MAD20 MSP1-F	AAATGAAGGAACAAGTGGAACAGCTGTTAC	Dd2=205,D10=168	MSP-1	3D7 + **Dd2**, 3D7 + **D10**	Chen *et ai*. 2014
MAD20 MSP1-R	ATCTGAAGGATTTGTACGTCTTGAATTACC		MSP-1		Chen *et al.* 2014
Dd2 MSP6-F	AAGAGCCAACATCGGAGGAATATC	116	MSP-6	**Dd2** + D10	Newly designed
Dd2 MSP6-R	CGGTTTTGAGTATTTGGTCTGGT		MSP-6		Newly designed
D10 MSP6-F	ATCAATATACTGGTACATCTATATCAGGTAT	148	MSP-6	Dd2 + **D10**	Newly designed
D10 MSP6-R	CAGAGAAGTTGTAGTACTATTACTATGAGA		MSP-6		Newly designed

Table 7.1. Allele-specific qPCR assay primer information for assessment of *in vitro* competitive growth.

melt-curve analysis across all assays, indicative of specific, single product generation as previously performed for *Plasmodium* allele-specific assays (Cheesman *et al.* 2003). Moreover, all assays were found not to amplify the non-target *P. falciparum* clone alleles needed to perform allele-specific absolute quantitation in the presence of such non-target alleles. To ensure allele-specific quantitation accuracy, copy number mixtures of extracted non-target and target allele clone DNA were set up in triplicate for all primers but the pan-genotype Pfs25 assay in the ratios 99:1, 90:10, 50:50, 10:90, 1:99. Competitive growth assays were set up using a modification of the standardised growth assay (section 2.6.4) for the six possible combinations in triplicate erythrocyte donors alongside monoculture controls. DNA extractions were subsequently performed (section 2.7.1) and PMR of each clone in competition was determined (section 2.7.3).

For transfection experiments, laboratory clones 3D7A, Dd2 and D10 were used and the pEFGFP plasmid was modified to contain either a mCherry or mKate2 encoding sequence (section 2.8.3). Subsequently, the laboratory clones were transfected (section 2.8.7), live parasite images were taken to confirm parasite fluorescence of the appropriate fluorophores (section 2.8.9) and when integration at the Pf47 locus was detectable by PCR (section 2.8.6) following successive rounds of on/off drug selection, limiting dilution cloning was performed (section 2.8.8). Concordance between allele-specific qPCR and fluorescent microscopy counting was subsequently assessed (section 2.8.10).

7.3 Results

7.3.1 Validation of allele-specific qPCR assays for *in vitro* competition:

From the melt-curve analysis, only single peaks were seen across all assays indicative of specific, single product generation, whilst all assays were also confirmed to not amplify the non-target *P. falciparum* clone alleles needed to perform allele-specific quantitation of pairwise competition (de Roode *et al.* 2003) (Figure 7.1). In addition, these assays were further validated by quantifying known *in vitro* mixtures of target and non-target alleles for each allele-specific qPCR assay over a proportional ratio range of 99:1, 90:10, 50:50, 10:90 and 1:99 (Figure 7.2), which found an extremely high degree of concordance between the observed and expected proportions of quantified DNA (R^2 values for all 6 allele-specific qPCR assays > 0.99). Standard curves produced during the qPCR run for each assay showed efficient levels of amplification across a 1 x 10^6 genome copies to 10 genome copies per µL genomic DNA dilution series (R^2 values were consistently above 0.99 and amplification efficiency above 0.90), whilst negative template controls did not pass the amplification threshold. In addition, for each assay a limit of detection of at least 10 genome copies per µL was recorded.

7.3.2 Quantifying *in vitro* competitive growth rates of *P. falciparum*:

Using the allele-specific qPCR assays, PMR of the four laboratory clones in monoculture and pairwise competition with the three other clones was performed. Importantly, to ensure that parasites within control and experimental flasks experienced the same overall parasite density conditions, initial starting parasitemia of each competition flask was half that of the monoculture control, with the other 50% coming from the presence of the additional competing laboratory clone. Across the three intraerythrocytic cycles from which PMR was

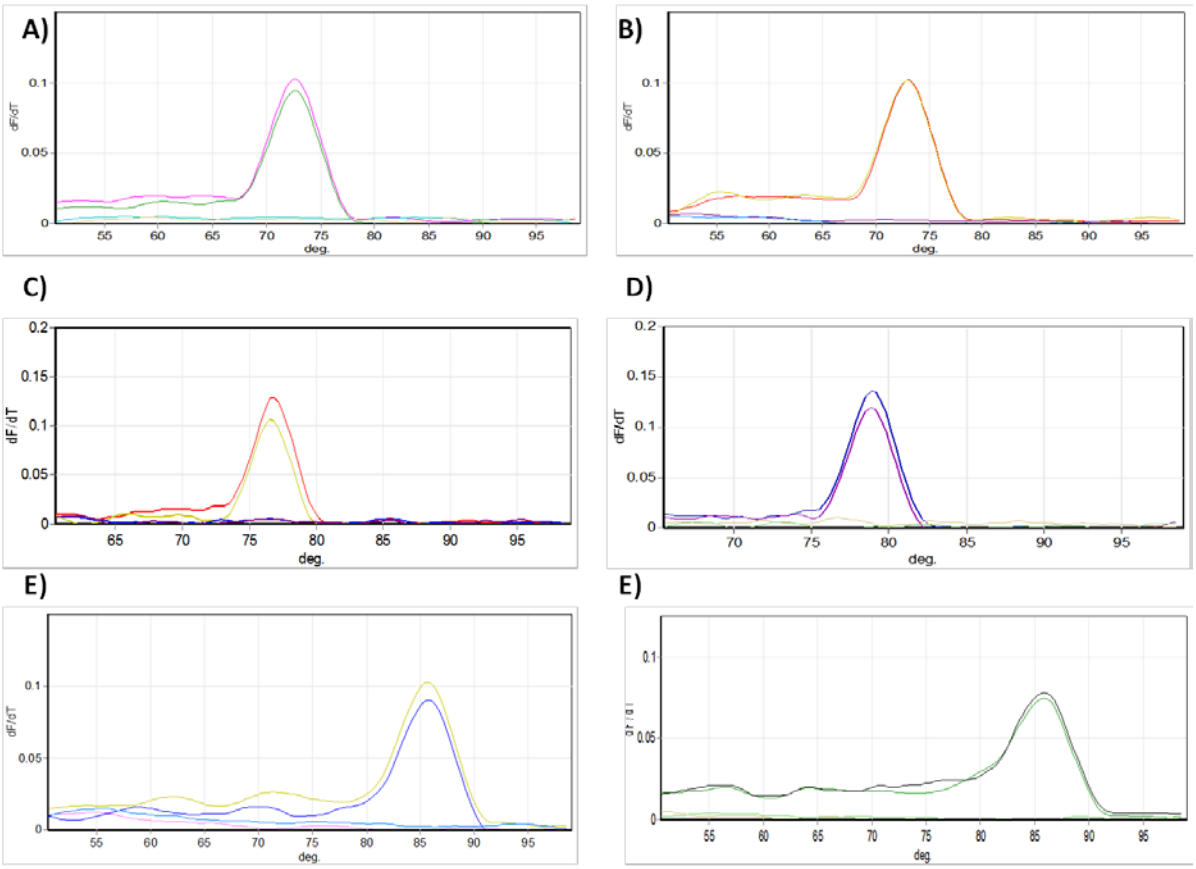


Figure 7.1 Fluorescent melt curve analysis confirms allele-specificity of qPCR assays used for *in vitro* competition studies. Melt curve peaks are shown for each of the amplicons produced by the 6 allele-specific qPCR assays used. Primer sets are labelled as A) HB3 MSP1, B) Dd2 MSP1, C) D10 MSP6, D) Dd2 MSP6, E) K1 MSP1 and F) MAD20 MSP1. Panel E and F does show increased levels of background fluorescence, but this did not affect allele-specific quantitation (Figure 7.2). Within each panel, the target clone was run in duplicate alongside a duplicate of the non-target clone to be distinguished from.

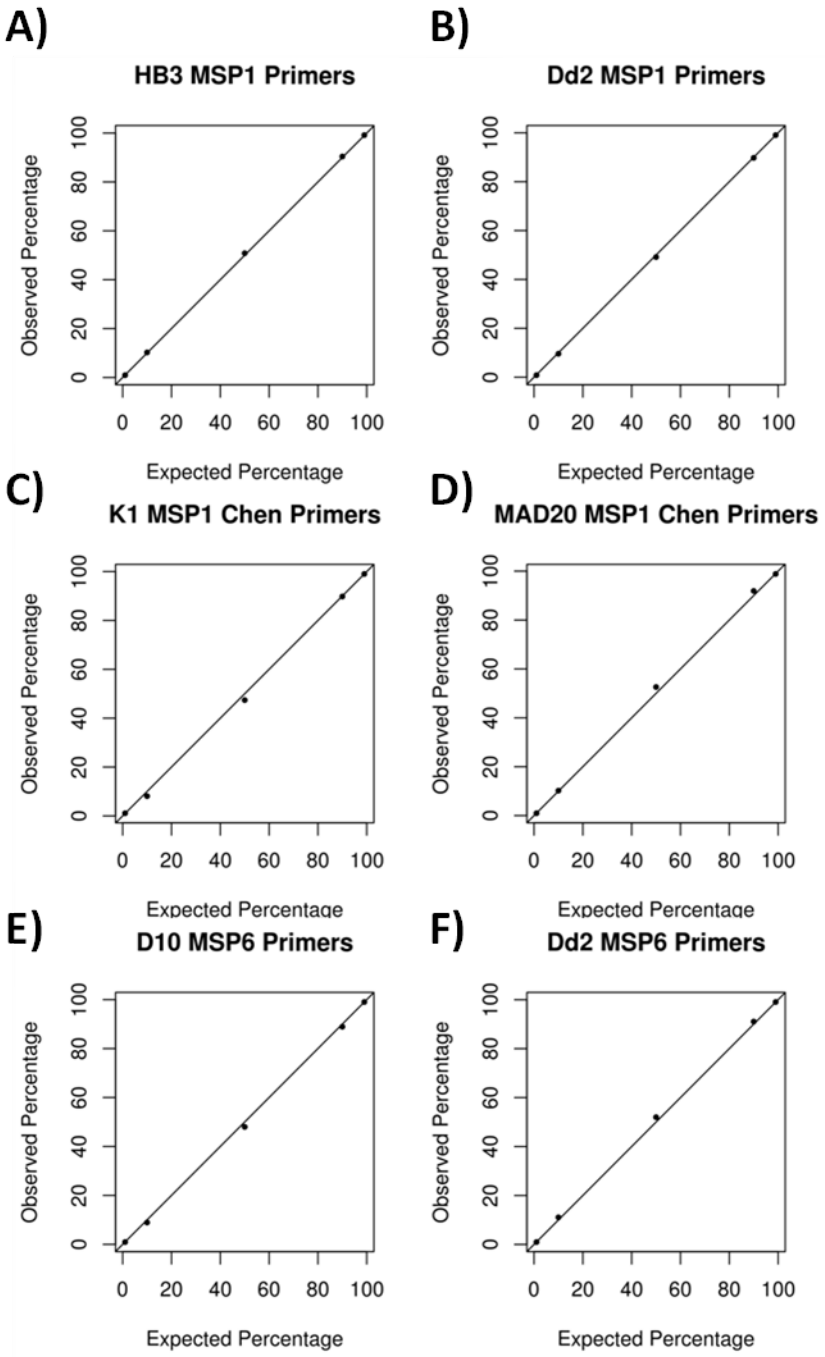


Figure 7.2 Allele-specific quantitation of controlled mixtures of genomic DNA of four laboratory clones. For each allele-specific primer used, control mixtures containing ratios of the two clones to be distinguished between were made up of 99:1, 90:10, 1:1, 10:90 and 1:99. Points represent the mean of three technical replicates, whilst diagonal lines represent where the observed percentage would equal the expected percentage quantified for each clone. Primer sets are labelled as A) and B) quantitation of mixtures of HB3 and Dd2 by HB3 MSP1 and Dd2 MSP1 primers, C) and D) quantitation of mixtures of 3D7A and MAD20-like D10 by K1 MSP1 and MAD20 MSP1 primers and for E) and F) quantitation of mixtures of D10 and Dd2 by D10 MSP6 and Dd2 MSP6 primers, with all R^2 values >0.99.

calculated, no difference was detected between monoculture PMR and PMR of any clone when in competition with any of the three other laboratory clones (Figure 7.3, Table 7.2). Thus, there was no evidence of a significant phenotypic growth response of a clone when placed into *in vitro* growth with another clone at the range of densities experienced over the course of the assay. The absence of this process, termed competitive suppression, within the dataset is in disagreement with a previous report of HB3 being suppressed by Dd2 (Wacker *et al.* 2012), which was determined using comparisons of endpoint genome copy density after 11 days of competition.

7.3.3 Generation of differentially fluorescent *P. falciparum* parasites

From transfection of the three different fluorescent proteins GFP, mCherry and mKate2, fluorescence was monitored by live parasite microscopy. Using an initial drug selection regime of 2.5 μM for pEFGFPLM1 and pEFmCherryLM2, only the 3D7A-mCherry parasite line was found to show evidence of integration at the Pf47 locus. The second round of transfections, used pEFGFP and pEFmKate2LM, since mKate2 was purportedly a brighter fluorescent protein within the red spectral overlap and were subject to a 5 nM WR99210 concentration. Subsequently, 6 parasite lines were confirmed to express the appropriate fluorescent protein (Figure 7.4) and also to show evidence of the Pf47 locus by PCR (Figure 7.5). Two of these lines were successfully cloned to produce 3D7A-mCherry-CL3 and D10-mKate2-CL4. Whilst 3D7A-mKate2-CF11 also did not show PCR evidence of the wild-type Pf47 locus, but would need repeating with a higher parasitemia culture due to the faint integration PCR band recorded (Figure 7.5B, Lane 7). Although the other three lines also

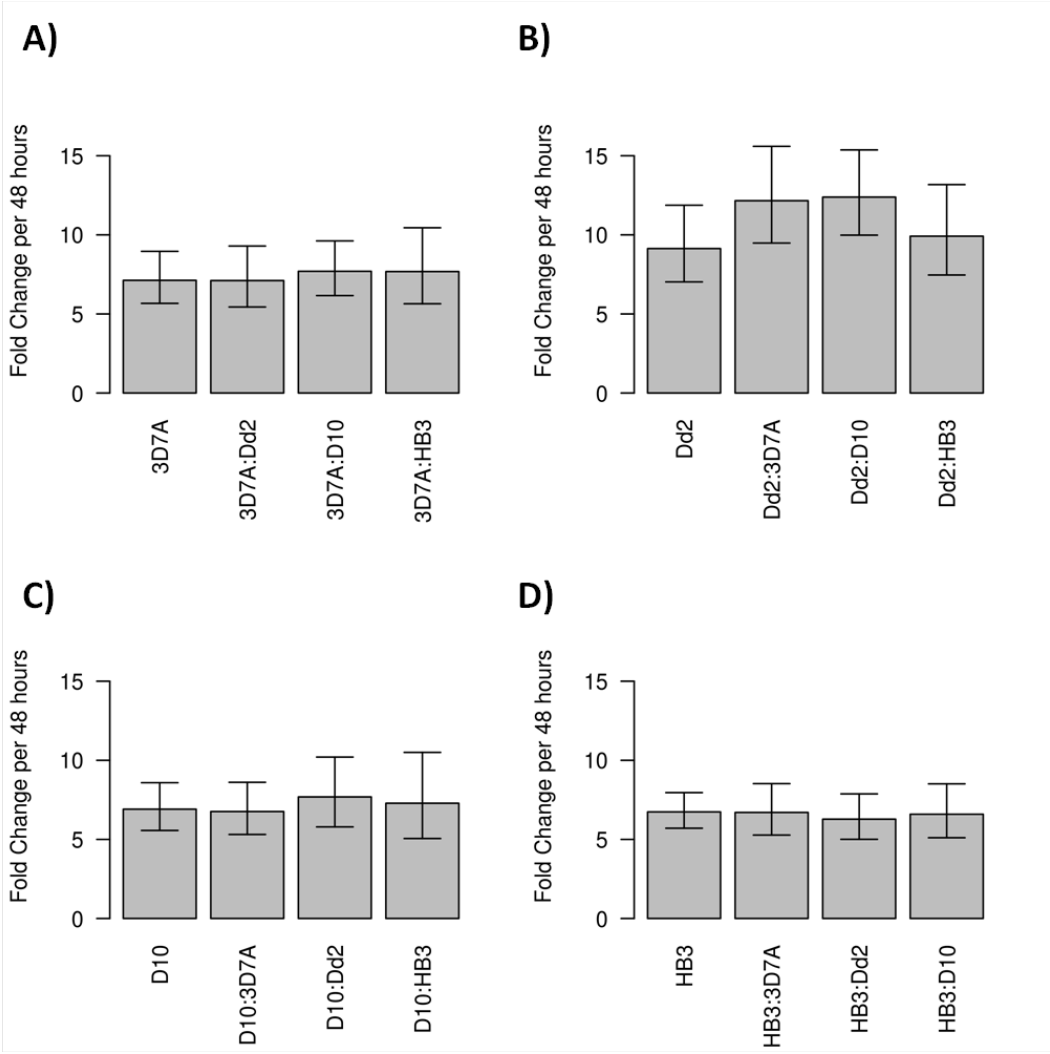


Figure 7.3. Quantitation of PMR for four laboratory clones across three intraerythrocytic cycles for all six possible pairwise *in vitro* competitive scenarios. Error bars are 95% confidence intervals around mean PMR values derived by general linear modelling for A) 3D7A in monoculture and pairwise competition, B) Dd2 in monoculture and pairwise competition, C) D10 in monoculture and pairwise competition and D) HB3 in monoculture and pairwise competition.

Culture scenario	PMR	Upper 95% CI	Lower 95% CI
3D7A-GFP Monoculture	7.13	8.96	5.67
3D7A-GFP vs Dd2	7.11	9.29	5.44
3D7A-GFP vs D10	7.70	9.62	6.16
3D7A-GFP vs HB3	7.68	10.04	5.64
Dd2 Monoculture	9.13	11.87	7.03
Dd2 vs 3D7A-GFP	12.16	15.60	9.48
Dd2 vs D10	12.39	15.37	9.98
Dd2 vs HB3	9.92	13.18	7.46
D10 Monoculture	6.91	8.58	5.56
D10 vs 3D7A-GFP	6.76	8.61	5.31
D10 vs Dd2	7.69	10.20	5.79
D10 vs HB3	7.29	10.50	5.06
HB3 Monoculture	6.74	7.96	5.71
HB3 vs 3D7A-GFP	6.70	8.52	5.27
HB3 vs Dd2	6.28	7.88	5.00
HB3 vs D10	6.59	8.51	5.10

Table 7.2. Parasite multiplications rates (PMR) calculated for the four laboratory clones in pairwise competitive growth experiments with upper and lower confidence intervals taken from growth across three consecutive 48 hour periods. All rates are given per 48 hour period for each monoculture and PMR scenario, with bold letter typing indicating which of the clones within competition the rates of growth and confidence intervals correspond to.

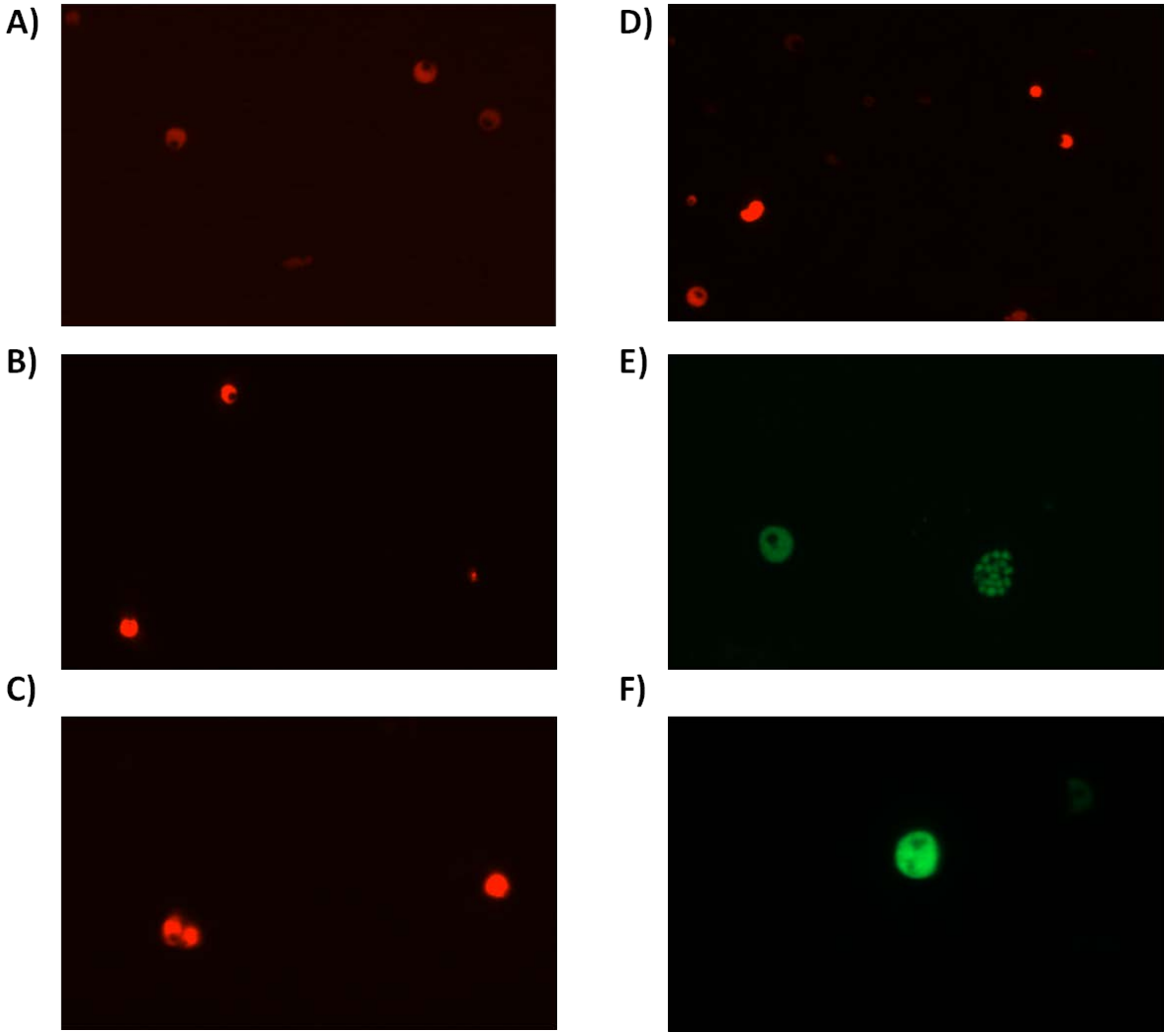


Figure 7.4 Confirmation of fluorescent protein expression in *P. falciparum*. Imaged under a 40x lens of an EVOS FL Cell Imaging System, fluorescent images are shown for A) 3D7A-mCherry, B) 3D7A-mKate2, C) D10-mKate2, D) Dd2-mKate2, E) Dd2-GFP and F) D10-GFP.

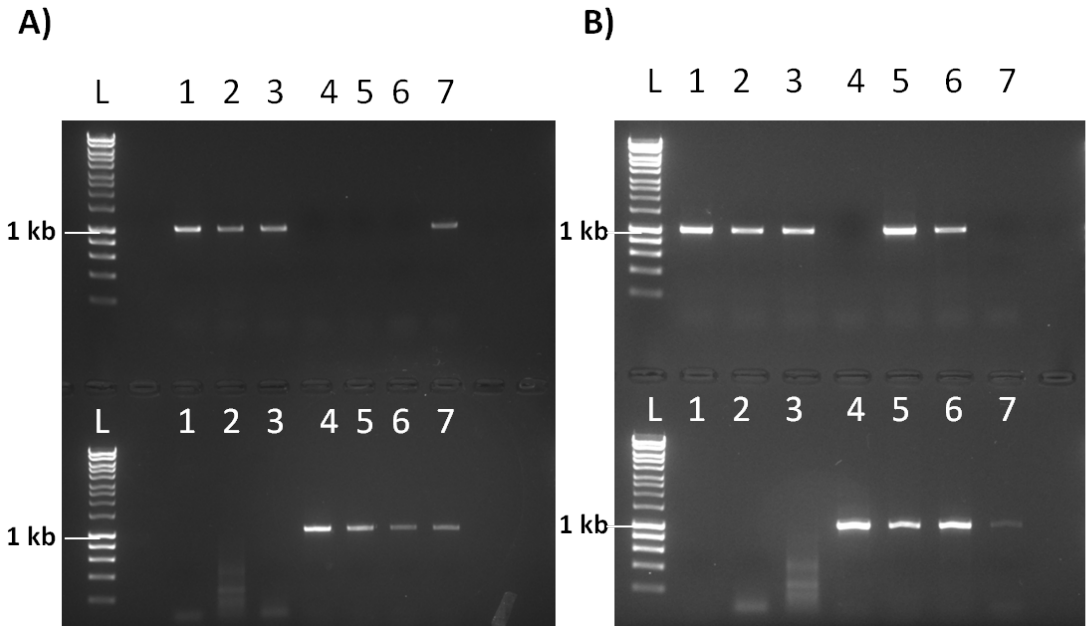
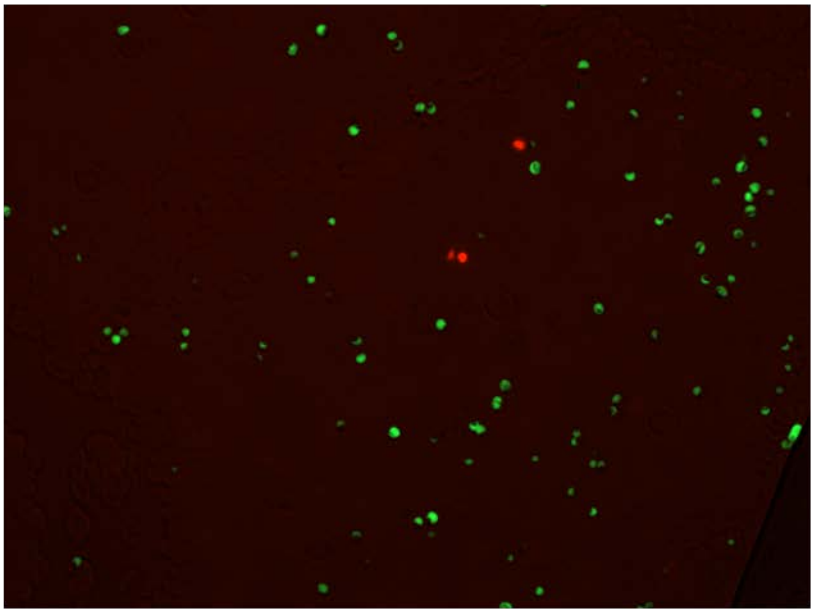


Figure 7.5 PCR-based integration genotyping for the fluorescent parasites at the Pf47 locus. In both gels A and B the top half of the gel shows genotyping for the wild type (WT) Pf47 locus using primers AT1 and AT2, whereas the bottom half shows genotyping for the integrated Pf47 locus, using primers AT1 and ST509. On gel A, L= 1 kb ladder, 1= 3D7A WT, 2= D10 WT, 3= Dd2 WT, 4=3D7A-GFP integration control, 5= 3D7A-mCherry-CL3, 6= D10-mKate2-CL4 and 7= Dd2-GFP. On gel B, L=1 kb ladder, 1= 3D7A WT, 2= Dd2 WT, 3=D10 WT, 4= 3D7A-GFP integration control, 5= D10-GFP, 6= Dd2-mKate2, 7= 3D7-mKate2 CF11. Both the WT and integration bands are similarly sized amplicons at approximately 1 kb.

underwent cloning after they showed evidence of integration cloning, clones were not successfully obtained and this would need to be repeated to generate clonal lines lacking wild-type parasites possessing episomal plasmid copies. Moreover, clonal identity after several months of drug selection was confirmed by the appropriate allele-specific qPCR assay. In order to confirm that the use of fluorescent protein expressing parasites could potentially be a useful tool for analyses of competition *in vitro* for *P. falciparum*, live parasite counts of controlled mixtures of fluorescent parasites were compared against proportions of genotypes within the mixture using allele-specific qPCR assays. This was performed using combinations of 3D7A-GFP, D10-mKate2-CL4, Dd2-GFP and 3D7A-mCherry-CL3, which allowed for the 3 possible *in vitro* competition scenarios between these genotypes; 3D7A-mCherry with Dd2-GFP, 3D7A-GFP with D10-mKate2-CL4 and Dd2-GFP with D10-mKate2-CL4. Through comparison of mixtures of sorbitol synchronised early trophozoite stage parasites for the three clones across a range of ratios from 20:1 to 1:20 using live imaging microscopy (Figure 7.6) and allele-specific qPCR provided good concordance between the two methods (Figure 7.7, Dd2-GFP and D10-mKate2-CL4 R^2 = 0.984; 3D7A-GFP and D10-mKate2-CL4 R^2 = 0.994; 3D7A-mCherry-CL3 and Dd2-GFP R^2 = 0.998).

A)



B)

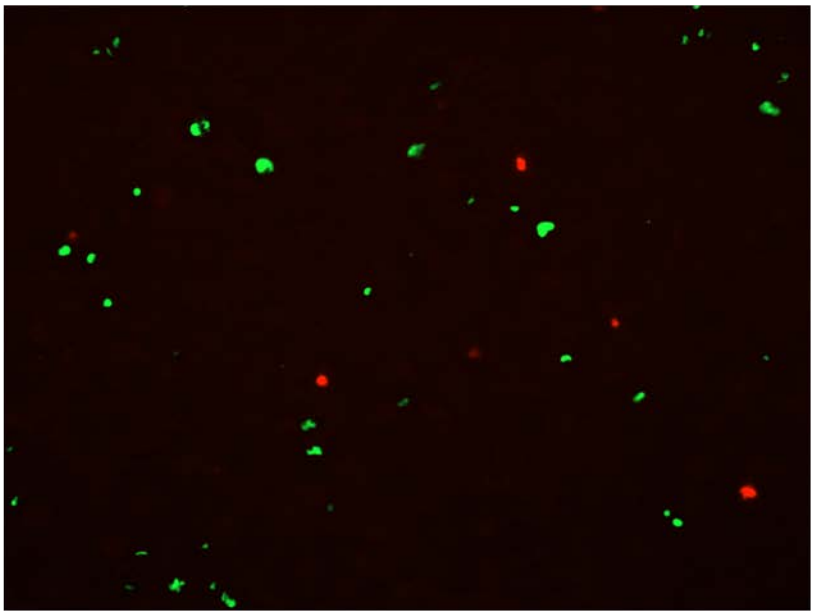


Figure 7.6 Examples of fluorescent overlay live parasite images for correlation with genome copy numbers by allele-specific qPCR. Panel A) shows an approximate 20:1 ratio mixture of Dd2-GFP and D10-mKate2, whilst Panel B) shows an approximate 10:1 ratio mixture of 3D7A-GFP and D10-mKate2.

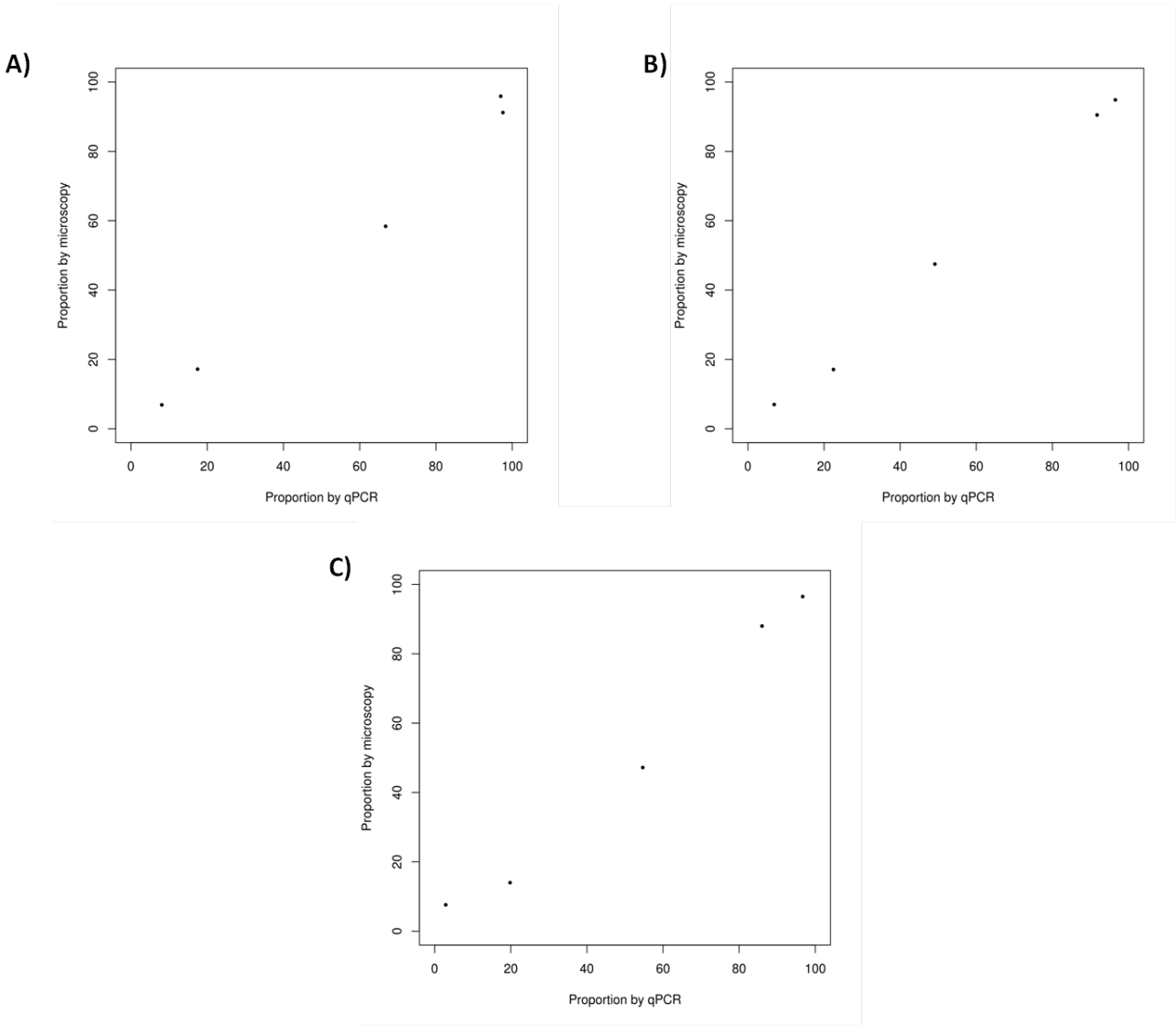


Figure 7.7 Concordance of measurement of controlled mixtures of parasites expressing different fluorescent proteins through live imaging microscopy and allele-specific qPCR. Microscopic proportions of mixes of fluorescent parasites were calculated from counts of a minimum of 200 parasites, with allele-specific qPCR assays for A) 3D7A-GFP with D10-mKate2-CL4, B) 3D7A-mCherry-CL3 with Dd2-GFP and C) Dd2-GFP with D10-mKate2-CL4. Proportions shown with each panel are for the first named parasite line within each mixture, all R^2 values >=0.98.

7.4 Discussion:

Within this chapter, an allele-specific qPCR based approach has been used to assess the growth rates of four laboratory clones in the presence of a competitor, which showed no significant differences in PMR recorded within monoculture conditions and pairwise competition assays from low starting parasitemias. Previous studies that have quantified *in vitro* competitive growth of different *P. falciparum* genotypes using either allele-specific qPCR or pyrosequencing approaches under different conditions, have reached different conclusions. One study reported a significant effect of asymmetrical competitive suppression of Dd2 over HB3 based on an endpoint measurement of density (Wacker *et al.* 2012). Whilst data within this study would support that finding if endpoint densities at Day 6 were compared for these two clones, another study which measured PMR of Dd2 and HB3 using deep sequencing of SNPS located within the PfCRT gene over a 2 month period found Dd2 to have only a 2.2% increased PMR over HB3 in co-culture per cycle (Petersen *et al.* 2015), thus suggesting that there was no evidence of the strong suppressive effect reported by Wacker *et al.* (2012).

The assay format used within this chapter was adopted to provide a quantitative measurement of genotype-specific PMR across multiple intraerythrocytic cycles from a low starting parasitemia to minimise any potential density-dependent effect on parasite growth rate. It should be noted however that, any genotype-specific competitive effect is likely also to be density-dependent and another study of *in vitro* competition between parasites genotypes reported that the total density of parasite genotypes was the key factor that influenced genotype-specific PMR (Zervos *et al.* 2012). Thus, further assays across a range of starting densities and ratio concentrations of different genotypes could be

utilised to ascertain whether genotype-specific competitive effects on PMR are present within *P. falciparum*. Furthermore, laboratory clones that have been adapted to culture and grown for long periods of time in genotypic isolation may have lost any potential phenotypic response to competition, whilst the laboratory clones used within this thesis appear to have similar rates of growth within the adopted assay format. Data on the *ex vivo* PMR exhibited by clinical isolates, as provided by this thesis (section 6), will aid the design of assays to test for competitive growth interactions between clinical isolates that exhibit different PMR. It would be expected that if two clinical isolates genotypes were co-cultured, that if there was no parasite interactive effect on PMR, that the genotype with the higher rate of growth would increase its proportion within the culture at a rate that could be predicted from the monoculture PMR seen by each isolate. Subsequently, a departure from this model could then be tested for to look for evidence of any potential interaction.

Using the parasite lines produced within this chapter, competitive interaction studies would be able to be performed through microscopy over longer-time periods than 6 days and be also tested for simultaneously tested for using the allele-specific qPCR assays. In addition, an *in vitro* genotype specific differential fluorescence system would also uniquely allow for physical separation of *P. falciparum* genotypes through cell sorting approaches (Miao *et al.* 2010; Miao & Cui 2011), that could offer a system to investigate potential plastic phenotypic responses to genotype competition such as gametocytogenesis (Reece *et al.* 2008; Neal & Schall 2014).

8. Discussion

The analyses within this thesis investigate the genomic diversity of clinical infections at the within-host level within West Africa, apply *in vitro* methods for the dissection of competition between multiple genotypes and focus on characterising baseline parasitological information of two clinically and epidemiologically relevant traits which have previously been reported as showing evidence of plastic responses to the presence of co-infecting genotypes within *Plasmodium* (Pollitt *et al.* 2014; Reece *et al.* 2008).

8.1 Whole genome sequence based investigation of within-host diversity

Initially within this thesis, measures of the within-host diversity of an endemic West African population from a whole genome sequence methodology (Manske *et al.* 2012; Auburn *et al.* 2012) were compared against those from a traditional microsatellite based assessment (Anderson *et al.* 2000; Mobegi *et al.* 2012). These analyses provide correlative and complementary information on the within-host genotypic structure of *P. falciparum* infections within this population. Whilst almost the microsatellite typing provided sensitive detection of distinct clones within almost all infections profiled, the quantitative whole-genome approach highlighted that the >50% of infections were consisted predominantly of genotypically similar parasites, in agreement with other studies (Druihle *et al.* 1998; Nkhoma *et al.* 2012). In addition, there appears to be no general consensus within *P. falciparum* population genetics as to what the definition of a single clone infection is, but within this thesis I generally have applied the terminology of having predominantly single-genotype infections.

Unlike the sensitive PCR based approaches, F_{WS} relies on the use of read count data to determine levels of within-host and within-population heterozygosity at biallelic SNPs. Within the filtering pipeline for calculation of F_{WS}, a minimum total coverage of 20 reads for an isolate and a minimum coverage of 5 reads for the minor population allele within an isolate were implemented. Thus, within isolates with coverage of 20 reads at a particular SNP, the minor allelic variant would need to have a frequency of 25% at that position to be considered within the analysis. Minor alleles are thus more likely to be detected at SNPs with greater sequencing coverage. Moreover, sampling of genome sequencing read count data would have some technical noise, or randomness, in counting alleles, with minor alleles where reads were counted closer to the 5 read limit, likely to be more greatly affected. In addition to this, any single timepoint measure of within-host diversity, will not overcome the issue of proportions of clones in peripheral blood of infected humans with asymptomatic cases vary over time, and can show marked differences between successive days (Farnert *et al*. 1997).

Thus, it is plausible that the within-host minor allele frequency within particular isolates at certain SNP loci be underestimated and contribute to a reduced measure of within-host heterozygosity and thus an elevated isolate or loci-specific F_{WS} score. To date there has yet to be a direct comparison of the sensitivity of detecting minor alleles through PCR and Illumina-based short read sequencing (Auburn *et al.* 2012). A potential approach to do this would be to make controlled mixtures of laboratory clones at a range of starting concentrations and perform both whole-genome sequencing and allele-specific PCR typing to compare their sensitivity.

Within the Guinean population analysis presented in this thesis however, no correlation was seen between isolate F_{WS} score and mean sequence coverage depth, whilst average genome-wide coverage for all 95 isolates was greater than 80 reads per SNP (Figure 3.3C). In addition, within the genome-wide scans of the F_{WS} index, the approach focused on identifying loci with consistently high levels of within-host homozygosity, such as within a scenario where a SNP has a high population minor allele frequency, but both biallelic variants are never found together within the same isolate in the population and the two alleles are effectively being 'segregated' within different hosts. Thus, the sensitivity of counting the minor allele within-host would have no effect on being able to identify SNPs where the two alleles are consistently being segregated amongst different hosts within the population.

An additional confounding factor in interpreting F_{WS} is the presence of multiple clones at different frequencies (section 3.1) and the spectrum of relatedness seen between these clones present within an infection (Figure 1.5). F_{WS} does not distinguish between separate clones that are present within a host, instead providing a measure of relatedness and relative proportions on a scale between 0 and 1. In order to precisely characterise the relative contributions of multiple clones within an infection in producing an F_{WS} score, there are two potential options. Firstly, quantitative allele-specific genotyping, as previously applied to *in vitro* protist experiments (Minter *et al.* 2015) could be designed to sensitively characterise and quantify contributions of different clones at any desired loci. Alternatively, to fully elucidate the relative frequencies of multiple clones across several SNP positions, potentially to the length of individual chromosomes, would require haplotype reconstruction or phasing. At present, it is difficult to reconstruct such haplotypes, as for example, across five biallelic SNPs for

which reads are counted for both alleles, the number of potential clones or haplotypes present across that loci could theoretically be up to 2^5, or 32 distinct clones (Nkhoma *et al.* 2012). Deep sequencing approaches do offer the potential to link together separate reads to enable haplotype reconstruction, but this is aided by the availability of long-read sequencing fragments, potentially greater than 10kb in length, that ensure haplotypes are generated for the majority of genes within a *Plasmodium* genome and offer greater overlap between adjacent sequence reads for computational reconstruction of individual haplotypes from the same infection (Kuleshov *et al.* 2014; Vembar *et al.* 2016). Furthermore, only repeated longitudinal sequencing or genotyping of an infection would be able to detect any clones within an infection that may be undetectable from a single time point estimate, due to sequestration.

8.2 Identification of molecular targets of within-host selection

Furthermore, implementation of an allele-frequency based scan of within-host heterozygosity across the genome within this thesis did not identify loci that are consistently homozygous at the within-host level within multiple populations in West Africa. Potential loci under strong within-host selective forces could offer *in vivo* growth advantages to particular alleles, or potentially lead to strain-specific immune suppression or clearance, as has previously been reported within the genetic cross progeny within *Plasmodium* parasites. There are however alternatives to further examine potential evidence of within-host selection of *P. falciparum* from clinical isolates using sequencing based analyses (section 4.4), through application of long-read sequencing technologies, haplotype based phasing or transcriptomic analyses.

Within other pathogen species, within-host selection of genomic loci and phenotypes such as virulence is classically interrogated using an experimental evolution approach (Wright *et al.* 2011; Didelot *et al.* 2016). Within the genus *Plasmodium*, a combination of *in vitro* adaptation and whole genome sequencing has monitored genomic changes over long periods of time to profile baseline rates of mitotic evolution in *P. falciparum* (Bopp *et al.* 2013; Claessens *et al.* 2014), however this would only ethically be possible to replicate within clinical isolates within asymptomatic cases that would allow for repeated sampling (Farnert *et al.* 1997), which would need to be combined with sequencing pipelines using only limited quantities of genomic DNA (Oyola *et al.* 2014; Nair *et al.* 2014). In addition, a comparative genomics and *in vitro* adaptation approach has also been utilised to identify a normocyte binding protein that is significantly advantageous for the invasion of human erythrocytes by the zoonotic parasite *P. knowlesi* (Moon *et al.* 2016). An experimental study assessing within-host genomic evolution would however be possible using a rodent malaria parasite system, which have so far only been used to assess phenotypic responses to pressures such as drug treatment, selection for virulence, immunisation and competition from other genotypes (Nuralithia *et al.* 2015; Abkallo *et al.* 2015; Huijben *et al.* 2013; Barclay *et al.* 2008; Mackinnon & Read 1999; Bell *et al.* 2006; de Roode *et al.* 2005).

None of these studies have however sought to uncover the molecular targets of selection, but a potential avenue of research would be to subject a genetically modified *P. berghei* parasite line which has been generated as a 'super mutator' through destruction of the DNA proofreading activity domain of DNA polymerase δ to various selection pressures and monitor genomic selection. This line has been determined to have an approximately 90-fold higher mutation

rate than that of the parental *P. berghei* parasite (Honma *et al.* 2014), which would significantly facilitate temporal within-host evolution experiments. Alternatively, it could also be possible to adopt the same transfection approach targeting the orthologous gene within other rodent malaria models such as *P. yoelli* or *P. chabaudi*, or adapted to *P. falciparum* for *in vitro* experimental evolution studies.

Within this thesis, I also apply and develop methodologies that allow for further investigation of a putative mediator of selection within the host, which is the genetic composition of an infection. If kin discrimination mechanisms (Reece *et al.* 2008) are present within *P. falciparum,* then this could influence other parasite phenotypes to gametocyte production strategies. An *in vitro* system allows for dissection of such potential parasite-parasite interactions (Mantel *et al.* 2013; Regev-Rudzki *et al.* 2013) without the confounding presence of immune-mediated responses. Even with potential physical separation of clones from *in vitro* mixed infections using fluorescent methods, it is possible that long-term laboratory adapted parasites have lost such phenotypes and thus co-culture studies using recently adapted clinical isolates would be ideal, which will require sensitive versatile assays to monitor PMR of parasites within competitive scenarios. Furthermore, if transcriptional responses to co-culture were pursued, these could potentially avoid the need for physical separation by calling SNPs from a deep RNA-Seq profile of the co-culture (Yamagishi *et al.* 2014), but this would also need to overcome difficulties in distinguishing genuine differential expression from differences in intraerythrocytic cycle stage (Lemieux *et al.* 2009).

8.3 Disparity between the availability of genomic and phenotypic information

This thesis underlines what I believe to be one of the clearest dichotomies within malariology research at present; the discrepancy between having genomic sequence data available for thousands of clinical isolates taken from a global range of populations and the limited quantitative phenotypic information available on some key parasitological phenotypes such as clinical isolate growth rate *ex vivo* and transcriptional expression levels of gametocytogenesis markers *in vivo.* Profiling of the *ex vivo* PMR of nineteen clinical isolates within this study remarkably represents the largest biologically replicated survey undertaken. Whilst laboratory adapted isogenic lines offer tractable, flexible systems for reverse genetic based approaches to probe gene function, phenotypic characterisation of *ex vivo* isolates is more representative of natural parasite population than long term laboratory adapted clones, due to genomic and phenotypic adaptations occurring during the culture adaptation process. These include the reported deletion of chromosome regions that are essential for the production of gametocytes (Alano *et al.* 1995) and the identification of gene copy number amplification and deletion events throughout the genome of laboratory adapted clones (Cheeseman *et al.* 2009).

To complement further phenotyping of *ex vivo* clinical isolates further, two key developments are required. Firstly, high throughput protocols, potentially through adaptation of the approaches designed within this thesis, would significantly increase the production of isolate phenotyping data and ideally would be performed within a plate based format. Although gametocyte production methodologies are well established for laboratory clones (Fivelman *et al.* 2007; Brancucci *et al.* 2015; Delves *et al.* 2016), monitoring of *ex vivo*

gametocyte production has rarely been reported. *Ex vivo* attempts of gametocytogenesis of clinical isolates have described differences in *ex vivo* gametocyte production and sex ratio phenotypes of clones that had been propagated both within and between parental isolates (Ponnudurai *et al.* 1982; Graves *et al.* 1984). In addition, another study has reported that 100% of 44 clinical isolates which underwent *ex vivo* culture adaptation continued to produce gametocytes after a period of between 4 and 11 weeks (Nakazawa *et al.* 2011). Despite the observations of phenotypic differences in gametocytogenesis, no study has characterised any genomic or transcriptional differences in parasites that show evidence of different gametocyte commitment strategies.

Moreover, the culture adaptation process remains a limiting factor on the successful phenotyping of clinical isolates *ex vivo*, with significant nutritional and physiological differences between the *in vitro* and *in vivo* environments (Le Roux *et al.* 2009), in addition to the challenges associated with cryopreserving clinical isolates. Furthermore, the lower reported initial *ex vivo* PMR seen within African isolates (Deans *et al.* 2006) is likely to make adaptation of parasites from these populations more challenging than South-East Asian parasite populations (Chotivanich *et al.* 2000). Presently, it is still uncertain how *ex vivo* phenotypes of isolates correlate with *in vivo* phenotypes, particularly due to the phenotypic and transcriptional plasticity that has been widely reported within *P. falciparum* of both laboratory clones and clinical isolates. Establishment of clinical grade cell banks of *ex vivo* clinical isolates is a possibility for controlled human malaria infection studies (Stanisic *et al.* 2015), which would allow for such correlative studies of baseline phenotypes *in vivo* and *ex vivo*.

8.4 Identifying genomic determinants of phenotypic and transcriptional variation from recently adapted *P. falciparum* clinical isolates

The principles of identifying loci that are highly significantly associated with a particular trait of *P. falciparum* recorded within natural populations has been pioneered through using a combination of parasite phenotyping and genetic differentiation and haplotype-based genome scans to identify regions that were associated with a slow artemisinin clearance phenotype exhibited by isolates within certain populations (Anderson *et al.* 2010; Cheeseman *et al.* 2012). Such an approach has been further advanced by the developments of pooled sequencing methods to scan for genomic determinants of a phenotype (Cheeseman *et al.* 2015). Within yeast genetics, due to the availability of high throughput methods for phenotyping and robust *ex vivo* adaptation possible, it has been possible to quantify heterogeneity of over 200 phenotypes and identify significantly associated genomic variants for 89 of them within *Saccharmoyces pombe* (Jeffares *et al.* 2015). These studies have all relied on using large sample size numbers of to associate genomic loci with a particular phenotype, whilst the broad range phenotyping applied to *S. pombe* is not presently feasible for application to *P. falciparum*.

Within this thesis, two important phenotypes of clinical isolates are investigated; *ex vivo* PMR and gametocyte transcript expression. An alternative forward genetics based approach that would require only small numbers of isolates, limited whole-genome sequencing and enable rapid discovery of genes associated with particular phenotypes of *P. falciparum* would be a linkage group selection approach (Culleton *et al.* 2005). This would require the use of a humanized mouse within which harvest of recombinant progeny is possible (Vaughan *et al.* 2015) and *ex vivo* gametocytogenesis of clinical isolates

(Graves *et al.* 1984; Nakazawa *et al.* 2011). Other than the previous necessity to use a chimpanzee within a cross that is now avoided by the use of the humanized mouse, another current limit on a linkage group selection approach for *P. falciparum* is the development of phenotype-specific selection assays that are applied to pooled recombinant progeny. For investigation of potential loci associated with PMR (section 6), two *ex vivo* isolates could be selected for crossing by exhibiting very different replication rate measures and subject to growth selection (Pattaradilokrat *et al.* 2009), followed by genome-wide sequencing of the selected progeny pool to identify loci that show evidence of increased homozygosity compared to the unselected progeny pool.

This assay design, which would potentially be applicable to any phenotype which is selectable *in vitro*, consists of three stages; firstly selecting isolates showing marked variation in a particular phenotype, secondly gametocyte production of these clonal parasite lines and subsequent crossing within the humanized mouse model and thirdly a phenotype-specific selection assay on the recombinant progeny produced. So far only loci associated with growth rates, drug resistance and strain-specific immunity have been identified using a linkage group selection approach (section 1.8). A number of other phenotype-selection assays are possible *in vitro*, but would be likely to result in the production of relatively low parasite densities, thus requiring the application of limited-cell sequencing approaches (Oyola *et al.* 2014). Another example of a phenotype-specific selection assay to identify associated genomic loci would be to induce gametocytogenesis *in vitro* for the pooled cross progeny, using a pan-gametocyte or sex-specific selection assay, through the use of antibodies (Barr *et al.* 1991; Quakyi *et al.* 1987 and cell-sorting or magnetic separation (Ribaut *et*

al. 2008) to ensure a high purity population of gametocytes are selected and subsequently sequenced.

In addition, the high levels of transcriptional variation that are seen within *P. falciparum* could also be subject to a similar pooled progeny approach to identify genome regions polymorphisms that influence expression of a specific transcript, as performed using genome-wide expression quantitative trait loci (eQTL) mapping of a genetic cross between HB3 and Dd2, which required culturing and microarray transcript profiling for 36 different clones at 18 hours post invasion (Gonzales *et al.* 2008). Clinical isolates display a spectrum of variation in transcript expression, most notably for transcripts expressed at the schizont stage (section 1.7). Antibody-based selection of differentially expressed transcripts (Nery *et al.* 2006) in a recombinant progeny pool and cell-sorting could reveal genomic variants that regulate the expression of invasion genes (Josling *et al.* 2015), of which several are vaccine candidates for *P. falciparum* (Conway 2015).

References

Abkallo, H. M., Tangena, J. A., Tang, J., Kobayashi, N., Inoue, M., Zougrana, A., Colegrave, N., Culleton, R. (2015) Within-host competition does not select for virulence; studies with *Plasmodium yoelli*. *PLoS Pathog.*, 11, e1004628.

Alano, P., Roca, L., Smith, D., Read, D., Carter, R., Day, K. (1995) *Plasmodium falciparum*: parasites defective in early stages of gametocytogenesis. *Exp. Parasitol.*, 81, 227-235.

Alexa, A., Rahnenfuhrer, J., Lengauer, T. (2006) Improved scoring of functional groups from gene expression date by decorrelating GO graph structure. *Bioinformatics*, 22, 1600-1607.

Amambua-Ngwa, A., Tetteh, K. K., Manske, M., Gomez-Escobar, N., Stewart, L. B., Deerhake, M., Cheeseman, I. H., Newbold, C. I., Holder, A. A., Knuepfer, E., Janha, O., Jallow, M., Campino, S., MacInnis, B., Kwiatkowski, D. P., Conway, D. J. (2012) *PLoS Genet.*, 8, e1002992.

Anderson, T. J. (2004) Mapping drug resistance genes in *Plasmodium falciparum* by genome-wide association. *Curr. Drug Targets Infec. Disord.,* 4, 65-78.

Anderson, T. J., Haubold, B., Williams, J. T., Estrada-Franco, J. G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakuwa, S., French, N.,Whitworth, J.,Velez, I. D., Brockman, A. H., Nosten, F., Ferreira, M. U. & Day, K. P. 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Mol. Biol. Evol.,* 17**,** 1467-82.

Anderson, T. J., Nair, S., Nkhoma, S., Williams, J. T., Imwong, M., Yi, P., Socheat, D., Das, D., Chotivanich, K., Day, N. P., White, N. J., Dondorp, A. M. (2010) High heritability of malaria parasite clearance rate indicates genetic basis for Artemisinin resistance in Western Cambodia. *J. Infect. Dis.*, 201, 1326 -1330.

Anderson, T. J., Su, X. Z., Bockarie, M., Lagog, M., Day, K. P. (1999) Twelve microsatellite markers for characterisation of *Plasmodium falciparum* from finger-prick blood samples. *Parasitology*, 119, 113-125.

Angrisano, F., Tan, Y. H., Sturm, A., McFadden, G. I., Baum, J. (2012) Malaria parasite colonisation of the mosquito midgut—placing the *Plasmodium* ookinete centre stage. *Int. J. Parasitol.*, 42, 519-527.

Anthony, T. G., Conway, D. J., Cox-Singh, J., Matusop, A., Ratnam, S., Shamsul, S., Singh, B. (2005) Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. *J. Infect. Dis.*, 191, 1558-1564.

Anthony, T. G., Polley, S. D., Vogler, A. P., Conway, D. J. (2007) Evidence of non-neutral polymorphism in *Plasmodium falciparum* gamete surface protein genes Pfs47 and Pfs48/45. *Mol. Biochem. Parasitol.*, 156, 117-123.

Asante, K. P., Zandoh, C., Dery, D. B., Brown, C., Adjei, G., Antwi-Dadzie, Y., Adjuik, M., Tchum, K., Dosoo, D., Amenga-Etego, S., Mensah, S., Owusu-Sekyere, K. B., Anderson, C., Krieger, G., Owusu-Agyei, S. (2011) Malaria epidemiology in the Ahafo area of Ghana. *Malar. J.*, 10, 211.

Ashley, E. A., White, N. J. (2014) The duration of *Plasmodium falciparum* infections. *Malar. J.*, 13, 500.

Assefa, S. A., Preston, M. D., Campino, S., Ocholla, H., Sutherland, C. J., Clark, T. G. (2014) estMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinformatics*, 30, 1292-1294.

Auburn, S., Campino, S., Miotto, O., Djimde, A. A., Zongo, I., Manske, M., Maslen, G., Mangano, V., Alcock D., MacInnis, B., Rockett, K. A., Clark, T. G., Doumbo, O. K., Ouedraogo, J. B., Kwiatkowski, D. P. (2012) Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS One*, 7, e32891.

Babbitt, S. E., Altenhofen, L., Cobbold, S. A., Istvan, E. S., Fennell, C., Doerig, C., Llinás, M., Goldberg, D. E. (2012) *Plasmodium falciparum* responds to amino acid starvation by entering into a hibernatory state. *Proc. Natl. Acad. Sci. U. S. A*, 109, E3278-3287.

Bailey, J. A., Mvalo, T., Aragam, N., Weiser, M., Congdon, S., Kamwendo, D., Martinson, F., Hoffman, I., Meshnick, S. R., Juliano, J. J. (2012) Use of massively parallel pyrosequencing to evaluate the diversity of and selection on *Plasmodium falciparum csp* T-Cell epitopes in Lilongwe, Malawi. *J. Infect. Dis.*, 206, 580-587.

Bannister, L., Mitchell, G. (2003) The ins, outs and roundabouts of malaria. *Trends Parasitol*., 19, 209-213.

Barclay, V. C., Chan, B. H., Anders, R. F., Read, A. F. (2008) Mixed allele malaria vaccines: host protection and within-host selection. *Vaccine*, 26, 6099-6107.

Barr, P. J., Green, K. M., Gibson, H. L., Bathurst, I. C., Quakyi, I. A., Kaslow, D. C. (1991) Recombinant Pfs25 protein of *Plasmodium falciparum* elicits malaria transmission-blocking immunity in experimental animals. *J. Exp. Med.*, 174, 1203-1208.

Baton, L. A., Ranford-Cartwright, L. C. (2005) Spreading the seed of million-murdering death: metamorphoses of malaria in the mosquito. *Trends Parasitol.*, 21, 573-580.

Baum, J., Pinder, M, Conway, D. J. (2003) Erythrocyte invasion phenotypes of *Plasmodium falciparum* in The Gambia. *Infect. Immun.*, 71, 1856-1863.

Baum, J., Thomas, A. W., Conway, D. J. (2003) Evidence for diversifying selection on erythrocyte-binding antigens of *Plasmodium falciparum* and *P. vivax*. *Genetics*, 163, 1327-1336.

Bei, A. K., Duraisingh, M. T. (2012) Functional analysis of erythrocyte determinants of *Plasmodium* infection. *Int. J. Parasitol.*, 42, 575-582.

Bejon, P., Andrews, L., Andersen, R. F., Dunachie, S., Webster, D., Walther, M., Gilbert, S. C., Peto, T., Hill, A. V. (2005) Calculation of liver-to-blood inocula, parasite growth rates, and preerythrocytic vaccine efficacy, from serial quantitative polymerase chain reaction studies of volunteers challenged with malaria sporozoites. *J. Infect. Dis.*, 191, 619-626.

Bell, A. S., de Roode, J. C., Sim, D., Read, A. F. (2006) Within-host competition in genetically diverse malaria infections: parasite virulence and competitive success. *Evolution*, 60, 1358-1371.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, 27, 573-580.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvilli, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E., Catenazzi, M., Chang, S., Cooley, R. N., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Furey, W. S., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Ivanov, D. V., Johnson, M. Q., James, T., Jones, T. A. H., Kang, G. D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ng, B. L., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Pinkard, D. C., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Rodriguez, A. C., Roe, P. M., Rogers, J., Bacigalupo, M. C. R., Romanov, N., Romieu, A.,

Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Sohna Sohna, J. E., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worlsey, G. J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J. C., Hurles, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R., Smith, A. J. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-59.

Bhasin, V. K., Trager, W. (1984) Gametocyte-forming and non-forming gametocyte clones of *Plasmodium falciparum*. *Am. J. Trop. Med. Hyg.*, 33, 534-537.

Bhatia, G., Patterson, N., Sankararaman, S., Price, A. L. (2013) Estimating and interpreting FST: the impact of rare variants. *Genome Res.*, 23, 1514-1521.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briet, O., Penny, T. A., Smith, A., Bennett, J., Yukich, T. P., Eisele, T. P., Griffin, J. T., Fergus, C. A., Lynch, M., Lindgren, F., Cohen, J. M., Murray, C. L. J., Smith, D. L., Hay, S. I., Cibulskis, R. E., Gething, P. W. (2015) The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*, 526, 207-211.

Bopp, S. E., Manary, M. J., Bright, A. T., Johnston, G. L., Dharia, N. V., Luna, F. L., McCormack S., Plouffe, D., McNamara, C. W., Walker, J. R., Fidock, D. A., Denchi, E. L., Winzeler. E. A. (2013) Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. *PLoS Genet.*, 9, e1003293.

Bousema, J. T., Schneider, P., Gougana, L. C., Drakeley, C. J., Tostmann, A., Houben, R., Githure, J. J., Ord., R., Sutherland, C. J., Omar, S. A., Sauerwein, R. A. (2006) Moderate effect of artemisinin-based combination therapy on transmission of *Plasmodium falciparum*. *J. Infec. Dis.*, 193, 1151-1159.

Bowyer, P. W., Stewart, L. B., Aspeling-Jones, H., Mensah-Brown, H. E., Ahouidi, A. D., Amambua-Ngwa, A., Awandare, G. A., Conway, D. J. (2015) Variation in *Plasmodium falciparum* erythrocyte invasion phenotypes and merozoite ligand gene expression across different populations in areas of malaria endemicity. *Infect. Immun.*, 83, 2575-2582.

Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J., DeRisi, J. L. (2003) The transcriptome of the intraerythrocytic development cycle of *Plasmodium falciparum*. *PLoS Biol.*, 1, E5.

Brancucci, N. M., Goldowitz, I., Buckholz, K., Werling, K., Marti, M. (2015) An assay to probe *Plasmodium falciparum* growth, transmission stage formation and early gametocyte development. *Nat. Protoc.*, 10, 1131-1142.

Bronner, I. F., Otto, T. D., Zhang, M., Udenze, K., Wang, C., Quail, M. A., Jiang, R. H., Adams, J. H., Rayner, J. C. (2016) Quantitative insertion-site sequencing (QIeq) for high throughput phenotyping of transposon mutants. *Genome Res.*, 26, 980-989.

Browning, S. R., Browning, B. L. (2012) Identity-by-descent between distant relatives: detection and applications. *Annu. Rev. Genet.*, 46, 617-633.

Bruce, M. C., Alano, P., Duthie, S., Carter, R. (1990) Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology*, 100,191-200.

Burkot, T. R., Williams, J. L., Schneider, I. (1984) Infectivity to mosquitoes of *Plasmodium falciparum* clones grown *in vitro* from the same isolate. *Trans. R. Soc. Trop. Med. Hyg.*, 78, 339-341.

Campbell, C. C., Collins, W. E., Nguyen-Dinh, P., Barber, A., Broderson, J. R. (1982) *Plasmodium falciparum* gametocytes from culture *in vitro* develop to sporozoites that are infectious to primates. *Science*, 217, 1048-1050.

Carter, R., McGregor, I. A. (1973) Enzyme variation in *Plasmodium falciparum* in the Gambia. *Trans. R. Soc. Trop. Med. Hyg.*, 67, 830-837.

Cheeseman , I. H., Gomez-Escobar, N., Carret, C. K., Ivens, A., Stewart, L. B., Tetteh, K. K., Conway, D. J. (2009) Gene copy number variation throughout the *Plasmodium falciparum* genome. *BMC Genomics*, 10, 353.

Cheeseman, I. H., McDew-White, M., Phyo, A. P., Sriprawat, K., Nosten, F., Anderson, T. J. (2015) Pooled sequencing and rare variant association tests for identifying the determinants of emerging drug resistance in malaria parasites. *Mol. Biol. Evol.*, 32, 1080-1090.

Cheeseman, I. H., Miller, B. A., Nair, S., Nkhoma, S., Tan, A., Tan, J. C., Al Saai, S., Phyo, A. P., Moo, C. L., Lwin, K. M., McGready, R., Ashely, E., Imwong, M., Stepniewska, K., Yi, P., Dondorp, M., Mayxay, M., Newton, P. N., White, N. J., Nosten, F., Ferdig, M. T., Anderson, T. J. (2012) A major genome region underlying artemisinin resistance in malaria. *Science*, 336, 79-82.

Cheesman, S.J., de Roode, J.C., Read, A.F., Carter, R. (2003) Real-time quantitative PCR for analysis of genetically mixed infections of malaria parasites: technique validation and applications. *Mol. Biochem. Parasitol,* 131, 83-91.

Cheesman, S. J., O'Mahony, E., Pattaradilokrat, S., Degnan, K., Knott, S., Carter, R. (2010) A single parasite gene determines the strain-specific

protective immunity against malaria: the role of the merozoite surface protein I. *Int. J. Parasitol.*, 40, 951-961.

Chen, P., Lamont, G., Elliott, T., Kidson, C., Brown, G., Mitchell, G., Stace, J., Alpers, M. (1980) *Plasmodium falciparum* strains from Papua New Guinea: culture characteristics and drug sensitivity. *Southeast Asian J. Trop. Med. Public Health*, 11, 435-440.

Chen, K., Sun, L., Lin, Y., Fan, Q., Zhao, Z., Hao, M., Feng, G., Wu, Y., Cui, L., Yang, Z. (2014) Competition between *Plasmodium falciparum* strains in clinical infections during *in vitro* culture adaptation. *Infect. Genet. Evol.*, 24, 105-110.

Chotivanich, K., Udomsangpetch, R., Simpson, J. A., Newton, P., Pukrittayakamee, S., Looareesuwan, S., White, N. J. (2000) Parasite multiplication potential and severity of *Falciparum* malaria*. J. Infec. Dis.*, 181, 1206-1209.

Claessens, A., Hamilton, W. L., Kekre, M., Otto, T. D., Faizullabhoy, A., Rayner, J. C., Kwiatkowski, D. P. (2014) Generation of antigenic diversity in *Plasmodium falciparum* by structured rearrangement of Var genes during mitosis. *PLoS Genet.*, 10, e1004812.

Conway, D. J. (2007) Molecular epidemiology of malaria. *Clin. Microbiol. Rev.*, 20, 188-204.

Conway, D. J. (2015) Paths to a malaria vaccine illuminated by parasite genomics. *Trends Genet.*, 31, 97-107.

Conway, D. J., Cavanagh, D. R., Tanabe, K., Roper, C., Mikes, Z. S., Sakihama, N., Bojang, K. A., Oduola, A. M., Kremsner, P. G., Arnot, D. E., Greenwood, B. M., McBride, J. S. (2000) A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nat. Med.*, 6, 689-692.

Conway, D. J., Greenwood, B. M., McBride, J. S. (1991) The epidemiology of multiple-clone *Plasmodium falciparum* infections in Gambian patients. *Parasitology*, 103, 1-6.

Conway, D. J., Machado, R. L., Singh, B., Dessert, P., Mikes, Z. S., Povoa, M. M., Oduola, M., Roper, C. (2001) Extreme geographical fixation of variation in the *Plasmodium falciparum* gamete surface protein gene Pfs48/45 compared with microsatellite loci. *Mol. Biochem. Parasitol.*, 115, 145-156.

Conway, D. J., Roper, C., Oduola, A. M., Arnot, D. E., Kremsner, P. G., Grobusch, M. P., Curtis, C. F., Greenwood, B. M. (1999) High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.*, 96, 4506-4511.

Cornet, S., Nicot, A., Rivero, A., Gandon, S. (2014) Evolution of plastic transmission strategies in avian malaria. *PLoS Pathog.*, 10, e1004308.

Culleton, R. L., Abkallo, H. M. (2015) Malaria parasite genetics: doing something useful. *Parasitol. Int.*, 64, 244-253.

Culleton, R., Martinelli, A., Hunt, P., Carter, R. (2005) Linkage group selection: rapid gene discovery in malaria parasites. *Genome Res.*, 15, 92-97.

Daniels, R., Chang, H. H., Sene, P. D., Park, D. C., Neafsey, D. E., Schaffner, S. F., Hamilton, E. J., Lukens, A. K., Van Tyne, D., Mboup, S., Sabeti, P. C., Ndiaye, D., Wirth, D. F., Hartl, D. L., Volkman, S. K. (2013) Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PLoS One*, 8, e60780.

Daniels, R., Volkman, S. K., Milner, D. A., Mahesh, N., Neafsey, D. E., Park, D. J., Rosen, D., Angelino, E., Sabeti, P. C., Wirth, D. F., Wiegand, R. C. (2008) A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar. J.*, 7, 223.

Deans, A. M., Lyke, K. E., Thera, M. A., Plowe, C. V., Koné, A., Doumbo, O. K., Kai, O., Marsh, K., Mackinnon, M. J., Raza, A., Rowe, J. A. (2006) Low multiplication rates of African *Plasmodium falciparum* isolates and lack of association of multiplication rate and red blood cell selectivity. *Am. J. Trop. Med. Hyg.*, 74, 554-563.

Deitsch, K., Driskill, C., Wellems, T. (2001) Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. *Nucleic Acids Res.*, 29, 850-853.

de Koning-Ward, T. F., Fidock, D. A., Thathy, V., Menard, R., van Spaendonk, R. M., Waters, A. P., Janse, C. J. (2000) The selectable marker human dihydrofolate reductase enables sequential genetic manipulation of the *Plasmodium berghei* genome. *Mol. Biochem. Parasitol.*, 106, 199-212.

Delves, M. J., Straschil, U., Ruecker, A., Miguel-Blanco, C., Marques, S., Baum, J., Sinden, R. E. (2016) Routine *in vitro* culture of *P. falciparum* gametocytes to evaluate transmission-blocking interventions. *Nat. Protoc.*, 11, 1668-1680.

de Roode, J. C., Pansini, R., Cheesman, S. J., Helinski, H. M., Huijben, S., Wargo, A. R., Bell, A. S., Chan, B. H., Walliker, D., Read, A. F. (2005) Virulence and competitive ability in genetically diverse malaria infections. *Proc. Natl. Acad. Sci. U. S. A.*, 102, 7624-7628.

Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W., Wilson, D. J. (2016) Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.*, 14, 150-162.

Dondorp, A. M., Desakorn, V., Pongtavornpinyo, W., Sahassananda, D., Silamut, K., Chotvanich, K., Newton, P. N., Pitisuttithum, P., Smithyman, A. M.,

White, N. J., Day, N. P. (2005) Estimation of the total parasite biomass in acute falciparum malaria from plasma PfHRP2. *PLoS Med.*, 2, e204.

Drakeley, C. J., Duraisingh, M. T., Póvoa, M., Conway, D. J., Targett, G. A., Baker, D. A. (1996) Geographical distribution of a variant epitope of Pfs48/45 *Plasmodium falciparum* transmission-blocking vaccine candidate. *Mol. Biochem. Parasitol.*, 81, 253-257.

Druihle P., Daubersies, P., Patarapotikul, J., Gentil, C., Chene, L., Chongsuphajaisiddhi, T., Mellouk, S., Langsley, G. (1998) A primary malaria infection is composed of a very wide range of genetically diverse but related parasites. *J. Clin. Invest.*, 101, 2008-2016.

Duffy, C. W., Assefa, S. A., Abugri, J., Amoako, N., Owusu-Agyei, S., Anyorigiya, T., MacInnis, B., Kwiatkowski, D. P., Conway, D. J., Awandare, G. A. (2015) Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country within high malaria endemicity. *BMC Genomics*, 16,527.

Dye, C., Williams, G. (1997) Multigenic drug resistance among inbred malaria parasites. *Proc. Biol. Soc.*, 264, 61-67.

Dyer, M., Day, K. P. (2003) Regulation of the rate of asexual growth and commitment to sexual development by diffusible factors from *in vitro* cultures of *Plasmodium falciparum*. *Am. J. Trop. Med. Hyg.*, 68, 403-409.

Eksi, S., Morahan, B. J., Haile, Y., Furuya, T., Jiang, H., Ali, O., Xu, H., Kiattibutr, K., Suri, A., Czesny, B., Adeyemo, A., Myers, T. G., Sattabongkot, J., Su, X. Z., Williamson, K. C. (2012) *Plasmodium falciparum* gametocyte development 1 (Pfgdv1) and gametocytogenesis early gene identification and commitment to sexual development. *PLoS Pathog.*, 8, e1002964.

Farnert, A., Arez, A. P., Babiker, H. A., Beck, H. P., Benito, A., Bjorkman, A., Bruce, M. C., Conway, D. J., Day, K. P., Henning, L., Mercerau-Puijalon, O., Ranford-Cartwright, L. C., Rubio, J. M., Snounou, G., Walliker, D., Zwetyenga, J., do Rosario, V. E. (2001) Genotyping of *Plasmodium falciparum* infections by PCR: a comparative multicentre study. *Trans. R. Soc. Trop. Med. Hyg.*, 95, 225-232.

Farnert, A., Snounou, G., Rooth, I., Bjorkman, A. (1997) Daily dynamics of *Plasmodium falciparum* subpopulations in asymptomatic children in a holoendemic area. *Am. J. Trop. Med. Hyg.*, 56, 538-547.

Feagin, J. E. (1994) The extrachromosomal DNAs of apicomplexan parasites. *Annu. Rev. Microbiol.*, 48: 81-104.

Fidock, D. A., Nomura, T., Talley, A. K., Cooper, R. A., Dzekunov, S. M., Ferdig, M. T., Ursos, L. M. B., bir Singh Sidhu, A., Naude, B., Deitsch, K., Su, X. Z., Wootton, J. C., Roepe, P. D., Wellems, T. E. (2000) Mutations in the *P.*

falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell*, 6, 861-871.

Fivelman, Q. L., McRobert, L., Sharp, S., Taylor, C. J., Saeed, M., Swales, C. A., Sutherland, C. J., Baker, D. A. (2007) Improved synchronous production of *Plasmodium falciparum* gametocytes *in vitro*. *Mol. Biochem. Parasitol.*, 154, 119-123.

Francis, S. E., Sullivan Jr, D. J., Goldberg, D. E. (1997) Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu. Rev. Microbiol.*, 51, 97-123.

Galinsky, K., Valim, C., Salmier, A., de Thoisy, B., Musset, L., Legrand, E., Faust, A., Baniecki, M. L., Ndiaye, D., Daniels, R. F., Hartl, D. L., Sabeti, P. C., Wirth, D. F., Volkman, S. K., Neafsey, D. E. (2015) COIL: a methodology for evaluating complexity of infection using likelihood from single nucleotide polymorphism data. *Malar. J.*,14, 4.

Gallup, J. L., Sachs, J. D. (2001) The economic burden of malaria. *Am. J. Trop. Med. Hyg.*, 64, 85-96.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J, Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C. I., Davis, R. W., Fraser, C. M., Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419, 498-511.

Gerald, N., Mahajan, B., Kumar, S. (2011) Mitosis in the human malaria parasite *Plasmodium falciparum*. *Eukaryot. Cell.*, 10, 474-482.

Gething, P. W., Patil, A. P., Smith, D. L., Guerra, C. A., Elyazar, I. R., Johnston, G. L., Tatem, A. J., Hay, S. I. (2011) A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.*, 10, 378.

Ghorbal, M., Gorman, M., Macpherson, C. R., Martins, R. M., Scherf, A., Lopez-Rubio, J. J. (2014) Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat. Biotechnol.*, 32, 819-821.

Gibson, W., Peacock, L., Ferris, V., Williams, K., Bailey, M. (2008) The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*. *Parasit. Vectors*, 1, 4.

Golubchik, T., Batty, E. M., Miller, R. R., Farr, H., Young, B. C., Larner-Svensson, H., Fung, R., Godwin, H., Knox, K., Votintseva, A., Everitt, R. G.,

Street, T., Cule, M., Ip, C. L., Didelot, X., Peto, T. E., Harding, R. M., Wilson, D. J., Crook, D. W., Bowden, R. (2013) Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage. *PLoS One*, 8, e61319.

Gomez-Escobar, N., Amambua-Ngwa, A., Walther, M., Okebe, J., Ebonyi, A., Conway, D. J. (2010) Erythrocyte invasion and merozoite ligand gene expression in severe and mild *Plasmodium falciparum* malaria. *J. Infect. Dis.,* 201, 444-452.

Gonzales, J. M., Patel, J. J., Ponmee, N., Jiang, L., Tan, A., Maher, S. P., Wuchty, S., Rathod, P. K., Ferdig, M. T. (2008) Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Biol*., 6, e238.

Graves, P. M., Carter, R., McNeill, K. M. (1984) Gametocyte production in cloned lines of *Plasmodium falciparum*. *Am. J. Trop. Med. Hyg.*, 33, 1045-1050.

Guinet, F., Dvorak, J. A., Fujioka, H., Keister, D. B., Muratova, O., Kaslow, D. C., Aikawa, M., Vaidya, A. B., Wellems, T. E. (1996) A developmental defect in *Plasmodium falciparum* male gametogenesis. *J Cell Biol.*, 135, 269-278.

Haasl, R. J., Payseur, B. A. (2016) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Mol. Ecol.*, 25, 5-23.

Hadley, T. J., Flotz, F. W., Pasvol, G., Haynes, J. D., McGinniss, M. H., Okubo, Y., Miller, L. H. (1987) Falciparum malaria parasites invade erythrocytes that lack glycophorin A and B (MkMk). Strain differences indicate receptor heterogeneity and two pathways for invasion. *J. Clin. Invest.*, 80, 1190-1193.

Hamilton, W. D. (1967) Extraordinary sex ratios. A sex ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, 156, 477-488.

Hastings, I. M. (1996) Population genetics and the detection of immunogenic and drug-resistant loci in *Plasmodium*. *Parasitology*, 112, 155-164.

Hastings, I. M. (2006) Complex dynamics and stability of resistance to antimalarials drugs. *Parasitology*, 132, 615-624.

Hastings, I. M., Kay, K., Hodel, E. M. (2015) How robust are malaria parasite clearance rates as indicators of drug effectiveness and resistance? *Antimicrob. Agents Chemother.*, 59, 6428-6436.

Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., Kabaria, C. W., Manh, B. H., Elyazar, I. R., Brooker, S., Smith, D. L., Moyeed, R. A., Snow, R. W. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.*, 6, e1000048.

Hayton, K., Gaur, D., Liu, A., Takahashi, J., Henschen, B., Singh, S., Lambert, L., Furuya, T., Bouttenot, R., Doll, M., Nawaz, F., Mu, J., Jiang, L., Miller, L. H.,

Wellems, T. E. (2008) Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host Microbe*, 4, 40-51.

Hill, D. L., Eriksson, E. M., Schofield, L. (2014) High yield purification of *Plasmodium falciparum* merozoites for use in opsonizing antibody assays. *J. Vis. Exp.*, 89.

Hill, W. G., Babiker, H. A., Ranford-Cartwright, L. C., Walliker, D. (1995) Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genet. Res.*, 65, 53-61.

Honma, H., Hirai, M., Nakamura, S., Hakimi, H., Kawazu, S., Palacpac, N. M., Hisaeda, H., Matsuoka, H., Kawai, S., Endo, H., Yasunaga, T., Ohashi, J., Mita, T., Horii, T., Furusawa, M., Tanabe, K. (2014) Generation of rodent malaria parasites with a high mutation rate by destructing proofreading activity of DNA polymerase δ. *DNA Res.*, 21, 439-446.

Huijben, S., Bell, A. S., Sim, D. G., Tomasello, D., Mideo, N., Day, T., Read. A. F. (2013) Aggressive chemotherapy and the selection of drug resistant pathogens. *PLoS Pathog.*, 9, e1003578.

Illingworth, C. J., Fisher, A., Mustonen, V. (2014) Identifying selection in the within-host evolution of influenza using viral sequence data. *PLoS Comput. Biol.*, 10, e1003755.

Iwagami, M., Rivera, P. T., Villacorte, E. A., Escueta, A. D., Hatabu, T., Kawazu, S., Hayakawa, T., Tanabe, K., Kano, S. (2009) Genetic diversity and population structure of *Plasmodium falciparum* in the Philippines. *Malar. J.*, 8, 96.

Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Prevorovsky, M., Mourier, T., Marsellach, F. X., Iqbal, Z., Lau, W., Cheng, T. M., Mulleder, M., Lawson, J. L., Chessel, A., Bala, S., Hellenthal, G., O'Fallon, B., Keane, T., Simpson, J. T., Bischof, L., Tomiczek, B., Bitton, D. A., Sideri, T., Codlin, S., Hellberg, J. E., van Trigt, L., Jeffrey, L., Li, J. J., Atkinson, S., Thodberg, M., Febrer, M., McLay, K., Drou, N., Brown, W., Hayles, J., Carazo Salas, R. E., Ralser, M., Maniatis, N., Balding, D. J., Balloux, F., Durbin, R., Bahler J. (2015) The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nat. Genet.*, 47, 235-241.

Jiang, H., Li, N., Gopalan, V. Zilversmit, M. M., Varma, S., Nagarajan, V., Li, J., Mu, J., Hayton, K., Henschen, B., Yi, M., Stephens, R., McVean, G., Awadalla, P., Wellems, T. E., Su, X. Z. (2011) High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. *Genome Biol.*, 12, R33.

Joice, R., Hilsson, S. K., Montgomery, J., Dankwa, S., Egan, E., Morahan, B., Seydel, K. B., Bertuccini, L., Alano, P., Williamson, K. C., Duraisingh, M. T., Taylor, T. E., Milner, D. A., Marti, M. (2014) *Plasmodium falciparum*

transmission stages accumulate in the human bone marrow. *Sci. Transl. Med.*, 6, 244re5.

Josling, G. A., Petter, M., Oehring, S. C., Gupta, A. P., Dietz, O., Wilson, D. W., Schubert, T., Langst, G., Gilson, P. R., Crabb, B. S., Moes, S., Jenoe, P., Lim, S. W., Brown, G. V., Bozdech, Z., Voss, T. S., Duffy, M. F. (2015) A *Plasmodium falciparum* bromodomain protein regulates invasion gene expression. *Cell Host Microbe*, 17, 741-751.

Juliano, J. J., Porter, K., Mwapasa, V., Sem, R., Rogers, W. O., Ariey, F., Wongsrichanalai, C., Read, A. F., Meshnick, S. F. (2010) Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proc. Natl. Acad. Sci. U. S. A.*, 107, 20138-20143.

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

Kapulu, M. C., Da, D. F., Miura, K., Li, Y., Blagborough, A. M., Churcher, T. S., Nikolaeva, D., Williams, A. R., Goodman, A. L., Sangare, I., Turner, A. V., Cottingham, M. G., Nicosia, A., Straschil, U., Tsuboi, T., Gilbert, S. C., Long, C. A., Sinden, R. E., Draper, S. J., Hill, A. V., Cohuet, A., Biswas, S. (2015) *Sci. Rep.*, 5, 11193.

Kasasa, S., Asoala, V., Gosoniu, L., Anto, F., Adjuik, M., Tindana, C., Smith, T., Owusu-Agyei, S., Vounatsou, P. (2013) Spatio-temporal malaria transmission patterns in Navrongo demographic surveillance site, northern Ghana. *Malar. J.*, 12, 63.

Katzer, F., Lizundia, R., Ngugi, D., Blake, D., McKeever, D. (2011) Construction of a genetic map for *Theileria parva*: identification of hotspots of recombination. *Int. J Parasitol.*, 41: 669-675.

Klein, E. Y., Smith, D. L., Laxminarayan, R., Levin, S. (2012) Superinfection and the evolution of resistance to antimalarials drugs. *Proc. Biol. Sci.*, 279, 3834-3842.

Konaté, L., Zwetyenga, J., Rogier, C., Bischoff, E., Fontenille, D., Tall, A., Spiegel, A., Trape, J. F., Mercereau-Puijalon, O. (1999) Variation of *Plasmodium falciparum* msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. *Trans. R. Soc. Trop. Med. Hyg.*, 93, 21-28.

Kouyos, R. D., Metcalf, J. E., Birger, R., Klein, E. Y., Abel zur Wiesch, P., Ankomah, P., Arinaminpathy, N., Bogich, T. L., Bonhoeffer, S., Brower, C., Chi-Johnston, G., Cohen, T., Day, T., Greehouse, B., Huijben, S., Metlay, J., Mideo, N., Pollitt, L. C., Read, A. F., Smith, D. L., Standley, C., Wale, N., Grenfell B.

(2014) The path of least resistance: aggressive or moderate treatment? *Proc. Biol. Soc.*, 281, 20140566.

Kozarewa, I., Ning, Z., Quail, M. A., Sanders, M. J., Berriman, M., Turner, D. J. (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods.*, 6, 291-295.

Kuleshov, V., Xie, D., Chen, R., Pushkarev, D., Ma, Z., Blauwkamp, T., Kertesz, M., Snyder, M. (2014) Whole-genome haplotyping using long reads and statistical methods. *Nat. Biotechnol.*, 32, 261-266.

Kwiatkowski, D., Nowak, M. (1991) Periodic and chaotic host-parasite interactions in human malaria. *Proc. Natl. Acad. Sci. U. S. A.*, 88, 5111-5113.

Lambros, C., Vandenberg, J. P. (1979) Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J. Parasitol.*, 65, 418-420.

Lasonder, E., Rijpma, S. R., van Schaik, B. C., Hoeijmakers, W. A., Kensche, P. R., Gresnigt, M. S., Italiaander, A., Vos, M. W., Woestenenk, R., Bousema, T., Mair, G. R., Khan, S. M., Janse, C. J., Bártfai, R., Sauerwein, R. W. (2016) Integrated transcriptomic and proteomic analyses of *P. falciparum* gametocytes: molecular insight into sex-specific processes and translational repression. *Nucleic Acids Res.*, 44, 6087-6101.

Leggett, H. C., Brown, S. P., Reece, S. E. (2014) War and peace: social interactions in infections. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 369, 20130365.

Lemieux, J. E., Gomez-Escobar, N., Feller, A., Carret, C., Amambua-Ngwa, A., Pinches, R., Day, F., Kyes, S. A., Conway, D. J., Holmes, C. C., Newbold, C. I. (2009) *Proc. Natl. Acad. Sci. U. S. A*., 106, 7559-7564.

Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., Winzeler, E. A. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, 301, 1503-1508.

Le Roux, M., Lakshmanan, V., Daily, J.P. (2009) *Plasmodium falciparum* biology: analysis of *in vitro* versus *in vivo* growth conditions. *Trends. Parasitol.,* 25, 474-481.

Machado, R. L., Povoa, M. M., Calvosa, V. S., Ferreira, M. U., Rossit, A. R., dos Santos, E. J., Conway, D. J. (2004) Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. *J. Infect. Dis.*, 190, 1547-1555.

Mackinnon, M. J., Read, A. F. (1999) Genetics relationships between parasite virulence and transmission in the rodent malaria *Plasmodium chabaudi*. *Evolution*, 53, 689-703.

Mackinnon, M. J., Read, A. F. (2004) Virulence in malaria: an evolutionary viewpoint. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 359, 965-986.

Mackintosh, C. L., Beeson, J. G., Marsh, K. (2004) Clinical features and pathogenesis of severe malaria. *Trends Parasitol.*, 20, 597-603.

MalariaGEN Plasmodium falciparum Community Project. (2016) Genomic epidemiology of artemisinin resistant malaria. *eLife*, 5, e08714.

Mancio-Silva, L., Lopez-Rubio, J. J., Claes, A., Scherf, A. (2013) Sir2a regulates rDNA transcription and multiplication rate in the human malaria parasite *Plasmodium falciparum*. *Nat. Commun.*, 4, 1530.

Manske, H. M., Kwiatkowski, D. P. (2009) SNP-o-matic. *Bioinformatics*, 25, 2434-2435.

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., O'Brien, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J. B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S. M., Marsh, K., Jiang, H., Su, X. Z., Amaratunga C., Fairhurst, R., Socheat, D., Nosten, F., Imwong, M., White, N. J., Sanders, M., Anastasi, E., Alcock, D., Drury, E., Oyola, S. O., Quail, M. A., Turner, D. J., Ruano-Rubio, V., Jyothi, D., Amenga-Etego, L., Hubbart, C., Jeffreys, A., Rowlands, K., Sutherland, C., Roper, C., Mangano, V., Modiano, D., Tan, J. C., Ferdig, M. T., Amambua-Ngwa, A., Conway, D. J., Takala-Harrison, S., Plowe, C. V., Rayner, J. C., Clark, T. G., Newbold, C. I., Berriman, M., MacInnis, B., Kwiatkowski, D. P. (2012) Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature,* 487, 375-379.

Mantel, P. Y., Hoang, A. N., Goldowitz, I., Potashnikova, D., Hamza, B., Vorobjev, I., Ghiran, I., Toner, M., Irimia, D., Ivanov, A. R., Barteneva, N., Marti, M. (2013) Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system. *Cell Host Microbe*, 13, 521-534.

Metzker, M. L. (2010) Sequencing technologies – the next generation. *Nat. Rev. Genet.*, 11, 31-46.

Miao, J., Cui, L. (2011) Rapid isolation of single malaria parasite-infected red blood cells by cell sorting. *Nat. Protoc.*, 6, 140-146.

Miao, J., Li, X., Cui, L. (2010) Cloning of *Plasmodium falciparum* by single-cell sorting. *Exp. Parasitol.*, 126, 198-202.

Mideo, N., Bailey, J. A., Hathaway, N. J., Ngasala, B., Saunders, D. L., Lon, C., Kharabora, O., Jamnik, A., Balasubramanian, S., Bjorkman, A., Martensson, A., Meshnick, S. R., Read, A. F., Juliano, J. J. (2016) A deep sequencing tool for partitioning clearance rates following antimalarials treatment in polyclonal infections. *Evol. Med. Public. Health*., 2016, 21-36.

Milani, K. J., Schneider, T. G., Taraschi, T. F. (2015) Defining the morphology and mechanism of the haemoglobin transport pathway in *Plasmodium falciparum*-infected erythrocytes. *Eukaryot. Cell*, 14, 415-426.

Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., Gould, K., Mead, D., Drury, E., O'Brien, J., Ruano Rubio, V., MacInnis B., Mwangi, J., Samarakoon, U., Ranford-Cartwright, L. C., Ferdig, M., Hayton, K., Su, X. Z., Wellems, T., Rayner, J., McVean, G., Kwiatkowski, D. P. (2016) Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Res.,* Epub, doi:10.1101/gr.203711.115.

Miller, L. H. (1969) Distribution of mature trophozoites and schizonts of *Plasmodium falciparum* in the organs of *Aotus trivirgatus,* the night monkey. *Am. J. Trop. Med. Hyg.*, 18, 860-865.

Miller, L. H., Ackerman, H. C., Su, X. Z., Wellems, T. E. (2013) Malaria biology and disease pathogenesis: insights for new treatments. *Nat. Med.*, 19, 156-167.

Minter, E. J. A., Lowe, C. D., Brockhurst, M. A., Watts, P. C. (2015) A rapid and cost-effective quantitative microsatellite genotyping protocol to estimate intraspecific competition in protist microcosm experiments. *Methods Ecol. Evol.*, 6, 315-323.

Miotto O., Almagro-Garcia, J., Manske, M., MacInnis, B., Campino, S., Rockett, K. A., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J. M., Duong, S., Nguon, C., Chuor, C. M., Saunders, D., Se, Y., Lon, C., Fukuda, M. M., Amenga-Etego, L., Hodgson, A. V., Asoala, V., Imwong, M., Takala-Harrison, S., Nosten, F., Su, X. Z., Ringwald, P., Ariey, F., Dolecek, C., Hien, T. T., Boni, M. F., Thai, C. Q., Amambua-Ngwa, A., Conway, D. J., Djimde, A. A., Doumbo, O. K., Zongo, I., Ouedraogo, J. B., Alcock, D., Drury, E., Auburn, S., Koch, O., Sanders, M., Hubbart, C., Maslen, G., Ruano-Rubio, V., Jyothi, D., Miles, A., O'Brien, J., Gamble, C., Oyola, S. O., Rayner, J. C., Newbold, C. I., Berriman, M., Spencer, C. C., McVean, G., Day, N. P., White, N. J., Bethell, D., Dondorp, A. M., Plowe, C. V., Fairhurst, R. M., Kwiatkowski, D. P. (2013) Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat. Genet.*, 45, 648-655.

Mitchell, G. H., Hadley, T. J., McGinniss, M. H., Klotz, F. W., Miller, L. H. (1986) Invasion of erythrocytes by *Plasmodium falciparum* malaria parasites: evidence for receptor heterogeneity and two receptors. *Blood*, 67, 1519-1521.

Mitri, C., Theiry, I., Bourgoin, C., Paul, R. E. (2009) Density-dependent impact of the human malaria parasite *Plasmodium falciparum* gametocyte sex ratio on mosquito infection rates. *Proc. Biol. Soc.*, 276, 3721-3726.

Mobegi, V. A., Duffy, C. W., Amambua-Ngwa, A., Loua, K. M., Laman, E., Nwakanma, D. C., MacInnis, B., Aspeling-Jones, H., Murray, L., Clark, T. G., Kwiatkowski, D. P., Conway, D. J. (2014) Genome-wide analysis of selection on

the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol. Biol. Evol.*, 31, 1490-1499.

Mobegi, V. A., Loua, K. M., Ahouidi, A. D., Satoguina, J., Nwakanma, D. C., Amambua-Ngwa, A., Conway, D. J. (2012) Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. *Malar. J.*, 11, 223.

Mok, S., Ashley, E. A., Ferreira, P. E., Zhu, L., Lin, Z., Yeo, T., Chotivanich, K., Imwong, M., Pukrittayakamee, S., Dhorda, M., Nguon, C., Lim, P., Amaratunga, C., Suon, S., Hien, T. T., Htut, Y., Faiz, M. A., Onyamboko, M. A., Mayxay, M., Newton, P. N., Tripura, R., Woodrow, C. J., Miotto, O., Kwiatkowski, D. P., Nosten, F., Day, N. P., Preiser, P. R., White, N. J., Dondorp, A. M., Fairhurst, R. M., Bozdech, Z. (2015) Drug resistance. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. (2015) *Science*, 347, 431-435.

Molineaux, L., Gramiccia, G. (1980) The Garki Project. Research on the epidemiology and control of malaria in the Sudan savanna of West Africa. World Health Organization, (Geneva).

Moon, R. W., Sharaf, H., Hastings, C. H., Ho, Y. S., Nair, M. B., Rchiad, Z., Knuepfer, E., Ramaprasad, A., Mohring, F., Amir, A., Yusuf, N. A., Hall, J., Almond, N., Lau, Y. L., Pain, A., Blackman, M. J., Holder, A. A. *Proc. Natl. Acad. Sci. U. S. A.*, 113, 7231-7236.

Murray, L., Mobegi, V. A., Duffy, C. W., Assefa, S. A., Kwiatkowski, D. P., Laman, E., Loua, K. M., Conway, D. J. (2016) Microsatellite genotyping and genome-wide single nucleotide polymorphism-based indices of *Plasmodium falciparum* diversity within clinical infections. *Malar. J.*, 15, 275.

Mzilahowa, T., McCall, P. J., Hastings, I. M. (2007), “Sexual” population structure and genetics of the malaria agent *P. falciparum*. PLoS One., 18, e613.

Nair, S., Nkhoma, S. C., Serre, D., Zimmerman, P. A., Gorena, K., Daniel, B. J., Nosten, F., Anderson, T. J., Cheeseman, I. H. (2014) Single-cell genomics for dissection of complex malaria infections. *Genome Res.*, 24, 1028-1038.

Nakazawa, S., Culleton, R., Maeno, Y. (2011) *In vivo* and *in vitro* gametocyte production of *Plasmodium falciparum* isolates from Northern Thailand. *Int. J. Parasitol.*, 41, 317-323.

Neal, A. T., Schall, J. J. (2014) Testing sex ratio theory with the malaria parasite *Plasmodium mexicanum* in natural and experimental infections. *Evolution*, 68, 1071-1081.

Nery, S., Deans, A. M., Mosobo, M., Marsh, K., Rowe, J. A., Conway, D. J. (2006) Expression of *Plasmodium falciparum* genes involved in erythrocyte

invasion varies among isolates cultured directly from patients. *Mol. Biochem. Parasitol.*, 149, 208-215.

Nilsson, S. K., Childs, L. M., Buckee, C., Marti, M. (2015) Targeting human transmission biology for malaria elimination. *PLoS Pathog.*, 18, e1004871.

Nkhoma, S. C., Nair, S., Cheeseman, I. H., Rohr-Allegrini, C., Singlam, S., Nosten, F., Anderson, T. J. (2012) Close kinship within multiple-genotype malaria parasite infections. *Proc. Biol. Soc.*, 279, 2589-2998.

Noor, A. M., Kinyoki, D. K., Mundia, C. W., Kabaria, C. W., Mutua, J. W., Alegana, V. A., Fall, I. S., Snow, R. W. (2014) The changing risk of *Plasmodium falciparum* malaria infection in Africa 2000-2010: a spatial and temporal analysis of transmissioin intensity. *Lancet*, 383, 1739-1747.

Nsobya, S. L., Kiggundu, M., Joloba, M., Dorsey, G., Rosenthal, P. J. (2008) Complexity of *Plasmodium falciparum* clinical samples from Uganda during short-term culture. *J. Infec. Dis.*, 198, 1554-1557.

Ntoumi, F., Ngoundou-Landji, J., Lekoulou, F., Luty, A., Deloron, P., Ringwald, P. (2000) Site-based study on polymorphism of *Plasmodium falciparum* MSP-1 and MSP-2 genes in isolates from two villages in Central Africa. *Parassitologia*, 42, 197-203.

Nuralitha, S., Siregar, J. E., Syafruddin D., Roelands, J., Verhoef, J., Hoepelman, A. I., Marzuki, S. (2015) Within-host selection of drug resistance in a mouse model of repeated incomplete malaria treatment: Comparison between atovaquone and pyrimethamine. *Antimicrob. Agents Chemother.*, 60, 258-263.

Nwakanma, D. C., Duffy, C. W., Amambua-Ngwa, A., Oriero, E. C., Bojang, K. A., Pinder, M., Drakeley, C. J., Sutherland, C. J., Milligan, P. J., MacInnis, B., Kwiatkowski, D. P., Clark, T. G., Greenwood, B. M., Conway, D. J. (2014) Changes in malaria parasite drug resistance in an endemic population over a 25-year period with resulting genomic evidence of selection. *J. Infect. Dis.*, 209, 1126-1135.

Nwakanma, D. C., Gomez-Escobar, N., Walter, M., Crozier, S., Dubovsky, F., Malkin, E., Locke, E., Conway, D. J. (2009) Quantitative detection of *Plasmodium falciparum* DNA in saliva, blood and urine. *J. Infect. Dis.*, 199, 1567-1574.

Oakley, M. S., Gerald, N., McCutchan, T. F., Aravind, L., Kumar, S. (2011) Clinical and molecular aspects of malaria fever. *Trends Parasitol.* 27, 442-449.

Ochola, L. I., Tetteh, K. K., Stewart, L. B., Riitho, V., Marsh, K., Conway, D. J. (2010) Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. *Mol. Bio. Evol.*, 27, 2344-2351.

Ocholla, H., Preston, M. D., Mipando, M., Jensen, A. T., Campino, S., MacInnis, B., Alcock, D., Terlouw, A., Zongo, I., Oudraogo, J. B., Djimde, A. A., Assefa, S., Doumbo, O. K., Borrman, S., Nzila, A., Marsh, K., Fairhurst, R. M., Nosten, F., Anderson, T. J., Kwiatkowski, D. P., Craig, A., Clark, T. G., Montgomery, J. (2014) Whole-genome scans provide evidence of adaptive evolution of Malawian *Plasmodium falciparum* isolates. *J Infect. Dis.*, 210, 1991-2000.

Okell, L. C., Ghani. A. C., Lyons, E., Drakeley, C. J. (2009) Submicroscopic infection in *Plasmodium falciparum*-endemic populations: a systematic review and meta-analysis. *J Infect Dis.*, 200, 1509-1517.

Osier, F. H., Murungi, L. M., Fegan, G., Tuju, J., Tetteh, K. K., Bull, P. C., Conway, D. J., Marsh, K. (2010) Allele-specific antibodies to *Plasmodium falciparum* merozoite surface protein-2 and protection against clinical malaria. *Parasite Immunol.*, 32, 193-201.

Otto, T. D., Rayner, J. C., Bohme, U., Pain, A., Spottiswoode, N., Sanders, M., Quail, M., Ollomo, B., Renaud, F., Thomas, A. W., Prugnolle, F., Conway, D. J., Newbold, C., Berriman, M. W. (2014) Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat. Commun.*, 5, 4754.

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

Owusu-Agyei, S., Asante, K. P., Adjuik, M., Adjei, G., Awini, E., Adams, M., Newton, S., Dosoo, D., Dery, D., Agyeman-Budu, A., Gyapong, J., Greenwood, B., Chandramohan, D. (2009) Epidemiology of malaria in the forest-savannah transitional zone of Ghana. *Malar. J.*, 8, 220.

Oyola, S. O., Manske, M., Campino, S., Claessens, A., Hamilton, W. L., Kekre, M., Drury, E., Mead, D., Gu, Y., Miles, A., MacInnis, B., Newbold, C., Berriman, M., Kwiatkowski, D. P. (2014) Optimized whole-genome amplification strategy for extremely AT-biased template. *DNA Res.*, 21, 661-671.

Oyola, S. O., Otto, T. D., Gu, Y., Maslen, G., Manske, M., Campino, S., Turner, D. J., MacInnis B., Kwiatkowski, D. P., Swerdlow, H. P., Quail, M. A. (2012) Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics*, 13, 1.

Park, D. J., Lukens, A. K., Neafsey, D. E., Schaffner, S. F., Chang, H. H., Valim, C., Ribacke, U., Van Tyne, D., Galinsky, K., Galligan, M., Becker, J. S., Ndiaye, D., Mboup, S., Wiegand, R. C., Hartl, D. L., Sabeti, P. C., Wirth, D. F., Volkman, S. K. (2012) *Proc. Natl. Acad. Sci. U. S. A.*, 109, 13052-13057.

Pattaradilokrat, S., Culleton, R. L., Cheesman, S. J., Carter, R. (2009) Gene encoding erythrocyte binding ligand linked to blood stage multiplication rate

phenotype in *Plasmodium yoelli yoelli. Proc. Natl. Acad. Sci. U. S. A.*, 106, 7161-7666.

Paul, R. E., Coulson, T. N., Raibaud, A., Brey, P. T. (2000) Sex determination in malaria parasites. *Science*, 287, 128-131.

Paul, R. E., Packer, M. J., Walmsley, M., Lagog, M., Ranford-Cartwright, L. C., Paru, R., Day, K. P. (1995) Mating patterns in malaria parasite populations of Papua New Guinea. *Science*, 269, 1709-1711.

Petersen, I., Gabryszewski, S. J., Johnston, G. L., Dhingra, S. K., Ecker, A., Lewis, R. E., de Almeida M. J., Straimer, J., Henrich, P. P., Palatulan, E., Johnson, D. J., Coburn-Flynn, O., Sanchez, C., Lehane, A. M., Lanzer, M., Fidock, D. A. (2015) Balancing drug resistance and growth rates via compensatory mutations in the *Plasmodium falciparum* chloroquine resistance transporter. *Mol. Microbiol.*, 97, 381-395.

Peters, J. M., Fowler, E. V., Krause, D. R., Cheng, Q., Gatton, M. L. (2007) Differential changes in *Plasmodium falciparum var* transcription during adaptation to culture. *J. Infec. Dis.*, 195, 748-755.

Polley, S. D., Conway, D. J. (2001) Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics*, 158, 1505-1512.

Pollitt, L. C., Huijben, S., Sim, D. G., Salathe, R. M., Jones, M. J., Read, A. F. (2014) Rapid response to selection, competitive release and increased transmission potential of artesunate-selected *Plasmodium chabaudi* malaria parasites. *PLoS Pathog.*, 10, e1004019.

Pollitt, L. C., Mideo, N., Drew, D. R., Schneider, P., Colegrave, N., Reece, S. E. (2011) Competition and the evolution of reproductive restraint in malaria parasites. *Am. Nat.*, 177, 358-367.

Ponnudurai, T., Leeuwenberg, A. D., Meuwissen, J. H. (1981) Chloroquine sensitivity of isolates of *Plasmodium falciparum* adapted to *in vitro* culture. *Trop. Geogr. Med.*, 33, 50-54.

Ponnuduruai, T., Lensen, A. H., van Gemert, G. J., Bolmer, M. G., Meuwissen, J. H. (1991) Feeding behaviour and sporozoite ejection by infected *Anopheles stephensi*. *Trans. R. Soc. Trop. Med. Hyg.*, 85, 175-180.

Ponnuduruai, T., Meuwissen, J. H., Leeuwenberg, A. D., Verhave, J. P., Lensen, A. H. (1982) The production of mature gametocytes of *Plasmodium falciparum* in continuous cultures of different isolates infective to mosquitoes. *Trans. R. Soc. Trop. Med. Hyg.*, 76, 242-249.

Prudencio, M., Rodriguez, A., Mota, M. M. (2006) The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nat. Rev. Microbiol.*, 4, 849-856.

Pumpaibool, T., Arnathau, C., Durand, P., Kanchanakhan, N., Siripoon, N., Suegorn, A., Sitthi-Amorn, C., Renaud, F., Harnyuttanakorn, P. (2009) Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country. *Malar. J.*, 8, 155.

Quakyi, I. A., Carter, R., Rener, J., Kumar, N., Good. M. F., Miller, L. H. (1987) The 230-kDa gamete surface protein of *Plasmodium falciparum* is also a target for transmission-blocking antibodies. *J. Immunol.*, 139, 4213-4217.

Raberg, L., de Roode, J. C., Bell, A. S., Stamou, P., Gray, D., Read, A. F. (2006) The role of immune-mediate apparent competition in genetically diverse malaria infections. *Am. Nat.*, 168, 41-53.

Ranford-Cartwright, L. C., Balfe, P., Carter, R., Walliker, D. (1991) Genetic hybrids of *Plasmodium falciparum* identified by amplification of genomic DNA from single oocysts. *Mol. Biochem. Parasitol.*, 49, 239-243.

Ranford-Cartwright, L. C., Balfe, P., Carter, R., Walliker, D. (1993) Frequency of cross-fertilization in the human malaria parasite *Plasmodium falciparum*. *Parasitology*, 107, 11-8.

Ranford-Cartwright, L. C., Mwangi, J. M. (2012) Analysis of malaria parasite phenotypes using experimental genetic crosses of *Plasmodium falciparum*. *Int. J. Parasitol.*, 42, 529-534.

Read, A. F., Narara, A., Nee, S., Keymer, A. E., Day, K. P. (1992) Gametocyte sex ratios as indirect measures of outcrossing rates in malaria. *Parasitology*, 104, 387-395.

Reece, S. E., Drew, D. R., Gardner, A. (2008) Sex ratio adjustment and kin discrimination in malaria parasites. *Nature*, 453, 609-614.

Regev-Rudzki, N., Wilson, D. W., Carvalho, T. G., Sisquella, X., Coleman, B. M., Rug, M., Bursac, D., Angrisano, F., Gee, M., Hill, A. F., Baum, J., Cowman, A. F. (2013) Cell-cell communication between malaria-infected red blood cells via exosome-like vesicles. *Cell*, 153, 1120-1133.

Reilly, H. B., Wang, H., Steuter, J. A., Marx, A. M., Ferdig, M. T. (2007) Quantitative dissection of clone-specific growth rates in cultured malaria parasites. *Int. J. Parasitol.*, 37, 1599-1607.

Reilly Ayala, H. B., Wacker, M. A., Siwo, G., Ferdig, M. T. (2010) Quantitative trait loci mapping reveals candidate pathways regulating cell cycle duration in *Plasmodium falciparum*. *BMC Genomics,* 11, 577.

Ribaut, C., Berry, A., Chevalley, S., Reybier, K., Morlais, I., Parzy, D., Nepveu, F., Benoit-Vical, F., Valentin, A. (2008) Concentration and purification by magnetic separation of the erythrocytic stages of all human *Plasmodium* species. *Malar. J.*, 7, 45.

Rita Gomes, A., Bushell, E., Schwach, F., Girling, G., Anar, B., Quail, M. A., Herd, C., Pfander, C., Modrzynska, K., Rayner, J. C., Billker, O. (2015) A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite. *Cell Host Microbe*, 17, 404-413.

Robert, V., Read, A. F., Essong, J., Tchuinkam, T., Mulder, B., Verhave, J. P., Carnevale, P. (1996) Effect of gametocyte sex ratio on infectivity of *Plasmodium falciparum*. *Trans. R. Soc. Trop. Med. Hyg.*, 90, 621-624.

Robert, V., Sokhna, C. S., Rogier, C., Ariey, F., Trape, J. F. (2003) Sex ratio of *Plasmodium falciparum* gametocytes in inhabitants of Dielmo, Senegal. *Parasitology*, 127, 1-8.

Robinson, T., Campino, S. G., Auburn, S., Assefa, S. A., Polley, S. D., Manske, M., MacInnis, B., Rockett, K. A., Maslen, G. L., Quail, M. A., Chiodini, P. L., Kwiatkowski, D. P., Clark, T. G., Sutherland, C. J. (2011) Drug-resistant genotypes and multi-clonality in *Plasmodium falciparum* analysed by direct genome sequencing from peripheral blood of malaria patients. *PLoS One*, 6, e23204.

Rosenberg, R., Wirtz, R. A., Schenider, I., Burge, R. (1990) An estimation of the number of malaria sporozoites ejected by a feeding mosquito. *Trans. R. Soc. Trop. Med. Hyg.*, 84, 209-212.

Rousset, F., Roze, D. (2007) Constraints on the origin and maintenance of genetic kin recognition. *Evolution*, 61, 2320-2330.

Rovira-Graells, N., Gupta, A. P., Planet, E., Crowley, V. M., Mok, S., Ribas de Pouplana, L., Preiser, P. R., Bozdech, Z., Cortes, A. (2012) Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res.*, 22, 925-938.

Sama, W., Killeen, G., Smith, T. (2004) Estimating the duration of *Plasmodium falciparum* infection from trials of indoor residual spraying. *Am. J. Trop. Med. Hyg.*, 70, 625-634.

Santhanam, J., Raberg, L., Read, A. F., Savill, N. J. (2014) Immune-mediate competition in rodent malaria is most likely caused by induced changes in innate immune clearance of merozoites. *PLoS Comput. Biol.*, 10, e1003416.

Schall, J. J. (2009) Do malaria parasites follow the algebra of sex ratio theory? *Trends. Parasitol.*, 25, 120-123.

Schneider, P., Reece, S. E., van Schaijk, B. C., Bousema, T., Lanke, K. H., Meaden, C. S., Gadalla, A., Ranford-Cartwright, L. C., Babiker, H. A. (2015) Quantification of female and male gametocytes by reverse transcriptase quantitative PCR. *Mol. Biochem. Parasitol.*, 199, 29-33.

Schultz, L., Wapling, J., Mueller, I., Ntsuke, P. O., Senn, N., Nale, J., Kiniboro, B., Buckee, C. O., Tavul, L., Siba, P. M., Reeder, J. C., Barry, A. E. (2010) Multilocus haplotypes reveal variable levels of diversity and population structure of *Plasmodium falciparum* in Papua New Guinea, a region of intense perennial transmission. *Malar. J.*, 9, 336.

Schwach, F., Bushell, E., Rita Gomes, A., Anar, B., Girling, G., Herd, C., Rayner, J. C., Billker, O. (2014) *Plasmo*GEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites. *Nucleic Acids Res*., 43, D1176-D1182.

Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N., Palmer, A. E., Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.*, 22, 1567-1572.

Shcherbo, D., Murphy, C. S., Ermakova, G. V., Solovieva, E. A., Chepurnykh, T. V., Shcheglov, A. S., Verkhusha V. V., Pletnev, V. Z., Hazelwood, K. L., Roche, P. M., Lukyanov, S., Zaraisky, A. G., Davidson, M. W., Chudakov, D. M. (2009) Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.*, 418, 567-574.

Simpson, J. A., Aarons, L., Collins, W. E., Jeffery, G. M., White, N. J. (2002) Population dynamics of untreated *Plasmodium falciparum* malaria within the adult human host during the expansion phase of infection. *Parasitology*, 124, 247-263.

Sinden, R. E. (2009) Malaria, sexual development and transmission: retrospect and prospect. *Parasitology*, 136, 1427-1434.

Sinden, R. E., Hartley, R. H. (1985) Identification of the meiotic division of malarial parasites. *J. Protozool.*, 32: 742-744.

Sisya, T. J., Kamn'gona, R. M., Vareta, J., Fulakeza, J. M., Mukaka, M. F. J., Seydel, K. B., Laufer, M. K., Taylor, T. E., Nkhoma, S. C. (2015) Subtle change in *Plasmodium falciparum* infection complexity following enhanced intervention in Malawi. *Acta Tropica*, 142, 108-114.

Soulard, V., Bosson-Vanga, H., Lorthois, A., Roucher, C., Franetich, J. F., Zanghi, G., Bordessoulles, M., Thellier, M., Morosan, S., Le Naour, G., Capron, F., Suemizu, H., Snounou, G., Moreno-Sabater, A., Mazier, D. (2015) *Plasmodium falciparum* full life cycle and *Plasmodium ovale* liver stages in humanized mice. *Nat. Commun.*, 6, 7690.

Snounou, G., Viriyakosol, S., Zhu, X. P., Jarra, W., Pinheiro, L., do Rosario, V. E., Thaithong, S., Brown, K. N. (1993) High sensitivity of detection of human malaria parasites by the use of the nested polymerase chain reaction. *Mol. Biochem. Parasitol.*, 61, 315-320.

Snounou, G., Zhu, X., Siripoon, N., Jarra, W., Thaithong, S., Brown, K. N., Viriyakosol, S. (1999) Biased distribution of msp1 and msp2 allelic variants in *Plasmodium falciparum* populations in Thailand. *Trans. R. Soc. Trop. Med. Hyg.*, 93, 369-374.

Sowunmi, A., Gbotosho, G. O., Happi, C. T., Folarin, O. A., Balogun, S. T. (2009) Population structure of *Plasmodium falciparum* gametocyte sex ratios in malarious children in an endemic area. *Parasitol. Int.*, 58, 438-443.

Stanisic, D. I., Liu, X. Q., De, S. L., Batzloff, M. R., Forbes, T., Davis, C. B., Sekuloski, S., Chavchich, M., Chung, W., Trenholme, K., McCarthy, J. S., Li. T., Sim, B. K., Hoffman, S. L., Good, M. F. (2015) Development of cultured *Plasmodium falciparum* blood-stage malaria cell banks for early phase *in vivo* clinical trial assessment of anti-malaria drugs and vaccines. *Malar. J.*, 14, 143.

Stepniewska, K., Price, R. N., Sutherland, C. J., Drakeley, C. J., von Seidlin, L., Nosten, F., White, N. J. (2008) *Plasmodium falciparum* gametocyte dynamics in areas of different malaria endemicity. *Malar. J.*, 7:249.

Stresman, G. H., Stevenson, J. C., Ngwu, N., Marube, E., Owaga, C., Drakeley, C., Bousema, T., Cox, J. (2014) High levels of asymptomatic and subpatent *Plasmodium falciparum* parasite carriage at health facilities in an area of heterogeneous malaria transmission intensity in the Kenyan highlands. *Am. J. Trop. Med. Hyg.*, 91, 1101-1108.

Sundararaman, S. A., Plenderleith, L. J., Liu, W., Loy, D. E., Learn, G. H., Li, Y., Shaw, K. S., Ayouba, A., Peeters, M., Speede, S., Shaw, G. M., Bushman, F. D., Brisson, D., Rayner, J. C., Sharp, P. M., Hahn, B. H. (2016) Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat. Commun.*, 7, 11078.

Su, X. Z., Ferdig, M. T., Huang, Y., Huynh, C. Q., Liu, A., You, J., Wootton, J. C., Wellems, T. E. (1999) A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science*, 286, 1351-1353.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.

Takala-Harrison, S., Clark, T. G., Jacob, C. G., Cummings, M. P., Miotto, O., Dondorp, A. M., Fukuda, M. M., Nosten, F., Noedl, H., Imwong, M., Bethell, D., Se, Y., Lon, C., Tyner, S. D., Saunders, D. L., Socheat, D., Ariey, F., Phyo, A. P., Starzengruber, P., Fuehrer, H. P., Swoboda, P., Stepniewska, K., Flegg, J., Arze, C., Cerqueira, G. C., Silva, J. C., Ricklefs, S. M., Porcella, S. F., Stephens, R. M., Adams, M., Kenefic, L. J., Campino, S., Auburn, S., MacInnis, B., Kwiatkowski, D. P., Su, X. Z., White, N. J., Ringwald, P., Plowe, C. V. (2013) Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. *Proc. Natl. Acad. Sci. U. S. A.*, 110, 240-245.

Talman, A.M., Blagborough, A.M. and Sinden, R.E. (2010) A *Plasmodium falciparum* strain expressing GFP throughout the parasite's life-cycle. *PLoS One*, 5, e9156.

Talman, A. M., Domarle, O. Mckenzie, F. E., Ariey, F., Robert, V. (2004) Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malar. J.*, 3, 24.

Tanriverdi, S., Blain, J. C., Deng, B., Ferdig, M. T., Widmer, G. (2007) Genetic crosses in the apicomplexan parasite *Cryptosporidium parvum* define recombination parameters. *Mol. Microbiol.*, 63, 1432-1439.

Tao, D., Ubaida-Mohien, C., Mathias, D. K., King, J. G., Pastrana-Mena, R., Tripathi, A., Goldowitz, I., Graham, D. R., Moss, E., Marti, M., Dinglasan, R. R. (2014) Sex-partitioning of the *Plasmodium falciparum* stage V gametocyte proteome provides insight into falciparum-specific cell biology. *Mol. Cell. Proteomics*, 13, 2705-2724.

Tetteh, K. K., Stewart, L. B., Ochola, L. I., Amambua-Ngwa A., Thomas, A. W., Marsh, K., Weedall, G. D., Conway, D. J. (2009) Prospective identification of malaria parasite genes under balancing selection. *PLoS One*, 4, e5568.

Trager, W., Jensen, J. B. (1976) Human malaria parasites in continuous culture. *Science*, 193, 673-675.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., Rozen, S. G. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Res.*, 40, e115.

Vaughan, A. M., Mikolajczak, S. A., Wilson, E. M., Grompe, M., Kaushansky, A., Camargo, N., Bial, J., Ploss, A., Kappe, S. H. (2012) Complete *Plasmodium falciparum* liver-stage development in liver-chimeric mice. *J Clin Invest.,* 122, 3618-3628.

Vembar, S. S., Seetin, M., Lambert, C., Nattestad, M., Schatz, M. C., Baybayan, P., Scherf, A., Smith, M. L. (2016) Complete telomere-to-telomere de novo assembly of the *Plasmodium falciparum* genome through long-read (>11kb), single-molecule, real-time sequencing. *DNA Res.*, 23, 339-351.

Volkman, S. K., Sabeti, P. C., DeCaprio, D., Neafsey, D., Schaffner, S. F., Milner, D. A. Jr., Daily, J. P., Sarr, O., Ndiaye, D, Ndir, O., Mboup, S., Duraisingh, M. T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R. Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D. B., Zainoun, J., Wiegand, R. C., Birren, B. W., Hartl, D. L., Galagan, J. E., Lander, E. S., Wirth, D. F. (2007) A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.*, 39, 113-119.

Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T., Corcoran, L. M., Burkot, T. R., Carter. R. (1987) *Science*, 236, 1661-1666.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189-1191.

Weedall, G. D., Conway, D. J. (2010) Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Trends Parasitol*., 26, 363-369.

Weir, B. S., Cockerham, C. C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 13.

Wilson, D.W., Crabb, B.S., Beeson, J.G. (2010) Development of fluorescent *Plasmodium falciparum* for *in vitro* growth inhibition assays. *Malar. J.*, 9, 152.

Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W., Williamson, D. H. (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 261, 155-172.

Wacker, M.A., Turnbull, L.B., Walker, L.A., Mount, M.C., Ferdig, M.T. (2012) Quantification of multiple infections of *Plasmodium falciparum in vitro*. *Malar. J,.* 11, 180.

Waters, A. P. (2016) Epigenetic roulette in blood stream *Plasmodium*: Gambling on sex. *PLoS Pathog.*, 12, e1005353.

Wellems, T. E., Panton, L. J., Gluzman, I. Y., do Rosario, V. E., Gwadz, R. W., Walker-Jonah, A., Krogstad, D. J. (1990) Chloroquine resistance not linked to mdr-like genes in a *Plasmodium falciparum* cross. *Nature*, 345, 253-255.

Wellems, T. E., Walker-Jonah, A., Panton, L. J. (1991) Genetic mapping of the chloroquine-resistance locus on *Plasmodium falciparum* chromosome 7. *Proc. Natl. Acad. Sci. U. S. A.*, 88, 3382-3386.

White, N. J. (2010) Artemisinin resistance—the clock is ticking. *Lancet*, 376, 2051-2052.

White, J. 3rd, Mascarenhas, A., Pereira, L., Dash, R., Walke, J. T., Gawas, P., Sharma, A., Manoharan, S. K., Guler, J. L., Maki, J. N., Kumar, A., Mahanta, J., Valecha, N., Dubhashi, N., Vaz, M., Gomes, E., Chery, L., Rathod, P. K. (2016) In vitro adaptation of *Plasmodium falciparum* reveal variations in cultivability. *Malar. J.*, 15, 33.

Wright, C. F., Morelli, M. J., Thébaud, G., Knowles, N. J., Herzyk, P., Paton, D. J., Haydon, D. T., King, D. P. (2011) Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *J. Virol.*, 85, 2266-2275.

World Health Organization (2015). World Malaria Report (Geneva).

Yamagishi, J., Natori, A., Tolba, M. E., Mongan, A. E., Sugimoto, C., Katayama, T., Kawashima, S., Makalowski, W., Maeda, R., Eshita, Y., Tuda, J., Suzuki, Y. (2014) Interactive transcriptome analysis of malaria patients and infecting *Plasmodium falciparum*. *Genome Res.*, 24, 1433-1444.

Zervos, T.M., Hernandez, J.N., Sutton, P.L. and Branch, O.H. (2012) Quantification of *Plasmodium falciparum* malaria from complex infections in the Peruvian Amazon using quantitative PCR of the merozoite surface protein 1, block 2 (PfMSP1-B2): *in vitro* dynamics reveal density-dependent interactions. *Parasitology*, 139, 701-708.

Zhang, Q., Zhang, Y., Hang, Y., Xue, X., Sun, X., Wang, J., Mccutchan, T. F., Pan, W. (2011) From *in vivo* to *in vitro*: Dynamic analysis of *P. falciparum var* gene expression patterns of patient isolates during adaptation to culture. *PLoS One*, 6, e20591.

List of Appendices

As listed on attached CD-Rom:

Appendix 1 Example R script for calculation of isolate F_{WS} scores

Appendix 2 Primers used in fluorescent parasite generation

Appendix 3 COIL genotype barcodes for Guinean populations

Appendix 4 Guinean F_{WS}, microsatellite typing, COIL and estMOI scores

Appendix 5 Isolate F_{WS} scores for comparison of West African populations

Appendix 6 to 9 Gene F_{WS} scores for West African populations (Guinea, Gambia, Northern Ghana, Central Ghana respectively)

Appendix 10 to 13 GO Term analysis results for West African populations (Guinea, Gambia, Northern Ghana, Central Ghana respectively)