On the molecular evolution of the *Plasmodium falciparum* genome

Pierre Marcel Durand

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa, in fulfillment of the requirements for the degree of Doctor of Philosophy. This thesis is presented as a series of publications and unpublished manuscripts.

Johannesburg, 2009

DECLARATION

I declare that this thesis is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg, and has not been submitted before for any degree or examination at any other university. Where applicable, the co-author contributions to each publication are listed under "Author Contributions", unless stated otherwise.

…….
Pierre Durand

……. day of …….. 2009

Ethics clearance was obtained from the University of the Witwatersrand for culturing malaria parasites in the blood of human volunteers (clearance number: M03-11-06).

DEDICATION

To my mother and father

and Nazir Hoosen

on science

“I think that only daring speculation can lead us further and not accumulation of facts.”

on humanity

“The true value of a human being is determined primarily by the measure and the sense in which he has attained liberation from the self.”

These quotes are attributed to Albert Einstein.

ACKNOWLEDGEMENTS

I express my sincere appreciation to

- my supervisor, Prof. Theresa L. Coetzer, for her guidance, expertise, sincerity and professional ethics,
- my family, David, La Clanche, Samantha, Daniel, Jean and Silvia,
- colleagues, students and friends, who have advised, inspired, or otherwise assisted with helpful conversations.

I gratefully acknowledge the financial support from

- The National Health Laboratory Service
- The University of the Witwatersrand Medical Faculty Research Endowment Fund
- The National Research Foundation
- The National Bioinformatics Network

ABSTRACT

Research in the *Plasmodium falciparum* molecular evolution field has predominantly comprised three distinct areas: phylogenetics, host-parasite co-evolution and evolutionary genomics. These areas have greatly enhanced our understanding of the early origins of the phylum Apicomplexa, the emergence of *P. falciparum*, and the co-evolution between parasite and human hereditary erythrocyte disorders. In addition, the genome sequencing projects have elucidated the complexity and extremely unusual nature of the parasite genome. Some aspects of parasite molecular evolution, however, are controversial, such as human pyruvate kinase (PK) deficiency and *P. falciparum* virulence co-evolution. Other aspects, like *Plasmodium* whole genome evolution have remained unexplored.

This thesis includes a collection of manuscripts that address aspects of the broad field of *P. falciparum* molecular evolution. The first deals with the limitations of bioinformatic methods as applied to *P. falciparum*, which have arisen due to the unusual nature of the parasite genome, such as the extreme nucleotide bias. Although conventional bioinformatics can partially accommodate and compensate for the genome idiosyncrasies, these limitations have hampered progress significantly. A novel alignment method, termed FIRE (Functional Inference using the Rates of Evolution) was therefore developed. FIRE uses the evolutionary constraints at codon sites to align sequences and infer domain function and overcomes the problem of poor sequence similarity, which is commonly encountered between *P. falciparum* and other taxa. A second aspect addressed in this thesis, is the host-parasite relationship in the context of PK deficiency. It was demonstrated that PK deficient erythrocytes are dramatically resistant to parasite infection, providing *in vitro* evidence for this phenomenon and confirming this aspect of host-parasite co-evolution.

The unexplored field of parasite genome evolution was initiated in this thesis by investigating two major role-players in genome dynamics, mobile genetic

elements (MGEs) and programmed cell death (PCD). MGEs were absent in *P. falciparum*, possibly due to a geno-protective mechanism, which increased the AT nucleotide bias. Interestingly, the parasite telomerase reverse transcriptase, which is a domesticated MGE, was identified. In addition, there is genomic evidence for the second determinant, a classical PCD pathway. Intriguingly, functional and structural evidence for a p53-like DNA-binding domain, which plays a key role in genome evolution, was obtained. Using MGEs and PCD as examples, a theoretical framework for investigating genome dynamics was developed. The framework proposes an ecological approach to genome evolution, in which a trade-off exists between two opposing processes: the generation of diversity by factors such as MGEs and the maintenance of integrity by factors like PCD. The framework is suggested for proposing and testing hypotheses to investigate the origins and evolution of the *P. falciparum* genome.

Finally, a novel approach, termed Evolutionary Patterning (EP), was developed to limit the problem of parasite drug resistance and demonstrates the value of employing molecular evolution to address biomedical challenges.

Some of this work, such as the FIRE method, the host-parasite co-evolution studies, the PCD findings and the EP approach have been incorporated in grant proposals and adopted in future projects. It is hoped that this research will be used to further our understanding of *P. falciparum* evolution and advance the efforts to control this deadly pathogen.

PUBLICATIONS PRESENTED IN THIS THESIS

(in the order presented in this thesis)

Durand PM and Coetzer TL. **Utility of computational methods to identify the apoptosis machinery in unicellular eukaryotes.** Bioinformatics and Biology Insights 2008, 2: 101-117.

Durand PM, Hazelhurst S and Coetzer TL. **Evolutionary rates at codon sites may be used to align sequences and infer protein domain function.** BMC Bioinformatics 2009, a revised manuscript is in second review.

Hazelhurst S and Durand PM. ***FIRE:*** **Functional Inference using the Rate of Evolution.** To be submitted following acceptance of the above manuscript.

Durand PM and Coetzer TL. **Pyruvate kinase deficiency in a South African kindred caused by a 1529A mutation in the PK-LR gene.** South African Medical Journal 2008, 98: 456-457.

Durand PM and Coetzer TL. **Human pyruvate kinase deficiency protects against malaria.** Haematologica 2008, 93: 939-940.

Durand PM and Coetzer TL. **Hereditary red cell disorders and malaria resistance.** Haematologica 2008, 93: 961-963.

Durand PM, Oelofse AJ and Coetzer TL. **An analysis of mobile genetic elements in three *Plasmodium* species and their potential impact on the nucleotide composition of the *P. falciparum* genome.** BMC Genomics 2006, 7: 282.

Durand PM and Coetzer TL. **Genomic evidence for elements of an apoptosis pathway and a p53 DBD-like domain in *Plasmodium*: implications for parasite programmed cell death.** BMC Biology, manuscript to be submitted.

Durand PM, Naidoo K and Coetzer TL. **Evolutionary patterning: a novel approach to the identification of potential drug target sites in *Plasmodium falciparum*.** *PLoS* ONE 2008, 3(11): e3685.

INVITED SEMINARS

On the molecular evolution of the *Plasmodium falciparum* genome. Department of Medical Genetics, National Health Laboratory Service, Johannesburg, South Africa. March 2009.

A study on molecular evolution in *Plasmodium falciparum*. Fred Hutchinson Cancer Research Center, Seattle, USA, January 2008.

A study on molecular evolution in *Plasmodium falciparum*. Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, USA, November 2008.

Determinants of *P. falciparum* evolution using a genome-centered approach. Department of Physiology and Biophysics, Weill Cornell Medical College, New York, USA, April 2008.

CONFERENCE PRESENTATIONS

Coetzer TL, Naidoo K and Durand PM. **A Novel Evolutionary Approach Provides Hope for Limiting Future Drug Resistance in Malaria.** Abstract 420. 50th Annual Meeting of the American Society of Hematology. San Francisco, USA. December 2008. (Abstract publication: Blood 2008, 112: 160).

Durand PM and Coetzer TL. **Pyruvate kinase deficiency protects against malaria in humans.** Federation of South African Societies of Pathology 48th Annual Congress, July 2008, Cape Town, South Africa, and Faculty of Health

Sciences Research Day, University of the Witwatersrand, Johannesburg, South Africa. October 2008.

Durand PM and Coetzer TL. **Determinants of *P. falciparum* evolution using a genome-centered approach.** Keystone Symposium: Molecular evolution as a driving force in infectious disease, Breckenridge, USA. April 2008.

Durand PM, Oelofse AJ and Coetzer TL. **Potential mobile genetic elements and their derivatives in two *Plasmodium* species.** South African Society of Biochemistry and Molecular Biology, Pietermaritzburg, South Africa. July 2006.

Durand PM and Coetzer TL. **Reverse transcriptase-encoding genes in *Plasmodium* species.** South African Society of Biochemistry and Molecular Biology, Stellenbosch, South Africa. February 2005.

PREFACE

The work presented here falls under the umbrella term “molecular evolution”. The body of the thesis is a collection of six publications, one manuscript under review, two manuscripts to be submitted and one conference presentation, which cover several aspects of molecular evolution in *Plasmodium falciparum*. To facilitate the flow and overall structure, manuscripts are not presented in chronological order. Chapter 1 introduces the reader to the field of *Plasmodium* molecular evolution and the specific aspects covered in this thesis. Chapters 2 to 6 include the manuscripts and conference presentation that comprise the major thrust of this thesis. A brief introduction to each aspect of molecular evolution and the contribution to the field by each publication are provided at the beginning of every chapter. Copies of the publications and links to the journal manuscript URLs are provided, except for the three manuscripts under review or for submission, which are .pdf copies of the word documents. Manuscripts are not included in the page numbering system at bottom right. The publications in chapter 2 include a review of the available computational methods and a new method developed by the author for use in molecular evolution research. Chapter 3 covers the work done in the area of host-parasite co-evolution. Chapters 4 and 5 deal with the evolution of *P. falciparum* at a genome level, specifically the roles played by mobile genetic elements, telomerase and programmed cell death. In chapter 6, a novel evolutionary approach to the problem of drug resistance in *P. falciparum* is described and demonstrates the benefits of applying evolutionary data to biomedical challenges. Chapter 7 discusses the most important findings and their relevance to the field, current research conducted by the author in this field, and proposes hypotheses for future testing.

A number of issues are addressed in this thesis and it is hoped that the thoughts of the evolutionary biologist Theodosius Dobzhansky act as an undertone throughout: “Nothing in biology makes sense except in the light of evolution” (American Biology Teacher, 1973).

PD, Johannesburg, 2009

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Figures and tables in manuscripts are not included.

ABBREVIATIONS

ACA – ancient common ancestor

DBD – DNA-binding domain

DNA – deoxyribonucleic acid

EP – Evolutionary Patterning

FIRE – functional inference using the rates of evolution

GSC – Gerstein-Sonnhamer-Chothia

HMM – hidden Markov model

Indels – insertions / deletions

MGE – mobile genetic element

MLE – maximum likelihood estimate

OWM – old world monkeys

PCD – programmed cell death

PK – pyruvate kinase

RCA – recent common ancestor

RIP – repeat-induced point mutation

RNA – ribonucleic acid

SNP – single nucleotide polymorphism

TERT – telomerase reverse transcriptase

CHAPTER 1
INTRODUCTION

The causative agents of malaria are species of the genus *Plasmodium*, which represent a highly successful group of parasites and there are currently more than 200 known species infecting mammals, reptiles, and birds. There are at least five species that cause disease in humans: *P vivax*, *P. malariae*, *P. ovale*, *P. knowlesi* and *P. falciparum*, of which the latter is responsible for >95% of fatalities. Malaria remains the most lethal protozoan disease of humans and infects over 300 million people each year, leading to between one and three million deaths (Snow *et al.*, 2005).

1.1 Origin and evolution of *Plasmodium*

1.1.1 Early origin of a *Plasmodium* ancestor

The current classification of *P. falciparum* in the NEWT taxonomy database, which is maintained by the UniProt consortium, is: Superkingdom: Eukaryota with rank Alveolata; Phylum: Apicomplexa; Class: Aconoidasida; Order: Haemosporida; and Genus: Plasmodium (Phan *et al.*, 2003). Although the deep roots of eukaryote evolution remain unresolved, mainly due to limitations in the availability of genome data and the phylogenetic methods used in the analyses, there is consensus regarding the early evolution of the *Plasmodium* ancestor (Embley and Martin, 2006, Raven and Allen, 2003). The most parsimonious explanation involves three endosymbiotic events, which led to the nuclear, mitochondrial and plastid genomes present in *Plasmodium* parasites today (Figure 1). The most controversial issue has been the origin of the plastid genome in apicoplasts, which resulted from a third endosymbiotic event (Delwiche, 1999) involving either red (rhodophytes) (Williamson *et al.*, 1994) or green

(chlorophytes) (Kohler *et al.*, 1997) algae. The presence of chlorophyte-like nuclear and plastid genes in the nuclear genome of apicoplasts (which presumably migrated there from the engulfed alga) lent early support to a symbiotic event involving a green alga. However, limited taxon sampling and poor statistical support cast doubts on this conclusion and subsequent work provided stronger evidence for a non-green algal origin (McFadden *et al.*, 1997). The most recent finding of an ancient alveolate linking Apicomplexan plastids with rhodophytes, has swung the debate firmly in favor of the red alga hypothesis (Moore *et al.*, 2008).

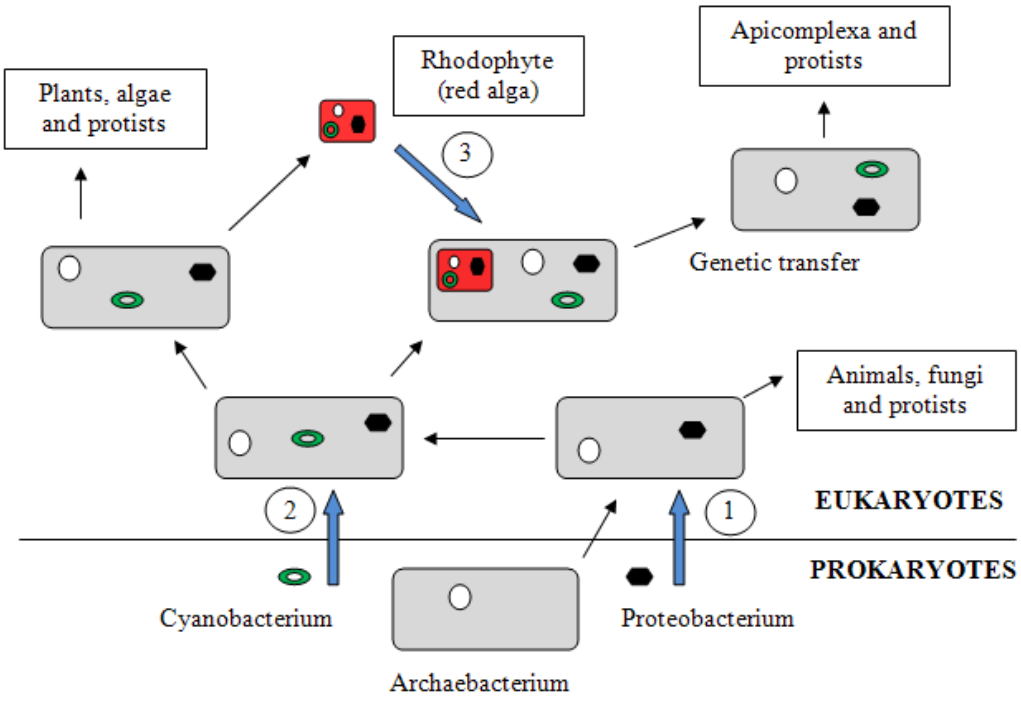

Figure 1: Early origin of the Apicomplexa.
The Apicomplexa evolved following three key endosymbiotic events: 1) uptake of a proteobacterium by an archaebacterium, 2) uptake of a cyanobacterium by the archaebacterium-proteobacterium symbiont, and 3) the phagocytosis of a primordial rhodophyte by another primitive eukaryote. Subsequent genetic transfer and disintegration of the primordial rhodophyte led to the emergence of the Apicomplexa.

1.1.2 The “big bang” in *Plasmodium* evolution

Investigations in the phylogeny of Apicomplexa, and in particular *Plasmodium*, have regularly yielded conflicting results. Nevertheless, the consensus is that Apicomplexan lineages are undoubtedly ancient, and based on the analysis of the small subunit rRNA genes from 60 species belonging to three classes, it has been proposed that *Plasmodium* diverged from other genera well before the appearance of vertebrates, possibly as early as the Cambrian period several hundred million years ago (Escalante and Ayala, 1995). Following this initial split it was assumed that, due to the parasite-host-vector specificity, Plasmodia diverged along with speciation events in their vertebrate hosts. Figure 2 describes the vertebrate host and vector stages of the *P. falciparum* life-cycle.

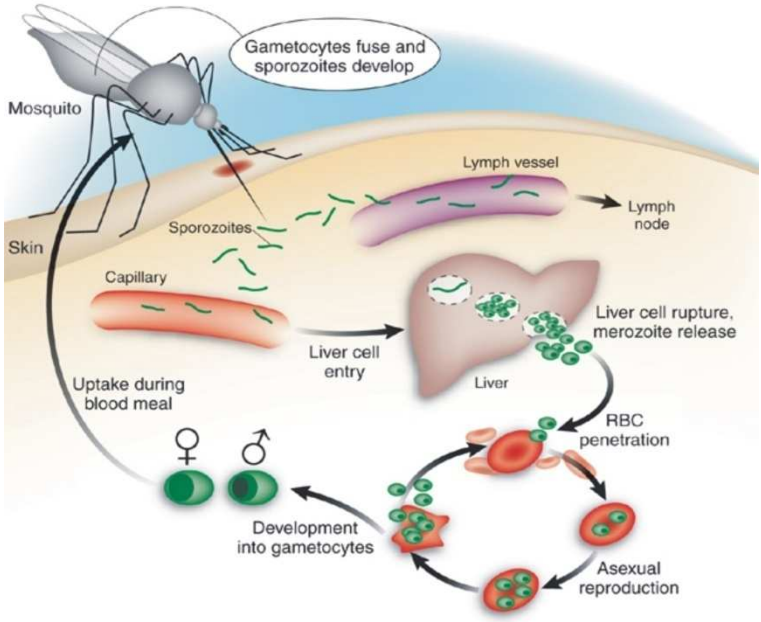

Figure 2: The *P. falciparum* life cycle.
The parasite has a complex life-cycle involving human and mosquito hosts. The disease pathogenesis is closely related to the asexual erythrocyte stage, where the parasite undergoes numerous rounds of replication and re-invasion, leading to the signs and symptoms of malaria. (Taken from Jones and Good, 2006).

The earlier assumption that Plasmodia species co-evolved with host speciation events has recently been proved incorrect following a phylogenetic analysis of whole mitochondrial genome sequences from large numbers of *Plasmodium* species (Hayakawa *et al.*, 2008, Roy and Irimia, 2008). The "big bang" in *Plasmodium* evolution occurred 20-40 million years ago during a period of rapid speciation and well after the divergence times of most of their hosts, which occurred between 50 and 280 million years earlier. These data indicate that parasite diversification, with the notable exception of *P. falciparum* and *P. reichenowi*, was not caused by codivergence along host lineages. Instead, this period of accelerated radiation of extant parasites coincides with a burst in mammalian diversification, suggesting that the emergence of an abundance of new host species and subsequent host-switch events led to an explosion in parasite diversity (Hayakawa *et al.*, 2008). The phylogenetic relationships between the 21 *Plasmodium* species used in the Hayakawa study (2008) are shown in Figure 3.

1.1.3 The rise of *P. falciparum*

Perhaps the most controversial topic in *Plasmodium* evolution has been the dating and phylogenetics of *P. falciparum*, which have been characterized by frequent revisions (Hartl *et al.*, 2002). Two opposing camps emerged that presented alternate hypotheses to explain the level of genetic diversity in *P. falciparum* populations. The recent common ancestor (RCA) hypothesis suggested that, although *P. falciparum* may have existed for nearly 100 000 years, parasite populations have passed through frequent evolutionary bottlenecks and that current populations arose from a common ancestor only ~10 000 years ago (Joy *et al.*, 2003). This hypothesis explained the relatively low genetic diversity in most genes in the genome. Support for the RCA hypothesis was provided by two other important evolutionary events, which coincided with this timeframe: (i) the development of slash-and-burn agriculture in human populations, and (ii) the diversification of the highly anthropophilic mosquito vectors *A. gambiae* and *A.*

funestus. Both these events would have favoured the expansion of parasite populations. In contrast, the ancient common ancestor (ACA) hypothesis sought to explain the highly divergent sequences in antigenic genes like merozoite surface proteins, MSP1 and MSP2, as well as the high frequency of synonymous mutations in some non-coding regions and suggested an origin for *P. falciparum* of ~100 000 years ago.

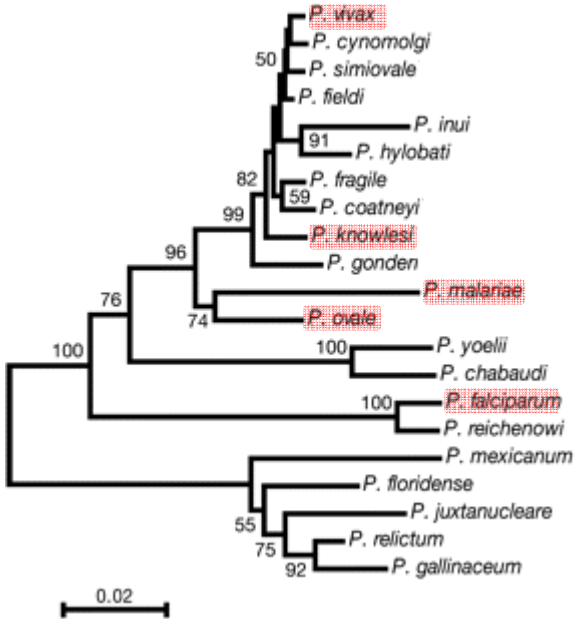

Figure 3: Phylogenetic relationships within *Plasmodium*.
Relationships between the five human parasites (shaded pink) suggest that there have been multiple host switch events. Dating of the *P. falciparum* and *P. reichenowi* (chimpanzee parasite) divergence (Hayakawa *et al.*, 2008) indicated that these species diverged along with their human and chimpanzee hosts. Bayesian posterior probability bootstrap support values are indicated and the scale bar refers to nucleotide changes per site. (Adapted from Hayakawa *et al.*, 2008).

The ACA hypothesis received considerable support following a single nucleotide polymorphism (SNP) analysis of the entire chromosome 3, which included 204 genes and 403 polymorphic sites (Mu *et al.*, 2002). This analysis, which used the rates of synonymous substitutions in coding sequences and substitutions in non-

coding regions to estimate the minimum time elapsed since the parasite's origin, concluded that *P. falciparum* emerged at least 100 000 to 180 000 years ago.

The most recent approach to date the origin of *P. falciparum* made use of the finding that amino acid substitution rates in the mitochondrial genes and a nuclear house-keeping gene in a subgroup of *Plasmodium* species (that included *P. falciparum*) were constant (Hayakawa *et al.,* 2008). This finding of a constant molecular clock rate enabled researchers to estimate a date for the emergence of *P. falciparum*. The divergence between African and Asian old world monkeys (OWMs) was assumed to match the divergence of *P. gonderi* and Asian OWM parasites and was used as a calibration point. Phylogenetic analysis with a UPGMA (Unweighted Pair Group Method with Arithmetic mean) tree pushed the origin of *P. falciparum* back even further, in the order of millions of years. Due to the assumptions and potential weaknesses inherent in the computational methods, this may be an exaggeration of the time to the most recent common ancestor, but certainly provides more support for an ancient origin of the parasite. Furthermore, dating of the *P. falciparum* / *P. reichenowi* split (the latter is a pathogen of chimpanzees) coincides with the human / chimpanzee divergence, suggesting that *P. falciparum* codiverged with the emergence of *H. sapiens*, making it the most ancient pathogen of humans (Hayakawa *et al.,* 2008). This is in agreement with an earlier independent phylogenetic analysis of whole mitochondrial genomes from 21 *Plasmodium* species, which demonstrated multiple host-switch events including humans (Roy and Irimia, 2008).

1.2 The *P. falciparum* genome

1.2.1 General features of the nuclear genome

An in depth analysis of the complete *P. falciparum* genome sequence uncovered a wealth of information regarding parasite evolution and genomics (Gardner *et al.,* 2002). The nuclear genome comprises 14 chromosomes, encodes ~5 300

genes, of which 54% contain introns, and is 22.8Mb in size. In comparison to the model single celled eukaryote *S. cerevisiae*, there is approximately the same number of genes but gene density is halved. Genes are found on both positive and negative DNA strands, occasionally leading to overlapping reading frames. Coding regions in *P. falciparum* are generally much longer than other eukaryotes and frequently contain long stretches of low complexity sequences. At the time of complete genome sequencing, >60% of the predicted proteome had no identifiable homologues. There may be several reasons for this, including: (i) the low complexity regions disrupt evolutionary conserved domains, (ii) the greater evolutionary distances between *Plasmodium* species and other eukaryotes, and (iii) nucleotide and codon usage biases. These factors impact negatively on the sensitivity and specificity of conventional similarity-based methods for homologue identification.

The most striking feature of the nuclear genome is the extreme nucleotide bias. Overall AT composition is 80.6%, far more than any other organism sequenced to date, and increases to >90% in non-coding regions (Gardner *et al.*, 2002). The AT bias of 77.4% in *P. y. yoelii* is slightly less (Carlton *et al.*, 2002), while the remaining two completed genomes, *P. knowlesi* and *P. vivax*, are more balanced (Carlton *et al.*, 2008a, Pain *et al.*, 2008). The *P. falciparum* overall genome content (Gardner *et al.*, 2002) and analyses of the small subunit ribosome (Baldauf *et al.*, 2000) reveal it is more similar to the plant *A. thaliana* than other taxa. However, this is misleading and results from multiple gene migrations to the *P. falciparum* nuclear genome from the nuclear and plastid genomes of the tertiary symbiont (Figure 1), which was the primordial ancestor of both algae and plants (Embley and Martin, 2006, Raven and Allen, 2003).

1.2.2 Plastid and mitochondrial genomes

Due to the initial difficulties in separating mitochondrial and plastid genomes with density gradient centrifugation, both genomes were erroneously believed to

be of mitochondrial origin (Feagin *et al.*, 1991). Once the existence of the plastid genome was accepted, it became a defining feature of the Apicomplexa and was demonstrated to have a common origin across the phylum (Lang-Unnasch *et al.*, 1998). The plastid genome is located in the apicoplast, an organelle homologous to the chloroplasts of algae and plants. It is ~35kb in size, codes for 30 proteins (Wilson *et al.*, 1996), and transcription in the plastid is essential for parasite survival (McConkey *et al.*, 1997). The plastid encodes several transcription and translation proteins required for its own replication, as well as key enzymes involved in fatty acid, isoprenoid and haem biosynthesis (Ralph *et al.*, 2004). An additional 551 nuclear-encoded proteins are predicted to reside in the apicoplast, providing more evidence of extensive transfer of genetic material from organelles to the nucleus.

The mitochondrial genome is a ~6kb tandemly arrayed circular chromosome coding for two mitochondrial proteins, cytochrome b and cytochrome c oxidase subunit I and two rRNA-like genes (Vaidya *et al.*, 1989). The relatively small size, tandem array genomic organization and presence of only two protein coding regions make it an unusual form of mitochondrial DNA.

Apicoplast and mitochondrial genomes have proved very valuable for resolving problems in Apicomplexan and *Plasmodium* phylogenetics (Hayakawa *et al.*, 2008, Lang-Unnasch *et al.*, 1998, Moore *et al.*, 2008, Roy and Irimia, 2008, Vaidya *et al.*, 1989). A further application that was recognized early in this area of genomics, is the potential for developing parasite-targeted chemotherapies, particularly with regard to the plastid-derived proteins (Fichera and Roos, 1997).

1.3 The study of molecular evolution in *P. falciparum*

Evolutionary theory is a cornerstone of all biology and molecular evolution applies the principles of evolution at the molecular level. Traditionally, this has referred to phylogenetic analyses, the estimation of substitution rates in nucleic

acid and protein sequences, and the evolutionary important phenomena such as recombination and DNA duplication events (Graur and Li, 2000) that result from the physico-chemical properties of biological molecules. However, the field of molecular evolution is itself evolving, absorbing and accommodating new discoveries, and includes the evolution of macromolecules such as chromosomes, whole genomes, and genetic conflict such as host-parasite co-evolution. Molecular evolution research in *Plasmodium* species has comprised mainly phylogenetic analyses, host-parasite co-evolution, and comparative studies of complete genome sequences.

1.3.1 Molecular phylogenetics of *P. falciparum*

Molecular phylogenetic studies of malaria parasites have been ongoing for more than three decades and an extensive body of work exists, which has contributed significantly to the understanding of the origins of Apicomplexa, the evolutionary relationships between *Plasmodium* species (Rich and Ayala, 2003) and the recent dating of the emergence of *P. falciparum* (Hayakawa *et al.*, 2008). The genomic era has brought a welcome addition to conventional phylogenetics, where tree reconstructions can be performed with whole genomes rather than single genes or concatenated sequences (McCann *et al.*, 2008) and this whole genome phylogeny approach has been applied to Apicomplexa (Kuo *et al.*, 2008) and Plasmodia (Martinsen *et al.*, 2008). Findings from these analyses are in keeping with the currently accepted views of Apicomplexan origins and *Plasmodium* phylogenetics introduced in section 1.1.

1.3.2 Host-parasite co-evolution

The co-evolutionary relationship between malaria and humans was first suggested by Haldane 60 years ago (Haldane, 1949) and since then a wealth of data are available demonstrating the evolutionary association between humans and malaria, in particular *P. falciparum* malaria (Weatherall, 2008). The

laboratory evidence for this relationship has predominantly taken the form of demonstrating a resistance to malaria infection by specific human genetic erythrocyte disorders *in vitro*, or making use of animal models to demonstrate the association *in vivo*. At a population and epidemiological level, the evidence indicates that the abnormal allele would be expected to occur at greatly reduced frequencies in human populations had it not been for a selective pressure such as a relative resistance to malaria.

The long and intimate association between these two organisms has left a lasting impact on both genomes. The parasite life-cycle and malaria pathogenesis is closely associated with the red blood cell (Figure 2), and it is not surprising therefore, that most of the human genetic conditions that protect against malaria are related to the erythrocyte and comprise a range of disorders and polymorphic conditions (Table 1).

The impact of the host on the evolution of the parasite genome is most obvious when one considers the complex nature by which the pathogen evades the host's immune response. The parasite's immune-evasion strategy centers on its ability to switch antigens. Gene duplication events have led to a vast array of virulence genes, which include the *rif*, *stevor* and *var* genes whose expression is tightly controlled such that there is an ongoing switch between antigenic determinants (Freitas-Junior *et al.*, 2000). At a chromosomal level, the location of the antigenic gene families in subtelomeric regions facilitates recombination events leading to further variation (Figueiredo *et al.*, 2000).

1.3.3 Comparative evolutionary genomics in *P. falciparum*

The completion of four complete *Plasmodium* genome sequencing projects, several other partial genome sequences and numerous within-species field isolates have launched a new era of comparative genomics in malaria research (Carlton *et al.*, 2008b). The comparative approach has provided new insights into

numerous aspects of *Plasmodium* molecular evolution, including parasite diversity (Jeffares *et al.*, 2007), evolutionary origins (Mu *et al.*, 2002) and host adaptation (Martinsen *et al.*, 2008). Whole genome data are still relatively recent (the first genome sequences were published seven years ago, (Carlton *et al.*, 2002, Gardner *et al.*, 2002)), and it is expected, therefore, that this area of *Plasmodium* molecular evolution will grow rapidly in the future.

Table 1: Hereditary human erythrocyte disorders and polymorphisms that protect against *P. falciparum* malaria.
Evidence that these disorders and polymorphisms protect against malaria has come from *in vitro* and/or clinical and/or epidemiological data. For items with an asterisk (*), there is only *in vitro* evidence available. (Taken from Durand and Coetzer, 2008d).

Condition	*Protein conferring a protective effect*
Hemoglobinopathies	
Sickle cell trait	Hb S
Alpha thalassemia	α-Hb
Beta thalassemia	β-Hb
Hemoglobin C	Hb C
Hemoglobin E	Hb E
Red cell membrane proteins	
Hereditary spherocytosis*	Spectrin, band 3, protein 4.2
Hereditary elliptocytosis*	Spectrin, protein 4.1
Hereditary pyropoikilocytosis*	Spectrin
South-east Asian ovalocytosis	Band 3
Blood group O	Glycosyl tansferase
Other blood group antigens	Glycophorin A, B and C
Complement receptor	CR-1
Red cell enzymes	
G6PD deficiency	Glucose-6-phosphate dehydrogenase
PK deficiency	Erythrocyte pyruvate kinase

1.3.4 Whole genome evolution

Many of the biological phenomena that are implicated in genome evolution have been elucidated, for example gene and whole genome duplications, recombination events, exon shuffling, and chromosome fusions (Graur and Li, 2000, Patthy, 1999) and some examples of these processes have been identified in *P. falciparum* (Kyes *et al.*, 2007). In addition to these biological processes,

which are typically passive in nature and the result of the physico-chemical properties of nucleic acids and proteins, there are genes and proteins that are more directly involved in genome evolution, such as those which actively generate new genetic material (for example, mobile genetic elements (MGEs), Brosius, 1999) or preserve genome architecture (for example, telomerase, Meyer and Bailis, 2008). The "active" factors encourage genome variation, such as the creation of novel functions or changes in genome architecture (MGEs). Opposing factors (telomerase) maintain genome integrity and lead to genomic change only if their functions are disrupted. The resultant diversity facilitates the process of evolution by natural selection at the genome level. The discovery of elements implicated in genome evolution is an ongoing endeavour. In *P. falciparum*, homologues of most of these elements are still undiscovered, with the notable exception of telomerase (Figueiredo *et al.*, 2005).

It has long been known that genomes are not static (McClintock, 1929), however, an understanding of the dynamic interactions between the active role-players in genome evolution is leading to the emergence of a new area of investigation: the ecology of the genome (Mauricio, 2005, Ungerer *et al.*, 2008). Currently, a generic appreciation of this field is still in its infancy, and for *Plasmodium* researchers this aspect of molecular evolution has hardly begun. The primary reason for this is that the groundwork, such as the identification of the key role-players, must be performed before the ecological dynamics of the *P. falciparum* genome can be investigated.

1.4 Molecular evolution methods for the investigation of *P. falciparum*

The field of molecular evolution, which frequently requires the processing of large data sets, relies heavily on computational methods (Yang, 2006). The two fundamental arms of molecular evolutionary analyses are (i) phylogenetic methods, which are typically used to analyze evolutionary relationships, and (ii) nucleotide and amino acid substitution rate methods, which test evolutionary

models and make predictions concerning the mechanisms of evolution. Both require the accurate identification of homologous (orthologous and paralogous) sequences and the accurate alignment of multiple sequences, which in turn depend on sequence similarity. In *P. falciparum*, this can be problematic due to the unusual nucleotide, codon and amino acid usage biases, low complexity regions, and frequent insertions and deletions. These genomic peculiarities have been a limitation from the outset when ~60% of putative proteins had no identifiable homology (Gardner *et al.*, 2002) and continues to plague researchers, the most recent example of which has been the limitation in automated reconstruction of metabolic pathways in the parasite (Ginsburg, 2009).

Nevertheless, the careful selection of data, prudent use of traditional computational methods and careful analysis of results have significantly advanced the field of *P. falciparum* evolution and genomics (Aravind *et al.*, 2003). In addition, modifications to traditional methods can improve sensitivity and specificity (Cawley *et al.*, 2001, Le and Gascuel, 2008) and novel approaches, some of which are non-homology based, are constantly being developed and provide additional tools for *P. falciparum* genome analysis (Brehelin *et al.*, 2008, Marcotte, 2000). The drive to develop *Plasmodium*-friendly algorithms will have a major impact on *P. falciparum* molecular evolution.

1.5 Research aims and objectives

The broad aim of this work was to provide further insights into the molecular evolution of *P. falciparum*. To achieve this, the following specific objectives were set:

- review the bioinformatic methods that are available to molecular evolution researchers, and develop a novel approach that addresses some of the limitations,

- examine the potential co-evolutionary relationship between human pyruvate kinase deficiency and *P. falciparum* virulence,
- identify an active role-player, which promotes genome diversity in *P. falciparum*,
- identify an active role-player, which limits genome diversity in *P. falciparum*,
- demonstrate the benefit of molecular evolution research to the medical problem of drug resistance in *P. falciparum*.

These five aspects of *P. falciparum* molecular evolution are presented as a collection of six publications, one manuscript under review, two manuscripts for submission, and one conference presentation in five chapters. To facilitate the flow of the thesis, publications are not in chronological order. In each chapter, the relevance and originality of the publication(s) to the field are introduced. In the final chapter, a synopsis of this work and its contribution to the field are highlighted and the future studies are discussed.

CHAPTER 2
BIOINFORMATIC METHODS IN MOLECULAR EVOLUTION

Durand PM and Coetzer TL. **Utility of computational methods to identify the apoptosis machinery in unicellular eukaryotes.** Bioinformatics and Biology Insights, 2008: 2, 101-117. (ISI impact factor for 2008: unavailable).
URL: http://www.la-press.com/article.php?article_id=605

Durand PM, Hazelhurst S and Coetzer TL. **Evolutionary rates at codon sites may be used to align sequences and infer protein domain function.** BMC Bioinformatics, 2009. Revised manuscript in second review. (ISI impact factor for 2008: 3.78).

Hazelhurst S and Durand PM. ***FIRE:*** **Functional Inference using the Rate of Evolution.** Manuscript to be submitted.

2.1 Introduction to publications

A review of the bioinformatic methods available to molecular evolution researchers was performed, using the apoptosis machinery in unicellular eukaryotes to demonstrate their respective strengths and weaknesses (Durand and Coetzer, 2008a). The limitations of similarity-based methods, particularly for *P. falciparum* investigators are discussed, and to address this problem a novel alignment algorithm was developed (Durand *et al.*, 2009). The algorithm, termed FIRE (Functional Inference using the Rate of Evolution), provides researchers with a method for aligning sequences without the need for significant sequence similarity. FIRE has many potential applications, the software is freely available at http://dept.ee.wits.ac.za/~scott/fire, and an “Applications Note” has been

prepared for submission, guiding users on its potential applications (Hazelhurst and Durand, 2009).

A misuse of the term “homology” has been noted in the publication “Utility of computational methods to identify the apoptosis machinery in unicellular eukaryotes” and the author of this thesis would like the reader to be aware of this. On page 105, Table 3, the word “homology” has not been used accurately to describe the multiple sequence alignment, and should be replaced with the word “similarity”.

2.2 Contribution to *P. falciparum* molecular evolution

The novelty of this work lies in the uniqueness of the alignment strategy, which makes use of evolutionary rates, rather than residues, to align sequences. In addition, it was demonstrated that the alignment may be used to infer protein domain function. This provides malaria researchers with an alignment method to complement conventional tools.

To build on this work, Profs Fourie Joubert (University of Pretoria), Scott Hazelhurst and Theresa Coetzer (University of the Witwatersrand) and the author of this thesis have applied for a National Bioinformatics Network grant for a MSc student to further evaluate and develop the FIRE method for researchers in *Plasmodium* bioinformatics.

CHAPTER 3
HUMAN AND *P. FALCIPARUM* CO-EVOLUTION

Durand, P.M. and Coetzer, T.L. **Pyruvate kinase deficiency in a South African kindred caused by a 1529A mutation in the PK-LR gene.** South African Medical Journal, 2008: 98, 456-7. (ISI impact factor for 2005: 1.07)
URL: http://www.sabinet.co.za/abstracts/m_samj/m_samj_v98_n6_a14.xml

Durand PM and Coetzer TL. **Human pyruvate kinase deficiency protects against malaria.** Haematologica / The Hematology Journal, 2008: 93, 939-940. (ISI impact factor for 2008: 5.51).
URL: http://www.haematologica.org/cgi/content/full/93/6/939

Durand PM and Coetzer TL. **Hereditary red cell disorders and malaria resistance.** Haematologica / The Hematology Journal, 2008: 93, 961-963. (ISI impact factor for 2008: 5.51).
URL: http://www.haematologica.org/cgi/content/full/93/7/961

3.1 Introduction to publications

For several decades, the co-evolutionary relationship between human pyruvate kinase (PK) deficiency and *P. falciparum* virulence was questioned (Weatherall, 2008). Researchers speculated that, as with many other genetic erythrocyte disorders, the prevalence of PK deficiency in some human populations may have been maintained by a relative resistance to malaria. Significant evidence for the relationship initially came from the mouse model (Min-Oo *et al.*, 2003), however no direct evidence for this had been found in humans. Erythrocyte cell cultures using blood from a homozygous PK deficient patient (Durand and Coetzer, 2008b) provided *in vitro* evidence that PK deficient human erythrocytes are

dramatically resistant to *P. falciparum* infection (Durand and Coetzer, 2008c). An independent group working at McGill University, Montreal, Canada obtained similar findings with different PK mutations (Ayi *et al.*, 2008).

3.2 Contribution to *P. falciparum* molecular evolution

The second publication presented here (Durand and Coetzer, 2008c) and the Canadian manuscript were published two weeks apart, and provided the first direct evidence for this co-evolutionary relationship in humans. Due to the significance of the findings, a “Perspectives / Editorial” article was invited on the subject (Durand and Coetzer, 2008d).

CHAPTER 4
THE ROLE OF MOBILE GENETIC ELEMENTS IN *P. FALCIPARUM* GENOME EVOLUTION

Durand PM, Oelofse AJ and Coetzer TL. **An analysis of mobile genetic elements in three *Plasmodium* species and their potential impact on the nucleotide composition of the *P. falciparum* genome.** BMC Genomics, 2006: 7(282). (ISI impact factor for 2008: 4.18).
URL: http://www.biomedcentral.com/1471-2164/7/282

4.1 Introduction to publication

Mobile genetic elements (MGEs) have had a major impact on genome evolution and comprise a significant percentage of most genomes (Brosius, 1999, Frost *et al.*, 2005). The publication presented in this chapter provides a comprehensive analysis of all classes of MGEs in three complete *Plasmodium* genome sequences. The absence of these elements in *P. falciparum*, led to the identification of a putative geno-protective mechanism in the parasite that protects against invading MGEs and may have contributed to the extreme AT bias.

4.2 Contribution to *P. falciparum* molecular evolution

This was the first comparative genome analysis of MGEs in *Plasmodium* and provided a partial explanation for the nucleotide bias in *P. falciparum*. The information in this publication was subsequently used by other investigators and contributed to an understanding of the pattern of intron loss during the evolution of the Apicomplexa (for example Roy and Penny, 2007).

CHAPTER 5
PROGRAMMED CELL DEATH AND *P. FALCIPARUM* GENOME EVOLUTION

Durand PM and Coetzer TL. **Genomic evidence for elements of an apoptosis pathway and a p53 DBD-like domain in *Plasmodium*: implications for parasite programmed cell death.** Manuscript to be submitted to BMC Biology.

Conference presentation:
Durand PM and Coetzer TL. **Determinants of *P. falciparum* evolution using a genome-centered approach.** Abstract 117. Keystone Symposium: Molecular evolution as a driving force in infectious disease, Breckenridge, USA. April 2008.

5.1 Introduction to manuscript and conference presentation

Programmed cell death (PCD) is essential for maintaining genome integrity and has played an integral role in molecular evolution (Ameisen, 2002, Koonin and Aravind, 2002). While there has been laboratory evidence for PCD in *Plasmodium* species (Al-Olayan *et al.*, 2002, Meslin *et al.*, 2007), the genes and proteins involved have largely remained undetected. This manuscript reports the presence of elements of a PCD pathway in *Plasmodium*, and includes functional and structural evidence for a p53-like DNA-binding domain.

The conference presentation uses data from manuscripts in chapters 4 and 5 to provide a theoretical framework for a genome-centered approach to evolution that may be used to investigate the idiosyncrasies of the *P. falciparum* genome. A copy of the conference abstract is provided.

5.2 Contribution to *P. falciparum* molecular evolution

Genomic evidence for a PCD pathway in *Plasmodium* was reported, which provides a mechanism for understanding the apoptosis phenotype observed in this genus. In addition, structural and functional evidence for a p53 DBD-like domain in *Plasmodium* was presented, suggesting that PCD may be important for limiting parasite reproduction and play a role in maintaining genome integrity.

The genome-centered approach to evolution provides a framework for studying the processes implicated in genome evolution in *P. falciparum*.

Conference abstract 117, presented at Keystone Symposium: Molecular evolution as a driving force in infectious disease, Breckenridge, USA. April 2008.

Determinants of *P. falciparum* evolution using a genome-centered approach.

Pierre M Durand and Theresa L Coetzer. Department of Molecular Medicine and Haematology, University of the Witwatersrand and National Health Laboratory Service, Johannesburg, South Africa.

The genome of the malaria parasite *P. falciparum* has several strikingly unique features, suggesting an unusual evolutionary history. To investigate this phenomenon, we have identified three major determinants in this organism that are intimately involved in genome evolution. These are mobile genetic elements (MGEs), telomerase reverse transcriptase, and a transcription factor involved in programmed cell death (PCD). MGEs are a major driver of genome diversity. Our data indicated an almost complete absence of these elements in *P. falciparum* and we have identified the putative genes for a repeat-induced geno-protective mechanism which extinguishes MGE activity and may have contributed to the AT bias in this organism. Telomerase reverse transcriptase is essential for maintaining genome integrity by protecting the ends of linear chromosomes. A *P. falciparum* ortholog of this gene was identified using homology-based methods. PCD is an essential process for the maintenance of genome integrity. Laboratory evidence exists for this mechanism in *Plasmodium* but very little is known about the genes and proteins involved. Using a number of computational methods, we have identified and investigated a transcription factor that could play a central role in PCD in *P. falciparum*. The three determinants mentioned above may be incorporated into a genome-centered model of evolution that describes an interaction between two opposing forces. Firstly, the mechanisms (of which MGEs are one example) that generate genome diversity; and secondly, the opposing mechanisms that maintain genome integrity such as telomerase activity and PCD.

CHAPTER 6
APPLICATIONS OF *P. FALCIPARUM* MOLECULAR EVOLUTION RESEARCH

Durand PM, Naidoo K and Coetzer TL. **Evolutionary patterning: a novel approach to the identification of potential drug target sites in *Plasmodium falciparum*.** *PLoS* ONE, 2008: 3, e3685. (ISI impact factor for 2008: unavailable) URL: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0003685

6.1 Introduction to publication

Concepts of health and disease are slowly being transformed by our understanding of molecular evolution (Spocter and Strkalj, 2007), however, to date this transformation has yielded few direct benefits. In this publication, a novel evolutionary approach, termed “Evolutionary Patterning” (EP), to the problem of drug resistance in malaria parasites was conceptualized.

6.2 Contribution to *P. falciparum* molecular evolution

The EP approach employs evolutionary data to identify the most suitable drug target sites so that the risk for the emergence of resistance mutations is minimized. The approach was implemented using *P. falciparum* glycerol kinase as an example and demonstrates the value of using molecular evolution to combat biomedical problems.

This manuscript has been well received, locally and internationally. Two invitations to present the work at international conferences were received and a

presentation on EP won the Faculty of Health Sciences Research Day (2008), University of the Witwatersrand, best presentation prize in its category. In addition, the EP approach was accepted for oral presentation at the 50th Annual Meeting of the American Society of Hematology in San Francisco, USA (December, 2008).

CHAPTER 7
DISCUSSION

Molecular evolution research has facilitated many advances in the field of *Plasmodium* biology (Carlton *et al.*, 2008b), not least of all the discovery of the complex evolutionary processes that gave rise to the Apicomplexa (Delwiche, 1999, Embley and Martin, 2006). In *P. falciparum* specifically, the study of molecular evolution has focused predominantly on host-parasite co-evolution, phylogenetics and genomics.

An understanding of the *Plasmodium*-human co-evolutionary relationship has led to a greater appreciation of human hereditary erythrocyte disorders (Weatherall, 2008) and parasite virulence factors (Penman and Gupta, 2008), and brings with it the hope for new therapeutic strategies. Phylogenetics continues to play an essential role in current *P. falciparum* research, particularly comparative genomics and genome annotation. In this thesis, three aspects of *P. falciparum* molecular evolution were investigated: (i) the co-evolution between human erythrocyte PK deficiency and malaria, (ii) the elements that play an active role in genome evolution, and (iii) the benefits of molecular evolution for malaria research. In addition, a major impediment for malaria researchers has been the limited application of computational approaches to *P. falciparum* molecular sequence data, and a new alignment method was developed to address this.

7.1 Bioinformatics and *P. falciparum* molecular evolution research

The problems posed by the unusual nature of the *P. falciparum* genome are evident from the slow progress in gene annotation. The reason for this is that computational methods are not tailor-made to deal with idiosyncratic features

such as extreme nucleotide, codon and amino acid biases. These problems are compounded by the great evolutionary distances between *P. falciparum* and most other taxa. Nevertheless, by employing novel methods such as FIRE, addressing the weaknesses of conventional homology-based methods like the model bias in HMMs, and using a multi-pronged approach to homologue identification, these limitations may be diminished.

7.1.1 The FIRE approach to sequence alignment

The FIRE approach (Durand *et al.*, 2009), described in this thesis, aligns sequences by using the dN/dS ratio at codon sites to find the codon alignment that maximizes the similarity metric. The significant advantage of this strategy is that it is not limited by sequence similarity and provides a method for analyzing evolutionary distant or fast evolving genes, and sequences with nucleotide biases. FIRE was also used to test the hypothesis that domains with similar functions are subject to similar evolutionary pressures and it was found that this approach has the potential to predict domain function based on its evolutionary constraints. This has particular relevance for gene annotation in *P. falciparum*, where the lack of sequence similarity has limited the identification of homologous domains.

The FIRE approach offers a promising new framework for sequence analysis and domain function prediction. The current recommendation is that FIRE is used in conjunction with conventional computational methods to limit the possibility of false positives. Following more extensive usage, strengths and weaknesses will emerge that may lead to improvements in the algorithm and ultimately more diverse applications.

7.1.2 Hidden Markov models: suggestions for limiting model bias

HMMs (Eddy, 1998) form the basis for the annotation of domains, such as those in the protein family database Pfam (Finn *et al.*, 2008), and play a pivotal role in

homology detection. At least three features of the *P. falciparum* genome, however, have specifically limited the efficacy of this approach: (i) the frequent low complexity sequences and indels found in coding regions, (ii) the extreme nucleotide, codon and amino acid biases, and (iii) the evolutionary distant nature of the genome (Gardner *et al.*, 2002). The first limitation has been addressed by algorithms that filter out low complexity regions from protein sequences (for example, Wootton and Federhan, 1993). The second concern is currently being addressed by researchers at LIRMM (Le Laboratoire d' Informatique, de Robotique et de Microélectronique de Montpellier, www.lirmm.fr), where the amino acid bias in *P. falciparum* proteins has been used to produce amino acid replacement matrices for incorporation into HMMs, and has led to the annotation of ~100 proteins (Olivier Gascuel and Eric Marechal, personal communications). Although these two approaches have improved gene annotation, the results have been limited. A possible reason for this is the rapid evolution of some parasite genes and the great evolutionary distance between *P. falciparum* and other organisms in general. To overcome this obstacle, it is suggested that (i) the complete genome sequences from organisms more closely related to *P. falciparum* may provide important "intermediate" genomes, and (ii) the HMM model bias should be addressed. Genome sequencing projects are ongoing and include closely related Apicomplexans (Liolios *et al.*, 2008), however, the problem of HMM bias has persisted.

Sequence weighting in HMMs

Several weighting algorithms have been developed to address sequence redundancy during HMM building (Karchin and Hughey, 1998). Failing to account for over-represented sequences leads to a model bias in favour of similar sequences. Weighting methods limit the bias during model training when under-represented sequences, which fit the model poorly, are given increased weighting. In this thesis, the Gerstein-Sonnhamer-Chothia (GSC) (Gerstein *et al.*, 1994) weighting method was employed and, in addition, consensus sequences of over-

represented taxa were used to minimize the bias. However, while this approach improves model sensitivity, none of the current methods employ a proportionate weighting system that accurately accounts for evolutionary distance.

It is hypothesized that an evolutionary distance weighting matrix is required to weight sequences proportionately during the model building process. This may be the only way to ensure that each sequence is weighted in proportion to its evolutionary distance from every other sequence, and not only overcomes the problem of sequence redundancy, but increases the weighting of evolutionary distant sequences.

This approach may encounter the problem of saturation, which occurs following multiple substitutions at the same site (nucleotide or amino acid) causing sequence dissimilarity to no longer reflect the 'true' evolutionary distance (the number of substitutions that have actually occurred since the divergence of two sequences). However, it is possible to limit the problem of saturation in nucleotide sequences by down-weighting transitions (Tang *et al.*, 1999) or excluding third position nucleotides (Nickrent *et al.*, 2000). In amino acid sequences, amino acid substitution probability matrices have been used successfully (van de Peer *et al.*, 2002). The problem of saturation is not unique to evolutionary distance estimation and is almost certainly an inherent problem of HMMs. Current amino acid or nucleotide HMM algorithms do not address the problem of sequence saturation or make use of evolutionary distance matrices.

It is likely that HMM sequences weighted in proportion to evolutionary distance and accounting for saturation, will increase model sensitivity and improve the identification of homologous sequences in *P. falciparum*. However, regardless of the weighting methods employed during HMM building, model sensitivity is still subject to data availability.

7.1.3 Current and future work

The current focus is the refinement and implementation of the FIRE algorithm. The author has developed a codon score weighting matrix to refine the FIRE algorithm. The matrix takes into account three additional parameters: (i) the distribution of positively selected sites; (ii) the probability of an ω MLE belonging to each decile between 0 and 1; and (iii) increased weighting of sites under positive selection (ω MLE>1.0). The first parameter accounts for the distribution of positively selected sites, since an even distribution across the coding sequence has different biological implications to a cluster of sites in a particular region. The second parameter corrects for the expected frequency of omega site classes between 0 and 1 and is analogous to correction for nucleotide frequencies employed by some evolutionary models such as the HKY85 model (Hasegawa *et al.*, 1985). The third parameter places increased importance on sites under positive selection.

Following the validation and implementation of these refinements, a FIRE analysis of all Pfam database (Finn *et al.*, 2008) entries will be performed to obtain a FIRE signature for each domain. This signature may then be used to identify similar domains in the *Plasmodium* database (Stoeckert *et al.*, 2006). This work will be explored with Profs. F. Joubert (African Centre for Genome Technology, University of Pretoria), S. Hazelhurst (School of Electrical and Information Engineering, University of the Witwatersrand) and T. L. Coetzer (Department of Molecular Medicine and Haematology, University of the Witwatersrand).

The combination of novel strategies like FIRE and refinements to existing HMM algorithms may be used as part of a comprehensive multi-pronged computational approach (Durand and Coetzer, 2008a) to address the problems posed by the *P. falciparum* genome.

7.2 Human-parasite co-evolution

The human-parasite co-evolution work presented in this thesis (Durand and Coetzer, 2008c, Durand and Coetzer, 2008d) and in an independent study (Ayi *et al.*, 2008) have provided the first direct evidence for the co-evolution of human hereditary erythrocyte PK deficiency and *P. falciparum* virulence. Future objectives and considerations include (i) clinical case control studies, (ii) the investigation of PK deficiency in Africans in malaria endemic areas, and (iii) a search for possible positive selection in the PK-LR gene in hominid lineages.

7.2.1 Clinical case control studies

The author was recently consulted on a Dutch immigrant living in Tanzania who suffered haemolytic episodes of unknown aetiology. The patient's history was strongly suggestive of PK deficiency; however, the quantitative PK assay was inconclusive due to a high reticulocytosis (>15%). Mutation analyses are currently being performed to confirm the diagnosis. The patient has contracted *P. falciparum* malaria on at least seven occasions in Tanzania. Each time the infection was mild and, in some instances, self-limiting, which is in keeping with the *in vitro* findings of a relative resistance to *P. falciparum* infection by PK deficient erythrocytes. Once the diagnosis of PK deficiency has been confirmed in this patient, *in vitro* parasite culture studies will be performed to corroborate the clinical findings. This work is being conducted in conjunction with Dr. R van Wijk (Department of Clinical Chemistry and Haematology at the University Medical Center, Utrecht, Holland).

7.2.2 PK deficiency in Africans: do these mutations confer malaria resistance?

PK deficiency has been assumed to be extremely rare in Africans; however, more data are emerging to suggest that this is not true (Alli *et al.*, 2008, Mohrenweiser, 1987, van Wijk *et al.*, 2009). In collaboration with Dr. R. van Wijk the novel

mutations identified in Africans (van Wijk *et al.*, 2009) will be investigated *in vitro* to determine the potential relationship between these mutations and malaria resistance.

7.2.3 Is there evidence for positive selection in the PK-LR gene in hominids?

The finding that PK deficiency confers malaria resistance in humans raises the possibility that there are sites in the PK-LR gene that are under positive selection. This question can be addressed using the branch-site models of PAML and, depending upon the future availability of funds and interests of graduate students, is an appealing line of enquiry.

7.3 *P. falciparum* genome evolution

Historically, genome evolution has been considered from the position of inter-genomic conflict, which results from competition within genome populations and, in the case of parasites and their hosts, the co-evolution between genomes. In both cases, molecular variation leads to variable fitness, which is acted on by natural selection to facilitate evolution. Our understanding of life, however, is being transformed by the realization that the evolutionary units that make up a population of "individuals", exist at multiple levels and take the form of single genes, gene networks, genomes, prokaryotic cells, eukaryotic cells, multicellular organisms, sexually reproducing pairs, populations and communities (Michod and Roze, 1997). Some researchers suggest that the genome-centric approach is set to take centre stage in this paradigm shift, leading to a resynthesis of evolutionary theory (Heng, 2009). Despite this, however, the concept of the genome as an "individual" or evolutionary unit, resulting from intra-genomic conflict between interacting subunits, is relatively new (Brookfield, 2005, Mauricio, 2005, Ungerer *et al.*, 2008). The investigation of genome evolution using this framework will typically require a complex systems approach (Testa and Lemont, 2000). However, prior to this, hypothetical arguments to explain the

evolutionary transition in individuality from lower evolutionary units (individual genes) to a higher level evolutionary unit (the genome), as well as the emergence of individuality within genome populations, are required. The hypothetical argument depends on the identification of key determinants that comprise lower evolutionary units and an understanding of the interactions between them, which was one of the aims of this thesis.

The author has proposed a theoretical framework for studying the genome as an evolutionary unit (Durand and Coetzer, 2008e), using *P. falciparum* as an example. It is rooted in the understanding that the fitness of evolutionary units comprises two fundamental components: reproduction and survival (Roff, 1992, Southwood, 1988), and that the trade-off between these components drives the evolution of individuality (Michod, 2006), which in this case, is the genome. This is analogous to the approach used to investigate the evolutionary transition from single cells to multicellularity, where it was demonstrated that the trade-off between specialized lower level units (germ and somatic cells) may have led to the emergence of a higher order evolutionary unit, a multicellular organism (Michod, 2005, Michod, 2006). It is the opinion of the author, that MGEs are representative of selfish gene reproduction, and PCD genes represent survival. At the level of the genome, MGEs and PCD are examples of determinants that are part of two fundamental processes: (i) the generation of diversity, and (ii) the maintenance of integrity (Figure 4). It is argued that a trade-off between these processes represents a theoretical framework for understanding the emergence of new genomes, which in the case of *P. falciparum* will provide insights into the biology of this extraordinary genome.

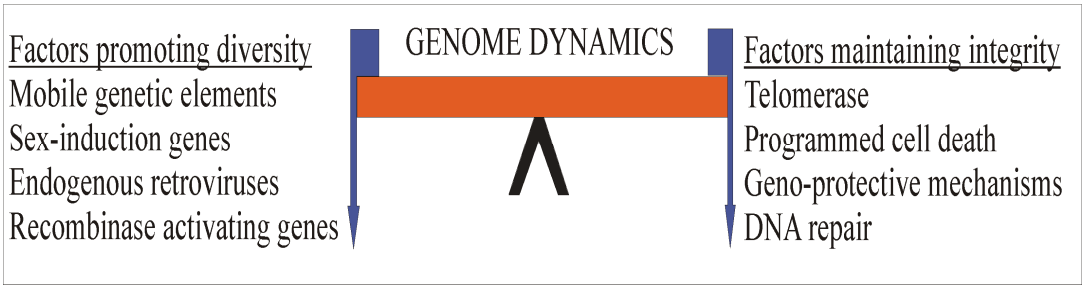

Figure 4: Ecological trade-offs in genome evolution.
The hypothetical model for genome dynamics involves two fundamental processes: the promotion of diversity and the maintenance of integrity. A trade-off between elements implicated in these two phenomena drives genome evolution.

7.3.1 Genome evolution: an ecological trade-off

A number of factors are implicated in the generation of genome diversity, including the induction of sexual reproduction, endogenous retroviruses, recombinases and MGEs, which was one of the key role-players investigated in this thesis. MGE activity generates a spectrum of diversity, ranging from point mutations to whole genome restructuring. The initial result typically induces genome damage (Le Rouzic and Capy, 2005), however, over time the co-evolution of parasitic MGE and the host genome leads to a symbiotic relationship. These dynamics have played a fundamental role in genome evolution in all major phylogenetic lineages (Kazazian, 2004, Kidwell and Lisch, 2001), including protozoan parasites (Bhattacharya *et al.*, 2002).

The genome has evolved numerous mechanisms to deal with the damaging effects of MGE activity, such as the RIP (repeat-induced point mutation) geno-protective mechanism and PCD (see 7.3.2 and 7.3.3 below). Occasionally, or in cases of uncontrolled MGE activity, significant genome restructuring may occur, leading to the emergence of new “individuals” or competitors (Brosius, 1999). A

life-history strategy, therefore, that controls extensive genome variation resulting from damage would have a selective advantage for the individual and provides a theoretical explanation for the emergence of PCD as a trade-off against MGE activity. Several PCD pathways have been described, of which p53-dependant PCD is of particular interest in this context. p53, described as "the guardian of the genome" (Lane, 1992), is essential for detecting disruptions to genome integrity and p53-dependant PCD specifically, is likely to be intimately associated with genome evolution.

A second determinant implicated in genome evolution is the telomerase reverse transcriptase (TERT) ribonucleoprotein enzyme, which has evolved from a domesticated MGE and is responsible for protecting the ends of linear chromosomes (Nakamura and Cech, 1998). An absence of normal TERT activity has been shown to cause chromosome damage, chromosome fusions and severe genome instability in model organisms (for example Blasco *et al.*, 1997). In chapter 4 a *Plasmodium* TERT was reported and in the genome evolution framework it comprises one of the elements implicated in genome maintenance.

In *P. falciparum*, three factors implicated in genome evolution were investigated: MGEs, TERT and PCD. Two of these determinants, MGEs and PCD, provide a convenient approach for investigating a hypothetical trade-off between the gene level fitness components of reproduction and viability. At the genome level, these two components represent a balance between the selection for elements that generate diversity and the mechanisms that maintain integrity.

7.3.2 An absence of MGEs in *P. falciparum*: implications for genome evolution

The unexpected finding that active MGEs, which are a key source of genome variation (Frost *et al.*, 2005), are absent in *P. falciparum* (Durand *et al.*, 2006) provided insight into the parasite's life-history (chapter 4). During *P. falciparum* evolution, it appears that a geno-protective mechanism extinguished

retrotransposon activity in the *Plasmodium* ancestor (Roy and Penny, 2007) and the responsible mechanism was a RIP (repeat-induced point mutation) pathway, which contributed to the AT nucleotide bias (Durand *et al.*, 2006). It is interesting to note that the two other *Plasmodium* species analyzed, *P. vivax* and *P. yoelii*, contain numerous MGEs belonging to all classes, and lack some of the essential genes required for the RIP mechanism, which correlates with their lower AT content. The specificity of the RIP mechanism for the *P. falciparum* genome is puzzling; however, it appears that protection against parasitic MGEs was paramount for the evolution of this species.

The elimination of MGEs has removed a major source of genome variation, which may, theoretically at least, incur a fitness cost. In addition, highly specialized organisms occasionally run the risk of becoming evolutionary dead-ends (Mayr, 1954), however, the genome protection mechanism and specialized life-history strategy in *P. falciparum* has undoubtedly contributed to the reproductive and survival success, and associated virulence in this organism.

7.3.3 The role of PCD in *P. falciparum* genome evolution

The evidence presented in chapter 5 strongly suggests that an ancient PCD pathway is responsible for apoptosis in *Plasmodium*. While the central role for PCD in the maintenance of genome integrity is widely accepted (Chipuk and Green, 2006, Stergiou and Hengartner, 2004), the ecological implications for genome evolution remain largely unexplored. In *P. falciparum*, there are good reasons for the continued presence of PCD at the level of individual genomes, as well as parasite populations, and these are dealt with in chapter 5. However, while PCD pathways may be important for limiting the parasite burden in the host, there are additional implications for genome evolution.

The presence of a geno-protective mechanism, as well as a PCD pathway in *P. falciparum*, suggests that the limitation of diversity and maintenance of genome

integrity have been critically important during the evolution of this species. This is supported by the finding of a potential p53-like protein, which is intimately associated with genome integrity PCD pathways, in *Plasmodium*. These data indicate that the *P. falciparum* life history strategy has been one of extreme specialization, which is in keeping with the complex parasite life-cycle as well as the genome idiosyncrasies, in particular the extreme AT bias. This specialization, which is absent from other *Plasmodium* genomes, correlates with the extreme virulence associated with this species.

7.3.4 Current work in *P. falciparum* genome evolution

The work in this area is currently focused on the laboratory verification of some of the findings in chapters 4 and 5. A spin-off of the search for MGEs in *Plasmodium* was the identification of telomerase reverse transcriptase (TERT) orthologues (Durand *et al.*, 2006), which have evolved from non-LTR retrotransposons (a class of MGEs) and serve an essential role in the maintenance of genome integrity (Belancio *et al.*, 2008). Five constructs of the two *P. falciparum* putative TERTs identified in chapter 4 have been cloned into the expression vector pGEX-4T2 and protein expression studies in an *E. coli* Rosetta bacterial system have been initiated (Appendices I and II).

7.4 The application of molecular evolution studies to malaria research

The benefits of molecular data to applied *P. falciparum* research have been enormous and have included major advances in almost every aspect of the malaria field. The use of evolutionary data, however, has been concentrated mainly in the area of comparative genomics (Carlton *et al.*, 2008b) and phylogenetics (Martinsen *et al.*, 2008, Rich and Ayala, 2003), with limited benefits for applied research. The Evolutionary Patterning (EP) approach (Durand *et al.*, 2008) described in chapter 6, demonstrated that one of the applied fields in which evolutionary data may be of significant benefit, is the area of drug

and vaccine development. Drug resistance is a major problem (Hyde, 2005) and optimizing the drug development pipeline to maximize the lifespan of chemotherapeutic agents may be possible using an evolutionary approach to drug target identification.

EP aims to minimize the emergence of drug resistance and potentially limit side-effects by identifying the most suitable drug target sites based on evolutionary constraints, however, there are a few caveats that should be considered. First, compensatory mutations may occur that change the selective constraints at a particular target site and second, the method does not account for sites that may have evolved adaptive mutations at other points in the phylogenetic tree (Emes and Yang, 2008). A third caveat is that *Plasmodium* sequences are highly divergent leading to sequence saturation, which typically increases errors in ω MLEs. Saturation was identified by *dS* values of between 3.8 and 48.2, which were relatively high (Yang and Nielsen, 2000) and the effects that saturation may have on the ω MLEs approach should be borne in mind, particularly when studying *Plasmodium* sequences. The EP approach, however, provides an example of how molecular evolution can be applied to biomedical challenges in *P. falciparum* research.

7.5 Contributions to *P. falciparum* molecular evolution research resulting from this thesis

The research presented in this thesis has advanced the field of *P. falciparum* molecular evolution in a number of ways. An evolutionary rates approach to sequence analysis and inference of protein domain function, termed FIRE, was developed in chapter 2 to address the major impediment of poor sequence similarity associated with *P. falciparum* coding regions. FIRE effectively removes the requirement for similarity to align sequences accurately and it is envisaged that with further use and subsequent refinements this method will have

broad bioinformatic applications in *P. falciparum* specifically, and computational biology in general.

A second advance was the discovery of a co-evolutionary relationship between human PK deficiency and *P. falciparum* infection. This has been a controversial point for almost 20 years and the publications in chapter 3 and data from a similar study (Ayi *et al.*, 2008) provided the first direct evidence for this relationship. The data highlight the enormous impact *P. falciparum* has had on human evolution.

A theoretical framework for a dynamic model of genome evolution was developed and represents the first step towards an understanding of the genome as an individual evolutionary unit. The model may be used to investigate the evolution of any genome and in *P. falciparum* provides an opportunity to uncover the reasons for the extraordinary features encountered in this genome. Two key determinants implicated in a genome trade-off in *P. falciparum*, MGEs and PCD, were investigated in chapters 4 and 5 and produced insights into the parasite's life-history strategy.

Finally, a novel approach, termed EP, to limit the problem of drug resistance in *P. falciparum* was developed in chapter 6. It is hoped that this approach will minimize the evolution of drug resistance by identifying the most suitable drug target sites and demonstrates the value of applying evolutionary data to biomedical problems.

7.5.1 Concluding remarks

The exponential increase in molecular data and parallel advances in statistics, applied mathematics and computational software, which have facilitated more sophisticated analyses, have propelled the field of molecular evolution into an era of rapid expansion. It is hoped that these advances, and the contributions from the

work presented here, will be used to improve our understanding of molecular evolution in *P. falciparum* and encourage researchers to apply this knowledge in the battle against the most lethal infection in human history.

APPENDIX I: Cloning of *P. falciparum* TERT.

3D7 *P. falciparum* parasites were cultured *in vitro* using fresh human erythrocytes at 5% hematocrit and 10% AB plasma (Trager and Jensen, 1976). Cultures were incubated at 37°C in 93% N_2, 5% CO_2 and 2% O_2 (Afrox, South Africa). *P. falciparum* DNA was extracted using phenol-chloroform and ethanol precipitation (Ljungstrom *et al.*, 2004). Five domains from the two putative *P. falciparum* TERTs (PF13_0080 and PFE1555c) identified in chapter 4 (Durand *et al.*, 2006) were amplified from 100ng DNA via PCR using 2.5U Expand High Fidelity *Taq* polymerase (Roche, Germany) and 10pmol of primers. The five domains were F1 (605bp), P1 (533bp) and FL1 (2100bp), which correspond to the finger and palm domains and the full-length PF13_0080, and F2 (670bp) and P2 (550bp), which correspond to the finger and palm domains of PFE1555c. The finger domain is implicated in RNA binding and the palm domain is responsible for reverse transcriptase activity. The full length region contained both finger and palm domains. Forward and reverse primers, containing *Bam*HI (forward) and *Xho*I (reverse) restriction sites (underlined), are below. Primers are written in 5` to 3` direction and amplicon names are given in parentheses.

Forward primers for PF13_0080

GTCATGGATCCATATATAAAATAAAAAATATTATAGAGAAAAGAAA (F1 and FL1);

GTCATGGATCCAAAATTATTAGTAATATATATGGC (P1).

Reverse primers for PF13_0080

AGCGTCTCGAGATCAAAGAAATTTTTTAAAATCTTG (F1);

AGCGTCTCGAGAGAATCATTTTTTTGTATTGTT (P1)

AGCGTCTCGAGATTATTTATAAAATCAAATGTGTATGAATAATTTAA (FL1).

Forward primers for PFE1555c

GACATGGATCCGAGTTATATAATAAGAAATATACAAATGACA (F2);

GACATGGATCCATAAACAATAAAAATTTAAACAACAATC (P2).

Reverse primers for PFE1555c

CTGCACTCGAGATATTTTTCTTTGTCATTCTCATTAG (F2);

CTGCACTCGAGACATAGGATAATTTTATATTCTACGTATC (P2).

All PCR reactions included an initail denaturation step at 94°C for 2 min and a final extension step at 72°C for 10 min. PCR was performed under the following conditions:

F1: 5 cycles: 94°C, 1 min; 40°C, 1 min; 72°C, 2 min;
39 cycles: 94°C, 1 min; 55°C, 1 min; 72°C, 2 min;
P1: 5 cycles: 94°C, 1 min; 45°C, 1 min; 72°C, 2 min;
39 cycles: 94°C, 1 min; 56°C, 1 min; 72°C, 2 min;
FL1: 5 cycles: 94°C, 1 min; 45°C, 1 min; 72°C, 2 min;
34 cycles: 94°C, 1 min; 59°C, 1 min; 72°C, 2 min;
F2: 5 cycles: 94°C, 1 min; 45°C, 1 min; 72°C, 2 min;
30 cycles: 94°C, 1 min; 58°C, 1 min; 72°C, 2 min;
P2: 5 cycles: 94°C, 1 min; 45°C, 1 min; 72°C, 2 min;
30 cycles: 94°C, 1 min; 55°C, 1 min; 72°C, 2 min.

The PCR amplicons and pGEX-4T-2 (Amersham, UK) expression vector were digested with *Bam*HI and *Xho*I (Fermentas, Lithuania) and domains were ligated downstream of the Glutathione S-transferase (GST) gene sequence in the vector (Figure A1). Competent DH5α *E. coli* (Invitrogen, USA) were transformed with the plasmid construct and positive colonies were selected on Luria Bertani (LB) / agar supplemented with 100µg/ml ampicillin (Roche, Germany). Inserts were verified using gene specific PCR, plasmid extraction, restriction enzyme analysis and DNA sequencing.

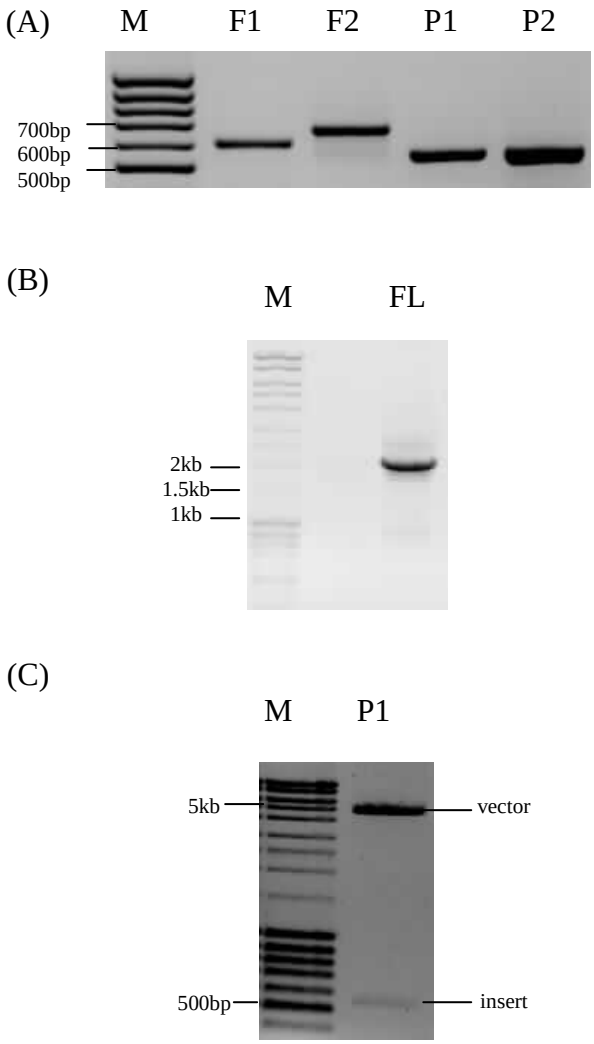

Figure A1: Cloning of PfTERT.

The 1% agarose gels display: (A) PCR amplicons of finger (F) and palm (P) domains from two putative TERTs. Expected sizes in bp were F1=605; F2=670; P1=533; and P2=550; M: 50bp ruler (Fermentas). (B) Full length (FL) PCR product of PF13_0080. The expected size was 2100bp; M: 1kb marker (Fermentas). (C) Each of the five pGEX-4-T2 plasmid constructs was extracted from positive colonies and inserts were verified with *Bam*H1 and *Xho*I digestion. A digestion of the P1 recombinant plasmid, which was used in the protein expression experiments in Appendix II, is shown. Expected sizes: linearized pGEX-4-T2 vector = 4.9kb; P1 insert = 533bp. M: 1kb marker (Fermentas).

APPENDIX II: Expression of recombinant PfTERT P1 domain.

The P1 construct was expressed in Rosetta2 (DE3) *E. coli* (Novagen, USA) as a fusion protein with an N-terminal GST tag. Transformed cells were selected in LB medium containing 100µg/ml ampicillin and 50µg/ml chloramphenicol (Roche, Germany). 1/100 (v/v) of the antibiotic-selected cultures were added to the Overnight Express™ Instant TB Medium (Novagen, USA) auto-induction system and allowed to grow to an OD_{600nm} ~1.5 at room temperature. Cells were lysed with BugBuster® HT (Novagen, USA) supplemented with the Protease Inhibitor Cocktail Set III (Novagen, USA). GST-P1 recombinant protein was purified by affinity chromatography with GST◦Mag™ Agarose beads (Promega, USA) and eluted from the beads with 100mM reduced glutathione in 50mM Tris-HCl, pH 8.0. Protein fractions were obtained from lysed cells (total), the supernatant (soluble) and pellet (insoluble) following centrifugation, and the eluted sample (purified). Samples were analyzed by SDS-PAGE (Laemmli, 1970) and immunoblotting with a 1:10000 (v/v) diluted anti-GST HRP-conjugated antibody (Amersham Biosciences, UK) and visualized with the SuperSignal® West Pico Chemiluminescent Substrate (Pierce, USA). The GST tag migrates at ~24kD and the GST-P1 recombinant protein at ~35kDa (Figure 2A). Immunodetection revealed the presence of recombinant protein in total, insoluble, soluble and purified fractions of the induced sample (labeled in italics). There is leaky expression of both GST and GST-P1 recombinant proteins in the uninduced samples. The absence of a visible band on the nitrocellulose membrane indicates that very small quantities of the protein are present, however, the GST tag was still detected by antibody. Recombinant protein production will be scaled up to produce greater quantities for functional analyses. Functional investigations will include RT (reverse transcriptase) and TRAP (telomere repeat amplification protocol) assays.

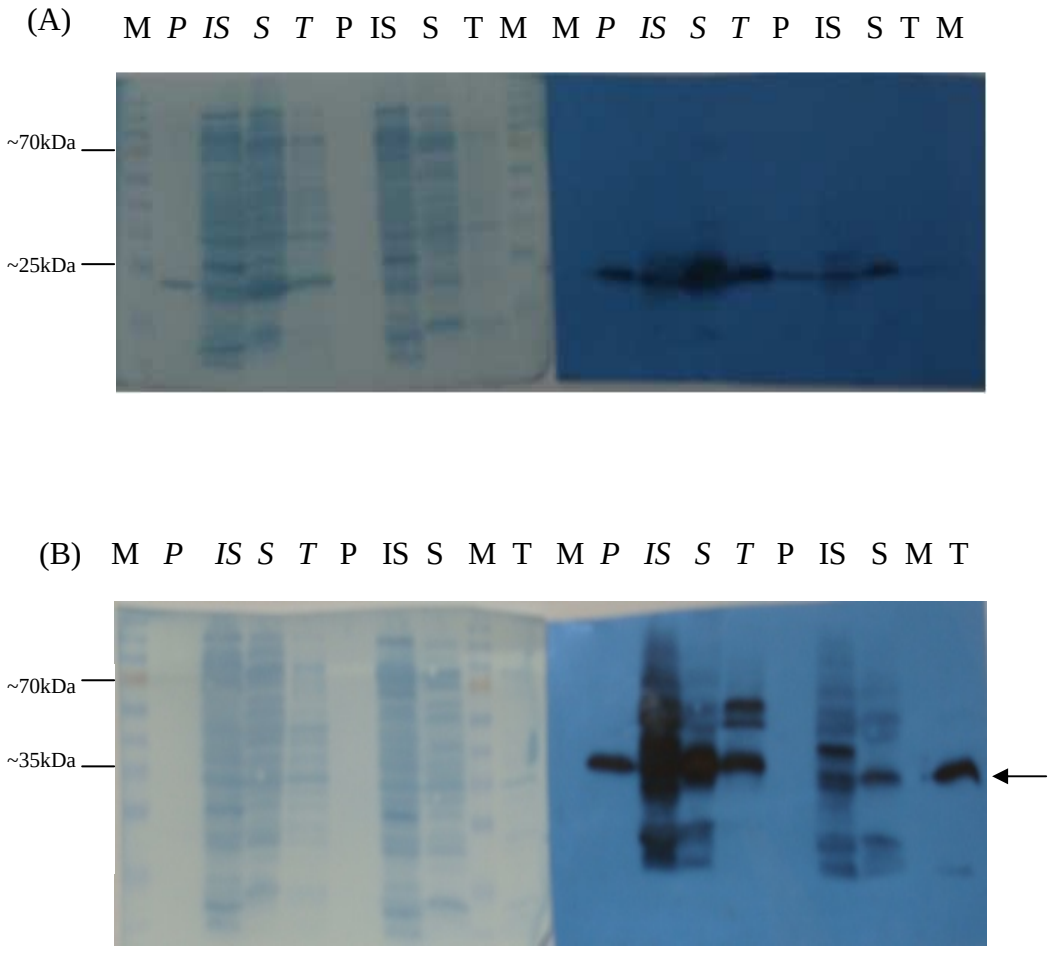

Figure A2: Expression of recombinant PfTERT P1 domain.
Amido-black stained nitrocellulose membranes and 10 second exposure autoradiographs of (A) GST, and (B) GST-P1 recombinant protein. M: PageRuler Prestained Protein Ladder (Fermentas); protein fractions: T-total; S-soluble; IS-insoluble and P-purified; induced sample fractions are shown in italics. GST migrates at ~25kDa and the GST-P1 fusion protein (arrow) at ~35kDa.

REFERENCES

Al-Olayan, E.M., Williams, G.T. and Hurd, H. (2002) Apoptosis in the malaria protozoan, *Plasmodium berghei*: a possible mechanism for limiting intensity of infection in the mosquito. *Int J Parasitol*, **32**, 1133-43.

Alli, N., Coetzee, M., Louw, V., van Rensburg, B., Rossouw, G., Thompson, L., Pissard, S. and Thein, S.L. (2008) Sickle cell disease in a carrier with pyruvate kinase deficiency. *Hematology*, **13**, 369-72.

Ameisen, J.C. (2002) On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Cell Death Differ*, **9**, 367-93.

Aravind, L., Iyer, L.M., Wellems, T.E. and Miller, L.H. (2003) *Plasmodium* biology: genomic gleanings. *Cell*, **115**, 771-85.

Ayi, K., Min-Oo, G., Serghides, L., Crockett, M., Kirby-Allen, M., Quirt, I., Gros, P. and Kain, K.C. (2008) Pyruvate kinase deficiency and malaria. *N Engl J Med*, **358**, 1805-10.

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I. and Doolittle, W.F. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, **290**, 972-7.

Belancio, V.P., Hedges, D.J. and Deininger, P. (2008) Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Res*, **18**, 343-358.

Bhattacharya, S., Bakre, A. and Bhattacharya, A. (2002) Mobile genetic elements in protozoan parasites. *J Genet*, **81**, 73-86.

Blasco, M.A., Lee, H.W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A. and Greider, C.W. (1997) Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell*, **91**, 25-34.

Brehelin, L., Dufayard, J.F. and Gascuel, O. (2008) PlasmoDraft: a database of *Plasmodium falciparum* gene function predictions based on postgenomic data. *BMC Bioinformatics*, **9**, 440.

Brookfield, J.F. (2005) The ecology of the genome - mobile DNA elements and their hosts. *Nat Rev Genet*, **6**, 128-36.

Brosius, J. (1999) Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica*, **107**, 209-38.

Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S.V., Merino, E.F., Amedeo, P., Cheng, Q., Coulson, R.M., Crabb, B.S., Del Portillo, H.A., Essien, K., Feldblyum, T.V., Fernandez-Becerra, C., Gilson, P.R., Gueye, A.H., Guo, X., Kang'a, S., Kooij, T.W., Korsinczky, M., Meyer, E.V., Nene, V., Paulsen, I., White, O., Ralph, S.A., Ren, Q., Sargeant, T.J., Salzberg, S.L., Stoeckert, C.J., Sullivan, S.A., Yamamoto, M.M., Hoffman, S.L., Wortman, J.R., Gardner, M.J., Galinski, M.R., Barnwell, J.W. and Fraser-Liggett, C.M. (2008a) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*, **455**, 757-63.

Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Pertea, M., Silva, J.C., Ermolaeva, M.D., Allen, J.E., Selengut, J.D., Koo, H.L., Peterson, J.D., Pop, M., Kosack, D.S., Shumway, M.F., Bidwell, S.L., Shallom, S.J., van Aken, S.E., Riedmuller, S.B., Feldblyum, T.V., Cho, J.K., Quackenbush, J., Sedegah, M., Shoaibi, A., Cummings, L.M., Florens, L., Yates, J.R., Raine, J.D., Sinden, R.E., Harris, M.A., Cunningham, D.A., Preiser, P.R., Bergman, L.W., Vaidya, A.B., van Lin, L.H., Janse, C.J., Waters, A.P., Smith, H.O., White, O.R., Salzberg, S.L., Venter, J.C., Fraser, C.M., Hoffman, S.L., Gardner, M.J. and Carucci, D.J. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*, **419**, 512-9.

Carlton, J.M., Escalante, A.A., Neafsey, D. and Volkman, S.K. (2008b) Comparative evolutionary genomics of human malaria parasites. *Trends Parasitol*, **24**, 545-50.

Cawley, S.E., Wirth, A.I. and Speed, T.P. (2001) Phat: a gene finding program for *Plasmodium falciparum*. *Mol Biochem Parasitol*, **118**, 167-74.

Chipuk, J.E. and Green, D.R. (2006) Dissecting p53-dependent apoptosis. *Cell Death Differ*, **13**, 994-1002.

Delwiche, C.F. (1999) Tracing the Thread of Plastid Diversity through the Tapestry of Life. *Am Nat*, **154**, S164-S177.

Durand, P.M. and Coetzer, T.L. (2008a) Utility of computational methods to identify the apoptotic machinery in unicellular eukaryotes. *Bioinform Biol Insights*, **2**, 101-117.

Durand, P.M. and Coetzer, T.L. (2008b) Pyruvate kinase deficiency in a South African kindred caused by a 1529A mutation in the PK-LR gene. *S Afr Med J*, **98**, 456-7.

Durand, P.M. and Coetzer, T.L. (2008c) Pyruvate kinase deficiency protects against malaria in humans. *Haematologica*, **93**, 939-40.

Durand, P.M. and Coetzer, T.L. (2008d) Hereditary red cell disorders and malaria resistance. *Haematologica*, **93**, 961-3.

Durand, P.M. and Coetzer, T.L. (2008e) Determinants of *P. falciparum* evolution using a genome-centered approach. Abstract 117. Keystone Symposium: Molecular evolution as a driving force in infectious disease, Breckenridge, USA.

Durand, P.M., Hazelhurst, S. and Coetzer, T.L. (2009) Evolutionary rates at codon sites may be used to align sequences and infer protein domain function. *BMC Bioinformatics*, under second review.

Durand, P.M., Naidoo, K. and Coetzer, T.L. (2008) Evolutionary patterning: a novel approach to the identification of potential drug target sites in *Plasmodium falciparum*. *PLoS ONE*, **3**, e3685.

Durand, P.M., Oelofse, A.J. and Coetzer, T.L. (2006) An analysis of mobile genetic elements in three *Plasmodium* species and their potential impact on the nucleotide composition of the *P. falciparum* genome. *BMC Genomics*, **7**, 282.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755-63.

Embley, T.M. and Martin, W. (2006) Eukaryotic evolution, changes and challenges. *Nature*, **440**, 623-30.

Emes, R.D. and Yang, Z. (2008) Duplicated paralogous genes subject to positive selection in the genome of Trypanosoma brucei. *PLoS One*, **3**, e2295.

Escalante, A.A. and Ayala, F.J. (1995) Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. *Proc Natl Acad Sci U S A*, **92**, 5793-7.

Feagin, J.E., Gardner, M.J., Williamson, D.H. and Wilson, R.J. (1991) The putative mitochondrial genome of *Plasmodium falciparum*. *J Protozool*, **38**, 243-245.

Fichera, M.E. and Roos, D.S. (1997) A plastid organelle as a drug target in apicomplexan parasites. *Nature*, **390**, 407-9.

Figueiredo, L.M., Pirrit, L.A. and Scherf, A. (2000) Genomic organisation and chromatin structure of *Plasmodium falciparum* chromosome ends. *Mol Biochem Parasitol*, **106**, 169-74.

Figueiredo, L.M., Rocha, E.P., Mancio-Silva, L., Prevost, C., Hernandez-Verdun, D. and Scherf, A. (2005) The unusually large *Plasmodium* telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus. *Nucleic Acids Res*, **33**, 1111-22.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, S.J., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L. and Bateman, A. (2008) The Pfam protein families database. *Nucleic Acids Res*, **36**, D281-8.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E. and Scherf, A. (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature*, **407**, 1018-22.

Frost, L.S., Leplae, R., Summers, A.O. and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol*, **3**, 722-32.

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., Davis, R.W., Fraser, C.M. and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498-511.

Gerstein, M., Sonnhammer, E.L. and Chothia, C. (1994) Volume changes in protein evolution. *J Mol Biol*, **236**, 1067-78.

Ginsburg, H. (2009) Caveat emptor: limitations of the automated reconstruction of metabolic pathways in *Plasmodium*. *Trends Parasitol*, **25**, 37-43.

Graur, D. and Li, W.H. (2000) *Fundamentals of molecular evolution*. Second edition. Sinauer Associates Inc., Sunderland, USA.

Haldane, J.B.S. (1949) The rate of mutation of human genes. Proceedings of the VIII International Congress of Genetics. *Hereditas*, **35**, 267-273.

Hartl, D.L., Volkman, S.K., Nielsen, K.M., Barry, A.E., Day, K.P., Wirth, D.F. and Winzeler, E.A. (2002) The paradoxical population genetics of *Plasmodium falciparum*. *Trends Parasitol*, **18**, 266-72.

Hasegawa, M., Kishino, H. and Yano, T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*, **22**, 160-74.

Hayakawa, T., Culleton, R., Otani, H., Horii, T. and Tanabe, K. (2008) Big bang in the evolution of extant malaria parasites. *Mol Biol Evol*, **25**, 2233-9.

Hazelhurst, S. and Durand, P.M. (2009) *FIRE*: Functional Inference using the Rate of Evolution. *Bioinformatics*, manuscript prepared.

Heng, H.H. (2009) The genome-centric concept: resynthesis of evolutionary theory. *Bioessays*, **31**, 512-25.

Hyde, J.E. (2005) Drug-resistant malaria. *Trends Parasitol*, **21**, 494-8.

Jeffares, D.C., Pain, A., Berry, A., Cox, A.V., Stalker, J., Ingle, C.E., Thomas, A., Quail, M.A., Siebenthall, K., Uhlemann, A.C., Kyes, S., Krishna, S., Newbold, C., Dermitzakis, E.T. and Berriman, M. (2007) Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nat Genet*, **39**, 120-5.

Jones, M.K. and Good, M.F. (2006) Malaria parasites up close. *Nat Med*, **12**, 170-1.

Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., Beerli, P. and Su, X.Z. (2003) Early origin and recent expansion of *Plasmodium falciparum*. *Science*, **300**, 318-21.

Karchin, R. and Hughey, R. (1998) Weighting hidden Markov models for maximum discrimination. *Bioinformatics*, **14**, 772-82.

Kazazian, H.H., Jr. (2004) Mobile elements: drivers of genome evolution. *Science*, **303**, 1626-32.

Kidwell, M.G. and Lisch, D.R. (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution*, **55**, 1-24.

Kohler, S., Delwiche, C.F., Denny, P.W., Tilney, L.G., Webster, P., Wilson, R.J., Palmer, J.D. and Roos, D.S. (1997) A plastid of probable green algal origin in Apicomplexan parasites. *Science*, **275**, 1485-9.

Koonin, E.V. and Aravind, L. (2002) Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death Differ*, **9**, 394-404.

Kuo, C.H., Wares, J.P. and Kissinger, J.C. (2008) The Apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. *Mol Biol Evol*, **25**, 2689-98.

Kyes, S.A., Kraemer, S.M. and Smith, J.D. (2007) Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the var multigene family. *Eukaryot Cell*, **6**, 1511-20.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-5.

Lane, D.P. (1992) p53, guardian of the genome. *Nature*, **358**, 15-6.

Lang-Unnasch, N., Reith, M.E., Munholland, J. and Barta, J.R. (1998) Plastids are widespread and ancient in parasites of the phylum Apicomplexa. *Int J Parasitol*, **28**, 1743-54.

Le Rouzic, A. and Capy, P. (2005) The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics*, **169**, 1033-43.

Le, S.Q. and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol Biol Evol*, **25**, 1307-20.

Liolios, K., Mavromatis, K., Tavernarakis, N. and Kyrpides, N.C. (2008) The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res*, **36**, D475-9.

Ljungstrom, I., Perlmann, H., Schlichtherle, M., Scherf, A. and Wahlgren, M. (2004) *Methods in malaria research*. Fourth. MR4 / ATCC, Manassas.

Marcotte, E.M. (2000) Computational genetics: finding protein function by nonhomology methods. *Curr Opin Struct Biol*, **10**, 359-65.

Martinsen, E.S., Perkins, S.L. and Schall, J.J. (2008) A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Mol Phylogenet Evol*, **47**, 261-73.

Mauricio, R. (2005) Can ecology help genomics: the genome as ecosystem? *Genetica*, **123**, 205-9.

Mayr, E. (1954) *Change of genetic environment and evolution.* In: Evolution as a process. Huxley, J and Hardy AC eds, p157-180, George Allen and Unwin, London.

McCann, A., Cotton, J.A. and McInerney, J.O. (2008) The tree of genomes: an empirical comparison of genome-phylogeny reconstruction methods. *BMC Evol Biol*, **8**, 312.

McClintock, B. (1929) A cytological and genetical study of triploid maize. *Genetics*, **14**, 180-222.

McConkey, G.A., Rogers, M.J. and McCutchan, T.F. (1997) Inhibition of *Plasmodium falciparum* protein synthesis. Targeting the plastid-like organelle with thiostrepton. *J Biol Chem*, **272**, 2046-9.

McFadden, G.I., Waller, R.F., Reith, M.E., Munholland, J. and Lang-Unnasch, N. (1997) Plastids in apicomplexan parasites. *Pl Syst Evol*, **Suppl 11**, 261-287.

Meslin, B., Barnadas, C., Boni, V., Latour, C., De Monbrison, F., Kaiser, K. and Picot, S. (2007) Features of apoptosis in *Plasmodium falciparum* erythrocytic stage through a putative role of PfMCA1 metacaspase-like protein. *J Infect Dis*, **195**, 1852-9.

Meyer, D.H. and Bailis, A.M. (2008) Telomerase deficiency affects the formation of chromosomal translocations by homologous recombination in *Saccharomyces cerevisiae*. *PLoS ONE*, **3**, e3318.

Michod, R.E. (2005) On the transfer of fitness from the cell to the multicellular organism. *Biol Philosoph*, **20**, 967-987.

Michod, R.E. (2006) The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. *Proc Natl Acad Sci U S A*, **103**, 9113-7.

Michod, R.E. and Roze, D. (1997) Transitions in individuality. *Proc Biol Sci*, **264**, 853-7.

Min-Oo, G., Fortin, A., Tam, M.F., Nantel, A., Stevenson, M.M. and Gros, P. (2003) Pyruvate kinase deficiency in mice protects against malaria. *Nat Genet*, **35**, 357-62.

Mohrenweiser, H.W. (1987) Functional hemizygosity in the human genome: direct estimate from twelve erythrocyte enzyme loci. *Human Genet*, **77**, 241-245.

Moore, R.B., Obornik, M., Janouskovec, J., Chrudimsky, T., Vancova, M., Green, D.H., Wright, S.W., Davies, N.W., Bolch, C.J., Heimann, K., Slapeta, J., Hoegh-Guldberg, O., Logsdon, J.M. and Carter, D.A. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature*, **451**, 959-63.

Mu, J., Duan, J., Makova, K.D., Joy, D.A., Huynh, C.Q., Branch, O.H., Li, W.H. and Su, X.Z. (2002) Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature*, **418**, 323-6.

Nakamura, T.M. and Cech, T.R. (1998) Reversing time: Origin of telomerase. *Cell*, **92**, 587-590.

Nickrent, D.L., Parkinson, C.L., Palmer, J.D. and Duff, R.J. (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol Biol Evol*, **17**, 1885-95.

Pain, A., Bohme, U., Berry, A.E., *et al.* (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*, **455**, 799-803.

Patthy, L. (1999) Genome evolution and the evolution of exon-shuffling--a review. *Gene*, **238**, 103-14.

Penman, B. and Gupta, S. (2008) Evolution of virulence in malaria. *J Biol*, **7**, 22.

Phan, I.Q., Pilbout, S.F., Fleischmann, W. and Bairoch, A. (2003) NEWT, a new taxonomy portal. *Nucleic Acids Res*, **31**, 3822-3.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, C.J., Roos, D.S. and McFadden, G.I. (2004) Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat Rev Microbiol*, **2**, 203-16.

Raven, J.A. and Allen, J.F. (2003) Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol*, **4**, 209.

Rich, S.M. and Ayala, F.J. (2003) Progress in malaria research: the case for phylogenetics. *Adv Parasitol*, **54**, 255-80.

Roff, D.A. (1992) *The Evolution of Life Histories: Theory and Analysis*. Chapman and Hall, New York. Chapters 3, 5-6.

Roy, S.W. and Irimia, M. (2008) Origins of human malaria: rare genomic changes and full mitochondrial genomes confirm the relationship of *Plasmodium falciparum* to other mammalian parasites but complicate the origins of *Plasmodium vivax*. *Mol Biol Evol*, **25**, 1192-8.

Roy, S.W. and Penny, D. (2007) Widespread intron loss suggests retrotransposon activity in ancient apicomplexans. *Mol Biol Evol*, **24**, 1926-33.

Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y. and Hay, S.I. (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*, **434**, 214-7.

Southwood, T.R.E. (1988) Tactics, strategies and templates. *Oikos*, **52**, 3-18.

Spocter, M.A. and Strkalj, G. (2007) Darwinian medicine: an evolutionary perspective on health and disease. *S Afr Med J*, **97**, 1044-6.

Stergiou, L. and Hengartner, M.O. (2004) Death and more: DNA damage response pathways in the nematode *C. elegans*. *Cell Death Differ*, **11**, 21-8.

Stoeckert, C.J., Jr., Fischer, S., Kissinger, J.C., Heiges, M., Aurrecoechea, C., Gajria, B. and Roos, D.S. (2006) PlasmoDB v5: new looks, new genomes. *Trends Parasitol*, **22**, 543-6.

Tang, K.L., Berendzen, P.B., Wiley, E.O., Morrissey, J.F., Winterbottom, R. and Johnson, G.D. (1999) The phylogenetic relationships of the suborder Acanthuroidei (Teleostei: Perciformes) based on molecular and morphological evidence. *Mol Phylogenet Evol*, **11**, 415-25.

Testa, B. and Lemont, B.K. (2000) Emergence and Dissolvence in the Self-organization of Complex Systems. *Entropy*, **2**, 1-25.

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

Ungerer, M.C., Johnson, L.C. and Herman, M.A. (2008) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity*, **100**, 178-83.

Vaidya, A.B., Akella, R. and Suplick, K. (1989) Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol Biochem Parasitol*, **35**, 97-107.

van de Peer, Y., Frickey, T., Taylor, J. and Meyer, A. (2002) Dealing with saturation at the amino acid level: a case study based on anciently duplicated zebrafish genes. *Gene*, **295**, 205-11.

van Wijk, R., Huizinga, E.G., van Wesel, A.C., van Oirschot, B.A., Hadders, M.A. and van Solinge, W.W. (2009) Fifteen novel mutations in PKLR associated with pyruvate kinase (PK) deficiency: structural implications of amino acid substitutions in PK. *Hum Mutat*, **30**, 446-53.

Weatherall, D.J. (2008) Genetic variation and susceptibility to infection: the red cell and malaria. *Br J Haematol*, **141**, 276-86.

Williamson, D.H., Gardner, M.J., Preiser, P., Moore, D.J., Rangachari, K. and Wilson, R.J. (1994) The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. *Mol Gen Genet*, **243**, 249-52.

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. and Williamson, D.H. (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol*, **261**, 155-72.

Wootton, J. and Federhan, S. (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Comput Chem*, **17**, 149-163.

Yang, Z. (2006) *Computational molecular evolution*. Oxford University Press, Oxford.

Yang, Z. and Nielsen, R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*, **17**, 32-43.