Global genomics analysis of the rodent malaria parasites *Plasmodium Berghei*, *Plasmodium Chabaudi* and *Plasmodium Yoelii* with emphasis on the identification of novel virulence factors in *Plasmodium Yoelii*

Liew, Kingsley Jiin Liang

2010

Liew, K. J. L. (2010). Global genomics analysis of the rodent malaria parasites *Plasmodium Berghei*, *Plasmodium Chabaudi* and *Plasmodium Yoelii* with emphasis on the identification of novel virulence factors in *Plasmodium Yoelii*. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/20818>

<https://doi.org/10.32657/10356/20818>

**GLOBAL GENOMICS ANALYSIS OF THE RODENT
MALARIA PARASITES PLASMODIUM BERGHEI,
PLASMODIUM CHABAUDI AND PLASMODIUM YOELII
WITH EMPHASIS ON THE IDENTIFICATION OF NOVEL
VIRULENCE FACTORS IN PLASMODIUM YOELII**

KINGSLEY LIEW JIIN LIANG

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2010

ACKNOWLEDGEMENTS

I would like to extend my thanks and gratitude to my supervisor A/Prof Peter Preiser for his valuable counsel and generosity in providing me with the opportunity to work in his lab. For that, I am indebted to his guidance and wisdom for driving my motivation and passion throughout the course of the project. I want to also thank A/Prof Zbynek Bozdech for helpful discussions and use of his laboratory facilities for the microarray experiments and analysis.

I would like to also express my thanks and appreciation to the past and present members of the Preiser Lab, especially: Dr. Amaladoss Anburaj, Dr. Jayasree Iyer and Ms. Amy Huang for their expertise in animal work and Dr. Annie Gao for her technical expertise, support and friendship. My gratitude also extends to past and present members of the Bozdech lab such as Dr. Jingguang Li for his technical expertise in microarray experiments and Mr. Guangan Hu for helpful discussions on data analysis and generating useful scripts for bulk data processing.

I want to also thank A/Prof Kevin Tan and Ms. Geok Choo Ng of NUS for the use of the microbiology laboratory facilities when our animals were first housed in NUS. I am also grateful for my dearest wife Carmen for her unwavering support, love and self-sacrifice. I wish to also thank my parents and grandmother for their encouragement and support through all these years.

TABLE OF CONTENTS

CONTENTS	PAGE
TITLE	1
ACKNOWLEDGEMENTS	2
TABLE OF CONTENTS	3
LIST OF TABLES	9
LIST OF FIGURES	10
CHAPTER 1: Summary	17
CHAPTER 2: Introduction	
2.1 Introduction to Malaria	19
2.2 Life cycle of the malaria parasite	20
2.3 Invasion of erythrocytes by malaria merozoites	22
2.4 Parasite – host genetic interactions	31
2.5 The Plasmodium Sequencing Project	33
2.6 A review on the recent Genomics and Proteomics studies	36
2.7 Tools for Proteomic investigations	41
2.8 Rodent models for malaria research	42
2.9 Virulence factors of <i>P. yoelii</i>	47
2.10 Challenges involved in rodent malaria research and the expected differences between the virulent and avirulent parasite lines	51
CHAPTER 3: Materials and Methods	
3.1 Propagation of rodent parasites in mice	55
3.2 Establishment of frozen parasites stocks	55

3.3	Parasite harvest and purification	56
3.4	Preparation and spotting of poly-L-lysine slides for microarray	57
3.5	Oligo array post-processing	58
3.6	Genomic DNA extraction from blood samples	60
3.7	Total RNA extraction	61
3.8	DNase Treatment of RNA samples prior to reverse transcription	62
3.9	Klenow reaction for genomic DNA microarray	62
3.10	First strand cDNA synthesis for microarray	63
3.11	Cy Dye Coupling	64
3.12	Microarray hybridization and analysis	64
3.13	Polymerase chain reaction (PCR)	65
3.14	Agarose gel electrophoresis	66
3.15	Purification of DNA from agarose gels	66
3.16	Purification of DNA samples	67
3.17	TOPO cloning of optimized PCR products	68
3.18	Cloning of restriction enzyme digested PCR products	69
3.19	Plasmid extraction and purification – Miniprep	71
3.20	Plasmid extraction and purification – Maxiprep	72
3.21	Electroporation transfection of <i>P. yoelii</i>	73
3.22	Southern Blot	74
3.23	Western Blot	77
3.24	Immunofluorescence Microscopy	79
3.25	Real-time RT-PCR	80

3.26	Selective Index (SI)	81
------	----------------------	----

CHAPTER 4: Comparative genome analysis of the three rodent malaria parasite species *Plasmodium berghei*, *Plasmodium chabaudi* and *Plasmodium yoelii*

4.1	Abstract	83
4.2	Introduction	
4.2.1	Rodent malaria parasites as useful models for <i>in vivo</i> studies	84
4.2.2	Genome organization of the human and rodent malaria parasites	86
4.2.3	Tools for elucidating genomic differences between closely related species	87
4.3	Results and discussion	
4.3.1	Validation of the pan-rodent long oligonucleotide array chip	90
4.3.2	Global analysis of gaps in the rodent malaria parasite genome and re-annotation of potential coding sequences via microarray hybridizations	100
4.3.3	Analysis of the remaining core syntenic gaps in the rodent malaria parasite genome with respect to the <i>P. falciparum</i> genes	107
4.3.4	Analysis of the rodent malaria parasite specific genes using <i>P. yoelii</i> as the index species	118
4.4	Conclusions	129

CHAPTER 5: Genomic differences between the virulent and avirulent strains of the rodent malaria parasite *Plasmodium yoelii*

5.1	Abstract	134
5.2	Introduction	135
5.3	Results and discussion	
5.3.1	Genomic differences between the avirulent YA strain and the virulent YM strain of <i>P. yoelii</i>	140
5.3.2	Identification of hypothetical proteins that are potentially involved in mediating host-parasite interactions	156
5.4	Conclusions	163

CHAPTER 6: Differential transcriptional profile of genes between the virulent YM and the avirulent YA strain of *Plasmodium yoelii* with focus on the SERA proteases

6.1	Abstract	168
6.2	Introduction	
6.2.1	The use of microarray as valuable tools for identification of differentially transcribed genes	169
6.2.2	The role of malarial proteases in invasion and egress	170
6.2.3	The serine repeat antigen (SERA) multigene family of the malaria parasites	171
6.3	Results and discussion	
6.3.1	Transcriptomic differences between the avirulent YA and virulent YM lines of <i>P. yoelii</i>	176
6.3.2	Validation of selected genes of interest via real-time RT-PCR	180
6.3.3	Real-time RT-PCR screen of the selected genes of interest to the different genetic lines of <i>P. yoelii</i> exhibiting different	

virulence phenotypes	184
6.3.4 Genomic location of the re-annotated <i>P. yoelii</i> SERA genes	195
6.3.5 Construction of GFP tagged PySERA1 and 2 proteins	197
6.3.6 Screening for transfected parasites expressing GFP tagged PySERA1 and PySERA2 proteins	200
6.3.7 Localization of PySERA1 and PySERA2	204
6.3.8 Targeted gene disruption of PySERA1 and PySERA2	212
6.3.9 Clonal isolation of transgenic <i>P. yoelii</i> parasites containing gene disruptions of PySERA1 and PySERA2	218
6.3.10 Phenotypic assessment of virulence of <i>P. yoelii</i> Δ PySERA1 clone	226
6.3.11 Phenotypic assessment of virulence of <i>P. yoelii</i> Δ PySERA2 clone	235
6.3.12 Real-time RT-PCR analysis of the transcriptional profile of the SERA genes in both Δ PySERA1 and Δ PySERA2 parasite lines	240
6.4 Conclusions	243

CHAPTER 7: Conclusions

7.1 Comparative genome analysis of the three rodent malaria parasite species <i>Plasmodium berghei</i> , <i>Plasmodium chabaudi</i> and <i>Plasmodium yoelii</i>	247
7.2 Genomic differences between the virulent YM strain and the avirulent YA strain of the rodent malaria parasite <i>Plasmodium yoelii</i>	249

7.3	Differential transcriptional profile of genes between the virulent YM and the avirulent YA strain of <i>Plasmodium yoelii</i>	250
7.4	Characterization of the SERA1 and SERA2 genes of <i>Plasmodium yoelii</i>	251
CHAPTER 8: References		254
CHAPTER 9: Appendices		
	Appendix 1: Buffers and reagents for rodent parasites	274
	Appendix 2: Buffers and reagents for proteomics	276
	Appendix 3: General buffers	280
	Appendix 4: Microarray buffers	283
	Appendix 5: Southern blot buffers	285
	Appendix 6: Primer sequences	288
	Appendix A: Identification of syntenic orthologs by microarray	CD-R
	Appendix B: <i>P. falciparum</i> genes with at least one rodent parasite ortholog	CD-R
	Appendix C: Orthology of the rodent malaria parasites	CD-R
	Appendix D: Polymorphic genes in YA as compared to 17x1.1	CD-R
	Appendix E: Gene expansions in YA as compared to 17x1.1	CD-R
	Appendix F: Polymorphic genes in YM as compared to 17x1.1	CD-R
	Appendix G: Gene expansions in YM as compared to 17x1.1	CD-R
	Appendix H: Polymorphic genes in YM as compared to YA	CD-R
	Appendix I: Polymorphic genes in YA as compared to YM	CD-R
	Appendix J: Differentially transcribed genes in YM with respects to YA	CD-R
	Appendix K: Differentially transcribed ring-stage genes in YM vs. YA	CD-R

LIST OF TABLES

TABLES		PAGE
2.1	Summary table of superfamilies of the erythrocyte-binding or Duffy-binding-like ligands (EBLs/DBLs) and the reticulocyte-binding protein homologs (RBLs) in the various <i>Plasmodium spp.</i>	24
2.2	Tabulated relationship between the different rodent malaria parasite strains and the typical experimental outcome in specific rodent host strains.	46
4.1	Summary of missing genes in the three rodent malaria parasite species via synteny.	104
4.2	Breakdown of the 37 additional <i>P. falciparum</i> genes found to be common to at least one rodent parasite species via orthology and tBLASTn searches.	110
4.3	Breakdown of the remaining annotated genes that have no orthologs in at least one other species.	111
4.4	Summary table of orthologs between the four <i>Plasmodium</i> species.	133
5.1	Summary tabulation of the repertoire of genes in YA that are likely to be significantly different from those in the reference 17x1.1 line.	146
5.2	Summary tabulation of the repertoire of genes in YM that are likely to be significantly different from those in the reference 17x1.1 line.	152
5.3	List of hypothetical proteins containing transmembrane domains with potential roles in host-parasite interactions.	157-158

5.4	List of polymorphic hypothetical proteins containing both transmembrane and signal peptide motifs.	162
6.1	A tabulated list showing the selected genes of interest.	183
6.2	Summary table of the SERA knockout phenotypes and localization studies in the malaria parasites.	246
9.1	PCR primers for verification of newly annotated genes.	288
9.2	PCR primers for verification of unique sets of genes.	289
9.3	Real-time RT-PCR primers for validating differentially transcribed genes.	290
9.4	PCR primers for SERA gene targeting.	291
9.5	Real-time RT-PCR primers for PySERA genes.	292

LIST OF FIGURES

FIGURES		PAGE
2.1	The life cycle of the malaria parasite showing both the asexual and sexual stages of development in the mammalian host and the mosquito vector.	21
2.2	Diagram showing the main features of a <i>P. falciparum</i> merozoite.	23
2.3	A model of the molecular events involved in host cell invasion.	27
2.4	A proposed model illustrating the process of selecting parasites that express appropriate ligands.	28
2.5	A schematic of the events leading to the genesis and of the apical complex.	30
2.6	Foci of infection of rodent malaria and the species of rodent <i>Plasmodia</i> isolated from this region.	43
4.1	The overall design schematics of the pan-rodent chip.	92
4.2	Venn diagram showing distribution of target rodent parasite gene hits of all oligonucleotides.	94
4.3	A schematic outlining the workflow for assigning specific oligonucleotides to their proper target gene.	95
4.4	Scatter plots of the median intensities of the individual oligonucleotides with the mean Cy3 signals plotted against the mean Cy5 signals.	97-98
4.5	An example of a typical scatter plot of signal intensity sorted in decending order against the number of <i>P. yoelii</i> specific	

	oligonucleotides.	99
4.6	Current scenarios depicting the probable scenarios of missing genes in the rodent malaria parasites.	101
4.7	A snapshot of the process whereby syntenic orthologs were identified using genomic DNA complementary hybridization experiments.	103
4.8	PCR screening of a random sample of newly discovered genes.	106
4.9	Graphical breakdown depicting the relationship between orthologous <i>Plasmodium falciparum</i> and the rodent malaria parasite (RMP) genes.	108
4.10	PCR screening of bioinformatics-filtered dataset with contig information.	114
4.11	PCR screening of bioinformatics-filtered dataset without contig information.	115
4.12	Distribution of the unique rodent malaria parasite (RMP) specific genes that are absent in <i>P. falciparum</i> .	121
4.13	Graphical plot showing the re-distribution of the rodent malaria parasite genes after removing the PIR genes.	122
4.14	Graphical plot showing the expression profile of the rodent malaria parasite genes after removing the PIR genes.	125-126
5.1	A schematic illustrating the possible scenarios involved in the comparison of genomic sequences from different parasite lines.	142
5.2	Semi-logarithmic curve depicting the distribution of genes with	

	respect to their similarity between <i>P. yoelii</i> YM line with respect to those in the 17x1.1 line.	143
5.3	Semi-logarithmic curve depicting the distribution of genes with respect to their similarity between <i>P. yoelii</i> YA line with respect to those in the 17x1.1 line.	144
6.1	Gene ordering of the SERA multigene family in eight representative <i>Plasmodium</i> species.	172
6.2	Gene structure of a typical SERA gene.	173
6.3	A schematic model of PfSERA5 processing in the role of egress from late blood stage parasites.	175
6.4	Distribution of genes found to be differentially expressed in YM as compared to YA.	177
6.5	Diagrammatic representation of the genetic cross between the virulent YM line and the avirulent AC line of <i>P. yoelii</i> .	185
6.6	Real-time RT-PCR screen of the progeny of the genetic cross between the virulent YM and avirulent AC lines of <i>P. yoelii</i> .	188-193
6.7	Genomic organization of the SERA genes in <i>P. yoelii</i> . The five known SERAs are flanked by 2 conserved hypothetical proteins.	196
6.8	Schematic depicting a single cross-over gene targeting construct strategy.	198
6.9	Schematic representation of the relative positions of the cysteine protease domain and the site of eGFP fusion with respect to the organization of the full-length SERA proteins.	199

6.10	Agarose gel electrophoresis profile of the PCR screen of eGFP tagged SERA genes in transfected parasites shown with the 1kb marker labeled as lane 'M'.	201
6.11	Molecular analysis of GFP-tagged PySERA proteins.	203
6.12	A representative live fluorescence microcopy image of a PySERA1-eGFP tagged parasite clone.	205
6.13	A representative live fluorescence microcopy image of a PySERA2-eGFP tagged parasite clone.	206
6.14	A representative live fluorescence microcopy image of a ruptured PySERA2-eGFP schizont clone.	208
6.15	A representative immunofluorescence microscopy stain of PySERA1	210
6.16	A representative immunofluorescence microscopy stain of PySERA2	211
6.17	Schematic representation of the relative positions of the cysteine protease domain and the site of eGFP fusion for a PySERA1 knockout with respect to the organization of the full length PySERA1 protein.	212
6.18	Targeted gene disruption of PySERA2 using a double crossover strategy.	214
6.19	PCR screen of the PySERA1 knockout parasites.	216
6.20	Detection of the PySERA2 knockout parasites.	217
6.21	Agarose gel electrophoresis profiles of PCR screens for the presence of correctly inserted targeting vector in recipient mice for the pure clone isolation of Δ PySERA1 parasites.	219

6.22	Agarose gel electrophoresis profiles of PCR screens for the presence of wild-type parasites in recipient mice for the pure clone isolation of Δ PySERA1 parasites.	220
6.23	Southern blot strategy for the detection of transgenic parasites that are pure clones of Δ PySERA1.	222
6.24	Agarose gel electrophoresis profiles of PCR screens for the presence of correctly inserted replacement targeting vector in recipient mice for the pure clone isolation of Δ PySERA2 parasites.	224
6.25	Agarose gel electrophoresis profiles of PCR screens for the presence of wild-type parasites in recipient mice for the pure clone isolation of Δ PySERA2 parasites.	225
6.26	Parasitemia curves of the sample populations of recipient mice infected with standardized doses of Δ PySERA1 and control wild-type parasites.	228
6.27	Average parasitemias growth curves of wild-type YM and Δ PySERA1 (291ko) parasites inoculated into 5 mice per sample population.	229
6.28	Giemsa-stained slides of the typical blood film smears of mice suffering from a chronic infection.	232
6.29	A scatter plot of the Selective Index (SI) values of both Δ PySERA1 and YM parasites calculated from infected mice with parasitemia loads of 5% to 10%.	234

6.30	Parasitemia curves of the sample populations of recipient mice infected with standardized doses of Δ PySERA2 and control wild-type parasites.	237
6.31	A scatter plot of the Selective Index (SI) values of both Δ PySERA2 and YM parasites calculated from infected mice with parasitemia loads of 5% to 10%.	238
6.32	Real-time RT-PCR screen of both Δ PySERA1 and Δ PySERA2 parasite lines for all five known SERA genes.	242

Summary

While the genome sequence of the human malaria parasite *Plasmodium falciparum* has already been completed, there are still sequence gaps and annotation errors in the genomes of the common rodent malaria parasite (RMP) species, namely *P. berghei*, *P. chabaudi* and *P. yoelii*. The primary aim is to establish a common 'core' *Plasmodia* gene set where the RMP genes can be re-annotated. Comparative genome hybridizations were performed on a microarray chip consisting of probes specific to all three RMP species. Unexpected hybridizations reveal novel orthologous genes in all species that are either due to missing sequence information or unannotated genes missed by current open reading frame prediction algorithms. Together with bioinformatics data mining and PCR verification, it was deduced that there is an expanded core set of *Plasmodia* genes that were actually common and present in the above-mentioned species. There is also a smaller subset of species-specific genes mainly limited to the subtelomeric regions. While the *P. falciparum* specific genes typically consist of the variant antigens such as the *var*, *stevor* and *rifin* multigene families, the elucidation of RMP specific genes was attempted by constructing a RMP orthology relationship table using genes from *P. yoelii* as an index and then removing genes that were orthologous to *P. falciparum*. RMP specific genes mainly consist of duplicated genes, hypothetical genes and the PIR multigene family, suggesting that the rodent parasite genomes are very similar. Comparative genomics study of different *P. yoelii* lines revealed significant polymorphisms in the *py235*, *yir* and hypothetical genes, thus identifying proteins under host immune selection pressure. In order to identify hypothetical genes that may function as novel parasite ligands, variant hypothetical genes containing at least a single transmembrane domain

Summary

were identified as putative targets for therapeutics as they are highly likely to be under immune selection pressure.

To identify novel genes associated with virulence, transcriptomic differences between the virulent YM and avirulent YA lines of *P. yoelii* were analyzed and the transcriptional activity of a selection of ten genes was confirmed via real-time RT-PCR. Of these, eight genes were consistently transcribed more abundantly in the other virulent lines of *P. yoelii*, thus implying a strong link to their potential role in virulence. Of these genes, two SERA proteases were selected for functional knockout studies to ascertain their role in virulence. PySERA1 knockout parasites showed an intermediate virulent phenotype, being more reticulocyte restricted and with surviving mice clearing the infection a full eight days earlier than the virulent wild-type YM line. However, PySERA2 knockout parasites did not show any significant decrease in virulence. Incidentally, the knockout lines did not exhibit any increase in transcriptional activity of the other SERA genes, thus reflecting a certain degree of redundancy amongst the members of this multigene family.

2.1 Introduction to Malaria

The term malaria which literally means bad air in Latin (mal-aria) was coined as the disease was casually linked to exposure with the poisonous vapors of swamps, marshes and even latrines. This historical association was not only a mere coincidence as these stagnant bodies of water are ideal breeding spots for the mosquito vector and they also typically contributed to ‘bad air’. It was only in 1880 that Dr. Alphonse Laveran who studied blood smears from febrile soldiers observed the developing parasite within infected erythrocytes. His pioneering work led to the discovery of the causative agent of malaria and was subsequently awarded the Medical Nobel prize in 1907 (NobelPrize.org). Malaria afflicts 300-600 million people annually in the tropical and subtropical areas of the world (Snow, R.W. *et al*, 2005) and is responsible for more than one million deaths. It is therefore still one of the most devastating and economically important diseases known to mankind.

Malaria is a parasitic disease caused by the species from the *Plasmodium* genus. These intracellular protozoan parasites can only be transmitted to a mammalian host via the *Anopheles* mosquito as demonstrated by Ronald Ross in 1897 who later received the Medical Nobel Prize in 1902 (NobelPrize.org). The complete life cycle of the parasite was finally deduced in 1948 when exo-erythrocytic schizonts were discovered in the livers of infected monkeys (Shortt, H.E. and Garnham, P.C.C., 1948) and in humans (Shortt, H.E., *et al*, 1951).

Humans are known to be infected with at least four species of the malaria parasite, namely: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovalae* and *Plasmodium malariae*. Interestingly, the simian parasite *Plasmodium knowlesi* is also

Introduction

known to infect humans. Knowles and Das Gupta (1932) have shown that experimental transfer of this simian parasite to man was permissible and perhaps the first reported case of *P. knowlesi* infection in man was described in 1965 (Chin, W. *et al*, 1965). Recently, reports of humans being infected with the primate parasite *P. knowlesi* in Southeast Asia have been shown by PCR analysis (Singh, B. *et al*, 2004). Of all the *Plasmodium* species that infect humans, *Plasmodium falciparum* is the most lethal and is responsible for most of the mortality and morbidity that are attributed to malaria. Patients afflicted with malaria commonly display a variety of symptoms such as: fever, chills, sweats, headaches, general malaise, body aches, nausea and vomiting. In *Plasmodium falciparum* malaria, additional symptoms include: mild jaundice, enlargement of the liver and increased respiratory rate (CDC Malaria).

2.2 Life cycle of the malaria parasite

The life cycle of the *Plasmodium falciparum* parasite (Figure 2.1, Good, M.F., 2001) begins when an infected female *Anopheles* mosquito feeds on a human host and the majority of the sporozoites are injected into dermis of the host since salivation occurs when the mosquito is actively probing for food. The sporozoites are capable of moving randomly in the dermis until they contact and penetrate blood capillaries to enter the blood circulation (Amino, R. *et al*, 2006). Sporozoites were slowly released from the skin to the circulation and a proportion was localized in the draining lymph nodes (Yamauchi, L. *et al*, 2007).

Figure 2.1. The life cycle of the malaria parasite showing both the asexual and sexual stages of development in the mammalian host and the mosquito vector. (Adapted from Good, M.F., 2001)

The sporozoites then rapidly invade hepatocytes where they multiply and form hepatic schizonts. The length of the exoerythrocytic schizogony and the number of hepatic merozoites released varies with different malaria parasite species. In *P. falciparum*, it takes five to seven days and each hepatic schizont can release around 40,000 merozoites; in *P. vivax*, it takes six to eight days to release around 10,000 merozoites. In the case of *P. ovale*, the liver stage lasts nine days and each hepatic schizont is capable of releasing around 15,000 merozoites. In contrast, *P. malariae* requires twelve to sixteen days for the hepatic schizonts to mature and releases around 2,000 merozoites per schizont (Despommier, D. *et al*, 2005).

Introduction

This asexual erythrocytic cycle is repeated as described and is relatively synchronous with the rupture of the infected cells followed by the release of nascent merozoites being responsible for the characteristic fever and other symptoms of malaria.

Some merozoites after invasion can differentiate to the sexual forms known as gametocytes. The gametocytes can then be ingested by another female *Anopheles* mosquito whereby fusion of the male and female gametes occurs to form the ookinete that invades the gut of the mosquito and then develops into an oocyst. As the oocyst matures, the released sporozoites move to the salivary glands and are able to infect another host once the mosquito takes another blood meal and the cycle continues. This mosquito cycle takes approximately 12 days and is temperature dependent (Gardiner, D.L. *et al*, 2004).

2.3 Invasion of erythrocytes by malaria merozoites

Apicomplexan parasites, inclusive of *Plasmodium*, contain a collection of organelles at the anterior ends. Proteins residing within these apical organelles bind to components of the host cell membranes and are responsible for the recognition, selection and penetration of the parasite into the host cell.

The apical organelles include: (i) rhoptries, a pair of electron-dense, membrane-bound pear-shaped ducted structures; (ii) micronemes, various small fusiform bodies attached to the rhoptry duct and (iii) dense granules (Figure 2.2, Preiser, P. *et al*, 2000). After the merozoites have been released from the schizont, they contain a surface coat of proteins such as the merozoite surface protein (MSP)-1, that mediates the initial contact with the erythrocyte. For invasion to proceed, the merozoite must somehow re-orientate

Introduction

itself such that its apical end contacts the host cell to create a tight junction between the parasite and erythrocyte membranes. Following apical contact, secretion of the microneme and rhoptry content occurs that lead to the invagination of the erythrocyte surface. The merozoite then moves into the newly created space, tunnels into the host cell and finally resides within a parasitophorous vacuole. During the final stages of invasion, the dense granules release their contents that cause further expansion of the membrane lining the vacuole.

Figure 2.2. Diagram showing the main features of a *P. falciparum* merozoite. (Adapted from Preiser, P. *et al*, 2000)

Merozoite invasion is a very complex, multi-step process and was reviewed extensively by Gaur and colleagues (2004). Basically, two distinct gene families that mediate parasite ligand-host receptor interactions such as the Duffy-binding-like (DBL) or

Introduction

erythrocyte binding like (EBL) and the reticulocyte-binding-like (RBL) protein families have been implicated in host cell specificity. The use of bioinformatics and direct experimental evidence has led to the discovery and putative host cell receptor targets of conserved members of these two protein families in the malaria parasites and is elegantly summarized as shown in Table 2.1 (Iyer, J. *et al*, 2007a)

Organism	Ligand (accession No.)	Receptor	Erythrocyte binding properties	Antibodies
<i>P. falciparum</i>	RH1 (PFD0110w)	Sialic acid residues trypsin-resistant receptor Y	Shown	Block binding
	RH2a (PF13_0196)	Unknown	Not shown	Block invasion
	RH2b (MAL13P1.176)	Chymotrypsin-sensitive receptor Z, independent of other members	Shown	Block invasion
	RH3 (PFL2520w)	Unknown	NA (pseudogene)	NA (pseudogene)
	RH4 (PFD1150c)	Sialic acid independent	Shown	Block binding
	RH5 (PFD1145c)	High trypsin sensitive	Shown	ND
	EBA175 (PF07_0128)	Glycophorin A	Sialic acid residues on Glycophorin A	Block binding and invasion
	EBA181/JESEBL (PFA0125c)	Trypsin-resistant, chymotrypsin-sensitive, sialic acid-dependent receptor W	Not shown	ND
	EBA165/PEBL (PFD1155w)	Unknown	NA (pseudogene)	NA (pseudogene)
	EBA140/BAEBL (MAL13P1.60)	Glycophorin C	Gerbich-positive erythrocytes	Block binding, inhibit ligand recognition
<i>P. vivax</i>	EBL1 (PFD1145c)	Unknown	Not shown	ND
	RBP-1 (Pv098585)	Unknown	Shown	ND
	RBP-2 (Pv090325)	Unknown	Shown	ND
<i>P. yoelii</i>	DBP (Pv110810)	Duffy antigen	Shown	Block invasion
	Py235 (PY01365)	Chymotrypsin- and trypsin-sensitive receptor	Shown	Block invasion
	Multigene family EBL (PY04764)	Unknown	Not shown	ND
<i>P. knowlesi</i>	DBP α (M90466)	Duffy antigen	Sialic acid independent	Block invasion of human, and partial inhibition of rhesus RBC
	DBP β (M90694)	Unknown	Not shown	ND
	DBP γ (M90964)	Unknown	Not shown	ND
<i>P. reichenowi</i>	RH	Unknown	Not shown	ND
	EBL	Unknown	Not shown	ND

Table 2.1. Summary table of superfamilies of the erythrocyte-binding or Duffy-binding-like ligands (EBLs/DBLs) and the reticulocyte-binding protein homologs (RBLs) in the various *Plasmodium spp.* Representative GenBank / PlasmoDB accession numbers are shown where available. (Modified from Iyer, J. *et al*, 2007a)

Introduction

It is clear that there is some redundancy in the utilization of the members from these two superfamilies as seen in *P. falciparum* where expression of PfRH4 correlates with a sialic acid-independent invasion phenotype when parasites were selected to invade neuraminidase-treated erythrocytes or by functional knock-out of the parasite *EBA175* gene (Stubbs, J. *et al*, 2005). This suggests some form of co-operation and interplay between the RBL and EBL protein families with regards to invasion pathways. Neuraminidase-adapted parasites were also able to switch back to their typical invasion pathway after re-adaptation to normal erythrocytes, with the silencing of the *PfRH4* gene. This reversible switching of erythrocyte receptors suggests the propensity of the parasite to adapt to host immune pressure and receptor polymorphisms and ensure the survival of the parasite. A functional disruption of the *PfRHI* gene also led to parasites that switched to a neuraminidase-resistant invasion phenotype, with compensation by other existing invasion molecules being likely to mediate host cell binding (Triglia, T. *et al*, 2005). A survey of *P. falciparum* field isolates demonstrated the natural ability of parasites to exhibit a much broader range of invasion pathways (Okoyeh, J.N. *et al*, 1999), thereby suggesting that parasites under constant immune pressure were more adaptable to switching to alternative invasion ligands than the typical laboratory strains.

Therefore, the most pertinent question that comes to mind is the molecular mechanism involved in sensing the repertoire of receptors on an erythrocyte leading to the committed invasion of a new host cell. A model of sensing by the RBLs followed by recruitment of EBLs (Figure 2.3) was proposed by Iyer and colleagues (2007a) based on the fact that *P. vivax* first senses the reticulocyte before recruitment of the DBL (Galinski

Introduction

M.R. and Barnwell, J.W., 1996) and the sensing of suitable host cells by the members of the *P. yoelii* PY235 multigene family (Khan, S.M. *et al*, 2001).

A model was proposed to explain the possible mechanism of parasites switching to alternative pathways of invasion (Iyer, J. *et al*, 2007a) where it was proposed that merozoites can express different repertoires of invasion ligands. Enzymatic treatment of erythrocytes to remove a certain subset of receptors would now select for merozoites that express an alternative ligand that is more specific to the resultant receptor profile of these enzymatically modified erythrocytes (Figure 2.4). Merozoites that succeed in invading these erythrocytes will thus be expanded and the progeny merozoites will develop some form of “genetic memory” such that a higher proportion will express ligands that are more suitable for invading such enzymatically modified erythrocytes. This same mechanism is also thought to come in play if a certain invasion ligand is targeted by antibodies. Merozoites expressing this ligand will be cleared while a small subset that express alternative ligands will then be selected for and subsequently expand and dominate. This model certainly fits very well with the PY235 story where individual merozoites in a single schizont are clonally distinct and express a distinct member of this multigene family (Preisner, P.R. *et al*, 1999). This implies that a ready pool of merozoites is available so as to rapidly change to a different invasion ligand.

Figure 2.3. A model of the molecular events involved in host cell invasion. Firstly, the RBLs may sense the receptors on the erythrocyte to determine its suitability as a host cell. After which, the high-affinity ligands such as EBA175 or other EBLs are recruited and leads to the release of the contents of the micronemes and junction formation to initiate the process of invasion. (Adapted from Iyer, J. *et al*, 2007a)

Introduction

Figure 2.4. A proposed model illustrating the process of selecting parasites that express appropriate ligands. A typical merozoite population will express different repertoires of invasion ligands and this is illustrated using different colors at the apical end of the merozoites. Procedures such as enzymatic treatment of erythrocytes or antibody directed against the erythrocyte receptor or parasite ligand lead to the loss of the dominant invasion pathway. A small subset of merozoites that express alternative invasion ligands can successfully invade with the generation of daughter merozoites that express the appropriate ligand at a high frequency to ensure the invasion potential. (Adapted from Iyer, J. *et al*, 2007a)

Introduction

Another phenomenon observed in *P. falciparum* is that most merozoites transcribe all or most of the RBL and EBL genes (Taylor, H.M. *et al*, 2002; Duraisingh, M.T. *et al*, 2003; Stubbs, J. *et al*, 2005; Cowman, A.F. and Crabb, B.S., 2006) but preferentially utilize only one invasion pathway as defined by a distinct subset of RBL and EBL genes. If these observations were to translate into equivalent amounts of protein, how does the parasite choose which RBL-EBL combination to use? It was proposed that only a small subset of EBL and/or RBL proteins can be exposed to interact with the erythrocyte surface (Duraisingh, M.T. *et al*, 2003; Baum, J. *et al*, 2005). Based on this proposed model, Iyer and colleagues (2007a) hypothesized the mechanism where parasites deficient in a particular invasion ligand would then present the opportunity for other ligands to occupy an 'active site' to interact with the erythrocyte surface, thus leading to a switch in invasion phenotype (Figure 2.5).

Figure 2.5. A schematic of the events leading to the genesis and of the apical complex. (A) A typical merozoite that expresses a wide repertoire of invasion ligands (RBLs and EBLs) in the apical organelles. (B) The loss of a particular ligand (blue) from the 'active site' that interacts with the erythrocyte then provides the opportunity for an alternative ligand (yellow) to occupy the 'active site'. (C) Only one ligand set of EBL and RBL (either blue or red spheres) can be exposed at the narrow 'active site' to interact with receptors on an erythrocyte. (Adapted from Iyer, J. *et al*, 2007a)

2.4 Parasite – host genetic interactions

In regions such as Africa and tropical countries where malaria is endemic and particularly chronic, the endogenous human population living in these areas exhibits a genetic drift whereby specific alleles that confer protection to severe malaria have been selected for and enriched throughout the community. For example, absence of the duffy blood group antigen receptor is so far the only mechanism that mediates protection to *P. vivax* infections. Individuals of an African descent have a higher proportion of a mutation that silences expression of the duffy antigen (Yazdanbakhsh, K. *et al*, 2000) whereas other populations (e.g. the Caucasians) accumulate an Arg89Cys mutation which leads to protein instability and hence reduced surface expression and loss of function of the receptor (Tamasauskas, D. *et al*, 2001).

Similarly, individuals with a heterozygous sickle cell anemia allele suffer less mortality and morbidity than individuals with homozygous normal trait. Consequently, this allele is maintained in high frequencies in endemic regions (Aidoo, M. *et al*, 2002). A key phenotype of the heterozygous sickle cell trait is a distinct malformed shape of the erythrocyte and it was thought that that made invasion somewhat less efficient. More recently, it has been shown that parasitized sickle-heterozygous erythrocytes do not display PfEMP-1 (a major ligand that is associated with sequestration) uniformly on the surface. Also, the typical knob structures look morphologically different and their distribution seems less dense than normal homozygous erythrocytes (Cholera, R. *et al*, 2008). Thus, these parasitized erythrocytes do not sequester efficiently and such patients often do not suffer life-threatening complications associated with cytoadherent parasites.

Introduction

Another factor is nitric oxide (NO) where high levels of this immune mediator seem to provide protection to severe complications such as cerebral malaria in African children (Anstey, N.M. *et al*, 1996). A G-954-C mutation in the nitric oxide synthase 2 promoter was consistently found in populations living in endemic regions but not in populations that do not usually encounter malaria (Kun, J.F. *et al*, 2001). The effect of this mutation led to elevated levels of NO that could potentially eliminate parasites in early infections. This mutation effectively raises the basal pool of NO to effectively deal with an initial infection without having to go through the typical signaling cascade for the induction of NO. Furthermore, localized release of high levels of NO following macrophage opsonization could have a cytotoxic effect on adjacent cytoadherent parasitized cells and thus reducing complications due to vascular sequestration. So far, another selectable mutation C-117-T in the same promoter also produced similar protection via elevated NO levels (Hobbs, M.R. *et al*, 2002).

Similarly, cellular receptors such as CD36 variants reduce sequestration of parasitized erythrocytes in severe malaria. CD36 is a major receptor on endothelial cells that is responsible for interacting with parasite proteins such as PfEMP1 and hence mediating cytoadherence. Therefore, any beneficial polymorphisms in this receptor is linked to protection to severe malaria and enriched in endemic regions (Pain, A. *et al*, 2001; Omi, K. *et al*, 2003). Another factor of note is the MHC class I B53 human leukocyte antigen. In this case, the presence of the B53 allele is thought to be associated with protection against malaria (Hill, A.V., 1992; Hill, A.V. *et al*, 1992). The fact that this MHC allele is found in high abundance in Africans suggests that an evolutionary pressure has presented endogenous populations living in malaria endemic areas with a

selective advantage of targeting pre-erythrocytic stages of parasite development. However, re-challenge experiments with sporozoites did not show any significant differences between populations containing the B53 allele and those that did not, thus suggesting that other factors may contribute to the actual protection (Dieye, A. *et al*, 1997).

2.5 The *Plasmodium* Sequencing Project

An effort was started in 1996 to sequence the *Plasmodium falciparum* genome in the hope that the sequence would provide new avenues for research (Hoffman, S.L. *et al*, 1997). Several independent groups have published sequences of single chromosomes such as chromosome 2 (Gardner, M.J. *et al*, 1999) and it was not until the year 2002 that the 23-megabase nuclear genome was deciphered (Gardner, M.J. *et al*, 2002).

The *P. falciparum* genome is organized into 14 chromosomes that encode about 5,300 genes and is the most (A+T)-rich genome ever sequenced. The malaria parasite expresses less enzymes and transporters as compared to other microbes. However, a large proportion of the genes are responsible for immune evasion and specific host-parasite interactions. Genes that are responsible for antigenic variation are commonly found in the subtelomeric regions of the chromosomes. Interestingly, only a third of the predicted proteins were found to be similar to known proteins in other organisms. This could perhaps reflect the huge evolutionary distance between *Plasmodium* and other eukaryotes (exaggerated by the (A+T)-rich genome) or the lack of current understanding on gene regulation and other biochemical processes of the malaria parasite.

Introduction

Also at the same time, the genome for the rodent parasite *Plasmodium yoelii* was deciphered (Carlton, J.M. *et al*, 2002). Rodent malaria parasites have been used extensively as models for malaria research ever since four species of the rodent parasites (*P. yoelii*, *P. berghei*, *P. chabaudi* and *P. vinckei*) have been adapted to grow in laboratory rodents (Carter, R. and Diggs, C.L., 1977). Experimental protocols were first developed in the rodent models prior to their use in the study of human parasites. An example is the establishment of stable transfectants in *P. berghei* (van Dijk, M.R. *et al*, 1995) and the subsequent success of having stable transfections in *P. falciparum* (Wu, Y. *et al*, 1996).

When compared to the reference *P. falciparum* 3D7 strain, the genome of *P. yoelii* has a high degree of synteny with respect to housekeeping genes that are predominantly located at the core body of the chromosomes. These include candidate drug targets such as dihydrofolate reductase, farnesyl transferase, dihydropteroate synthase, etc that could be targeted by specific inhibitors as well as evolutionarily conserved invasion-related antigens such as merozoite surface proteins, sporozoite surface antigen, apical membrane antigen 1, etc that could constitute potential candidate targets for vaccine development. Their conservation amongst the human and rodent parasite species suggests the involvement of specific invasion and metabolic processes that are conserved even in divergent species and provides a useful tool for *in vivo* studies on novel parasite-specific drugs/vaccines. Of note, synteny breaks between *P. falciparum* and *P. yoelii* occur almost exclusively to the rRNA loci (Gardner, M.J. *et al*, 2002; Carlton, J.M. *et al*, 2002) and thereby suggests that evolutionary phylogenetic changes between related malaria species are mainly due to synteny rearrangements at the rRNA

Introduction

regions. The availability of animal models would provide a means of assessing the bioavailability and pharmacokinetics of novel compounds in preliminary drug/vaccine trials and in areas of research into the development of resistance to existing drugs.

In contrast, the chromosomal regions responsible for antigenic variation and host immune evasion show the most divergence. The genome of *P. falciparum* contains species-specific subtelomeric genes involved in host cell invasion, adhesion and antigenic variation that are not found in the *P. yoelii* genome. For example, the *P. falciparum* genes that are located in the sub-telomeric regions include the *var*, *stevors* and *rifins* that are responsible for antigenic variation and hence immune evasion (Kyes, S. *et al*, 2001; Blythe, J.E. *et al*, 2003). In contrast, the *P. yoelii yir* gene family seems to be the largest multigene family found in the sub-telomeric regions and is absent in *P. falciparum*. Interestingly, this multigene family is common to all the rodent species and to *P. vivax* (Janssen, C.S. *et al*, 2002). A more recent study employing probabilistic modeling in conjunction with genomic organization and protein structure analysis bridges this gap and places the *P. falciparum rif/stevor* multigene family together into a conserved multigene superfamily of malaria parasites known as the *Plasmodium* interspersed repeats (PIRs) (Janssen, C.S. *et al*, 2004). This discovery further suggests that although genes are more conserved than previously thought; there is evidence of species specific divergence which is dependent on each species' interaction with the host immune system.

Recently, a review by Neil Hall and co-workers (2005) described the current status of the malaria sequencing projects with *P. falciparum* at 14.5x coverage, *P. yoelii* at 5x while *P. berghei* and *P. chabaudi* were at 4x coverage. While the *P. yoelii* sequence is not going to be improved, the Sanger Centre is currently aiming for at least an 8x

coverage for both *P. berghei* and *P. chabaudi* and the assembly and annotation of these two parasites species are in progress. Incidentally, the conservation of gene and amino acid sequences between the rodent parasites are above 80%. However, the average nucleotide sequences between the rodent parasites and *P. falciparum* are about 70% identical while the average amino acid sequence identity is around 60%. In short, 3,375 *Plasmodium* genes were found to be conserved between *P. falciparum* and *P. yoelii* while the differences could possibly arise from the diverse selection pressure from the host on genes that are responsible for host-parasite interactions.

2.6 A review on the recent Genomics and Proteomics studies

Genes in *Plasmodium* parasites are in a monocistronic organization and the first few promoters that have been identified are regulated by upstream enhancer elements (Horrocks, P. *et al*, 1998). However, only one element known as the G-box was identified upstream of several heat shock proteins and was conserved and functional across other species of the malaria parasite (Militello, K.T. *et al*, 2004). Other than this piece of information, little else is known about how the parasite regulates gene expression throughout its life cycle. With the advent of the genome sequence of *Plasmodium*, many studies on the parasite transcriptome and proteome have been done in order to elucidate the gene expression profile of the parasite.

An unexpected discovery lies in the paucity of metabolic and mitochondrial enzymes as compared to more primitive organisms like bacteria or yeasts (Gardner, M.J. *et al*, 2002) that either suggests an evolutionary adaptation or difficulties in gene prediction due to its AT-rich genome. Similarly, the proportion of transporters also

Introduction

seems to be low either due to the same reasons or perhaps the parasite just does not require that many different substrates for survival. Of interest, 10% of the predicted gene products are predicted to be imported into the apicoplast and are mainly responsible for important functions such as fatty acid and isoprenoid biosynthesis, iron metabolism and even membrane proteins, many of which are already drug targets (Waller, R.F. *et al*, 1998).

Classical methods for the determination of gene expression such as transient and stable transfections are time consuming. Thus, high-throughput genome wide transcriptome studies are especially valuable to identify, quantify and classify the expression of the predicted *P. falciparum* genes. Le Roch and co-workers (2003) characterized both mosquito and blood stages of the parasite life cycle and proposed that genes that were found to be correlated in terms of the levels and life cycle stage of the parasite had similar cellular function. Another study that followed the blood stage development of the parasite demonstrated the regulated transcription of stage-specific genes starting from general cellular processes like protein synthesis and finally to genes responsible for invasion (Bozdech, Z. *et al*, 2003). From this study, it was found that the induction of a particular gene occurred once per intra-erythrocytic cycle and only at a time when it was required.

Following the transcriptome studies, proteomic studies on the various stages of the parasite life cycle were performed. A complete understanding of the repertoire and relative amount of proteins that are expressed at a particular stage can only be fully understood with the proteome approach. Information from these data can be used to verify computational gene predictions and to re-annotate incorrectly predicted intro/exon

Introduction

junctions (Lasonder, E. *et al*, 2002). Unexpected data also came up with the proteomic approach. For example, the antigenically variant *var* and *rif* genes that are expressed on the surface of infected erythrocytes are also found in the proteome of the sporozoite stage of the parasite (Florens, L. *et al*, 2002). This suggests a novel survival mechanism for the non-replicative sporozoites. Functionally related genes have been shown to typically cluster in the yeast and human genomes (Cohen, B.A. *et al*, 2000; Caron, H. *et al*, 2001; Lercher, M.J. *et al*, 2002). Such ordered gene clustering has also been shown with genes involved in virulence and antigenic variation such as the *var*, *vir* and *rif* genes that are located in the sub-telomeric regions of the chromosomes (Hernandez-Rivas, R. *et al*, 1997; del Portillo, H.A. *et al*, 2001). This phenomenon is also seen in *P. falciparum* where ribosomal proteins, proteins involved in protein modification, and nucleotide metabolism were found in clusters on chromosome 3, 5 and 10 respectively. In addition, a cluster of four aspartic proteases in chromosome 14 were found to be expressed exclusively in the blood stages and not in the mosquito stages as these proteins function to degrade hemoglobin (Florens, L. *et al*, 2002). The occurrence of other co-expressed genes that often cluster and show highly ordered chromosomal organizations would provide a means to facilitate the search for stage-specific and cis-acting sequences and therefore yield more insight into the regulation of gene expression during the various stages of the parasite life cycle.

An important question to ask is whether the huge datasets from the microarray and proteomic studies correlate? On initial inspection, they seem to agree to a certain extent. Most genes show a high degree of correlation between mRNA and protein abundance and for those that do not, a common phenomenon was the accumulation of

Introduction

mRNA in the earlier stages of the parasite before the protein was detectable (Le Roch, K.G. *et al*, 2004). Another study has also found that the *Plasmodium* genome contains an unusually high proportion of proteins involved in mRNA stability and deadenylase activity. This, together with a significant lack of transcriptional regulators, would indicate that the effect of post-transcriptional regulatory controls throughout the different stages of the parasite is highly important and significant (Coulson, R.M. *et al*, 2004).

The relatively small repertoire of typical eukaryotic transcription factors in the face of tightly regulation transcription throughout the intraerythrocytic cycle of the parasite suggests that other factors or mechanisms are responsible for regulating transcription in the malaria parasite. To illustrate this disparity, a typical yeast has twenty-five to thirty genes under control of one transcription factor (Chervitz, S.A. *et al*, 1998; Aravind, L. *et al*, 2003) while in the case of parasitic apicomplexans, the range is between three hundred and fifty to eight hundred genes per detectable transcription factor (Aravind, L. *et al*, 2003). Bioinformatics analysis of the *Plasmodium falciparum* genome led to the identification of a novel protein (PF14_0633) that contains a small DNA-binding motif known as the AT-hook. A globular region was also discovered distal to the AT-hook motif and this was highly conserved in other *Plasmodia* species and to other apicomplexans such as *Theileria* and *Cryptosporidium* (Balaji, S. *et al*, 2005). A total of twenty-six members of this novel family of AP2-integrase DNA binding proteins (ApiAP2) were found in *P. falciparum* and twenty-two of these ApiAP2 proteins were expressed in different stages of the erythrocytic schizogony cycle. Further studies of these ApiAP2 proteins using protein-binding microarrays have led to the identification of their corresponding cis-regulating targets in the upstream region of developmentally regulated

Introduction

genes and also strongly implicates distinct subsets of these ApiAP2 members with regulation of gene expression in the intra-erythrocytic developmental cycle of the malaria parasite (de Silva, E.K. *et al*, 2008).

While post-transcriptional regulation, i.e. translational control has been postulated for the regulation of gene expression in *Plasmodia*, the molecular pathways have not been clearly demonstrated. Previous findings of distinct populations of ribosomes that dominate during different stages of development and differentiation (Waters, A.P. *et al*, 1989; McCutchan, T.F. *et al*, 1995; Li, J. *et al*, 1997) suggest some form of translational control in malaria parasites. It was only recently that the phenomenon of translational repression (TR) was demonstrated in the female gametocytes of *P. berghei*. A novel protein known as DOZI (development of zygote inhibited) is a homolog of a DDX6-family of RNA helicases that has been shown to exist in a complex with mRNA species that have been translationally repressed (Mair, G.R. *et al*, 2006). DOZI was thus postulated to be function in the storage of mRNAs in the female gametocyte that are required for the development of the zygote after fertilization. DOZI knock out clones were unable to store these developmentally required mRNAs and the obvious phenotype is the protracted development of the zygote as the required mRNAs were degraded. In addition, the target for DOZI was elucidated to be present in the 5' or 3'-UTR of the silenced *pb28* and *pb000245.02.0* genes and that a uridine-rich region of these DOZI-associated genes is implicated in TR (Braks, J.A. *et al*, 2008).

Thus, the malaria parasite employs a tightly regulated gene expression profile where stage-specific genes are transcribed just prior to when they are required, accumulates at a particular time point by which protein can then be found at detectable

levels. Hence, comparative studies based on phenotypic analysis can be inferred from both the proteomic and genomic analysis approaches.

The current concern with this incoming bulk of information is the verification of the annotated genes of the published parasite genomes. As the gene model for both *P. falciparum* and *P. yoelii* are both based on *in silico* gene prediction algorithms, they could still contain annotation mistakes such as over-prediction of genes (typically short ORFs) and missing legitimate genes due to the problems with an AT-rich genome. For example, there are orphan *P. yoelii* peptides identified by mass spectrometry that do not match to any annotated genes in the published genome sequence and many of them are partial hits, suggesting that intron-exon junction predictions are at fault (Carlton, J.M., *et al*, 2002). Since there is conserved synteny amongst the core genes, this presents an important tool for verification of cross-species gene models as a verified gene model from one species can be used as a template for verification of the corresponding gene in another species.

2.7 Tools for Proteomic investigations

The first generation of analytic tools for proteomics research has depended mainly on 2-D gel electrophoresis (2-DE) for separation of proteins according to their charge and mass and their subsequent identification via mass spectrometry (MS) (Barrett, J. *et al*, 2000). However, the choice of staining methods usually restricts the dynamic range of detection. The use of cyanine dyes to directly label lysine residues in 2-D fluorescence difference gel electrophoresis (2-D DIGE) enables different samples to be run on the same gel and has the advantages of eliminating inconsistencies in matching different gels

and also provides a means for fluorescence-based quantitation (Biron, D.G. *et al*, 2005). More recently, the isotope-coded affinity tag (ICAT) method and a fluorescence-based method (ITRAQ) can be used to label proteins prior to their separation by liquid chromatography (LC). Similarly to 2-D DIGE, different samples that have been distinctively labeled can be combined and analyzed by mass spectrometry (MS). The gel electrophoresis and liquid chromatography methods described above generally do correlate in respect of protein quantification. However, the 2-D/MS method seems to detect more hydrophobic proteins and lower molecular weight proteins whereas the LC/MS method gave better results with higher molecular weight proteins (Schmidt, F. *et al*, 2004). Hence, complementary use of both methods could prove useful while collaborations among research groups can increase accessibility to expensive equipment required for such analyses.

2.8 Rodent models for malaria research

The use of rodent malaria parasites as models for the human disease has been used extensively. Most of the laboratory strains have been isolated from the African thicket rat, *Grammomys surdaster* (Carter, R. and Diggs, C.L., 1977) and the shiny thicket rat *Thamnomys rutilans* (Landau, I and Boulard, Y., 1978). The detailed origins of the rodent malaria parasites are illustrated in Figure 2.6 and these parasites were predominantly found in the Central African Republic (Killick-Kendrick, 1978).

Plasmodium berghei was discovered in the blood of the thicket rat *Grammomys surdaster* in the Democratic Republic of Congo (Vincke and Lips 1948), now currently known as Zaire. The initial ANKA isolate was from the *Anopheles dureni millecampi*

Introduction

mosquito caught by Vinke and Bafort (Killick-Kendrick, 1974) and the initial stabilates from Killick-Kendrick were cloned in Edinburgh, sent out to Antwerp and then returned to Edinburgh where multiple deep freezes and mice passages were performed before the final stabiliate was made (Saul, A. *et al*, 1997). *P. berghei* ANKA preferentially invades reticulocytes and is typically lethal at parasitemias not exceeding 20% due to complications resulting from sequestration. The erythrocytic schizogony stage is twenty-one hours and is asynchronous.

Figure 2.6. Foci of infection of rodent malaria and the species of rodent *Plasmodia* isolated from this region. (Adapted from Landau, I. and Gautret, P., 1998).

Introduction

Plasmodium chabaudi was first isolated from the blood of a shiny thicket rat *Thamnomys rutilans* in the Central African Republic (Landau, I., 1965). The initial isolate AS (399BY) was cloned and the derivative clone AS(sens) was sensitive to pyrimethamine and chloroquine (Carter, R. and Walliker, D., 1975). *P. chabaudi* invades erythrocytes of all ages and the erythrocytic schizogony cycle lasts twenty-four hours and is tightly synchronous.

Plasmodium yoelii was discovered in the infected blood from the shiny thicket rat *Thamnomys rutilans* in La Maboke Field station in Central African Republic (Landau, I. and Chabaud, A.G., 1965). The initial isolate is termed 17x and was propagated in mice of different genetic backgrounds. Incidentally, the virulent YM line was obtained after thawing the original 17x isolate (Yoelii, M. and Hargreaves, B.J., 1975). Sub-clones were then obtained from this YM line (Knowles, G., and Walliker, D., 1980) and one of the virulent clones was termed YM and is deposited in MR4. Similarly, one avirulent clone was subsequently obtained and termed YA. *P. yoelii* has a high preference for invading reticulocytes, with an erythrocytic schizogony of only eighteen hours and infections are asynchronous.

It was also concurred that species that are naturally asynchronous, i.e. *P. berghei* and especially *P. yoelii* are more resistant to chloroquine and their merozoites are more persistent in circulation than synchronous species such as *P. chabaudi* (Beaute-Lafitte, 1994). While merozoite latency can explain the asynchronous erythrocytic schizogony of *P. berghei* and *P. yoelii*, this study elegantly linked the definition of speciation with drug resistance and merozoite persistence.

Introduction

Different species of rodent parasites exhibit different virulence phenotypes. For example, *Plasmodium berghei* K173 and *P. vinckei* and some strains of *P. chabaudi* and *P. yoelii* cause high parasitemias and are lethal in mice whereas infections with *P. yoelii*, *P. chabaudi chabaudi*, *P. chabaudi adami* and *P. vinckei petterei* typically do not cause high parasitemias and are eventually cleared (Table 2; Sanni, L.A. *et al*, 2002). Since they mirror the phenotype of human malaria species, rodent models have become extremely valuable as tools for malaria research as the principles of immune selection on the blood-stage parasites and the pathological effects of infections are similar in mice and humans.

In addition, the genomic data show a high degree of conservation between different *Plasmodium* species with the exception of the telomeric and subtelomeric regions that are responsible for species specific antigenic variation and evasion of the host immune surveillance (Waters, A.P., 2002; Carlton, J.M. *et al*, 2002). These differences could be harnessed to tie in a particular phenotype with a particular complement of genes. In addition, orthologues have been discovered in many distant species, thus the use of rodent models today is highly relevant especially with the need for *in vivo* transfection analysis experiments (Franke-Fayard, B. *et al*, 2004; Khan, S.M. *et al*, 2005) that cannot be performed in humans. In addition, the complete life cycle of the rodent parasites can be accomplished in labs with accredited insectariums (Carter, R. and Diggs, C.L. 1977) so as to generate sporozoites for the investigation of hepatic stage parasites as well as for the elucidation of ookinete interaction with the mosquito midgut (Ecker, A. *et al*, 2008). Sexual recombination between different parasite lines can also be accomplished using mosquitoes with their progeny inheriting traits in a Mendelian fashion (Walliker, D. and Sanderson, A., 1976).

Plasmodium	Strain/clone	Mouse strain	RBC infected	Lethal infection	Experimental model
<i>berghei</i>	ANKA	CBA/T6	Mature RBC and reticulocytes	Yes	Pathogenesis of cerebral malaria
		BALB/c			
		C57BL/6			
<i>yoelii</i>	K173	CBA/2J	Mature RBC and reticulocytes	Yes	Resolving cerebral malaria model
		CBA/T6			Non cerebral malaria control
		BALB/c			
	17x	CBA/2J	Mature RBC and reticulocytes	Yes	Pathogenesis of cerebral malaria
		C57BL/6			Pathogenesis of cerebral malaria
		Swiss	Reticulocyte	Yes	
		BALB/c			
	YM	CBA/Ca	Reticulocyte	No	Immune mechanisms
		BALB/c			Pathogenesis
		C57BL/6			
<i>vinckei vinckei</i> <i>chabaudi chabaudi</i>	AS	DBA/2J			
		C3H	Mature RBC and reticulocytes	Yes	Vaccine
		Swiss			
	CB	BALB/c	Mature RBC	Yes	Chemotherapy, immune mechanisms, pathogenesis
		C57BL/6	Mature RBC	No	Immune mechanisms
		NIH			
		CBA/Ca			
	556 KA	B10 series			
		BALB/c	Mature RBC	Yes	Immune mechanisms
		A/J DBA/2J	Mature RBC	Yes (up to 50%)	Immune mechanisms
<i>chabaudi adami</i> <i>vinckei petteri</i>	CR	BALB/c	Mature RBC	No	Immune mechanisms, pathogenesis
		C3H			
		C57BL/6			
<i>vinckei petteri</i>	CR	C57BL/6	Mature RBC	No	Chemotherapy
		BALB/c			Immune mechanisms

Table 2.2 .Tabulated relationship between the different rodent malaria parasite strains and the typical experimental outcome in specific rodent host strains. (Adapted from Sanni, L.A. *et al*, 2002)

Based on the genomic and phenotypic similarities, rodent models are invaluable tools in malaria research.

2.9 Virulence factors of *P. yoelii*

It has been established that newly released merozoites are capable of transient attachment to the surface of a new host cell. This initial contact need not be restricted to the apical end of the merozoite but instead can even involve any surface of the merozoite (Dvorak, J.A. *et al*, 1975). Based on this finding, research on merozoite surface proteins (MSPs) have gathered pace because inhibiting such interactions would certainly prevent the establishment of the intraerythrocytic cycle and thus the ultimate holy grail of malaria eradication. There are 10 characterized MSPs and perhaps the best studied amongst them is MSP-1 that is linked to the merozoite membrane by glycosyl-phosphatidylinositol (GPI) membrane anchors (Gerold, P. *et al*, 1996). The structure of MSP-1 consist of two epidermal growth factor (EGF)-like domains (Babon, J.J. *et al*, 2007) and these domains could potentially mediate protein-protein interactions. MSP-1 is synthesized at the schizont stage as a ~200kDa GPI-linked precursor and is then directed to the surface of merozoites (Gilson, P.R. *et al*, 2008). After localization to the merozoite surface and prior to schizont rupture (late schizont stage), this huge protein is then proteolytically cleaved into the 83kDa (N-terminus), 30 and 38 kDa (central region) and 42 kDa (C-terminus) fragments (Holder, A.A. *et al*, 1987). During invasion, the C-terminal 42kDa fragment is then cleaved to yield a 33kDa soluble fragment and the remaining 19kDa fragment remains bound to the merozoite surface (Blackman, M.J. *et al*, 1991). MSP-1 was also shown to be present in *P. yoelii* where it was localized to the merozoite surface (Oka, M.

Introduction

et al, 1984) thus providing a potential rodent model for further investigations. A link for its role in mediating invasion was shown when MSP-1 was identified as a potential vaccine candidate since inhibitory antibodies prevent processing of the 42kDa fragment (Blackman, M.J. *et al*, 1994) while immunization with this same 42kDa protein offers protection against an *in vivo* challenge (Singh, S. *et al*, 2003). Recently, a post-invasion function of the 19kDa MSP-1 fragment was shown to be involved in establishing the formation of a food vacuole in the early ring stages (Dluzewski, A.R. *et al*, 2008), hence elucidating its role in the establishment of a new infection.

After initial attachment mediated predominantly by the merozoite surface proteins, the apical pole of the merozoite is usually not in contact with the erythrocyte and must somehow reorientate itself as invasion can only occur at the apical pole. The apical membrane antigen-1 (AMA-1) in *P. falciparum* was identified as a potential virulence factor because the 83kDa protein initially accumulates in the rhoptries and is then proteolytically processed into a 66kDa form that relocates to the merozoite surface (Narum, D.L. and Thomas, A.W., 1994). It is also a vaccine candidate as antibodies proved successful in inhibiting merozoite invasion of new host cells (Hodder, A.N. *et al*, 2001). This protein is also highly conserved amongst the human and rodent parasites (Waters, A.P. *et al*, 1990) and is able to directly bind erythrocytes with the elucidation of a specific minimal binding domain (Fraser, T.S. *et al*, 2001). Quite recently, it was shown that addition of a monoclonal antibody directed against AMA-1 still allowed merozoites to bind to erythrocytes but these merozoites failed to reorientate their apical ends and were thus unable to invade (Mitchell, G.H. *et al*, 2004). Further evidence for the role of AMA-1 in invasion was provided when it was shown in *P. falciparum* that the gene

Introduction

cannot be knocked out but function can be complemented by a corresponding AMA-1 gene from *P. chabaudi*, thus indicating a conserved role in invasion even in more divergent species (Triglia, T. *et al*, 2000).

Perhaps the best known virulence factor in *P. yoelii* is the 235kDa rhoptry protein that is homologous to reticulocyte-binding protein-2 of *P. vivax* (Sinha, K. A. *et al*, 1996). PY235 is part of the reticulocyte binding like (RBL) protein family and this family also includes members in *P. falciparum* (Rayner, J.C *et al*, 2000; Rayner, J.C. *et al*, 2001; Triglia, T. *et al*, 2001; Taylor, H.M. *et al*, 2002). This multigene family, consisting of at least 14 members, has been implicated in host cell selection and immune evasion. In passive antibody transfer experiments, monoclonal antibodies against PY235 were able to modulate the infection and restrict invasion of parasites to reticulocytes (Freeman, R.R. *et al*, 1980); thus limiting the parasitemia and the ultimate survival of the host animal. A further example of its biological role has been shown where at least one of the members has been shown to bind specifically to erythrocytes (Ogun, S.A. and Holder, A.A., 1996; Ogun, S.A. *et al*, 2000). Incidentally, multiple members of *py235* transcripts were detected in a single schizont, with each merozoite expressing a distinct member of this multigene family (Preiser, P.R. *et al*, 1999). This offers a unique type of antigenic variation in which alternative forms of the protein are immediately made available at each round of asexual replication when the schizont ruptures and releases clonally distinct merozoites. It has also been shown that high levels of PY235 is consistent with increased virulence (increased host cell repertoire) while lower levels are expressed as a mechanism to evade host immunity and consequently leads to a decrease in virulence (decreased host

Introduction

cell repertoire) (Iyer, J.K. *et al*, 2007). This finding thus further provides a direct correlation of PY235 expression to host cell repertoire and hence virulence *per se*.

The *pir* (*Plasmodium* interspersed repeats) variant multigene superfamily is thought to be conserved in all *Plasmodia* (Janssen, C.S. *et al*, 2004) and is implicated in antigenic variation. Classical antigenic variation strategies commonly employed by pathogens expose only one or a few genes in the multigene family to the immune surveillance of the host. Hence, the pathogen is able to replicate until the host immune system catches up with the current repertoire of variant antigens and then the cycle continues where other variants now dominate, thus ensuring the chronicity of the infection, survival of the pathogen and hence increased opportunity for the pathogen to be transmitted to another host.

The presence of this *pir* multigene family in the rodent parasite species offers a golden opportunity to investigate its antigenic variation property. In *P. yoelii*, the *yir* multigene family is probably the largest of all *Plasmodia*, with its genome containing 838 *yir* genes (Carlton, J.M. *et al*, 2002). The protein is expressed on the surface of infected schizont-stage erythrocytes and its antigenic variation property was elegantly proven when a typical sub-set of *yir* genes were found to be differentially transcribed upon re-infection of mice (Cunningham, D.A. *et al*, 2005). More recently, it was discovered that not all *yirs* were transcribed in the blood stages, suggesting that a certain group of genes could possibly be involved in other stages and/or functions (Fonager, J. *et al*, 2007). Also, a unique and highly conserved *pir* specific promoter was also discovered and is common for all the members of this superfamily. Thus variable expression of the *pirs*

would possibly involve epigenetic silencing/activation and/or post-transcriptional regulation.

2.10 Challenges involved in rodent malaria research and the expected differences between the virulent and avirulent parasite lines

It is currently difficult to navigate through the genomes of the rodent malaria parasites. Sequence quality and coverage of *P. berghei*, *P. chabaudi* and *P. yoelii* are fraught with gaps and are poorly annotated. In the case of *P. chabaudi*, the over-prediction of gene models has also complicated attempts to understand and draw clear orthologous relationships to genes in other malaria parasites. The use of comparative genome microarray hybridization experiments as well as bioinformatics tools can potentially address these issues that can help re-annotate genes and reduce gaps in the rodent genomes. Species-specific genes can also be identified and these could address the different physiological and biochemical traits of individual parasite species.

To complement the qualitative data, quantitative comparison of the transcriptional activity of genes in the related virulent YM and avirulent YA lines of *P. yoelii* can be performed. By virtue of ‘guilt-by-association’, genes that are more abundantly transcribed in YM than YA could potentially be linked to virulence and these can be assayed via microarray and validated via real-time RT-PCR. These genes can be further linked to virulence with high confidence by assaying the progeny of genetic crosses of the virulent and avirulent parental stains of *P. yoelii*. For example, genes that were transcribed more abundantly in YM than in YA would also be expected to be more abundantly transcribed in the virulent progeny lines than the avirulent progeny lines.

Introduction

The ultimate goal of identifying and differentiating genes related to virulence and not just genes that are casually associated with virulence is for functional gene knock out studies. If these knock-out parasites can be shown to attenuate their invasiveness then it is extremely likely that that gene in question is directly related to virulence.

In order to assess and quantify virulence, this property has to be defined. Virulence can be defined as the range of host cells that the parasite can invade and is thus an important factor that is responsible for the morbidity and mortality of the disease. Parasite growth curves in an experimental rodent host and end-point experiments leading to the survival or death of the rodent host are useful tools to quantify virulence. A statistical measure of host cell selectivity known as the 'Selective Index' (SI) measures the parasite's potential to preferentially invade reticulocytes (Simpson, J.A. *et al*, 1999). This is based on the observation that reticulocyte-restricted parasites typically run out of suitable host cells and multiple merozoites tend to invade the same cell. Avirulent parasites such as *P. vivax* are reticulocyte-restricted and patients typically exhibit low parasite loads. Therefore, SI provides a quantifiable measure of host cell selectivity, with virulent parasites having a low SI (not reticulocyte restricted) while avirulent parasites having a significantly higher SI (Iyer, J. *et al*, 2007). Clearly, there must be differences in the recognition process between the merozoite ligands and the host cell receptors before they commit to the irreversible invasion process. Currently, two ligand families have been shown to be involved in invasion. The Duffy-binding-like (DBL) Superfamily consists of the *P. vivax* Duffy-binding protein (PvDBP) and its homologs. This group also consists of the erythrocyte-binding-like (EBL) proteins. Another family known as the reticulocyte-binding-like (RBL) proteins consists of the *P. vivax* reticulocyte binding

Introduction

protein and its homologs. The combined activity of proteins from the two families are thought to be responsible for invasion as shown by *P. falciparum* exhibiting alternative invasion pathways based on the repertoire of erythrocyte receptors as defined via enzymatic treatment and erythrocytes that are deficient of a particular receptor. As reviewed by Deepak Gaur and colleagues (2004), receptor redundancies confer immune evasion as other ligands can function in place of the immune targeted ligand. On another note, different ligands could define the ability of the parasite to invade erythrocytes of all ages. Interestingly, maturing reticulocytes continually lose plasma membrane proteins and receptors until they ultimately adopt the bi-concave shape as seen in mature normocytes (Koury, M.J. *et al*, 2005; Blanc, L. *et al*, 2005). Hence the repertoire of receptors available for the parasite to interact will be very much different and the reticulocyte restricted *P. yoelii* YA strain would possibly possess different proportions and/or composition of ligands as compared to the unrestricted YM strain. In addition, young reticulocytes that have just differentiated from erythroblasts by nuclear expulsion still contain lots of functional organelles such as the mitochondria and are able to manufacture hemoglobin and ATP (Geminard, C. *et al*, 2002). As compared to the rather metabolically inactive mature erythrocytes, the reticulocyte environment is certainly more enriched and as such, YA could have adjusted its metabolic requirements in order to benefit from the extra nutrients. Thus for YA, the repertoire of metabolic enzymes could be different from that of YM.

In order to establish a common ‘core’ set of *Plasmodia* genes and to improve on the coverage and annotation of the rodent malaria parasite genomes, comparative genomic microarray hybridizations of the rodent parasites will be performed. Via

Introduction

orthology, unexpected hybridizations would reveal novel orthologous genes that are either due to missing sequence information or unannotated genes missed by current gene prediction algorithms. In addition, bioinformatics tools can be employed to complement this process. In order to elucidate the repertoire of rodent parasite specific genes, genes that are orthologous to *P. falciparum* can be removed and the remaining gene set thus fulfils this requirement. In order to identify hypothetical genes that may function as novel parasite ligands, variant hypothetical genes from different *P. yoelii* lines that contain at least a single transmembrane domain were identified as putative targets for therapeutics as they are highly likely to be under immune selection pressure.

To identify novel genes associated with virulence, transcriptomic differences between the virulent YM and avirulent YA lines of *P. yoelii* will be analyzed via microarray and confirmed via real-time RT-PCR. Genes that were more abundantly transcribed in the YM and other virulent *P. yoelii* lines were highly likely to be implicated in virulence. A selection of these genes will then be short-listed for functional knockout studies to elucidate the phenotypic effects of such deletions.

3.1 Propagation of rodent parasites in mice

Balb/c mice (Laboratory Animals Centre, The National University of Singapore, Singapore) were infected with cryo-preserved parasite stocks or by syringe passage from a pre-existing infected mouse. Infections were monitored daily via blood smears fixed in methanol (Fluka, USA) in Giemsa's solution (Sigma-Aldridge, USA). *Plasmodium yoelii* infections were typically allowed to reach 20% parasitemia before the animals were sacrificed.

For typical expansions of parasites for use in molecular biology, infections were via intra-peritoneal (IP) injections from liquid nitrogen preserved stabilates. For studies on virulence and parasite growth in the animal, 1,000 infected erythrocytes were obtained from an infected donor mouse at 1- 4% parasitemia. At this parasitemia range, even the asynchronous non-reticulocyte restricted species (e.g. *P. yoelii*) were relatively synchronous due to their initial preference for reticulocytes. An inoculum of 1,000 parasites per recipient mouse was chosen via the intra-venous (IV) route as this dose was sufficient to administer with a high degree of confidence. Growth curves for this dose were not too high to overwhelm the mice and also enabled the study of host-parasite interactions, as we were able to track the parasite multiplication rate at low parasitemias where the parasites are forced to switch to normocytes.

3.2 Establishment of frozen parasite stocks

Infected mice were bled into 250ul Kreb's-Henseleit modified buffer (Sigma-Aldridge, USA) (Appendix 1) and heparin (25U/ml) and kept on ice. The blood was then centrifuged (Eppendorf, Germany) at 1,000g for 10min at 4°C. The supernatant was

removed and the packed cell volume estimated. Twice the volume of parasite freezing medium (Appendix 1) was added to the pellet by first adding 1/5 of the freezing medium drop wise with gentle mixing. The cells were left to stand for 5min and the remaining freezing medium added slowly with mixing. Aliquots of 250 μ l were put into each freezing vial (Nunc, USA) and were first placed in -80°C for 24hrs before being transferred to liquid nitrogen.

3.3 Parasite harvest and purification

Mice were bled into 250 μ l Kreb's-Henseleit modified buffer (Sigma-Aldridge, USA) (Appendix 1) and heparin (25U/ml) (Sigma-Aldridge, USA). The plasma was removed by swing-out centrifugation (Sigma, Germany) at 2,100rpm for 5min at ambient temperature and the cell pellet was washed with 10ml of RPMI 1640 (Invitrogen, USA) and re-spun as above. The cell pellet was then re-suspended to the original blood volume with fetal calf serum (FCS) (Invitrogen, USA).

An 80% to 50% discontinuous Histodenz gradient (Appendix 1) was layered and up to 2ml of parasitized blood in FCS layered on top of this gradient and centrifuged at 2,100rpm for 20min at 21°C. Schizonts appear at the FCS/50% interface, trophozoites appear at the 50%/60% interface, rings at the 60%/70% interface while uninfected erythrocytes are pelleted down at the bottom of the centrifuge tube. Thus, stage-specific parasitized erythrocytes can be isolated by pipetting out the interface of interest and the residual gradient material removed by washing with 10ml of RMPI 1640 as described above.

If merozoites are required, 2ml of re-suspended blood in FCS was then loaded onto a 50% Histodenz gradient (Appendix 1) and centrifuged at 2,100rpm for 20min at 21°C. The parasitized erythrocyte layer at the load/gradient interface was removed, diluted with Kreb's-Henseleit modified buffer to 10ml and centrifuged at 2,100rpm for 10min at 10°C.

The schizonts were then re-suspended in 10ml of RPMI 1640 and loaded onto an Amicon ultra-filtration cell (Millipore, USA) that contains a 2µm pore-sized IsoporeTM polycarbonate membrane (Millipore, USA). This cell chamber was then allowed to circulate on a magnetic stirrer platform (IKA Werke, Germany). The schizont suspension was allowed to pass through the filter by applying sufficient nitrogen gas pressure (Soxal, Singapore). The schizonts will then rupture and the released merozoites were collected in the effluent. Merozoites were collected by centrifugation at 1,000g for 8min in a swing-out centrifuge or in an eppendorf tube centrifuge (Eppendorf, Germany) for 5s at maximum speed. Merozoites were then washed once in phosphate buffered saline (PBS) (Appendix 3) as above and in 10mM Tris-HCL (Sigma-Aldridge, USA), pH7.4, to remove erythrocyte-derived β-haematin (Nirmalan, N. *et al*, 2004). The washed merozoite pellet was then stored at -80°C.

3.4 Preparation and spotting of poly-L-lysine slides for microarray

The following describes a protocol for coating 200 slides with poly-L-lysine. Gold Seal Microslides (Erie Scientific, USA) were inserted into slide racks (Shandon Lipshaw, USA) and up to 4 racks were placed into a 2L glass beaker. Next, wash solution (Appendix 4) was added to each 2L beaker and each beaker was covered in foil and

shaken gently at 60rpm for 4hrs. Next, the solution was decanted and the slides rinsed twice with distilled water followed by twice more with Milli-Q water. After which, the washed slides in the 4 racks were transferred into a 2L plastic beaker containing the poly-L-lysine solution (Appendix 4) and shaken at 60rpm for 45min. Next, decant the poly-L-lysine solution and rinse the racks with distilled water followed by Milli-Q water as described above. Finally, a further rinse with Milli-Q water was performed and beaker containing the submerged racks was carried to the centrifuge (Eppendorf, Germany) and the racks spun down at 600rpm for 5min to remove excess water. The slides were then transferred to a drying oven (Mettler, Germany) at 50°C for 20min to ensure that they are properly dried. The slides can then be stored in a clean plastic slide holder box for at least 14 days before the oligonucleotides can be spotted.

Oligonucleotides (Invitrogen, USA) were dissolved in 3x SSC (Appendix 3) at a concentration of 40 μ M for spotting on poly-L-lysine-coated glass slides using the VersArray ChipWriter™ Pro system (Bio-Rad, USA). The printed slides can be stored in a clean plastic slide holder at room temperature in a dessicator.

3.5 Oligo array post-processing

Recently printed slides were labeled with a sticker on opposite side of the slide for record purposes. Firstly, a slide warmer (Cole-Parmer, USA) was set to 45°C and a container filled with 3x SSC was allowed to warm up. Slides were placed oligo side down to be rehydrated by the evaporating buffer for 1.5min and then dried oligo side up on a heat block (Biometra, Germany) at 99°C till the condensation was removed. Next, the oligonucleotides were UV cross-linked in a Stratalinker (Stratagene, USA) by

Materials and Methods

exposing them to 80mJ of UV energy. The fixed slides (in slide rack) were then washed in a glass slide dish with a solution containing 3x SSC and 0.2% SDS that was first pre-warmed to 65°C. The slides were gently agitated for 5min and then rinsed with distilled water in another slide dish for 1min. This was followed by a final rinse in 95% ethanol in another slide dish and then spun down at 600rpm for 5min to dry.

Blocking solution (Appendix 4) was then poured into clean dry glass slide dish and the slides were vigorously plunged in and out of the blocking solution for 30s before letting them submerge in the slide dish and left on a horizontal shaker (Firstek, Taiwan) for 5min at 60rpm. Next, the slides were drained and transferred to a 4L beaker filled with distilled water such that the rack is submerged and then ‘swished’ in a circular wrist motion gently back and forth for a few seconds before resting under water for 1min. The slide rack was then removed from the water and plunge mixed in the same blocking buffer as described above before being placed into a new glass slide dish containing distilled water and plunge mixed for 2min. Following which, the slides were transferred to another glass slide dish of 95% ethanol and plunge mixed again for another 2min.

The submerged slides were then carried to a tabletop centrifuge (Eppendorf, Germany) and the racks spun down at 84g for 5min to remove excess water. The slides were then transferred to a drying oven (Mettler, Germany) at 50°C for 20min to ensure that they are properly dried. The slides can then be stored in a clean plastic slide holder box.

3.6 Genomic DNA extraction from blood samples

Firstly, host leukocytes were removed to eliminate potential interfering nucleic acids from downstream analysis. A Plasmodipur filter (Euro Diagnostica, Netherlands) that was designed to trap leukocytes was first equilibrated with 10ml of RPMI 1640 (Invitrogen, USA) via a 20ml syringe (BD Biosciences, USA). Next, total blood was diluted with at least an equivalent volume of RPMI 1640 and was allowed to pass through the filter and the filtrate collected. After which, 20ml of RPMI 1640 was then passed through the filter and filtered blood collected in a 50ml centrifuge tube (BD Biosciences, USA) together with the previous filtrate and centrifuged at 2100rpm for 5min. The supernatant was discarded and the erythrocyte pellet retained for nucleic acid extraction.

The Easy-DNA™ kit (Invitrogen, USA) was used for genomic DNA extraction. Briefly, 500µl of the kit's Solution A was added to 350µl of blood in a 2.0ml eppendorf tube and mixed by inverting several times. The mixture was incubated on a heat block (Biometra, Germany) at 65°C for 6min for cell lysis. Next, the tube was mixed by several inversions and 900µl of chloroform (Fisher, USA) was added and vortexed to mix. After which, 200µl of the kit's Solution B was added and the tube vortexed briefly till the mixture is uniformly viscous. The samples can then be centrifuged in a tabletop centrifuge (Eppendorf, Germany) at 16,000g for 10min at room temperature to allow for phase separation and the top aqueous layer was transferred into a fresh 2.0ml eppendorf tube.

DNA was precipitated by adding 1ml of room temperature ethanol (Merck, USA) and mixed by inversion till a precipitate is seen. The DNA was then pelleted by

centrifugation at 16,000g for 5min at room temperature. The supernatant was discarded and the DNA washed with 1ml of 70% ethanol before centrifuging at 13,000rpm for 2min at room temperature. The supernatant was discarded and the pellet was air-dried before 100µl of nuclease-free water was added to resuspend the pellet. The DNA was incubated at 65°C for 5min to allow for complete solubilization of the DNA before storage at 4°C or -20°C.

3.7 Total RNA extraction

Stage specific parasitized erythrocytes were obtained as described (Section 3.3) and contaminating host leukocytes removed (Section 3.6). The cell pellet (typically ~ 1ml) was flash frozen in liquid nitrogen and stored at -80°C. When required, they were thawed out in a 65°C water bath (Mettler, Germany) and then lysed with 10ml of TRIZOL[®] (Invitrogen, USA) with repetitive pipetting on ice. The samples were then incubated at room temperature for 5min to promote the dissociation of nucleoprotein complexes. Next, 2ml of chloroform (Fisher, USA) was added and the tubes shaken vigorously for 15s and then allowed to stand at room temperature for 3min. Phase separation was carried out by centrifuging (Eppendorf, Germany) the samples with low brake at 12,000g at 4°C for 15min. The top aqueous phase was then transferred into a fresh 14ml polypropylene round-bottom ultra centrifuge tube (Nalgene, USA) and 10µg of RNase-free glycogen (Invitrogen, USA) was added as a carrier to enhance yield. Next, 5ml of isopropanol (Merck, USA) was added and the tube mixed by inversion and allowed to stand at room temperature for 10min before centrifuging at 12,000g for 10min at 4°C (Beckman, USA).

The supernatant was discarded and the RNA pellet washed with 10ml of 75% ethanol by vortexing gently and centrifuged at 7,500g for 5min at 4°C. The supernatant was then carefully decanted and the pellet allowed to air dry and then resuspended in 100µl of RNases-free water (Invitrogen, USA) before storage at -80°C.

3.8 DNase Treatment of RNA samples prior to reverse transcription

Removal of genomic DNA contamination from purified total RNA was performed with the TURBO DNA-free Kit (Ambion, USA). Basically, 4µg of total RNA was treated in a 0.5ml eppendorf tube (Axygen, USA) with 2µl of 10X TURBO DNase Buffer and 1.5µl of TURBO DNase and topped up with RNase-free water (Invitrogen, USA) to a final volume of 20µl. This mixture was incubated on a thermocycler (Eppendorf, Germany) at 37°C for 15min and then topped up with an additional 1.5µl of TURBO DNase for another 15min of incubation. Next, 2µl of the resuspended DNase Inactivation Reagent was added and mixed by pipetting. The solution was then allowed to stand at room temperature for 2min and then centrifuged at 10,000g for 1.5min to pellet down the DNase-inactivation slurry complex. The supernatant that contains the RNA was then aspirated carefully.

3.9 Klenow reaction for genomic DNA microarray

Genomic DNA (3µg) was used as a template and mixed with 2.5µg of random nanomer primers (Invitrogen, USA) and topped up to a final volume of 15.25µl with molecular biology grade water (Eppendorf, Germany). The mixture was incubated at 100°C for 5min and snap cooled on ice for 5min. Next, 5µl of Buffer 2 (New England

Biolabs, USA) and 1.5µl of 10x aa-dUTP dNTPs (Appendix 4) were added followed by 4µl of Klenow Fragment (3' – 5' exo⁻, New England Biolabs, USA) and the reaction tube topped up to 50µl with molecular biology grade water. The mixture was allowed to stand at room temperature for 10min and then incubated overnight at 37°C.

The reaction was cleaned-up with the Qiagen MinElute PCR purification kit (Qiagen, Germany). Briefly, 10 volumes of Buffer PBI were added to 1 volume of sample to be cleaned-up. This mixture was then loaded onto the provided column and centrifuged at 16,000g for 1min. Next, the flow through was discarded and 750µl of Buffer PE was added to the column and allowed to stand for 5min. The buffer was removed by centrifugation at 16,000g for 1min and discarding the flow through followed by another round of centrifugation to completely remove the buffer. DNA was then eluted by adding 12µl of Buffer EB, letting the column stand for 2min before centrifuging at 16,000g for 2min.

3.10 First strand cDNA synthesis for microarray

In a 0.2ml flat cap PCR tube (Axygen, USA), 12µg of total RNA was added together with 2µg each of random nanomer primer (N₉) and oligo-dT primer (T₂₅) (Sigma-Aldridge, USA) to a volume of 30.5µl. This was then incubated at 70°C for 10min and snap cooled on ice for 10min. Next, 1.5µl of 30x aa-dUTP dNTPs (Appendix 4), 9µl of 5x reaction buffer (Fermentas, USA), 1µl of RNaseIN (Promega, USA) and 1.5µl of RevertAidTM H Minus M-MuLV reverse transcriptase (Fermentas, USA) were added and topped up with molecular biology water to a final volume of 45µl. The completed mix is then incubated at 42°C for 2hr and the reaction stopped and RNA

hydrolyzed by adding 10 μ l each of 0.5M EDTA (Sigma-Aldridge, USA) and 1.0M NaOH (Appendix 3) with incubation at 65°C for 15min.

The newly synthesized amino-allyl cDNA can then be purified using the Qiagen MinElute PCR purification kit as described above (Section 3.9).

3.11 Cy Dye Coupling

A vial (40nmol) of the appropriate Cy dye (Ge Healthcare, USA) was thawed, spun down and reconstituted with 25 μ l of DMSO (Sigma-Aldridge, USA). To each tube of purified aminoallyl incorporated nucleic acids (12 μ l), they were first incubated at 100°C for 5min and then snap cooled on ice for 5min. Next, Cy dye coupling was performed by the addition of 1 μ l of 0.5M sodium bicarbonate, pH9.0 (Appendix 3) and 2 μ l of reconstituted Cy dye. The remaining dye can be aliquoted in 2 μ l portions, dried down in a speed vac (Eppendorf, USA) and stored at -20°C for further use. This coupling mixture is then allowed to incubate at room temperature in the dark for 2hr.

Excess unincorporated dye was removed using the Qiagen MinElute PCR purification kit as described above (section 3.9) except that elution is now done with 13 μ l of buffer EB.

3.12 Microarray hybridization and analysis

Both reference (Cy3 labeled) and experimental (Cy5 labeled) samples were combined to a total of 26 μ l. To that, 6.4 μ l of 20x SSC (Appendix 3), 0.9 μ l of 10% SDS (Sigma-Aldridge, USA), 1 μ l of HEPES buffer, pH 7.5 (Appendix 3) and 10.7 μ l of ddH₂O. The mixture was heated to 100°C for 5min and then snap cooled on ice for 5min.

Hybridizations were carried out on a MAUI hybridization system (BioMicro, USA) at a temperature of 65°C for 16 hours.

After hybridization, the slides were removed from the MAUI station and placed in a slide rack that is sitting inside a glass slide dish containing a buffer of 340ml of ddH₂O, 10ml of 20x SSC and 1ml of 10% SDS. The slides were gently agitated for 5min and then the rack was transferred to another glass slide dish containing 349ml of ddH₂O and 1ml of 20x SSC and agitated gently for 5min. Next, the slide dish was carried to a tabletop centrifuge and dried by centrifuging at 84g for 5min.

The dried microarrays were scanned on the GenePix 4000B scanner and the images analyzed using GenePix Pro 3.0 (Axon Instruments, USA). The subsequent data was then normalized using NOMAD (<http://sourceforge.net/projects/ucsf-nomad/>).

3.13 Polymerase chain reaction (PCR)

The chosen primers will determine the annealing temperature (usually melting temperature - 5°C) of the PCR profile. Generally the thermocycler (Eppendorf, Germany) was programmed as: 1 cycle of 95°C for 2min; followed by 35 cycles of 95°C for 1min, 50°C for 1min and 68°C for an appropriate time (1kb/min). A final extension was set for 10min at 68°C and then the tubes were kept at 4°C.

For each 20µl reaction, 2ng of genomic DNA or cDNA was used as a template with 10pmol of the forward and reverse primers (Appendix 6), 200µM of dNTPs (iDNA, Singapore) with 1.5 units of Taq DNA polymerase (Kapa Biosystems, USA).

3.14 Agarose gel electrophoresis

PCR products were electrophoresed in a 1%w/v agarose (Bio-Rad, USA) gel containing 0.5 $\mu\text{g/ml}$ of ethidium bromide (Bio-Rad, USA) for the visualisation of DNA bands. 10 μl of the initial PCR products or 5 μl of a 0.1 $\mu\text{g}/\mu\text{l}$ DNA ladder stock (100bp / 1kb DNA Ladder Mix, New England Biolabs, USA) were mixed with 1 μl of 10x gel loading buffer (Eppendorf, USA) before loading into the wells. For the optimized PCR, 5 μl of the product was used instead. The gel was electrophoresed until the bromophenol blue dye front has reached about 1.5cm from the end of the gel. Next, the gel was then washed and cooled with ddH₂O before being viewed over a gel doc system (Bio-Rad, USA).

3.15 Purification of DNA from agarose gels

The Qiagen gel extraction kit (Qiagen, Germany) was used for this process. DNA fragments of 70bp to 10kb can be purified from cut out gel slices using this kit. DNA in the agarose gel was excised using Swann-Morton carbon steel sterile surgical blades. The blades and the transilluminator were first cleaned with 70% ethanol before use. A new blade or a newly wiped blade must be used for every band that is to be excised. The excised gel pieces were then weighed and 3 volumes of Buffer QG was added to 1 volume of gel slice (100mg ~ 100 μl).

The resultant mixture was incubated at 55°C until the gel pieces were completely dissolved. The color of the mixture must be checked that it is still yellow ($\text{pH} \leq 7.5$). If the color changed to yellow or violet (higher pH), 10 μl of 3M NaOAc, pH 5.0 must be added to bring the pH back to 7.5 or lower. Next, 1 volume of isopropanol (Merck, USA)

was added to the sample and mixed. The mixture was transferred to the QIAquick column and the column was placed in a 2ml collection tube. The resultant setup was then centrifuged at 16,000g for 1 min. The flow-through was discarded and the QIAquick column placed back in the same collection tube. An additional 500µl of Buffer QG was added to the QIAquick column and centrifuged at 16,000g for 1 min to ensure that all traces of agarose were removed. The flow-through is again discarded and the QIAquick column placed back into the collection tube. Next, 750µl of Buffer PE was added to the QIAquick column to wash the DNA and the mixture was allowed to stand for 5 minutes before centrifugation at 16,000g for 1 min. The flow-through was discarded and the QIAquick column was placed back into the same collection tube to be centrifuged at 13,000rpm for another 1 minute. The residual flow-through was discarded and the QIAquick column placed into a clean 1.5ml microfuge tube (Axygen, USA). 30µl of Buffer EB (or ddH₂O) was then added to the centre of the membrane of the column and allowed to stand for 2 min. Finally, DNA is eluted by centrifuging the above setup for 2 min at 16,000g. The flow-through contains the eluted DNA and it is stored at -20°C until it is required for use.

3.16 Purification of DNA samples

Nucleic acids such as PCR products and restriction digest products can be purified using the QIAGEN PCR purification kit (Qiagen, Germany). Similarly to the gel purification protocol, DNA fragments of 100bp to 10kb can be purified using this kit. DNA was mixed with 10 volumes of Buffer PBI and loaded onto the provided spin column that is attached to a 2ml collection tube. The color of the mixture must be

checked that it is still yellow ($\text{pH} \leq 7.5$). If the color changed to yellow or violet (higher pH), 10 μl of 3M NaOAc, pH 5.0 must be added to bring the pH back to 7.5 or lower. Next, the spin column was then centrifuged at 16,000g for 1 min. The flow-through was discarded and the QIAquick column placed back in the same collection tube. Next, 750 μl of Buffer PE was added to the QIAquick column to wash the DNA and the mixture was allowed to stand for 5 minutes before centrifugation at 16,000g for 1 min. The flow-through was discarded and the QIAquick column was placed back into the same collection tube to be centrifuged at 16,000g for another 1 minute. The residual flow-through was discarded and the QIAquick column placed into a clean 1.5ml microfuge tube (Axygen, USA). 30 μl of Buffer EB (or ddH₂O) was then added to the centre of the membrane of the column and allowed to stand for 2 min. Finally, DNA is eluted by centrifuging the above setup for 2 min at 16,000g. The flow-through contains the eluted DNA and it is stored at -20°C until it is required for use.

3.17 TOPO cloning of PCR products

T-A cloning can be achieved by ligation of PCR products to a vector containing thymine overhangs. Briefly, 0.5 to 4 μl of fresh PCR product was mixed with 1 μl of salt solution (Invitrogen, USA) and topped up with ddH₂O to a final volume of 5 μl . Next, 1 μl of TOPO[®] PCR1.2 vector (Invitrogen, USA) was added and the solution mixed gently and allowed to stand for 5min at room temperature. Next, the required aliquots of One Shot[®] TOP10 *E. coli* (Invitrogen, USA) chemically competent cells were thawed on ice for 5min and 2 μl of the above cloning reaction was added to each vial of competent cells

and mixed gently. The cells were allowed to rest on ice for 30min before being heat shocked at 42°C in water bath (Mettler, Germany) for 90s and then allowed to recover.

In order for the cells to recover, 250µl of room temperature S.O.C. Medium (Invitrogen, USA) was added and the tubes incubated at 37°C for 1hr with shaking at 225rpm. During this incubation, LB Ampicillin plates (Appendix 3) were removed from the cold room and allowed to reach room temperature before 40µl of 20mg/ml X-Gal Solution (Fermentas, USA) was spread evenly onto the plates. After recovery, plate 100µl of each transformation culture onto duplicate LB/ampicillin/X-Gal plates and the plates were incubated overnight (16 – 18 hours) at 37°C. White colonies generally contain inserts while blue colonies generally do not contain inserts.

White colonies were transferred with a sterile toothpick onto a master LB/ampicillin/X-Gal plate and the same toothpick was then put into a sterile universal bottle containing 3ml of LB medium (Appendix 3) and 3µl of ampicillin (100µg/ml) (Gibco, USA). The master plate was then incubated at 37°C for 16 to 18 hours while the universal bottles were incubated at 37°C for 16 to 18 hours with shaking (~225 rpm).

3.18 Cloning of restriction enzyme digested PCR products

Basically, restriction enzymes (New England Biolabs, USA) cut DNA under controlled conditions that include a particular buffer (1-4) and temperature. Template DNA (restriction enzyme digested vector / PCR product) of 1µg was added to an eppendorf tube followed by 2µl of a suitable 10x reaction buffer and 2µl of 10x BSA. Next, ddH₂O was added to top up the tube to 19µl and 1µl of the required enzyme was

added. The reaction was allowed to proceed overnight for complete digestion and then purified by the Qiagen PCR purification kit as described in the previous section.

Ligation of the vector and insert was typically performed using the LigaFast™ Rapid DNA Ligation System (Promega, USA). Typically, 100ng of insert DNA and 33ng of vector DNA was added into a fresh eppendorf tube followed by 5µl of 2x Rapid Ligation Buffer. This was followed by 1µl of T4 DNA ligase (3U/µl) and nuclease-free water to a final volume of 10µl. This reaction was incubated at room temperature for 5min for cohesive-end ligations.

Next, the required aliquots of One Shot® TOP10 *E. coli* (Invitrogen, USA) chemically competent cells from the -80°C freezer were thawed on ice for 5min and add 2µl of the above cloning reaction to each vial of competent cells and mix gently. The cells were allowed to rest on ice for 30min before being heat shocked at 42°C in water bath (Mettler, Germany) for 90s and then allowed to recover.

In order for the cells to recover, 250µl of room temperature S.O.C. Medium (Invitrogen, USA) was added and the tubes incubated at 37°C for 1hr with shaking at 225rpm. During this incubation, LB Ampicillin plates (Appendix 3) were removed from the cold room and allowed to reach room temperature. After recovery, 100µl of each transformation culture was plated onto duplicate LB Ampicillin plates and the plates were incubated overnight (16 – 18 hours) at 37°C.

Colonies were transferred with a sterile toothpick onto a master LB Ampicillin plate and the same toothpick was then put into a sterile universal bottle containing 3ml of LB medium (Appendix 3) and 3µl of ampicillin (100µg/ml) (Gibco, USA). The master

plate was then incubated at 37°C for 16 to 18 hours while the universal bottles were incubated at 37°C for 16 to 18 hours with shaking (~225 rpm).

3.19 Plasmid extraction and purification – Miniprep

High-copy plasmids up to 10kb can be purified from overnight cultures of *Escherichia coli* in Luria-Bertani (LB) medium. The Qiagen Miniprep kit (Qiagen, Germany) was used for the procedure. Briefly, pelleted bacteria cells were resuspended in 250µl of Buffer P1, containing RNase A, and transferred into another microfuge tube. 250µl of Buffer P2 was then added to the new tube and the tube is gently inverted for 6 times to mix until the solution becomes viscous and slightly clear. This lysis reaction should not proceed for more than 5 min. Next, 350µl of Buffer N3 was added and the tube was gently inverted 6 times to neutralize the lysis reaction and for precipitation. The resultant solution was centrifuged at 13,000rpm for 10 min and a white pellet will form. After centrifugation, decant the supernatant to the QIAprep column that was placed in a 2ml collection tube that comes with the kit. The QIAprep column was then centrifuged at 16,000g for 1 min and the flow-through was discarded. Next, 500µl of Buffer PB was added to the column and centrifuged at 16,000g for 1 min to remove trace nuclease activity and the flow-through is discarded. Following that, 750µl of Buffer PE was added to the QIAprep column to wash the DNA and the mixture was allowed to stand for 5 min before centrifugation at 16,000g for 1 minute. The flow-through was discarded and the QIAprep column was placed back into the same collection tube to be centrifuged at 16,000g for another 1 minute. The residual flow-through was discarded and the QIAprep column placed into a clean 1.5ml microfuge tube.

To elute the plasmid DNA, 50µl of Buffer EB (or ddH₂O) was then added to the centre of the membrane of the column and allowed to stand for 2 min. Finally, DNA is eluted by centrifuging the above setup for 2 min at 16,000g. The flow-through contains the eluted plasmid DNA and it is stored at -20°C until it is required for use.

3.20 Plasmid extraction and purification – Maxiprep

Large-scale purification of plasmid was performed using the Qiagen Maxiprep kit (Qiagen, USA) and all centrifugations performed in an ultracentrifuge (Beckman Coulter USA). On day 1, start a 5 ml culture in the morning from an isolated colony from an agar plate with the appropriate selective antibiotic. After 8hr of incubation at 37°C with shaking at 225rpm, the starter culture was used to inoculate 250 ml of LB medium with the required antibiotic and shaken for 18hr under the same conditions.

On the next morning, the cells were harvested by centrifugation at 6,000g for 15 min. The supernatant was decanted and the cell pellet was resuspended in 10 ml of Buffer P1. Next, the cells were lysed with 10 ml of Buffer P2 and mixed gently by inverting the tube 4-6 times before incubating at room temperature for 5 min. Cell lysis was stopped with 10 ml of Buffer P3 and mixed gently before incubating on ice for 15-20 min.

The mixture was then transferred to a fresh 30 ml Oakridge tube (Nalgene, USA) and cell debris was pelleted by centrifugation at 20,000g for 30 min at 4°C. The supernatant was then pipetted to a fresh 30 ml Oakridge tube (Nalgene, USA) and centrifuged at 20,000g for 15 min at 4°C. Next, the QIA-tip 500 was equilibrated by applying 10 ml of buffer QBT before applying supernatant to the column. Next, the

column was washed twice with 30 ml of Buffer QC and the plasmid DNA eluted with 15 ml of Buffer QF into a fresh 30 ml Oakridge tube (Nalgene, USA).

DNA was precipitated with 0.5ml of isopropanol (Merck, USA) at room temperature and then centrifuged at 15,000g for 30 min at 4°C. Next, the supernatant was removed by inverting the tube and the pellet was washed with 5ml of 70% ethanol and centrifuged at 15,000g for 30 min at 4°C. The supernatant was carefully decanted and the pellet allowed to air dry for 5 to 10min. The DNA was then resuspended in 150µl of molecular biology grade water (Eppendorf, Germany) and then stored at -20°C till required.

3.21 Electroporation transfection of *P. yoelii*

Infected mice (typically 2) were sacrificed and schizonts obtained by Nycodenz density gradient centrifugation as described above. Schizonts obtained in this manner were then further cultured at a hematocrit of 5% in a T-75 cell culture flask (Nunc, USA) with culture medium containing 30ml of RPMI 1640 (Invitrogen, USA), 10ml of FBS (Gibco, USA) and 1mg of gentamicin (Gibco, USA). The flask was gassed with mixed gas (90% N₂, 5%CO₂ and 5%O₂) (Soxal, Singapore), tightly capped and incubated at 37°C on a horizontal shaker (Barnstead International, USA) at 60rpm.

At suitable intervals, 10µl of culture was removed from the flask into an eppendorf tube and the sample spun down in a tabletop centrifuge (Eppendorf, Germany) at maximum speed for 5s. The culture medium was removed as much as possible, leaving just enough media to resuspend the cells to smear onto a glass slide (Sail Brand, China) and stained with Giemsa (Sigma-Aldridge, USA) for visualization under a light

microscope (Nikon, USA). Once the schizonts have fully matured and consist of fully segmented merzoites, the total number was then evaluated using a hemocytometer (Iwaki, Japan) and $1 \times 10^7 - 5 \times 10^7$ schizonts were mixed with $10\mu\text{g}$ (up to $10\mu\text{l}$) of appropriate DNA in supplemented Basic Parasite Nucleofactor Solution 1 (Amaza, USA). This mixture was transferred into an Amaza cuvette and electroporated in an Amaza nucleofector (Amaza, USA) using the U-33 program. About $100\mu\text{l}$ of culture media (as above) was added to the cuvette to resuspend the schizonts and the sample was then gently transferred into an eppendorf tube, taken up by an insulin syringe (BD-Biosciences, USA) and injected intravenously into recipient Balb/c mice.

After 24hrs, tail blood was taken from the mice and smeared to monitor the parasitemia and to ensure that viable parasites were introduced after electroporation. Transfectants were selected based on drug resistance by replacing the drinking water. Basically, pyrimethamine (Sigma-Aldridge, USA) was dissolved by vortexing in DMSO (Sigma-Aldridge) to a stock solution of 7mg/ml . This was then diluted down with ddH₂O to a working 'drink' of $70\mu\text{g/ml}$ and then the pH equilibrated with 1M HCl to between 3.0 and 5.0. Parasites that survived the drug treatment were harvested were used for DNA extraction and cryo-preserved and/or syringe passaged into naïve mice for further rounds(s) of drug selection.

3.22 Southern Blot

Probes were generated via polymerase chain reaction by incorporating digoxigenin (DIG), a sterol hapten, to label DNA probes and detection was performed similar to western blot by using a monoclonal anti-DIG antibody.

Materials and Methods

Firstly, 20nmol each of dATP, dGTP, dCTP and dTTP (GE Healthcare, USA) were mixed with 10nmol of dUTP-DIG (Roche, USA). Template DNA (2ng) was added together with gene specific primer pairs (1st Base, Singapore) of 20pmol each. Also, 10 μ l of Buffer A and 2.5U of DNA polymerase (Kapabiosystems, USA) were added and the mixture topped up to 100 μ l with ddH₂O. Thermal cycling was then programmed using established PCR conditions. After PCR, the DIG-labeled product was run on an agarose gel and the size should show a shift in size when compared to a normal PCR due to the additional DIG group. Probes were then purified using the Qiagen PCR purification kit (Qiagen, Germany) as described earlier.

Genomic DNA (5 μ g per reaction) was digested overnight with the appropriate restriction enzyme (New England Biolabs, USA). On the next day, the digested DNA was run on a 0.8% w/v agarose gel (Bio-Rad, USA) containing 0.5 μ g/ml of ethidium bromide (Bio-Rad, USA) and the 1kb ladder DNA molecular weight marker (New England Biolabs, USA) was captured on a gel doc system (Bio-Rad, USA) such that the position of the markers can be clearly seen.

The gel was then submerged in Depurination Solution (Appendix 5) on a horizontal shaker (Firstek, Taiwan) until the bromophenol blue dye front turned to yellow. Next, decant the solution and rinse the gel twice with ddH₂O before submerging the gel in Denaturation Solution (Appendix 5) for 60min shaking gently at room temperature. Following that, the solution was replaced with Neutralizing Solution (Appendix 5) and then allowed to shake gently for 60min at room temperature. A large glass tray was used to contain 10x SSC (Appendix 3) and a glass plate placed over the tray for a filter paper (Whatman, USA) to act as a wick such that it rests on the plate with both ends dipping

Materials and Methods

into the 10x SSC. Next, the gel was placed sample well side down, taking care to remove any air bubble. This was followed by a positively charged nitrocellulose membrane (GE Healthcare, USA) which was then overlaid with 6 pieces of filter paper (Whatman, USA) and then followed by a 15cm high stack of paper towels. Another glass plate was placed on top of the paper towels and a weight of 600g was added on top. This transfer was allowed to proceed overnight.

When the transfer is done, the nitrocellulose membrane was placed on filter paper (Whatman, USA) that was soaked with 10x SSC (Appendix 3) and UV cross-linked using the automatic program in a Stratalinker (Stratagene, USA). After that, the membrane was rinsed briefly in ddH₂O and allowed to air-dry. The dried membrane was then placed into a hybridization tube (Thermo, USA) with 20ml of DIG-Easy-Hyb solution (Roche, USA) for 2hr in a hybridization oven (Thermo, USA) set at 42°C. In the meantime, 500ng of purified probe was added to 20ml of DIG-Easy-Hyb solution (Roche, USA) and heated in a boiling water bath for 10min before being chilled on ice for 5min. The cooled probe was then used to replace the pre-hybridization buffer and hybridization was carried out overnight at 42°C.

On the next day, Washing Solution I (Appendix 5) and the hybridization oven were warmed to 55°C and the probe removed and kept in -20°C for future use. A volume of 20ml of Washing Solution I was used for stringency wash for 20min at 55°C for three times. Next, the hybridization oven was cooled to room temperature and the membrane incubated with 20ml of Blocking Solution (Roche, USA) for 30min. This was then discarded and replaced with 2µl of anti-DIG (Roche, USA) in 20ml of Blocking Solution and incubated for 30min. Next, the membrane was washed twice for 15min with Washing

Solution II, which is essentially 1x Maleic Acid Buffer (Appendix 5) with 0.3%v/v Tween-20 (Amresco, USA). After that, the membrane was equilibrated with 20ml of Detection Buffer (Appendix 5) for 5min before placing the membrane with DNA side facing up on a development folder (Kodak, USA). CSPD ready-to-use (Roche, USA) was added liberally to the membrane and the membrane was immediately covered with the second sheet of the folder to spread the substrate evenly without air bubbles. This was then left to incubate for 5min at room temperature. Excess liquid was then squeezed out and an X-ray film (Kodak, USA) was introduced for exposure for 5-20min and then developed on an automated developer (Kodak, USA) for visualization.

3.23 Western Blot

The mini-gel system from Bio-Rad was used for the PAGE run and subsequent blot. Firstly, the separating gel mix (10% polyacrylamide gel, Appendix 2) was allowed to set and the gel overlaid with isopropanol (Fluka, USA) to flatten the meniscus. After setting, the isopropanol was decanted and then the stacking gel mix (Appendix 2) was overlaid and the well combs inserted into the glass plates immediately.

Pelleted samples were mixed with an equivalent volume of 2x SDS sample buffer (Appendix 2) and homogenized in a water bath sonicator (Ultrawave, UK) for 2min. Next, they were boiled at 100°C in a heating block (Biometra, Germany) for 10min before being snap cooled on ice for 5min. The samples were then spun down in a tabletop centrifuge (Eppendorf, Germany) at 16,000g for 1min and then the supernatant was loaded into the wells of the PAGE gel. A pre-stained dual color 10-250kDa protein marker (Bio-Rad, USA) was used to track the gel run as well as the efficiency of the

subsequent blotting step. The samples were initially run at 100V until the dye front has left the stacking layer and entered the separating gel before being ramped up to 200V. Samples were run until the bromophenol blue dye front reached the end of the gel.

The semi-dry trans-blot system (Bio-Rad, USA) was used for blotting. Basically, 3 pieces of filter paper (Whatman, USA) were placed onto the base of the apparatus and then wetted with transfer buffer (80% 1x SDS running buffer (Appendix 2) and 20% methanol (Sigma-Aldridge, USA)). Next, the nitrocellulose membrane (Bio-Rad, USA) was overlaid, followed by the gel. Air bubbles were removed by gently pushing them to the edges. Following that, 3 more pieces of filter paper were overlaid and wet as above. The apparatus was then secured and the blot was allowed to proceed at 18V for 60min.

The blotted membrane was then transferred into a suitable container and blocked in PBS (Appendix 3) containing 5% w/v non-fat milk and 0.05v/v Tween (Amresco, USA) for at least 60min at room temperature. The blocking solution was then decanted and the membrane washed with PBS (Appendix 3) containing 0.05% Tween 20 (Amresco, USA). The membrane was then incubated with an appropriate primary antibody (0.25µg/ml) in blocking solution either for 60min at room temperature or overnight at 4°C. After which, the antibody solution was decanted and the membrane washed thrice with washing solution for 5min each. Next, an appropriate peroxidase conjugated secondary antibody (0.25µg/ml) in blocking solution was added to the membrane and incubated for 60min at room temperature. The membrane was then washed thrice with washing solution for 5min each and then transferred on a development folder (Kodak, USA). The SuperSignal® West Pico Chemiluminescent Substrate (Pierce, USA) was used as a detection system. Basically 1ml of luminol reagent was mixed with

1ml of peroxidase reagent and then spread liberally to the membrane. The membrane was immediately covered with the second sheet of the folder to spread the substrate evenly without air bubbles. This was then left to incubate for 5min at room temperature. Excess liquid was then squeezed out and an X-ray film (Kodak, USA) was introduced for exposure for an appropriate time to optimize signal intensity versus background noise and then developed on an automated developer (Kodak, USA) for visualization

3.24 Immunofluorescence Microscopy

Parasitized blood films were smeared onto glass slides (Sail Brand, China) and allowed to air dry. Dried smeared slides were then wrapped in aluminium foil and stored at -80°C until use. On the day of the experiment, the slides were allowed to warm up at room temperature before removing the foil. The slides were then fixed in ice-cold acetone (Sigma-Aldridge, USA) for 5min then allowed to air dry.

The appropriate primary antibody was then diluted in ice-cold Diluent Buffer (1x PBS with 1%w/v bovine serum albumin and 0.01%w/v sodium azide) (Sigma-Aldridge, USA) to an appropriate concentration. The fixed slides were then placed in a humidified chamber and then the diluted primary antibody was added and allowed to incubate at room temperature for 1hr. Next, the slides were transferred into a washing trough and washed for 5min in 1xPBS. After which, the PBS was discarded and the wash repeated for a total of three times before the slide was dried gently using a hairdryer on cold setting.

An appropriate concentration of secondary antibody diluted in Diluent buffer was then added to the slide and incubated at room temperature for 1hr before the washing

steps were performed as described above. DAPI or 4',6-diamidino-2-phenylindole (Molecular Probes-Invitrogen, USA) was used to counterstain nuclei and a 5mg/ml stock was first made in dimethylformamide (Sigma-Aldridge, USA) and stored at -20°C before it is diluted in 1x PBS to 100ng/ml and this is used as a working solution. Basically, the slides were now dipped into ice-cold working solution of DAPI for a few seconds and then washed in 1xPBS for 1min. This was repeated thrice and the slides allowed to dry with a hairdryer at cold setting. Mounting media (Citifluor Ltd, UK) was then added onto the dried slide before a cover slip (CellPath, UK) was allowed to cover the slide.

The slide can be viewed immediately on a fluorescence microscope (Olympus, USA) or stored in the dark at 4°C up to several weeks.

3.25 Real-time RT-PCR

Short regions of the genes of interest to be assayed by real-time RT-PCR were first TOPO-cloned to generate standard curves for quantitative real-time PCR. cDNA from parasites were amplified using the same conditions for generating the TOPO clones in the presence of the Sybr Green Master mix (Applied Biosystems, USA) and analyzed on an ABI 7500 thermocycler (Applied Biosystems, USA). The absolute amount of transcripts were quantified relative to the standard curve and normalized to parasite 18S rRNA. Consistent results were obtained whether 18S rRNA or β -actin were used as normalization housekeeping genes. The un-paired T-test was used to determine the significance of the differences between biological triplicates and a *p*-value of less than 0.05 is deemed to be significant.

3.26 Selective Index (SI)

Mice that were first matched with age, sex and weight with parasitemias from 5% to 10% were chosen in order to reduce any possible variations in terms of proportion of available reticulocytes and any physiological differences in the animals. The Selective Index was calculated as a ratio of the observed number of multiple-invaded cells to the expected number of multiple-infected cells and the statistical significance analyzed by the T-test (Simpson, J.A. *et al*, 1999). This test was based on the Poisson distribution where parasite invasion is assumed to be independent and random events. The rationale can be illustrated by considering a sample of 5% parasitemia, then μ (mean number of parasites per erythrocyte) = 0.05. By defining x_i as the number of parasites per i th erythrocyte, the probability of x parasites = $e^{-\mu} \mu^x / x!$

By using an example of a sample count of 300 erythrocytes, the probability of x parasites per erythrocyte at 5% parasitemia is 0.951 for 0 parasites; 0.0476 for 1 parasite; 0.0012 for 2 parasites; 0.00002 for 3 parasites, etc. Thus, the expected number of infected cells containing 1 parasite is calculated as the probability of having 1 parasite divided by the probability of having any parasites, multiplied by the total number of infected cells investigated. By choosing to investigate 300 parasitized erythrocytes, the expected number of infected cells having 1 parasite is $(0.0476 / (1 - 0.951)) \times 300 = 291$; 2 parasites is $(0.0012 / (1 - 0.951)) \times 300 = 7.35$; 3 parasites is $(0.00002 / (1 - 0.951)) \times 300 = 0.122$, etc.

The Selective Index (SI) was then calculated based on the observed number of multiple-infected cells (O) / expected number of multiple-infected cells (E). Using this statistical measure, the SI for *P. falciparum* and *P. vivax* at 1-2.0% parasitemia is 2.44

Materials and Methods

and 7.60 respectively and is significantly different with a T-test score of <0.001 (Simpson, J.A *et al*, 1999). The reproducibility and high confidence of this test in differentiating the reticulocyte restricted *P. vivax* from the non-restricted *P. falciparum* highlights the usefulness of this test in the determination of reticulocyte-restricted parasites. While SI is a statistical test of the multiplicity of infection, the high correlation of high SI with reticulocyte-restricted species of human, simian and rodent parasites is an indirect measure of red blood cell selectivity.

Comparative genome analysis of the three rodent malaria parasite species
Plasmodium berghei*, *Plasmodium chabaudi* and *Plasmodium yoelii

4.1 Abstract

Although the draft release of the genome sequences of the rodent malaria parasites *P. berghei*, *P. chabaudi* and *P. yoelii* have been released, they are fraught with gaps and are relatively poorly annotated. Our in-house pan-rodent cross genome long oligonucleotide microarray chip was designed to the predicted protein coding genes of the above three rodent parasite species and this was used as a main tool for re-annotating the incomplete genomes. Genomic DNA hybridizations to this chip reveal additional hits to oligonucleotides designed for other species, suggesting that these genes while absent in the draft sequences, are actually present. The use of bioinformatics tools (tBLASTn) and PCR screens confirmed a selection of these genes while also suggesting that other missing genes do exist and that a ‘core’ gene set should be present in all *Plasmodium spp.* A rodent malaria parasite specific orthology map was also constructed with *P. yoelii* as the index in order to elucidate the relationship of the rodent malaria parasite specific genes to one another. By using an identical workflow, the rodent malaria parasite specific genes were then analyzed across all three species. The results show that there is an expanded common gene set across the malaria parasite species and that the rodent parasite specific genes consist mainly of the PIRs, hypothetical genes and genes that have undergone expansion from the common ‘core’ set.

4.2 Introduction

4.2.1 Rodent malaria parasites as useful models for *in vivo* studies

Plasmodium falciparum being the main causative agent responsible for most of the mortality and morbidity associated with malaria is therefore the main species studied by many malaria research groups in the global effort to further understand the biology of the parasite and ultimately to the development of novel drugs and vaccines.

Although well-established culture and molecular techniques have been developed for *P. falciparum*, the use of rodent malaria parasites as models in the study of the human disease is still as relevant today because they are very similar to the human and primate parasites in terms of life cycle, physiology and structure (Carter, R. and Diggs, C.L., 1977). Studies on human malaria are restricted to the *in vitro* culture model and while other primate models have been used, perhaps the easiest and most accessible model for any form of *in vivo* studies have been rodent malaria parasites. Most of the laboratory strains have been isolated from the African thicket rat, *Grammomys surdaster* (Carter, R. and Diggs, C.L., 1977) and the shiny thicket rat *Thamnomys rutilans* (Landau, I and Boulard, Y., 1978) and they have all been adapted to proliferate successfully in other rodent hosts as well, particularly in the typical laboratory mice strains. Similarly to human malaria whereby different parasite species differ in the host cell selection, morbidity and mortality, different species of rodent parasites also exhibit different virulence phenotypes. More importantly, these parasite species have been used as models for studying different aspects of disease pathogenesis. For example, *P. berghei* has been used extensively to investigate the pathogenesis of cerebral malaria (Delahaye, N.F. *et al*, 2007; Oakley, M.S. *et al*, 2008) while *P. yoelii* has been used to investigate pathogenesis

linked to invasion using virulent and avirulent lines (Walliker, D., *et al*, 1976; Bungener, W., 1985; Iyer, J. *et al*, 2007). The synchronous growth pattern of *P. chabaudi* (Gautret, P. *et al*, 1995) can be used to elucidate molecular mechanisms involved in coordinating the start of a new round of erythrocytic cycle with the host's normal waking and sleeping hours. Also, *P. chabaudi* has been typically used to elucidate disease pathogenesis (Mideo, N. *et al*, 2008) and host immune mechanisms (Stephens, R. *et al*, 2009). Therefore, the use of rodent malaria parasites is certainly useful to address important parasite-host interactions that cannot be addressed using *in vitro* models.

For the purposes of this study, it is pertinent to address the origin and lineage of the three different rodent malaria parasite species. Parasites are typically discovered in rodent hosts and these first discoveries are termed as isolates that can be defined as a sample of parasites collected from a wild-caught animal on a unique occasion. Such isolates may contain more than one species of parasite, and more than one genetically distinct clone of a given species. From these initial isolates, distinct lines were made and they are defined as parasites that have undergone a particular passage or treatment. While such selected parasites usually have certain characteristics in common, they are not necessarily genetically identical. Genetically identical clones could then be obtained from a single haploid parasite from established lines.

As mentioned in the introduction chapter, our *P. berghei* ANKA clone was obtained from the Edinburgh clones while the drug sensitive *P. chabaudi* AS(sen) clone was cloned from the initial AS isolate. The *P. yoelii* 17x1.1 parasites were derived from a clone originally obtained by Dr. Walter Weiss from an uncloned 17x line (Weiss, W.R. *et al*, 1989).

4.2.2 Genome organization of the human and rodent malaria parasites

Since the release of the genome sequences of *P. falciparum* and the rodent malaria species, common features of these haploid genomes include a genome size of 22-26Mb that are arranged in 14 chromosomes ranging from 0.5-3.0Mb. In addition, the current genomic data show a high degree of conservation between different *Plasmodium* species with the exception of genes located in the telomeric and subtelomeric regions that are extremely variable due to their role in antigenic variation and immune evasion (Waters, A.P., 2002; Carlton, J.M. *et al*, 2002). These highly polymorphic genes are typically part of large multigene families such as the *Plasmodium* interspersed repeats (PIRs) (Janssen, C.S. *et al*, 2004) and their conservation amongst the *Plasmodium* species could potentially facilitate the elucidation of their regulatory control. In addition, orthologues have been discovered in many distant species, thus the use of rodent models today is highly relevant especially to study antigenic variation and validate *in vitro* experimental data.

A recent phylogenetic survey of rodent malaria parasites based on DNA sequences from multiple loci in the nuclear, mitochondrial and plastid genomes placed *P. berghei* and *P. yoelii* as sister species forming a distinct clade while *P. chabaudi* and another rodent parasite *P. vinckei* forming another group (Perkins, S.L. *et al*, 2007) thus suggesting that *P. berghei* and *P. yoelii* seemed to be more evolutionarily related to each other while *P. chabaudi* seemed to be a more distinct species.

However, differences between the rodent parasites especially at the genomic level have not been fully elucidated as yet and thus linking genomic differences to phenotypic traits of these parasites have been difficult. This has been the main stumbling block as the

genome sequence of *Plasmodium falciparum* has already been completed, annotated and partially validated while the rodent parasite genomes still contain gaps where contigs generated by shotgun sequencing failed to overlap. As a result, it is difficult to assemble these genomes with high confidence, with the draft genomes containing gaps in the sequences and poor gene prediction and annotation of the rodent parasite genomes.

4.2.3 Tools for elucidating genomic differences between closely related species

Genome-wide screens using DNA microarrays have been widely used in biological studies and an essential tool is the application of DNA microarray technology based on ‘spotting’ of long oligonucleotides onto glass slides (Kane, M.D. *et al*, 2000; Hughes, T.R. *et al*, 2001; Bozdech, Z. *et al*, 2003). In order to elucidate the similarities and differences between closely related species, whole genome comparisons using microarray hybridization have been used to compare genomic differences between similar species in both prokaryotes and eukaryotes (Daran-Lapujade, P. *et al*, 2003; Dong, Y. *et al*, 2001; Ong, C. *et al*, 2004). Besides analyzing sequence variation, the gain/loss of DNA can also be used as a tool for elucidating evolutionary divergence between species in their natural environment.

The utilization and confidence of microarray technology have recently gained momentum with the development of a robust program called ‘OligoRankPick’ which utilizes a weighted rank-sum strategy to optimize oligonucleotide selection even along genomes of many diverse organisms (Hu, G. *et al*, 2007). The resultant *P. falciparum* dataset generated from this strategy was shown to be highly reproducible and also led to

increased coverage of the *P. falciparum* genome as compared to earlier *P. falciparum* chip designs.

While the genome sequence coverage for the human parasite *P. falciparum* has been completed to a high degree of confidence, this cannot be said for *P. berghei* (4x), *P. chabaudi* (4x) and *P. yoelii* (5x) (Hall, N. *et al*, 2005). Of note, the Sanger Institute has already sequenced *P. berghei* and *P. chabaudi* to 8x coverage. However, sequence closure and annotation of the contigs are still in progress. To illustrate the confidence of sequence information, *P. falciparum* genes have been mapped with high confidence onto chromosomes while those for the rodent parasites are only organized into contigs of varying sizes due to difficulties in closing the gaps.

Although the incompleteness of these rodent parasite genomes could present difficulties in their comparative analysis, there is some evidence that they are highly conserved (Kooij, T.W. *et al*, 2005). A pan-rodent cross-genome oligonucleotide microarray chip for the rodent malaria species *P. yoelii*, *P. berghei* and *P. chabaudi* was designed based on the ‘OligoRankPick’ strategy (Hu, G. *et al*, 2007) whereby a rank-sum strategy automatically picks the best oligo amongst all possible combinations for a particular gene. In summary, oligos were designed to be able to overlap in all three or two species and where this was not possible, oligos were designed to be species-specific. In this analysis, the oligos were first validated for performance and then used to reannotate gaps in the synteny map in one or more of the rodent species based on complementary hybridization data, thus complementing the status quo. The remaining gaps were then screened using bioinformatics tools to evaluate their probability of being truly absent while manual screening of a subset of these genes were screened via PCR to confirm their

Chapter 4

status and the probability of their presence or absence. The aim here is to identify common and unique genes amongst a human and three rodent malaria parasite species

4.3 Results and discussion

4.3.1 Validation of the pan-rodent long oligonucleotide array chip

This long oligonucleotide microarray chip was designed based on the ‘OligoRankPick’ algorithm used to design probes that are specific to their target gene and robust enough for genomes such as *Plasmodium falciparum* that contains large fluctuations of GC content, and abundant gene duplications (Hu, G. *et al*, 2007). Similarly, this same algorithm was then modified (Guangan, Hu, NTU/SBS) to create a chip that can firstly target genes common to all 3 rodent malaria parasite (RMP) species (based on PlasmoDB version 4.4; <http://www.plasmodb.org>), namely *P. berghei* ANKA, *P. chabaudi* AS and *P. yoelii* 17x1.1. In the PlasmoDB database, there are about 7,861 predicted coding sequences in the *P. yoelii* genome, 12,216 in *P. berghei* and 15,095 in *P. chabaudi*. Since *P. yoelii* is the only RMP with a published draft genome (Carlton, J.M. *et al*, 2002), oligos were firstly designed against the predicted protein coding genes of *P. yoelii*. Next, all possible oligos of *P. yoelii* were used to query the homologous regions of the other two RMP species using the NCBI blastn program (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) (Figure 4.1A). Next, four parameters were used, namely: BLAST scores (first and second hit), GC content and self-annealing score to search the oligos (Figure 4.1B). Each score is then transformed into a rank and a weighted rank-sum is calculated for each oligonucleotide with the final oligonucleotide being selected based on the smallest rank-sum value. These oligos were then used to select for those that are optimal for all three species. Next, oligonucleotides for the remaining genes were designed with the strategy of pairing a *P. yoelii* gene with an ortholog from another rodent parasite species, i.e. oligos were designed for *P. yoelii* and

P. berghei; and then for *P. yoelii* and *P. chabaudi* (Figure 4.1A). The remaining sequences were then used to design oligos common to *P. berghei* and *P. chabaudi*. Next, oligonucleotides were designed based on the remaining species-specific genes. Thus, the principle design strategy was to maximize the potential of detecting homologous genes between the 3 rodent parasite species in order to assay the similarities and differences between them by having all 3 species' probes on 1 chip, thus maximizing experimental time and costs while also reducing normalization errors if 3 different species specific chips were used. The breakdown of the hybridization targets of the pan-rodent oligonucleotides is shown in Figure 4.2.

Figure 4.1. The overall design schematics of the pan-rodent chip. (A) Methodology of the chip design. Firstly, all possible oligonucleotides of *P. yoelii* were used to search in the homologous region of the other two species using NCBI blastn and were scored and ranked accordingly. The oligonucleotides were then filtered using three rules such they must have: (i) at least 90% homology to target sequences, (ii) less than 37.5% to non-target sequences and (iii) GC% tolerance of ±5%. Oligonucleotides for all three species were selected followed by oligonucleotides for *P. yoelii* and *P. berghei* and then for *P. yoelii* and *P. chabaudi*. Next, the remaining oligonucleotides were selected to be specific for both *P. berghei* and *P. chabaudi*. The remaining sequences unaccounted for were then

used to design oligonucleotides specific either to *P. yoelii*, *P. berghei* or *P. chabaudi*. (B) Rank-sum strategy. Oligonucleotides were scored accordingly to (i) first and second BLAST hits, (ii) GC content (T_m) and (iii) Smith-Waterman score (self-binding). The oligonucleotides are then ranked based on each parameter and ordinal rank number is given to all oligonucleotides in each parameter rank independently. The final weighted rank-sum (RS) is calculated for all oligonucleotides using multiple weight sets (not indicated) and the lowest value is considered. Finally, the optimal candidate is selected based on the lowest RS(k) amongst all oligonucleotides in the locus of interest (Hu, G. et al, 2007). (Figure used with kind permission from Guangan, Hu)

Figure 4.2. Venn diagram showing distribution of target rodent parasite gene hits of all oligonucleotides. All oligonucleotides are 60 bases long and the GC content is targeted at 30% and the allowable deviation is 5% for overlapping oligonucleotides. Complementary oligonucleotides to each rodent malaria parasite species was calculated from the sum of all possible combinations, i.e. oligonucleotides specific to itself and those that can hybridize to itself and to other rodent malaria parasite species. (Legend: *Pb* = *P. berghei* specific oligonucleotides only; *Pc* = *P. chabaudi* specific oligonucleotides only; *Py* = *P. yoelii* specific oligonucleotides only; *Pbc* = *P. berghei* & *P. chabaudi* specific oligonucleotides; *Pby* = *P. berghei* & *P. yoelii* specific oligonucleotides; *Pcy* = *P. chabaudi* & *P. yoelii* specific oligonucleotides; *Pbcy* = oligonucleotides specific to all 3 rodent parasite species).

The performance of the oligonucleotides can only be verified by actual microarray experiments and the chip was ‘printed’ by spotting all the oligonucleotides onto polylysine-coated microscopic slides as described previously (DeRisi, J.L. *et al*, 1997). Since the oligos were designed based on the older version of PlasmoDB 4.4, they were then subjected to a BLASTn search of the updated PlasmoDB version 5.4. This led to the observation of more annotated genes in all the three rodent parasite species. Therefore, all the oligonucleotides were matched with their respective target gene via BLASTn to PlasmoDB version 5.4 (<http://www.plasmodb.org>) so that an accurate normalized mean hybridization data with respect to a particular gene can be obtained for analysis (Figure 4.3).

Figure 4.3. A schematic outlining the workflow for assigning specific oligonucleotides to their proper target gene. Firstly, the oligonucleotide sequences were matched to their respective target gene via BLASTn to PlasmoDB version 5.4. This gives a linkage table whereby the normalized hybridization data from a repertoire of oligonucleotides that target a specific gene can thus be appended to this table and a mean value can then be obtained for that gene of interest.

With this new reference database, genome wide survey of each individual rodent parasite species was performed using total genomic DNA. Basically, equivalent amounts of genomic DNA (3 μ g) from the same species were labeled with Cy3 and Cy5 fluorescent dyes and each working oligonucleotide were expected to be fluorescent yellow in color, thereby indicating no bias in labeling (Figure 4.4). Further confidence in reproducibility can be drawn from a high Pearson correlation of duplicate slides of *P. berghei*, *P. chabaudi* and *P. yoelii* with values at 0.987, 0.987 and 0.964 respectively. A stringency filter of twice the background intensity and twice of its standard deviation was employed to rule out non-specific hybridization signals (Figure 4.5). Oligonucleotides that were designed to hybridize with their target genome and the percentage of successful oligonucleotides for *P. yoelii*-specific, *P. berghei*-specific and *P. chabaudi*-specific oligonucleotides were 91% (7,347), 90% (6,314) and 84% (6,356) respectively. A possible reason why the remaining oligonucleotides did not show significant hybridization to their intended target could be due to the poor genome coverage as compared to the more complete *P. falciparum* genome. There might be errors in the current sequence drafts of the rodent malaria parasites due to the propensity of random sequence rearrangements of an AT-rich genome in a sequencing vector, thus it is not surprising that a low proportion of oligonucleotides were unable to hybridize with their intended targets especially if they were designed in regions with low confidence in sequence quality. Thus in terms of numbers and representation, this repertoire of successful oligonucleotides would lie between the conservative minimum and the current number of over-predicted number of genes that reside in the public database. Hence, this

array is extremely useful in navigating the incompletely sequenced and poorly annotated rodent malaria parasite genomes.

Figure 4.4. Scatter plots of the median intensities of the individual oligonucleotides with the mean Cy3 signals plotted against the mean Cy5 signals. (A) Signals from *P. berghei* ANKA genomic DNA hybridizations were well correlated with the Cy3 and Cy5 signals correlating with each other at high confidence ($R^2 = 0.982$). (B) Signals from *P. chabaudi* AS genomic DNA hybridizations were well correlated with the Cy3 and Cy5 signals correlating with each other at high confidence ($R^2 = 0.983$). (C) Signals from *P. yoelii* 17x1.1 genomic DNA hybridizations were well correlated with the Cy3 and Cy5 signals correlating with each other at high confidence ($R^2 = 0.979$).

Figure 4.5. An example of a typical scatter plot of signal intensity sorted in descending order against the number of *P. yoelii* specific oligonucleotides. The Nomad software is set to filter off oligonucleotides that do not perform and a conservative lower limit of twice the background intensity and twice the background standard deviation was employed to rule out non-specific hybridizations. This same filter was also applied to raw data in subsequent experiments.

4.3.2 Global analysis of gaps in the rodent malaria parasite genome and re-annotation of potential coding sequences via microarray hybridizations

Since it has been established that the malaria parasite genomes are extremely conserved (Kooij, T.W. *et al*, 2005), the current status quo of genes that are missing in each species could potentially be identified based on the hybridization data from the array. Simply put, a species that currently does not possess a particular gene but now gives a signal on the array suggests the presence of this gene. Hence, this gene can then be re-annotated based on homology.

Missing genes could arise due to 2 scenarios: either the sequence information is missing, or that the genome sequence is present but the gene is missed by gene prediction algorithms (Figure 4.6). Since the oligonucleotides were designed based on annotated protein coding genes, a signal constitutes direct experimental evidence for the presence of a homologous gene.

Figure 4.6. Current scenarios depicting the probable scenarios of missing genes in the rodent malaria parasites. (A) The gene is not there due to sequencing and assembly gaps, or (B) the sequence is present but missed by gene prediction algorithms. (Genome browser figures adapted from PlasmoDB at <http://www.plasmodb.org>).

In order to quickly visualize the genomic location of where each gene lies and to ascertain their presence or absence, a synteny map of orthologous *Plasmodium* genes with *P. falciparum* as the index (Carlton, J.M. *et al*, 2008) was used. Using this linkage table, syntenic gaps can be filled in based on hybridization evidence of an orthologous gene found in another rodent malaria species that corresponds to the same locus with respect to *P. falciparum*. Therefore, by virtue of the significant conservation of synteny even amongst distantly related *Plasmodium* species (Kooij, T.W. *et al*, 2006), the process to reannotate genes that were not predicted by current gene prediction software or due to missing sequence information in the unfinished genomes of the rodent malaria parasite species can thus be attempted. Since *P. falciparum* was the index species, species-specific genes located in the telomeric and sub-telomeric regions such as the *var*, *rifin* and *stevor* gene families were excluded due to the absence of these genes in the rodent parasites. In essence, a missing gene for a particular species is only filled when there is genomic DNA hybridization data from that species for an oligonucleotide that is designed for an orthologous gene based on another species. By virtue of syntenic linkage to the conserved regions, it can then be deduced that that is indeed a missing gene and its function derived from the more complete annotation in *P. falciparum* (Figure 4.7).

Based on this approach, missing genes were filled in using hybridization data. This led to the discovery of 179 missing genes in *P. berghei*, 306 missing genes in *P. chabaudi* and 215 missing genes in *P. yoelii*. The majority of these genes are hypothetical proteins, as there are still many genes in *P. falciparum* that have not been fully described and have no function as yet assigned. For those whose function has been described, genes involved in biosynthesis, protein modifications, kinases and even invasion-related

proteins in the case of *P. chabaudi* and *P. berghei* have been discovered (Table 4.1; Appendix A).

Current syntenic relationship					New' syntenic relationship			
Pf#	PfGene	PbGene	PcGene	PyGene	Pb	Pc	Py	Pf Product
67	PFA0260c	PB001039.01 .0	PC000215.0 5.0	-	PB001039.01.0	PC000215.05.0	PB001039.01.0	hypothetical protein
69	PFA0270c	PB001363.02 .0	-	PY04768	PB001363.02.0	PB001363.02.0 + PY04768	PY04768	hypothetical protein
82	PFA0335w	PB000519.01 .0	-	PY01029	PB000519.01.0	PB000519.01.0 + PY01029	PY01029	<i>P. falciparum</i> GTP binding protein RAB5
84	PFA0345w	PB001391.02 .0	PC000332.0 5.0	-	PB001391.02.0	PC000332.05.0	PB001391.02.0	centrin, putative
100	PFA0425c	-	PC000290.0 5.0	PY00632	PY00632	PC000290.05.0	PY00632	hypothetical protein
101	PFA0430c	PB000915.00 .0	-	PY04869	PB000915.00.0	PB000915.00.0 + PY04869	PY04869	hypothetical protein
132	PFA0585w	-	-	PY01445	PY01445	PY01445	PY01445	hypothetical protein

Figure 4.7. A snapshot of the process whereby syntenic orthologs were identified using genomic DNA complementary hybridization experiments. The panel on the left shows the typical syntenic orthologous relationship between the *P. falciparum* and the three rodent malaria parasites where gaps in the core gene set are denoted by a dash. On the right panel, missing genes were appended in bold, with the description representing the accession numbers of the complementary oligonucleotide of a related rodent parasite species. In some cases, there is only 1 rodent parasite gene that matches a *P. falciparum* gene (PFA0585w) and the oligonucleotide designed to that gene in *P. yoelii* (PY01445) is able to pick up a related gene in the other two rodent parasite species. Therefore, although 700 rodent parasite genes were discovered, there is only an improvement of 583 syntenic orthologous relationships (Appendix a).

Gene / species	<i>P. berghei</i>	<i>P. chabaudi</i>	<i>P. yoelii</i>
Actin-binding			1
Adhesion		1	
Biosynthesis	7	6	8
Ca-binding			1
Catabolism		2	1
Cell cycle		1	1
Co-factor binding		1	1
DNA-binding	4	2	8
Electron transport / redox	1	1	1
GTP-binding	1	5	1
Hypothetical	141	223	161
Invasion	2	4	
Kinase	3	8	5
Motor activity	1	1	1
Nuclease / endonuclease	2	1	
Nucleotide-binding			1
Phosphatase		2	
Protease / proteosome associated		7	2
Protein synthesis / modifications	9	12	
Receptor	1		
Ribosomal	3	9	6
RNA-binding	1	7	4
Secretion	1	1	
Stage-specific proteins		2	
Structural		1	
Transport		5	7
Ubiquitin-like	2	2	4
Unknown		2	
Signaling			1
Total:	179	306	215

Table 4.1. Summary of missing genes in the three rodent malaria parasite species via synteny. The bulk of genes of these genes are hypothetical proteins. Of interest, *P. berghei*, protein synthesis/modification proteins include: 1 chaperon, 4 hydrolases, 2 protein synthesis, 1 protein-folding and 1 protein-binding protein. In *P. chabaudi*, these protein synthesis/modification proteins include: 3 hydrolases, 1 isomerase, 2 protein-folding, 1 polymerization and 5 synthesis proteins. Detailed information is available in Appendix A.

A few of the genes that have been filled by the array data were randomly selected for polymerase chain reaction (PCR) screening and direct sequencing in order to establish the confidence of this group of newly discovered genes. PCR primer pairs were designed flanking the oligonucleotide sequence to the species whereby there is known sequence information and these were used to amplify a newly predicted gene in another species where it is not annotated / predicted or where the sequence is absent due to the paucity of sequence information in that region (Figure 4.8). In this example, PCR primers based on *P. yoelii* were designed to flank the corresponding oligonucleotide region so as to amplify a syntenic ortholog in another species that we have now predicted to be present. All screens were performed on regions whereby sequence information for the new gene is not present (i.e. missing sequence) except for PY00632 and PY03414 whereby a corresponding *P. berghei* contig is present but the gene is not predicted. In summary, 7 of the 8 screens worked and the one that was PCR negative could possibly be due to sequence polymorphisms at the primer sites. Some bands also exhibit a shift in molecular weight, for example different PCR product sizes of PY00632 and its *P. berghei* ortholog are expected as sequence information is currently available. In this case, the automated ORF prediction algorithms missed the gene in *P. berghei* as analysis of the genome sequence reveals its existence. By aligning the *P. yoelii* and *P. berghei* genomic sequences, the difference in PCR product size can be attributed to variations in sequence length in the region bounded by the primers. The differences in PCR product size in the PY06972 screen of *P. yoelii* and *P. chabaudi* genomes are also due to the same reasons. The high congruence of PCR-positive screens showed the power of the array in detecting homologous sequences that are not yet found in the other species.

Figure 4.8. PCR screening of a random sample of newly discovered genes. Screenings were performed pair-wise with the PCR products of the species containing the known gene of interest loaded in odd-numbered wells while the corresponding PCR screen of the other species whereby sequence is absent or the gene is not predicted are in the even-numbered wells. (1&2): PY00632 screen with Py and Pb gDNA; (3&4):PY03414 screen with Py and Pb gDNA; (5&6): PY04600 screen with Py and Pb gDNA; (7&8): PY04485 screen with Py and Pb gDNA; (9&10): PY02086 screen with Py and Pc gDNA; (11&12): PY04869 with Py and Pc gDNA; (13&14): PY06972 screen with Py and Pc gDNA; (15&16): PY05482 screen with Py and Pc gDNA. (Legend: M = 100bp DNA ladder)

4.3.3 Analysis of the remaining core syntenic gaps in the rodent malaria parasite genome with respect to the *P. falciparum* genes

Although a total of 700 genes from all 3 rodent parasite species were re-annotated, there still remain a considerable number of gaps. Given the high degree of similarity and conservation of the common core genes between the malaria parasites (Kooij, T.W. *et al*, 2006; Kooij, T.W. *et al*, 2006; Hall, N. *et al*, 2005), there should be other means to make sense of and to further improve their relationship.

The core genes would be expected to be present in all *Plasmodia* and one strategy to further reduce the number of gaps in the genome is to not restrict their relationship by looking only at synteny, but also to consider known orthologs (from OrthoMCL DB) without any established syntenic relationship. These ‘orphan’ orthologs may be absent from the synteny map due to technical and computational limits of aligning and joining random shotgun sequences together to form a contig. Therefore it is technically challenging to match the fragmented sequences of the rodent parasite genomes and the success and confidence very much depends on the sequence quality and length of these contigs. In addition, chromosomal re-arrangements within each species due to evolutionary divergence would also complicate the alignment process. Thus, in the big picture of looking into whether genes are really absent or present in a global comparison of incomplete genomes, synteny would not give a clear picture. However, by considering orthologs, a known gene could then be filled in although it was previously considered ‘missing’ due to its absence in the synteny map. In addition, evidence that regulation of transcription of individual genes occurred independently without any constraints on chromosomal location further validates this survey (Bozdech, Z. *et al*, 2008).

Figure 4.9. Graphical breakdown depicting the relationship between orthologous *Plasmodium falciparum* and the rodent malaria parasite (RMP) genes. The blue bars (status quo) represent the current orthologous relationship while the red bars (hybridization) represent the syntenic orthologous gene relationship after re-annotation of the rodent parasite genomes by microarray. The green bars (tBLASTn) represent global orthologous gene relationship after considering known orthologs and tBLASTn scores with an expect value of $<10^{-15}$. (Legend: Pfb = *P. falciparum* - *P. berghei* specific; Pfc = *P. falciparum* - *P. chabaudi* specific; Pfy = *P. falciparum* - *P. yoelii* specific; Pfbc = *P. falciparum* - *P. berghei* - *P. chabaudi* specific; Pfby = *P. falciparum* - *P. berghei* - *P. yoelii* specific; Pfcy = *P. falciparum* - *P. chabaudi* - *P. yoelii* specific; Pfbcy = present in the human and all three rodent malaria parasite species).

Bioinformatics tools were employed to further fill in the gaps by looking at the other regions of the rodent malaria parasite genomes that may have been missed by automated gene prediction algorithms. The more complete *P. falciparum* amino acid sequences were used to query the rodent malaria parasite genomes (including the latest *P. chabaudi* shotgun reads at the Sanger Institute located at http://www.sanger.ac.uk/cgi-bin/blast/submitblast/p_chabaudi) as a tBLASTn search using an expect value of at least 10^{-15} as a cut-off threshold. By using this strategy, multigene families would be collapsed, as conserved domains would link such members together. Clearly, there is a significant increase in genes that were resolved to be present in all species, from 3,353 based on current status quo to 3,886 using just microarray data to 4,278 following bioinformatics search (Figure 4.9). While the total number of genes number at 4,543 in both the current status quo and hybridization-filled gene sets, the ortholog plus tBLASTn bioinformatics approach gave a total of 4,580 genes, i.e. an increase of 37 *P. falciparum* genes now deemed to contain at least one rodent parasite ortholog (Table 4.2; Appendix B).

Following this search, only 302 *P. falciparum* genes were left with at least one missing rodent parasite species as compared to the 4,278 genes that have been resolved to be present in the four parasite species. Of the remaining 302 genes, they contain at least a single species specific RMP gene orthologous to a *P. falciparum* gene. These genes were then grouped accordingly, i.e. 14 genes specific to *P. falciparum* and *P. berghei*, 27 genes specific to *P. falciparum* and *P. chabaudi*, 12 genes specific to *P. falciparum* and *P. yoelii*, 123 genes common between *P. falciparum*, *P. berghei* and *P. chabaudi*, 10 genes common between *P. falciparum*, *P. berghei* and *P. yoelii* and 116 genes common between *P. falciparum*, *P. chabaudi* and *P. yoelii*.

PI ID	Genomic Location	Position	Pb Gene	Pc Gene	Py Gene	PF product	Orthologs / Contigs
PF11_0283	MAL11: 1,062,716 - 1,082,967 (+)	C	-	-	-	hypothetical protein	PB104912.00.0 + (chab0578e12.q1k)
PF10700w	MAL12: 611,885 - 613,207 (+)	C	-	-	-	hypothetical protein	PB106359.00.0 + (chab0234b06.q1k)
PF14_0335	MAL14: 1,433,681 - 1,434,253 (-)	C	-	-	-	hypothetical protein, conserved	(PB_PH47413) + (chab0869b12.p1k)
PF10669c	MAL4: 656,876 - 657,714 (-)	C	-	-	-	hypothetical protein	(PC_PH0686)
PF10715w	MAL12: 616,775 - 617,380 (+)	C	-	-	-	hypothetical protein	PC107908.00.0
PF10400w	MAL12: 359,990 - 360,494 (+)	ST	-	-	-	50S ribosomal protein L29, putative	(MALPY01248) + (chab0658h03.q1k)
PF11_0298	MAL11: 1,113,994 - 1,115,475 (+)	C	-	-	-	GPI8p transamidase	PY03470 + (chab0403h02.p1k)
MAL13P1.305	MAL13: 2,458,402 - 2,459,994 (+)	T	-	-	-	hypothetical protein	(MALPY00289)
PF11_0135	MAL11: 488,480 - 489,045 (-)	C	-	-	-	hypothetical protein	(MALPY01751) + (chab1000h01.q1k)
PF11_0361	MAL11: 1,372,309 - 1,373,847 (+)	C	-	-	-	hypothetical protein	(MALPY00132)
PF14_0046	MAL14: 1,700,000 - 1,700,893 (+)	T	-	-	-	hypothetical protein	(MALPY03022) + (chab0979c02.p1k)
PF14_0129	MAL14: 523,890 - 525,305 (-)	ST	-	-	-	hypothetical protein	(MALPY01822) + (chab0536e04.q1k)
PF14_0138	MAL14: 553,639 - 555,772 (-)	ST	-	-	-	hypothetical protein	(MALPY01804) + (chab0763g10.p1k)
PF10692c	MAL4: 654,372 - 655,371 (-)	C	-	-	-	hypothetical protein	(MALPY02876) + (chab0657f08.p1k)
PF10140w	MAL12: 160,713 - 161,252 (+)	T	-	-	-	hypothetical protein	(MALPY01639)
PF10215c	MAL12: 208,366 - 208,784 (-)	T	-	-	-	hypothetical protein	(MALPY00265)
PF12435w	MAL12: 2,075,776 - 2,076,501 (+)	T	-	-	-	hypothetical protein	(MALPY02265) + (chab0598d12.q1k)
MAL13P1.212	MAL13: 1,669,725 - 1,670,987 (-)	C	-	-	-	hypothetical protein, conserved	PY07307 + (chab3b07.p1k)
MAL18P1.310	MAL8: 1,289,259 - 1,289,658 (-)	C	-	-	-	putative senescence-associated protein	(MALPY02241) + (chab00089b03.p1k)
MAL13P1.252	MAL13: 1,999,663 - 2,001,465 (-)	C	-	-	-	hypothetical protein	PB001585.02.0 + PC000752.02.0
PF11_0182	MAL11: 668,367 - 669,490 (+)	C	-	-	-	hypothetical protein	PB103706.00.0 + PC103008.00.0
PF11_0437	MAL11: 1,705,693 - 1,706,570 (-)	ST	-	-	-	hypothetical protein	PB001055.02.0 + PC301115.00.0
PF11_0494	MAL11: 1,927,913 - 1,928,032 (+)	T	-	-	-	hypothetical protein	PB000829.01.0 + PC300798.00.0
PF1E110w	MAL5: 901,409 - 902,168 (+)	C	-	-	-	hypothetical protein	PB108412.00.0 + (PC_RP4524)
PF10085c	MAL12: 103,861 - 104,604 (-)	T	-	-	-	hypothetical protein	PB105504.00.0 + PC405716.00.0
PF10365c	MAL12: 343,746 - 344,356 (-)	ST	-	-	-	hypothetical protein	PB405922.00.0 + PC103204.00.0
PF10155w	MAL12: 847,430 - 848,176 (+)	C	-	-	-	hypothetical protein	PB405795.00.0 + PC100644.00.0
PF12340w	MAL12: 2001973 - 2002257 (+)	ST	-	-	-	hypothetical protein	(PB_RP0214) + PC000521.02.0
PF13_0021	MAL13: 236,002 - 236,637 (-)	T	-	-	-	small heat shock protein, putative	(PB_RP1054) + (PC_RP4679)
MAL13P1.97	MAL13: 766,636 - 767,971 (+)	C	-	-	-	hypothetical protein	(PB_RP1128) + (MALPY00481) + (chab0732d06.p1k)
PF11_0214	MAL11: 784,591 - 785,349 (-)	C	-	-	-	hypothetical protein	PB101504.00.0 + (MALPY00553) + (chab0873f12.q1k)
PF14_0554	MAL14: 2,390,662 - 2,390,838 (+)	C	-	-	-	hypothetical protein	(PB_RP0593) + (MALPY00034) + (chab0568e02.q1k)
PF14_0654	MAL14: 2,816,036 - 2,821,807 (+)	ST	-	-	-	hypothetical protein	(PB_RP1775) + (MALPY00037)
MAL13P1.288	MAL13: 2,322,727 - 2,323,188 (-)	ST	-	-	-	hypothetical protein, conserved	(PB_RP2829) + (MALPY02361) + (chab0370f09.p1k)
MAL13P1.211	MAL13: 1,645,097 - 1,645,697 (+)	C	-	-	-	hypothetical protein	(PC_RP0116) + (MALPY00877)
PF11_0285	MAL11: 1,065,783 - 1,066,567 (+)	C	-	-	-	hypothetical protein	(MALPY01905) + (PC_PH4525)
PF14_0258	MAL14: 1,095,628 - 1,097,509 (+)	C	-	-	-	hypothetical protein	PC104734.00.0 + (MALPY02326)

Table 4.2. Breakdown of the 37 additional *P. falciparum* genes found to be common to at least one rodent parasite species via orthology and tBLASTn searches. These genes currently do not have any syntenic linkage to any of the rodent parasite species and predominantly consist of hypothetical genes with the exception of four genes – a ribosomal protein (PFL0400w), a GPI8p transamidase (PF11_0298), a putative senescence-associated protein (MAL13P1.252) and a heat shock protein (PF13_0021). Chromosomal locations are indicated as T: telomeric, ST: sub-telomeric and C: central / bounded by subtelomeric regions.

Pf-Pb	Pf-Pc	Pf-Py
F49C12.11-like protein	3'-5' exonuclease, putative	-
trophozoite stage antigen	u6 snRNA-associated sm-like protein, putative	
Pf-Pb-Pc	Pf-Pb-Py	Pf-Pc-Py
CDK-activating kinase assembly factor, putative	early transcribed membrane protein 11.1	50S ribosomal protein L29, putative
Cg7 protein		60S ribosomal protein L23a, putative
conserved protein, putative		cardiolipin synthetase, putative
glutathione peroxidase		chaperone protein
glycerol-3-phosphate acyltransferase, putative		early transcribed membrane protein, putative
hemolysin, putative		ethanolamine kinase, putative
mRNA cleavage factor-like protein, putative		GPI8p transamidase
nucleoside transporter 1		homologue of human HSPC025
phosphatidylinositol synthase, putative		prohibitin-like protein, putative
phosphoglycerate mutase, putative		putative senescence-associated protein
pre-mRNA splicing factor protein, putative		ribosomal protein L35, putative
Ribosomal protein L13, putative		ribosome recycling factor, putative
ribosomal protein L7Ae-related protein, putative		RNA binding protein, putative
ribosomal protein S3, putative		sexual stage-specific protein precursor
s-adenosylmethionine-dependent methyltransferase, putative		syntaxin, putative
small heat shock protein, putative		u3 small nucleolar ribonucleoprotein protein
translation initiation factor EF-1, putative		
tryptophanyl-tRNA synthetase, putative		
ubiquitin-like protein, putative		

Table 4.3. Breakdown of the remaining annotated genes that have no orthologs in at least one other species. The full details, including hypothetical proteins and other unannotated genes can be found in Appendix B. Legend: Pf-Pb = genes common to both *P. falciparum* and *P. berghei*; Pf-Pc: genes common to both *P. falciparum* and *P. chabaudi*; Pf-Py = genes common to both *P. falciparum* and *P. yoelii*; Pf-Pb-Pc = genes common to *P. falciparum*, *P. berghei* and *P. chabaudi*; Pf-Pb-Py = genes common to *P. falciparum*, *P. berghei* and *P. yoelii*; Pf-Pc-Py = genes common to *P. falciparum*, *P. chabaudi* and *P. yoelii*.

In the groups of genes where there is only a single RMP species associated with a *P. falciparum* gene, i.e. Pf-Pb, Pf-Pc and Pf-Py, they consist mainly of hypothetical proteins and the proportion of annotated genes is 14.29%, 7.41% and none respectively. However, the groups that are missing one of three RMP orthologous genes, i.e. Pf-Pb-Pc, Pf-Pb-Py and Pf-Pc-Py though also contain a fair share of hypothetical proteins; the representation of other functionally known proteins is marginally higher in these groups and they represent 15.45%, 10% and 13.79% respectively. These proteins are of diverse functions and include ribosomal proteins, transporters, kinases, metabolic enzymes, nucleotide binding proteins, protein modifying enzymes, etc (Table 4.3; Appendix B).

The next step is to verify the confidence where the RMP genes are now predicted to be present or absent in one species or another. Although hybridization data directly revealed the presence of homologous genes, genes filled in by bioinformatics needs to be screened across the three rodent parasite genomes to confirm these associations. Basically, *P. falciparum* genes with two or only one rodent parasite ortholog need to be confirmed before any strong statements about their possible roles can be mooted. While there are now 4,278 common ‘core’ genes between *P. falciparum* and the three rodent parasite species, there is a need to confirm the absence of orthologs from another species. This is because the possibility of these genes being present in the other species cannot be discounted without some form of direct experimental screen. Similarly, genes common in two of the three rodent parasite species also need to be screened as they could possibly be also found in all three species. Regions of low sequence accuracy and gaps could potentially confound even the most stringent thresholds using bioinformatics tools. PCR screening using primers flanking the region targeted by the microarray oligonucleotide

was performed similar to the description in the previous section, i.e. a small sample population was screened in order to obtain additional confidence for the dataset. One group of genes was selected based on whether a homologous contig for a particular species is present in a region whereby a gene is not predicted, thereby checking if this gene is truly deleted from this locus or still present, being either translocated to another locus or missed by the automated contig assembly (Figure 4.10). The other group consists of species specific gaps that do not have any homologous contig information in the corresponding coding region and these PCR primers were designed based on known sequences from the other species with annotated genes (Figure 4.11). This effect of this screen is to identify gaps or assembly errors that have excluded this gene. An exception to these 2 groupings was the PCR screen for genes orthologous to PFF1480w (so far only to PB000730.00.0 & PY03519) that were screened twice as the 3'-end of this gene does not have a corresponding match to a *P. chabaudi* contig while the 5'-end of this gene overlaps somewhat with a *P. chabaudi* contig. Thus, this region was selected twice so as to screen the two scenarios as described above.

Figure 4.10. PCR screening of bioinformatics-filtered dataset with contig information.

(A) Schematic depicting the scenario whereby a contig from 1 species containing the gene of interest (shown as a hatched block) is aligned with the best corresponding hit contig from another species. In this case, the gene of interest is lost in 1 species while the flanking sequences are still present. (B) PFI0535w-PB104921.00.0 orthologs screened with (1)Pb, (2)Pc and (3)Py gDNA. PF14_0473-PC0001359.02.0 orthologs screened with (4)Pb, (5)Pc and (6)Py gDNA. MAL13P1.345-PY04218 orthologs genes screened with (7)Pb, (8)Pc and (9)Py gDNA. PFL0595c-PB301230.00.0-PC000699.01.0 orthologs screened with (10)Pb, (11)Pc and (12)Py gDNA. PFF1480w(3'end)-PB000730.00.0-PY03519 orthologs screened with (13)Pb, (14)Pc and (15)Py gDNA. PF13_0131-PC000708.04.0-PY04599 orthologs screened with (16)Pb, (17)Pc and (18)Py gDNA. (Legend: M = 100bp DNA ladder)

Figure 4.11. PCR screening of bioinformatics-filtered dataset without contig information.

(A) Schematic depicting the scenario whereby a contig from 1 species containing the gene of interest (shown as a hatched block) is aligned with the best corresponding hit contigs from another species. In this case, the gene of interest is lost in 1 species due to missing sequence information while the flanking sequences are still present. (B) PFC0095c-PB000276.02.0 orthologs screened with (1)Pb, (2)Pc and (3)Py gDNA. PFL2450c-PC000344.03.0 orthologs screened with (4)Pb, (5)Pc and (6)Py gDNA. MAL8P1.310-PY06565 orthologs screened with (7)Pb, (8)Pc and (9)Py gDNA. PFB0645c-PB000193.00.0-PC000452.02.0 orthologs screened with (10)Pb, (11)Pc and (12)Py gDNA. PFF1480w(5'-end)-PB000730.00.0-PY03519 orthologs screened with (13)Pb, (14)Pc and (15)Py gDNA. PF13_0278-PC000860.02.0-PY04659 orthologs screened with (16)Pb, (17)Pc and (18)Py gDNA. (Legend: M = 100bp DNA ladder)

Based on the random sampling approach (Figure 4.10), the PF10535w-PB104921.00.0 pair seems to be exclusively found in *P. berghei* and not in the other 2 RMP species. For the PF14_0473-PC0001359.02.0 pair, it seems that this gene is also present in *P. berghei* while the MAL13P1.345-PY04218 pair looks to be also present in *P. berghei* as well. As for the other three pairs of genes (PFL0595c-PB301230.00.0-PC000699.01.0; PFF1480w(3'end)-PB000730.00.0-PY03519; PF13_0131-PC000708.04.0-PY04599) that seem to be missing in either one of the three RMP species, PCR screening suggests otherwise and that they are indeed common to all three RMP species.

When compared to the latest *P. chabaudi* Sanger shotgun reads, it was found that PF10535w, MAL13P1.345, PFF1480w contains their respective *P. chabaudi* ortholog but the results are PCR negative for PF10535w and MAL13P1.345, suggesting that the primer binding sites are polymorphic for these two gene pairs. However, the *P. chabaudi* ortholog for PFF1480w is positive by PCR, thus confirming the presence of this gene. What these results suggest are that rodent parasite genes that are seemingly absent on a contig are actually likely to be present with high confidence in all three rodent malaria parasite species, especially when the same gene exists in at least two rodent parasite species.

Almost all of the other panel's screenings (Figure 4.11) showed that all of those genes were present in the three RMP species barring the *P. chabaudi* exclusive gene (PFL2450c-PC000344.03.0) that seemed to be present also in *P. berghei* but seemingly not in *P. yoelii*. This suggests with high confidence that the missing genes due to incomplete sequence information are actually present. Overall, it seems that it is highly

likely that missing genes due to poor sequencing coverage are present and genes common to two rodent malaria parasite species are actually present in all three species.

It therefore seems that a significant number of genes that are currently deemed to have not been discovered either by experimental or bioinformatics evidence, are really present. Since the telomeric and sub-telomeric regions of the chromosomes are typically more varied, a survey of the genes that are missing in at least one rodent malaria parasite ortholog was undertaken to see if they cluster at the ends of the chromosomal arms. Since the gene organization of *P. falciparum* is the most complete to date, the corresponding *P. falciparum* chromosomal locations of these RMP genes was chosen to analyze the probable locations of the outstanding missing RMP genes. Taking 10% of the ends of chromosomal arms as telomeric and the following 10% as sub-telomeric regions (personal communication, Neil Hall), we sought to visualize the relative regions of the chromosome whereby these genes are still missing. Using this filter to survey the genes that are missing at least one rodent parasite ortholog, 42.86% (6/14), 44.44% (12/27) and 66.67% (8/12) of *P. berghei*, *P. chabaudi* and *P. yoelii* specific genes are respectively located at the telomeric and sub-telomeric regions. Conversely, telomeric and subtelomeric located genes common to *P. berghei* and *P. chabaudi* constitutes 34.15% (42/123), those common to *P. berghei* and *P. yoelii* makes up 50% (5/10) while only 33.62% (39/116) are common to *P. chabaudi* and *P. yoelii*. Clearly, these ‘missing’ genes are scattered all over the chromosome and are mostly located in the core regions. Of interest, *P. yoelii* specific genes were more abundant in these regions and this reflects the more complete nature of its genome even in the more variant telomeric and sub-telomeric regions. On the other hand, genes that are present in two but absent in one other species

seem to be somewhat uniformly distributed along the chromosomes, suggesting that there is no bias in how efficiently genes are annotated.

Hence, core orthologous genes should be present in all *Plasmodium* spp. surveyed here and that sequence errors and gaps are uniformly distributed along the chromosomes.

4.3.4 Analysis of the rodent malaria parasite specific genes using *P. yoelii* as the index species

Having already established the high congruence of the three rodent parasite genomes to *P. falciparum*, it is more informative to then analyze the similarities and differences of the remaining subset of genes that lie beyond the conserved ‘core’ set of genes. Since the array is designed specifically to the rodent malaria parasites, this can be useful for analyzing differences and similarities between the rodent parasite species. For this purpose, instead of using genes from *P. falciparum* as an index, the oligos were chosen as the index and then replaced with the best hit *P. yoelii* gene. Next, *P. falciparum* – *P. yoelii*, *P. berghei* – *P. yoelii* and *P. chabaudi* – *P. yoelii* orthologous genes were appended to the list. This strategy effectively creates a database of genes orthologous to *P. yoelii* and the reason for choosing this species as the index is due to it being more completely annotated than the other two rodent parasite species. This list was then made less cumbersome by removing matching orthologous genes from *P. falciparum* that were already present in the common ‘core’ set; thereby creating a filtered list that contains only genes that are specific to the rodent malaria parasites.

A small proportion of these rodent specific genes still contain a *P. falciparum* ortholog. These represent rodent parasite genes that have duplicated and therefore still

match to an identical *P. falciparum* ortholog. Similarly, a tBLASTn search using the *P. yoelii* amino acid sequence against the genome of *P. falciparum* also gave significant hits for a small proportion of *P. yoelii* genes to the non-coding or non-annotated regions of *P. falciparum* (Appendix C). These genes could represent novel ORFs in *P. falciparum* especially if they are found to be transcriptionally active in the rodent parasite species.

Based on this strategy, a total of 2,898 rodent parasite specific genes were extracted from the total set of genes. Of these, 1,548 were only found in *P. yoelii*, 548 were common between *P. yoelii* and *P. berghei*, 263 were common between *P. yoelii* and *P. chabaudi* and 539 were common amongst the three species (Figure 4.12).

To further complete the orthologous association, *P. yoelii* proteins were used to probe the genome sequences of *P. berghei* and *P. chabaudi* as a tBLASTn query using an expect value of at least 10^{-15} as a cut-off threshold. As described in the previous section, potential protein coding genes missed by standard prediction algorithms can now be filled in using this strategy. As a result, the distribution is now changed with 998 genes found only in *P. yoelii*, 211 common between *P. yoelii* and *P. berghei*, 91 common between *P. yoelii* and *P. chabaudi* and 1,598 common amongst all three species. Thus there are now less *P. yoelii* specific genes with more genes now being common to all 3 species, of which the PIRs dominate the transition (Figure 4.12).

Similarly to the workflow in the previous section involving the core gene set, hybridization data was also appended onto this rodent gene specific list thus filling in additional orthologous genes. Using this strategy, there is also a significant shift in gene distribution. There are now 454 *P. yoelii* specific genes, 252 genes common to *P. yoelii* and *P. berghei*, 210 genes common to *P. yoelii* and *P. chabaudi* while there are now

1,982 genes common to all three species (Figure 4.12). This represents a major shift where there are now less *P. yoelii* specific genes and consequently more genes are now common amongst all three rodent parasite species. There are instances where potential genes have been deduced to be present via bioinformatics (i.e. via tBLASTn) but lacking in hybridization data. Sequence polymorphisms could be biased to the 60-mer region that is targeted by the oligonucleotides and therefore reducing the efficiency of hybridization. However, a broader screen via bioinformatics looks at the overall sequence composition and it is therefore not surprising that this method pulls out more common genes and collapses multigene families, especially the PIRs. These genes are thus assumed to be present even where there is no hybridization data. The following discussions will focus on their differences, with the accession numbers and gene descriptions pegged to *P. yoelii*. In addition, the PIRs, which consist of the largest multigene family and dominate the rodent specific genes, will not be discussed due to their variability of expression and gene expansion amongst the rodent malaria parasites due to its main role in antigenic variation (Cunningham, D.A. *et al*, 2005) as well as its copy number being the most abundant in *P. yoelii*. Consequently, the huge reduction in genes common in all three rodent malaria parasite species is due to the removal of the PIR genes that could also confound relationships based on bioinformatics simply due to multiple matching of common conserved domains (Figure 4.13). Of note, the start and stop codons for these genes are located on the same contig, suggesting that they are intact genes.

Figure 4.12. Distribution of the unique rodent malaria parasite (RMP) specific genes that are absent in *P. falciparum*. The initial distribution of genes was formed when all RMP genes common to *P. falciparum* were removed as these constitute the core gene set (white bars). Next, a tBLASTn using *P. yoelii* was used to query the genomes of *P. berghei* and *P. chabaudi* (hatched bars). The completed final dataset appended with hybridization data (solid black bars). (Legend: py = *P. yoelii* specific genes; pyb = genes common to *P. yoelii* and *P. berghei*; pyc = genes common to both *P. yoelii* and *P. chabaudi*; pybc = genes common to all three rodent parasite species)

Figure 4.13. Graphical plot showing the re-distribution of the rodent malaria parasite genes after removing the PIR genes. (A) This chart represents the distribution of genes after removing the PIR multigene family. (B) This chart reflects the distribution of genes after searching through the Sanger Institute's latest *P. chabaudi* shot gun reads. Most of the reduction occurs in the group y-b-c in which all the three rodent malaria parasite species share a common gene. (Legend: Py = genes specific to *P. yoelii*; Py-Pb = genes common to both *P. yoelii* and *P. berghei*; Py-Pc = genes common to both *P. yoelii* and *P. chabaudi*; Py-Pb-Pc = genes common to all three species)

This analysis can also be further expanded to further segregate genes based on their transcriptional activity in the blood stages. Basically, total RNA from density gradient purified ring, trophozoite and schizont stages were reverse transcribed and labeled. For each species, each stage was then compared with a pool of all the blood stages, thereby generating a rough transcriptome of the intraerythrocytic cycle. Duplicate hybridizations were performed and Pearson correlation of each species' progress through the broad transcriptome were 0.913, 0.924 and 0.922 for *P. berghei*, *P. chabaudi* and *P. yoelii* respectively, hence emphasizing the reproducibility of the experiments. In order to ascertain the specific stage where the transcriptional activity of a particular gene is at its peak, an arbitrary filter was set to identify the maximum and minimum median signal ratios of each stage versus the intensity of the pooled stages. In order for a gene to be considered to be transcriptionally regulated, the stage with the highest signal ratio has to be at least twice the ratio of the stage containing the lowest ratio. The obvious limitation of this as compared to a strategy involving at least twelve discrete time points would certainly involve missing genes that are not strongly transcribed and have thus failed to pass the filter. It may also be difficult to ascertain the exact stage where the transcriptional activity of a gene is deemed to be at its peak, especially when its actual peak is in-between two stages. Although this data needs to be further evaluated, it is still valuable information while we are still working on a complete synchronous rodent parasite intraerythrocytic transcriptome profile. While evidence of transcription indicates a strong potential of gene function, absence of transcription could either suggest non-function in the blood stage or that this gene is complemented by another transcriptionally active gene of similar function.

It is obvious that most of these rodent parasite specific genes are not annotated and hence their functional roles are not immediately apparent. The dataset was further screened for potential host (rodent) and bacterial contamination while also being filtered with the criteria of possessing a coding length of at least 100 nucleotides long so as to afford a higher confidence that the predicted genes represent accurate gene models.

With these attributes in consideration, 208 of the *P. yoelii* specific genes are less than 100nt in length while one gene is of bacterial origin (PY07747). For the genes common in both *P. berghei* and *P. yoelii*, 56 genes were less than 100nt in length while 30 genes were of bacterial origin. Similarly, 34 genes common to *P. chabaudi* and *P. yoelii* were less than 100nt in length while 31 were bacterial sequences. As for genes common to all three species, 60 were less than 100nt long, 83 were of bacterial origins and 2 genes (PY07579 and PY07841) were host (rodent) contaminants (Appendix C). These inaccurate gene models will not be discussed in the following section with regards to species-specific differences.

Figure 4.14. Graphical plot showing the expression profile of the rodent malaria parasite genes after removing the PIR genes. (A) *P. yoelii* specific genes; (B) *P. yoelii* and *P. berghei* specific genes; (C) *P. yoelii* and *P. chabaudi* specific genes; (D) Genes common to all three rodent malaria parasite species. ‘Accurate’ genes are free of bacterial/host sequences and the coding sequences are longer than 100 nucleotides.

Of the 429 *P. yoelii* specific genes (Figure 4.14a; Appendix C), hypothetical proteins constitute the vast majority in which 215 genes were transcriptionally regulated while the remaining 214 genes were not. The only annotated gene is a predicted membrane protein (PY02547) that incidentally does not show any form of transcriptional regulation across the broad erythrocytic developmental stages.

Of the 179 genes common to *P. yoelii* and *P. berghei* (Figure 4.14b; Appendix C), 10 of these genes (all hypothetical proteins) were not transcriptionally regulated in *P. berghei* while 46 genes were not transcriptionally regulated in *P. yoelii*. The only annotated gene that was not transcriptionally regulated in both species is a nucleosome assembly protein-related gene (PY01604). Of the remaining 78 transcriptionally regulated genes common to both *P. yoelii* and *P. berghei*, the only annotated gene codes for a cyclophilin A-related protein (PY01524).

There are 234 genes common to *P. yoelii* and *P. chabaudi* (Figure 4.14c, Appendix C). Of these, 63 of them are not transcriptionally regulated in *P. chabaudi* and the annotated genes include a leucine rich repeat protein (PY06687), an asparagine-rich protein (PY03561), a *Plasmodium vivax* PV1H14195_P-related protein (PY02647) and a hydroxyproline-rich glycoprotein (PY07680). It is difficult to speculate about their function, being transcriptionally regulated in *P. yoelii* while not in *P. berghei*. Interestingly, the *P. vivax* related gene is indeed present in the Sal-I strain but it does not have any significant homology to any *P. falciparum* gene. While there are 22 genes that are not transcriptionally regulated in *P. yoelii*, there are 72 genes that are not transcriptionally regulated in both species. The remaining 77 genes that are transcribed in

both *P. yoelii* and *P. chabaudi* include annotated genes such as a latent nuclear antigen (PY06196) and a 40S ribosomal protein s14 (PY01635).

Now, the biggest group to be discussed is the set in which these 1,307 genes are common to all three rodent parasite species (Figure 4.14d, Appendix C). Of the common parasite specific genes, those that are not transcriptionally regulated in *P. chabaudi* include an immediate early protein homolog (PY05920), a synthetic antigen of *P. falciparum* (PY07540) and 3 copies of a 93kDa protein (PY01007, PY00799, PY04687).

Of the common annotated parasite specific genes that are not transcriptionally regulated in *P. berghei* and *P. chabaudi*, they include a dentin phosphoryn (PY06516) and a glutamine-asparagine rich protein (PY06496).

Of the common genes that are transcriptionally regulated only in *P. berghei*, they include an axenomal dynein heavy chain (PY02767), an exonuclease (PY01630) and a ring-infested erythrocyte surface antigen (PY05937).

4.4 Conclusions

The availability of our in-house long oligonucleotide array specific to the rodent malaria species *P. berghei*, *P. chabaudi* and *P. yoelii* has provided a golden opportunity to verify the genomes of these rodent malaria parasite (RMP) species. The definition of the core gene set of *Plasmodia* genes using the most completely annotated genome of *P. falciparum* as a reference index has enabled the use of direct hybridization and bioinformatics tools to expand the repertoire of common genes. PCR screens of a subset of genes without at least one RMP ortholog suggest that there is high confidence that any *P. falciparum* gene that contains at least an ortholog from one of the RMP species can potentially contain orthologs from all three RMP species. This clearly indicates that their genetic repertoires are indeed more common and that these genes should be present in the poorly assembled and annotated rodent parasite genomes.

While bioinformatics tools only provide an in-silico determination of similarity using a weighted score system, a conservative expect value of 10^{-15} was chosen to increase the specificity of an accurate hit. In this study, the tBLASTn query was chosen as amino acid sequences tend to be more conserved as compared to DNA sequences, especially where silent mutations are concerned. More importantly, functional domains are conserved and hence easier to identify proteins with common function.

The construction of a rodent specific orthology map using *P. yoelii* as the reference index was undertaken to study and survey rodent parasite specific genes that are distinct from the common 'core' set. Although this process would filter off genes specific to the other two RMP species, the *P. yoelii* genome was chosen as it is the most completely annotated rodent parasite genome while the gene set in *P. chabaudi* is over-

predicted thus making comparisons across species difficult. This associative table clearly demonstrates certain species specific metabolic differences and reveals gene duplications and expansions from both the core set of *Plasmodia* genes and the rodent parasite specific genes. Similarly, hybridization data reveals more orthologous genes than the current known set. Also, additional information such as the transcriptional activity of these genes in the erythrocytic cycle is revealed, thereby facilitating the understanding of the current set of rodent specific genes.

While it may be useful to just depend on syntenic linkage maps to predict missing genes in other species, this may not be true for some cases where species-specific recombination events may have duplicated, deleted or translocated genes to different chromosomal positions. It is also worthwhile to note that the RMP genomes are still organized into contigs and while some form of order can be established by an *in silico* method of aligning RMP contigs to *P. falciparum* chromosomes, this strategy is still subject to computational limits of aligning overlapping RMP sequences together and then attempting to order these contiguous sequences by direct linkage to *P. falciparum*. It is also tempting to fill in the sequence gaps by direct PCR and sequencing over these regions, but this would be a laborious process. The design strategy of the oligonucleotides aims to probe orthologs from as many species as possible and sometimes cover different regions of the same gene, thus increasing coverage and detection as compared to a typical PCR. Intergenic regions are often difficult to amplify due to the high frequency of sequence polymorphisms and repeats. In addition, while conserved domains amongst different species are typically highly conserved, this cannot be said of the other regions of the gene, especially where there is considerable pressure to

generate diversity in the case of exported and external membrane proteins. These concerns will be addressed in the following chapter.

If a stretch of sequence in the public domain database is present in one species (e.g. *P. yoelii* 17x1.1) and absent in the other two species, the array can also detect this sequence as every annotated gene is represented by at least one oligonucleotide. Therefore, its respective target sequence in *P. yoelii* can be detected, along with any potential conserved sequences in the other two species that potentially represent the corresponding ortholog. As for any strategy of detecting sequences based on hybridization, any sequence polymorphisms would render this method sub-optimal, with performance dependent on the number and cluster of nucleotide substitutions. This scenario can thus be employed to detect differences between virulent and avirulent lines of the same species, i.e. *P. yoelii* YM and YA respectively. A positive control to verify oligonucleotide performance using the reference clone 17x1.1 can be used to normalize the signal and hence accurately determine the absence and any sequence polymorphisms.

Data from both the core and rodent specific gene datasets suggest that differences in genetic repertoire between the RMP species cannot fully account for their differences in host cell selectivity and other pathophysiological traits such as sequestration. Due to the majority of these genes having unknown functions and the low proportion and repertoire of species specific annotated genes, it is thus difficult to qualitatively associate specific genes responsible for species specific traits. Differences could possibly lie in the regulation of gene expression either at the transcription and/or post-transcriptional levels.

Due to the high AT-bias in the genome of the *Plasmodia* species, sequencing the entire genome, joining overlapping sequences to form contigs and gene predictions have been a difficult process. While the majority of the rodent malaria parasite genes have been sequenced and annotated, many gaps still remain and the data from the pan-rodent cross-genome oligonucleotide microarray provides direct experimental evidence for this case. While the genomes from *P. berghei* and *P. chabaudi* continues to be refined at the Sanger Institute, improvements on the sequence coverage of the *P. yoelii* genome has not been taken up by any group. Therefore, any improvements in genome coverage for *P. yoelii* would be extremely valuable.

More importantly, the total number of orthologous gene pairs obtained in this survey (Table 4.4) is now significantly higher than the published dataset (Hall, N. *et al*, 2005) except for the *P. berghei* and *P. chabaudi* orthologous gene pair where they have been over-predicted. The total number of orthologous pairs is calculated by the sum of the core genes together with the remaining *P. falciparum* – RMP gene permutations as well as any RMP specific genes. For example, for the *P. yoelii* vs. *P. falciparum* orthologous gene pair, the sum is 4278 (core) + 12 (Pf-Py) + 10 (Pf-Pb-Py) + 116 (Pf-Pc-Py) = 4416. These calculations would still accommodate the low confidence gene models and the contaminated sequences as these genes were also included in the original comparison done by Hall and co-workers (2005). PCR screening also strongly suggests that a *P. falciparum* gene with at least a single RMP ortholog is also likely to contain orthologs from all three RMP species. Since these genes total up to 4580, the potential maximum of orthologous genes will be computed using this assumption where the 4278 core genes is replaced by the theoretical maximum of 4580 genes.

This work augments and complements the recent manual re-annotation of *P. yoelii* genes (Vaughan, A. *et al*, 2008) and the release of the dataset presented here is intended to provide a reference resource for researchers working on the *in vivo* rodent malaria species model, especially where predicted orthologs are currently unavailable.

	<i>Pb vs. Py</i>	<i>Pb vs. Pc</i>	<i>Pc vs. Py</i>	<i>Py vs. Pf</i>	<i>Pb vs. Pf</i>	<i>Pc vs. Pf</i>
‘Core’ set	4278	4278	4278	4278	4278	4278
<3 RMP orthologs	-	-	-	138	147	266
RMP specific	1486	-	1541	-	-	-
Total	5764	4278	5819	4416	4425	4544
Current set	3153	4641	3318	3375	3890	3842
Potential Total	6066	4580	6121	4580	4580	4580

Table 4.4. Summary table of orthologs between the four *Plasmodium* species. The ‘core’ set represent orthologs common in all four species defined via hybridization and bioinformatics information. Other gene set combinations were taken from the remaining permutations of genes generated from those with at least 2 RMP orthologs as well as the rodent parasite specific genes. The total number of orthologous pairs is the sum of the core genes with the other respective gene permutations while the current set denotes the status quo (Hall, N. *et al*, 2005). The potential total indicates the theoretical maximum of common genes as there is strong suggestion that a *P. falciparum* gene with at least one RMP ortholog can potentially be orthologous to all three RMP species.

Genomic differences between the virulent and avirulent strains of the rodent malaria parasite *Plasmodium yoelii*

5.1 Abstract

Rodent malaria parasite species are extremely valuable *in vivo* experimental models in the study of pathogenesis, immune mechanisms and vaccine development. Typically, the *Plasmodium yoelii* rodent parasite species is chosen where the effect of therapeutics on a strain that causes a lethal infection such as the YM line is compared to other less virulent strains such as the 17x and YA lines. With the availability of the draft *P. yoelii* genome, genomic differences between these related lines could thus be identified. Using the pan-rodent long oligonucleotide array, differential hybridization ratios of the YM or YA line against the reference 17x.1.1 line were used in this study. Since the probes were designed based on the reference 17x1.1 line, a low ratio implies significant sequence polymorphisms while a high ratio represents gene duplications. Based on this analysis, YA seem to be more divergent than YM with respect to the reference 17x1.1 line. Genes involved in antigenic variation such as members of the PIR multigene family and also genes coding for invasion-related proteins such as members of the PY235 multigene family, AMA-1 and MSP-8 are polymorphic as their coded proteins are constantly subjected to the host immune selection pressure. A proportion of polymorphic hypothetical proteins were found to contain at least one transmembrane domain and these could potentially constitute novel parasite ligands involved in mediating host-parasite interactions and therefore subject to host immune pressure as well.

5.2. Introduction

Rodent malaria models are useful for elucidating molecular interactions between the parasite and the host. There are clearly differences in virulence amongst the different species of rodent malaria parasites with *Plasmodium berghei* and *P. vinckei*, and certain strains of *P. yoelii* and *P. chabaudi* causing lethal infections in mice. Mice infected with *Plasmodium yoelii*, *P. chabaudi chabaudi*, *P. chabaudi adami* and *P. vinckei petterei* often leads to a short acute parasitemia load before being cleared by the host animal (Sanni, L.A. *et al*, 2002). Clearly, there are similarities to the human malaria parasites as *P. falciparum* indiscriminately invades erythrocytes of all ages often leading to mortality while *P. vivax* only invades reticulocytes leading to chronic infections with lower parasitemia and less potential for mortality (Carlton, J.M. *et al*, 2008b). As the typical reticulocyte count averages around 2% of the circulating repertoire of erythrocytes (Turowski, D. *et al*, 2000), parasites that are restricted to invading reticulocytes will typically exhibit lower parasitemia loads and hence reduced morbidity. Therefore, host cell selectivity is a key determinant of virulence as reticulocyte restricted parasites only invade a small subset of erythrocytes while non-restricted parasites can typically invade erythrocytes of all ages. This phenotype has been previously demonstrated in *P. yoelii* whereby the virulent YM strain invades erythrocytes of all ages while the avirulent YA strain predominantly invades reticulocytes and thus leading to lower parasitemia loads with a higher chance of survival for the host animal (Yoelii, M. *et al*, 1975; Walliker, D. *et al*, 1976; Bungener, W., 1985; Knowles, G. and Walliker, D., 1980).

Direct evidence of a parasite specific ligand as a virulence factor responsible for host cell selection was elucidated when protective monoclonal antibodies targeted against

the 235kDa rhoptry protein of *P. yoelii* was shown to reduce the parasite burden in infected mice (Freeman, R.R. *et al*, 1980). Mice immunized with the purified PY235 protein also developed protective immunity (Holder, A.A. and Freeman, R. R., 1981). It was also shown that a challenge dose with a lethal line of *P. yoelii* parasites led to the observation that these parasites do not indiscriminately invade erythrocytes but are instead reticulocyte restricted and displaying a phenotype more typical of a non-lethal line (Holder, A.A. and Freeman, R.R., 1984).

Basic interactions such as receptor-ligand binding for mediating the invasion of new host cells clearly mirror the same interactions as seen in human malaria. For example, the 235 kDa rhoptry proteins of *Plasmodium yoelii* (PY235 multi-gene family) has been shown to be homologous to the reticulocyte binding like (RBL) or reticulocyte binding protein homolog (RH) protein family that also contains members from *P. falciparum* and *P. vivax* (Rayner, J.C. *et al*, 2000; Triglia, T. *et al*, 2001). Just like the function of PY235 in mediating host cell adaptation and immune evasion (Iyer, J.K. *et al*, 2007), the RHs in *P. falciparum* are thought to function similarly (Triglia, T. *et al*, 2005).

Similarly, the Duffy blood group antigen was utilized as the sole receptor for *P. vivax* infections as individuals lacking this blood group antigen were completely resistant to infections from *P. vivax* (Miller, L.H. *et al*, 1976; Miller L.H. *et al*, 1978; Barnwell, J.W. *et al*, 1989). Based on studies in *P. knowlesi* (Miller, L.H. *et al*, 1979), members of this Duffy-binding-like (DBL) Superfamily seem to have a role in mediating junction formation in the invasion process and homologs have been found in *P. falciparum*

(Adams, J.H. *et al*, 1992; Adams, J.H. *et al*, 2001) and *P. yoelii* (Swardson-Olver, C.J. *et al*, 2002).

Other common parasite proteins involved in invasion include the merozoite surface proteins (MSPs), of which MSP-1 is perhaps the best studied and characterized of these proteins. Both the *P. falciparum* and *P. yoelii* MSP-1 proteins have been implicated in initial attachment and antibodies directed against Pf-MSP-1 (Singh, S. *et al*, 2003) and Py-MSP-1 (Daly, T.M. and Long, C.A., 1993) both conferred protection against a challenge infection.

These examples constitute only a small sampling of experimental evidence for the elucidation of invasion-related parasite proteins and the conservation of function in both the human and rodent malaria parasites species would provide opportunities to test and confirm hypotheses across species. In addition, animal models serve as an established *in vivo* model to evaluate potential chemotherapeutic drugs, vaccine candidates and the effect of the host animals' immune response to parasite survival and proliferation. The conservation of genome order (Carlton, J.M. *et al*, 2002) and the availability of common orthologs (Hall, N. *et al*, 2005) between the human and rodent malaria parasite species both serve as useful tools in malaria research whereby the potential for elucidating gene function and ultimately virulence can be easily translated from one model to another.

Therefore, the most pertinent question with regards to virulence in the *P. yoelii* model thus hinges on the experimental design and typical studies have always focused on using a virulent strain such as YM or 17XL to compare and contrast against an avirulent strain such as AC, 17XNL or YA (Walliker, D. *et al*, 1976; Iyer, J.K. *et al*, 2007). The elucidation of the PY235 rhoptry protein as a key determinant in host cell selection and

hence virulence is best illustrated when immunized mice were able to restrict a challenge dose of a lethal line of *P. yoelii* and restrict the invasion phenotype profile similar to the non-lethal lines (Holder, A.A. and Freeman, R.R., 1984). To date, the most significant virulence factor in terms of host cell selection and parasitemia load is the finding in which individual merozoites in a schizont are clonally distinct for a particular PY235 member protein (Preiser, P.R. *et al*, 1999) so as to ensure the survivability of the parasite under immune pressure; and that higher expression of this multigene family increases the potential for invading host cells of all ages (Iyer, J.K. *et al*, 2007). This phenomenon is also seen in *P. falciparum* whereby multiple RBLs were differentially expressed in different parasite lines (Duraisingh, M.T. *et al*, 2003) and that expression of PfRH4 was up regulated in parasites that have switched from a sialic acid-dependent to a sialic acid-independent invasion pathway (Gaur, D. *et al*, 2006; Stubbs, J. *et al*, 2005). However, it is more difficult to link expression of distinct RBL and EBL members of *P. falciparum* to virulence due to the lack of a human *in vivo* model.

Nonetheless, there is still significant evidence available to correlate virulence in terms of host cell selection between the human and rodent malaria parasite species. Recent work clearly supports the use of the virulent and avirulent strains of *P. yoelii* (Iyer, J.K. *et al*, 2007) as models for correlating virulence to a multigene family homologous to the reticulocyte binding protein homolog superfamily.

With the completion of the *P. yoelii* genome (Carlton, J.M. *et al*, 2002) and the presence of significant homology to the human malaria parasite proteins, this study aims to firstly determine differences at the genomic level between the avirulent YA strain of *P. yoelii* and the virulent YM strain. Microarray hybridizations using genomic DNA from

these 2 lines were used to differentiate gene polymorphisms and potential gene gain/loss and their implications will be discussed.

5.3 Results and discussion

5.3.1 Genomic differences between the avirulent YA strain and the virulent YM strain of *P. yoelii*

As mentioned in the previous chapter on the genomic differences between the three rodent malaria parasite species, the *P. yoelii*-specific probes from our in-house pan rodent oligonucleotide array was designed based on the reference *P. yoelii* 17x1.1 clone and was evaluated as described in the Chapter 4. Data from these oligonucleotides were used for this study. Since there are potential sequence polymorphisms amongst different parasite strains, any comparison between the avirulent YA and the virulent YM strains of *P. yoelii* have to be normalized against the signal generated by the reference 17x1.1 line. Basically, duplicate hybridizations of either YA or YM against 17x1.1 genomic DNA were performed and the data normalized across individual chips.

The sensitivity of using microarray for detecting polymorphisms is well documented especially with recent reports that single nucleotide polymorphisms (SNPs) were readily detectable where the net difference in intensity between identical and polymorphic sequences were at least five-fold (Zhang, G.Q. *et al*, 2008) and that a SNP near the middle of the oligonucleotide probe typically had the effect of halving the signal intensity as compared with a perfect hybridization, with the effect decreasing linearly as the SNP approaches either end of the oligonucleotide probe (Benovoy, D. *et al*, 2008). In addition, the discovery that a cut-off ratio of just 1.5-fold difference between identical and polymorphic sequences was sufficient to differentiate polymorphic genes clearly demonstrates the sensitivity of using such direct hybridization methods (Jiang, H. *et al*, 2008).

Genes with ratios under 0.5 would imply that these sequences would contain at least a single nucleotide mismatch (Benovoy, D. *et al*, 2008); while those with ratios above 2 would likely to constitute gene duplication / expansion with respect to the reference line (Figure 5.1). Only data from *P. yoelii* specific oligos were chosen for this analysis. As expected, most of the net signal lies around the scatter as determined by ratios of less than two-fold difference, i.e. ratios of either YM (Figure 5.2) or YA (Figure 5.3) against 17x1.1 of between 0.5 and 2. Although the original data was filtered accordingly to a threshold of twice the local background and twice of its standard deviation, this dataset was further filtered by considering only oligos that exhibit signal that was significantly above background in order to reduce comparison errors due to certain oligos that do not perform very well. Hence, an arbitrary cut-off value of 300 was chosen as this is at least thrice the signal of an oligo deemed to have poor hybridization with respect to 17x1.1 as this line is the reference to be used as a comparison.

Although the data generated is essentially genomic comparisons between different parasite lines with respect to 17x1.1, it provides a means to directly compare genomic differences between the YA and YM lines. The oligonucleotide probes on the array are designed to the sequence of 17x1.1 and therefore the competitive hybridizations with this reference line provides a means to normalize hybridization signals of the other lines. In summary, a *P. yoelii* specific probe must firstly be able to show significant hybridization (signal > 300) to 17x1.1 before it can be considered for this comparison. Subsequently, the bias in hybridization ratios between YA or YM to 17x1.1 can thus be used to elucidate genomic differences.

Figure 5.1. A schematic illustrating the possible scenarios involved in the comparison of genomic sequences from different parasite lines. (A) In the case where a particular gene from the query species (green) is identical to the one from the reference species (red) then they will both hybridize equally to the probe. (B) If a gene from the query species (green) contains sequence polymorphisms, this will hybridize less efficiently than the gene from the reference species (red). (C) When a gene is duplicated or has undergone gene expansion in the query species (green), multiple copies of this sequence out-compete with those in the reference species (red) and therefore hybridize more efficiently to the probe.

Figure 5.2. Semi-logarithmic curve depicting the distribution of genes with respect to their similarity between *P. yoelii* YM line with respect to those in the 17x1.1 line. Clearly the vast majority of the genes are very similar to each other in terms of sequence. The 134 genes that exhibit at least a two-fold increase in signal from that of 17x1.1 are likely to constitute gene duplications / expansions. Conversely, the other 198 genes that show at least a two-fold decrease in signal could possibly mean significant sequence polymorphisms or deletions between these 2 lines. The high quality and reproducibility of these comparative genome hybridizations is emphasized with a Pearson correlation score of 0.933 and 0.953 for duplicate slides representing 17x1.1 and YM respectively.

Figure 5.3. Semi-logarithmic curve depicting the distribution of genes with respect to their similarity between *P. yoelii* YA line with respect to those in the 17x1.1 line. Clearly the vast majority of the genes are very similar to each other in terms of sequence. The 86 genes that exhibit at least a two-fold increase in signal from that of 17x1.1 are likely to constitute gene duplications / expansions. Conversely, the other 637 genes that show at least a two-fold decrease in signal could possibly mean significant sequence polymorphisms or deletions between these 2 lines. The high quality and reproducibility of these comparative genome hybridizations is emphasized with a Pearson correlation score of 0.977 across duplicate slides for both 17x1.1 and YA.

A general consensus of this survey is the fact that most of the gene pairs (YM vs. 17x1.1 or YA vs. 17x1.1) do seem to be clustered around this region of parity (between 0.5 to 2), thus suggesting that a significant proportion of genes are indeed very similar and therefore do not have much sequence variation. Hence, outliers either suggest significant polymorphisms or gene expansions in either YA or YM with respect to 17x1.1. Before any comparisons can be made, we assume that the sequence data for 17x1.1 is accurate. This is because the oligonucleotides on the array are designed to this reference line and that genomic DNA from this line is used as means for normalization of data from YM and YA. It is with this assumption in mind that genes with hybridization ratios higher than 2.0 are indeed likely to represent gene duplications and conversely, ratios less than 0.5 represent significant sequence polymorphisms with respect to 17x1.1.

(A)

Gene function	No. of genes
PIRS	339
Hypothetical proteins	222
DNA binding	11
RNA binding	7
Invasion	4
Enzymes	13
Others	41

(B)

Gene function	No. of genes
PIRS	24
Hypothetical proteins	53
DNA binding	1
RNA binding	2
Transport	1
Unknown	2
Others	3

Table 5.1. Summary tabulation of the repertoire of genes in YA that are likely to be significantly different from those in the reference 17x1.1 line. Table (A) illustrates the repertoire of the 637 polymorphic genes in YA while (B) represents the repertoire of 86 duplicated genes in YA as compared to the reference 17x1.1 line. The invasion-related proteins in table (A) include two PY235 members, the apical membrane antigen 1 and a merozoite surface protein precursor / MSP8.

Of the 637 genes in YA that show significant sequence polymorphism (Table 5.1a; Appendix D), the majority consists of 222 hypothetical proteins (34.85%) and 336 PIRs (52.75%). While the exact functions of these hypothetical proteins are as yet unknown and unproven, they could potentially constitute actual genes with critical functions, as they seem to have undergone evolutionary pressure and have adapted as such. Similarly, the PIRs are expressed on the surface of the infected erythrocyte and this multigene family is thus implicated in antigenic variation so as to evade the host immune system (Cunningham, D.A. *et al*, 2005). Clearly, the PIRs would be expected to constitute the vast majority of genes under selection pressure so as to generate new combinations of sequences to outwit the host immune system. Their sub-telomeric location also suggests that recombination is a key mechanism for generating such sequence polymorphisms. Interestingly, a 235kDa rhoptry protein (PY04930) and another rhoptry protein (PY01365) were also found, though PY01365 was marginally included in this group with a ratio of only 0.48. A quick comparison with the PY01365 gene in YM (ratio 0.52) confirms that this dominant member of the PY235 multigene family (Iyer, J.K. *et al*, 2007) is under constant immune pressure in both YA and YM lines. Similarly, another member of the PY235 multigene family, PY04930, showed significant sequence polymorphism with a ratio of 0.19. This observation further confirms that it is also under constant immune pressure. Also, its transcriptional profile was shown to be higher in YA-infected rodents than in YM-infected rodents (Iyer, J.K. *et al*, 2007), thus implying a more involved role in YA infections and therefore also being subject to immune pressure. Sequence polymorphisms of invasion related ligands have also been observed in *P. falciparum* isolates. For example, nucleotide substitutions in region II of the BA erythrocyte binding

ligand (BAEBL) in *P. falciparum* were shown to exist in different clones from various parts of the world and that this variation enabled distinct parasite clones to bind to different subsets of erythrocyte receptors (Mayer, D.C. *et al*, 2002). This discovery thus demonstrates the potential of the parasite for developing genetic diversity in an erythrocyte binding ligand.

Interestingly, the apical membrane antigen-1 (PY01581) with a ratio of 0.26 was significantly different from the sequences in 17x1.1 and also to YM (ratio 1.08). This protein is involved in binding (Fraser, T.S. *et al*, 2001), apical reorientation (Mitchell, G.H. *et al*, 2004) and invasion of erythrocytes (Triglia, T. *et al*, 2000). Therefore, it is a universal ligand in all malaria parasitic infections and this change in YA suggests this molecule is under constant immune selection pressure. It has been recently shown that the apical membrane antigen-1 (AMA-1) gene in *P. falciparum* field isolates contained distinct genetic polymorphisms, particularly between Venezuelan and Southeast Asian isolates (Escalante, A.A. *et al*, 2001). This study reveals the effect of natural selection for maintaining the genetic diversity of for AMA-1. Similarly, sequence polymorphisms in a merozoite surface protein precursor, MSP-8 (PY06415), mature-parasite-infected erythrocyte surface antigen (PY02239) and a putative ring-infested erythrocyte surface antigen (PY02776) also suggest that the products of these genes are under constant immune selection pressure. It is not surprising that merozoite surface proteins are targets of immune surveillance and generation of variants are indeed advantageous to the parasite. Field samples of *P. falciparum* showed polymorphism of the merozoite surface protein-1 in different geographical locations (Mamillapalli, A. *et al*, 2007) while there is also evidence for polymorphisms in field isolates of the merozoite surface protein-3 of *P.*

vivax (Cristiano, F.A. *et al*, 2007). Similarly, parasite modified erythrocyte surface molecules such as Pf155/RESA are readily exposed to immune pressure, with the amino acid sequence of field samples being highly variable and thus casting doubts on its potential as a vaccine candidate (Sueblinvong, T. *et al*, 1996).

Of the other genes that are not related to invasion, their repertoire seems to be varied and most of the annotated genes are actually redundant (e.g. eight CCAAT-box DNA binding protein subunit B genes; three PfEMP3 genes) as they already appear in multiple copies in the genome and a high proportion of them are transcribed in the blood stage. It is not clear what the exact function of these PfEMP3-like genes are in *P. yoelii* though this annotation suggests that these proteins are likely to be localized on the surface of the parasitized erythrocyte and thus subject to immune pressure. As for the DNA binding proteins, the existence of polymorphic forms is even more puzzling if the annotation is to be believed. In fact, the only CCAAT-box DNA binding protein subunit B with a functional domain match is PY00689 and the predicted protein contains an AP2 transcription factor-like domain. This gene model seems to be accurate as syntenic orthologs are also present in *P. berghei*, *P. chabaudi*, *P. falciparum*, *P. vivax* and *P. knowlesi*. The existence of multiple copies of these genes and the lack of predicted or functional validation definitely hinders any sort of speculation on their actual function. Since most of the other annotated genes are responsible for cellular metabolism processes and not in invasion, we can speculate that these mutations are not at the active site and that these proteins are still functionally active. Alternatively, even if function is completely abrogated, the other duplicated genes would adequately fill the metabolic needs of the parasite. Furthermore, 419 of these genes contain at least a single

transmembrane domain, thus suggesting that these proteins might be exposed to the host immune system. Interestingly, only 57 of these genes contain at least a single transmembrane and signal peptide motif and these proteins would be of high confidence of being exposed to the host immune surveillance. Thus these genes are highly likely to be involved in some form of parasite-host interaction and/or modification of the parasite/host cell surface membranes. *P. yoelii* genes containing the *Plasmodia* protein-export motif were previously elucidated (Sargeant, T.J. *et al*, 2006) and only five hypothetical genes (PY00576, PY01602, PY03405, PY04734 and PY06581) possess this motif.

For the 86 genes that have potentially duplicated or expanded in YA (Table 5.1b; Appendix E), they are again mostly made up of 53 hypothetical proteins (61.63%) and also 24 PIRs (27.91%). These 2 groups of genes already constitute 90% of genes that have expanded in YA, while the remaining 10% consists of genes of diverse functions such as a transporter, nucleic acid binding proteins and a chloroquine resistance marker protein. The ‘unknown’ genes represent an uncharacterized protein (PY0902) that is present in *P. berghei* and *P. chabaudi* while an unknown-related protein (PY07738) is a bacterial contaminant. Genes annotated as ‘unknown’ typically code for proteins that have no known function either via bioinformatics or via experimental evidence. In contrast, genes annotated as ‘hypothetical’ represent predicted proteins whose expression has not been demonstrated. Exactly half of these genes (43/86) contain at least a single transmembrane domain, suggesting that these genes are likely to be exposed to the host immune system. However, none of these genes possess a signal peptide nor a PEXEL motif.

For YM, there are significantly less genes that are deemed to be polymorphic to the reference 17x1.1 line. There are only 198 polymorphic genes in YM (Table 5.2a; Appendix F) as compared to 637 polymorphic genes in YA. Similarly, hypothetical proteins make up the bulk and in this case, constituting exactly half (99) of this group. Again, the 79 PIRs dominate (39.90%) and are expected due to their role in antigenic variation. The remaining 10% includes genes of diverse function but an interesting one is a PY235 E5 protein (PY06381) with a ratio of 0.10, thus suggesting that this gene is highly polymorphic in YM. Other polymorphic genes are mainly involved in metabolic functions such as a protein phosphatase. Half of these genes (99/198) contain at least a single transmembrane domain while only 22 contain both transmembrane and signal peptide motifs, suggesting an increasing level of confidence that the respective proteins are likely to be exposed to the host immune surveillance. Only one hypothetical gene (PY05454) contains a PEXEL motif characteristic of exported parasite proteins.

While there are no overlapping genes from YM or YA that are both polymorphic with respect to 17x1.1, the PY01365 member of the PY235 rhoptry protein family comes close with a ratio of 0.48 in YA/17x1.1 and a ratio of 0.52 in YM/17x1.1. This is expected, as a dominant member of this multigene family is extremely likely to be constantly under immune pressure.

(A)

Gene function	Number of genes
PIRS	79
Hypothetical proteins	99
DNA binding	5
RNA binding	2
Invasion	1
Unknown	2
Others	10

(B)

Gene function	Number of genes
PIRS	26
Hypothetical proteins	81
DNA binding	3
Enzymes	3
Transport	2
Signaling	4
Others	15

Table 5.2. Summary tabulation of the repertoire of genes in YM that are likely to be significantly different from those in the reference 17x1.1 line. Table (A) illustrates the repertoire of the 198 polymorphic genes in YM while (B) represents the repertoire of the 134 duplicated genes in YM as compared to the reference 17x1.1 line. The invasion-related protein in table (A) is a PY235 E5 protein.

For the 134 genes that have expanded in YM (Table 5.2b; Appendix G), again the hypothetical proteins dominate, numbering in at 81 (60.45%) while the 26 PIRs (19.40%) also make up the majority of the genes in this group. In addition, genes that are known to exist in multiple copies also show up, namely the 93kDa proteins (PY01334; PY06147) and PfEMP3 (PY02872). This suggests that further gene expansions have occurred in YM. Furthermore, 62 genes contain at least a single transmembrane domain while 14 genes contain both a transmembrane domain as well as a signal peptide, suggesting an increasing level of confidence that the respective proteins are likely to be exposed to the host immune surveillance. Only one hypothetical gene (PY02662) contains a PEXEL motif characteristic of exported parasite proteins.

While these are differences between YA or YM to 17x1.1, these data are also essentially differences between YA and YM. While the normalized ratios of YM:17x1.1 and YA:17x1.1 can be combined to be expressed as a YM:YA ratio, we can only assume that any bias in hybridization is due to sequence polymorphisms since this method does not account for any potential gene expansions in either parasite lines. Similarly, a relative hybridization ratio of less than 0.5 (Appendix H) and more than 2.0 (Appendix I) represents significant genetic polymorphisms between these two lines. Basically, a low ratio of <0.5 represents genes from YA that preferentially bind to the probe and hence these genes are more polymorphic in YM. Conversely, a high ratio of >2 represents genes from YM that preferentially bind to the probe than YA and hence these genes are more polymorphic in YA. When compared to the gene sets obtained in the previous section where there are 198 polymorphic genes in YM and 637 polymorphic genes in YA with respect to 17x1.1 for a total of 835 polymorphic genes. Based on the gene sets in

Appendix H and I, a total of 892 genes were polymorphic between YA and YM in which 174 are polymorphic in YM and 712 are polymorphic in YA. This is also in agreement with the previous comparison with 17x1.1 in which YA is deemed to be the more variant line. In summary, almost all of the polymorphic genes present in the previous YA:17x1.1 or YM:17x1.1 comparisons are present while the extra genes are comprised of genes that have been filtered out but have now been included because of a subsequent round of normalization, i.e. a gene with a YM:17x1.1 ratio of 0.7 while having a YA:17x1.1 ratio of 1.5, while both not considered polymorphic, a YM:YA ratio ($0.7/1.5 = 0.47$) now picks this gene up as being significantly polymorphic (<0.5) between YA and YM.

In summary, 387 PIRs (43.39%) and 398 hypothetical proteins (44.62%) still dominate the total repertoire of genes with significant sequence polymorphisms between YA and YM. Incidentally, 202 genes (highlighted in bold in Appendices H and I) were newly included as compared to the previous tables that contained genes with low YM or YA:17x1.1 ratios. To note, the dominant member of the PY235 multigene family (PY01365) was previously shown to be polymorphic in YM and YA as compared to 17x1.1. This gene is now not picked up by this new round of filter as their ratios are now considered to be equivalent and therefore do not show any bias in competitive hybridizations. An interesting addition to this group is the *maebl* gene (PY03020) where the sequence in YA is more polymorphic than in YM with a YM:YA ratio of 2.21. This would make sense as an invasion-related parasite ligand is obviously targeted by the host immune system. However, it is not clear why this gene is more polymorphic in YA than in YM. Conservatively, 74 polymorphic YM genes and 438 polymorphic YA genes contain at least one transmembrane domain. More stringently, 22 polymorphic YM genes

and 73 polymorphic YA genes contain both transmembrane and signal peptide motifs. Of these genes six hypothetical genes (PY05454, PY00576, PY01602, PY03405, PY04734, PY06581 and PY07057) contain a PEXEL motif characteristic of exported parasite proteins.

5.3.2 Identification of hypothetical proteins that are potentially involved in mediating host-parasite interactions

The genomes of *Plasmodia* are composed of many hypothetical proteins. While many studies have focused on clearly annotated parasite genes such as those involved in invasion, metabolism and antigenic variation, functional studies on the hypothetical proteins were very much ignored.

The genomic comparisons that were performed in the previous section present a golden opportunity to prioritize the study of these hypothetical proteins. Based on the idea that variant genes are due to host selection pressure, hypothetical proteins were selected from the variant gene sets. Also, these variant *P. yoelii* hypothetical genes were then queried at PlasmoDB under the cellular location domains search to isolate genes that code for proteins that contain at least one transmembrane domain. The resultant list of proteins would then be of high confidence to be exposed to the host immune system. Functionally, these short-listed hypothetical proteins could possibly be involved in immune evasion or act as novel parasite ligands. Where available, *P. falciparum* orthologs were appended to the list to differentiate common or rodent parasite specific genes. Also, the peak transcriptional stage in the blood stage was added for additional information.

(A)

Py Acc. No.	av ya:17x	Predicted GO function	Pf ortholog	Peak Py transcription
PY00178	0.194108369	-	PFL0790w	R
PY00307	0.377734382	-	-	-
PY00330	0.336667749	-	-	-
PY00575	0.225694457	-	-	-
PY00576	0.381926774	-	-	-
PY00596	0.390510258	-	PF07_0070	R
PY00600	0.334071537	-	PF10_0199	R
PY00636	0.318472315	-	-	T
PY00890	0.297404577	metal ion transmembrane transporter activity	PF11_0210	-
PY01006	0.367198783	-	-	T
PY01160	0.29207903	-	-	R
PY01176	0.075542889	-	-	-
PY01236	0.187569472	-	-	-
PY01507	0.367737899	-	PF13_0202	S
PY01602	0.115745159	-	-	T
PY01650	0.09371254	-	-	-
PY01686	0.370226234	-	-	-
PY01809	0.363633976	-	PFL2390c	R
PY01975	0.239344087	-	-	-
PY01978	0.114356127	-	-	T
PY01985	0.140164506	-	-	-
PY02078	0.344163752	-	PF10_0290	T
PY02152	0.396349484	-	MAL8P1.13, PF11_0172	-
PY02184	0.053544365	-	-	-
PY02199	0.456421645	binding	PFD0720w	S
PY02934	0.427348591	-	PFE0260w	T
PY03061	0.485024321	-	-	-
PY03297	0.484509899	-	-	S
PY03322	0.267079504	-	-	S
PY03340	0.043442739	-	-	-
PY03369	0.44293253	-	PF11_0073	R
PY03405	0.243933825	-	-	S
PY03430	0.182946738	-	-	T
PY03484	0.48717702	protein binding	PFL0505c	S
PY03507	0.279273706	-	-	R
PY03657	0.093601043	-	-	R
PY03708	0.389063568	-	-	R
PY03848	0.086627988	-	-	-
PY03952	0.158979174	-	-	-
PY03968	0.271069025	-	-	R
PY03993	0.392984054	-	PFE0580w	T
PY04109	0.447354193	-	-	-
PY04123	0.268393569	-	-	-
PY04210	0.412658782	-	-	R
PY04316	0.337571052	-	-	R
PY04355	0.384750173	-	PF11295c, PFB0465c	S
PY04398	0.494097902	-	-	-
PY04420	0.260083489	-	-	-
PY04683	0.221740654	-	PF08_0090	-
PY04734	0.418129803	-	-	-
PY04866	0.335267466	-	PFD0555c	R
PY05202	0.284188317	-	PF10_0188	R
PY05240	0.445269976	-	-	-
PY05541	0.347113846	protein-S-isoprenylcysteine O-methyltransferase activity	PFL1780w	R
PY06021	0.182617005	-	-	-
PY06103	0.429702572	-	-	-
PY06159	0.308970931	-	-	R
PY06317	0.220874221	-	-	T
PY06581	0.125703094	-	-	R
PY06617	0.26386137	-	-	-
PY06664	0.345995212	-	PF10_0205	T
PY06731	0.322557754	-	-	-
PY06881	0.204395126	-	-	R
PY07087	0.407620455	-	-	-
PY07167	0.330304671	-	-	-
PY07217	0.241242804	-	-	T
PY07231	0.114978059	-	-	-
PY07242	0.474070092	-	-	-
PY07559	0.153392764	-	-	-
PY07562	0.287003482	-	-	T
PY07608	0.216201258	-	-	-
PY07633	0.217667375	-	-	-

(B)

Py Acc. No.	Mean YM:17x	Predicted GO function	Pf ortholog	Peak Py transcription
PY00330	0.33618515	-	-	-
PY00439	0.076385457	-	-	-
PY00440	0.24526414	-	-	T
PY01904	0.474621479	-	-	R
PY02429	0	-	PFL0555c	R
PY02985	0.192291066	-	PFB0600c	R
PY04315	0.09164679	-	-	R
PY04316	0.417417141	-	-	R
PY04317	0.140477275	-	-	R
PY04963	0.081994376	-	-	-
PY05408	0.42831017	-	-	R
PY05454	0.035839501	-	-	-
PY07279	0.114526325	-	-	-
PY07504	0.049167083	-	MAL7P1.142	T
PY07551	0.116545783	-	-	-
PY07592	0.030410584	-	-	T
PY07649	0.050537881	-	-	T
PY07665	0.0383869	-	-	-
PY07670	0.27977027	-	-	-
PY07671	0.039966286	-	-	-
PY07828	0.206175985	-	-	-
PY07837	0.047655973	-	MAL7P1.142	-
PY07845	0.187619143	-	-	-

(C)

Py Acc. No.	Mean YM:YA	Predicted GO function	Pf ortholog	Peak Py transcription
PY02299	0.492596749	-	-	R
PY04579	0.442542375	-	-	T
PY04938	0.287073979	-	PF11_0276	R
PY05096	0.440543927	-	MAL7P1.17	T
PY05506	0.454352046	-	-	-
PY07814	0.479394916	-	-	T
PY00311	2.552301779	-	-	-
PY00581	2.664054413	-	PFL1065c	R
PY00677	2.108088817	-	PFL1415w	T
PY01073	2.656131632	-	MAL8P1.130	T
PY01083	2.562563701	-	-	-
PY01141	2.136184479	-	-	-
PY01234	2.280017691	-	PF13_0235	R
PY01414	3.071843032	-	-	R
PY01729	2.485597093	-	-	R
PY01789	2.064912656	-	-	-
PY02043	2.063701085	-	-	-
PY02211	3.05954398	-	PFF1075w	R
PY02392	2.197184933	-	-	-
PY02399	2.528146419	-	-	-
PY02731	2.006309963	-	-	-
PY02989	2.453868773	-	PFB0580w	R
PY03129	2.160029526	-	PF10465c	T
PY03407	2.443249842	-	MAL13P1.87	S
PY03590	2.011107828	-	PFL1820w	R
PY03762	3.178675147	-	-	-
PY03959	2.052166214	-	PF14_0094	T
PY04022	2.713740119	-	-	R
PY04348	2.024665921	-	-	S
PY04401	2.857620214	-	PF11_0176	S
PY05323	2.19387098	-	PF0595w	T
PY05347	3.331712148	-	MAL13P1.49	T
PY05663	4.691738852	-	MAL8P1.28	S
PY05704	2.037269195	-	-	-
PY05723	2.739903367	-	-	R
PY06093	2.028908894	-	-	R
PY07034	2.604149852	-	-	T
PY07235	2.156657652	-	-	R

Table 5.3. List of hypothetical proteins containing transmembrane domains with potential roles in host-parasite interactions. (A) Genes more variant in YA than in 17x1.1; (B) Genes more variant in YM than in 17x1.1 (C) Additional genes more variant in YM (Mean YM:YA<0.5) or YA (Mean YM:YA>2.0)

Firstly, 72 hypothetical proteins were selected from the gene set that contains genes more polymorphic in YA than 17x1.1 (Table 5.3a). Interestingly, one gene (PY00890) is predicted to be a metal ion transporter with an ortholog in *P. falciparum* (PF11_0210). However, the failure to detect any regulated transcription in the blood stage may mean that this gene functions in the other stages. More importantly, both PY02199 and PY03484 are predicted to be involved in binding, contain a corresponding *P. falciparum* ortholog (PFD0720w and PFL0505c) and are with a transcriptional peak at the schizont stage. The predicted PY02199 protein is rich in ARM repeats and would be expected to be involved in protein-protein interactions while the predicted PY03484 protein is a SNARE protein and this class of proteins are believed to mediate most, if not all, cellular membrane fusion events. This suggests that these molecules are involved in late stage schizont maturation and might be subjected to some form of immune pressure as they might be exposed to immune pressure. The PY05541 hypothetical protein is predicted to function as a methyltransferase and has an ortholog in *P. falciparum* (PFL1780w) with a peak transcriptional profile at the ring stage. It would be interesting if this unique enzyme that is further characterized and shown to be exposed on the surface of the parasite. Of these 72 genes, 21 were transcriptionally active and are found specifically in *P. yoelii*.

Next, there are 23 hypothetical proteins that are more polymorphic in YM than in 17x1.1 (Table 5.3b). Three genes (PY02429, PY02985 and PY07504) are actively transcribed in the blood stage and possess a *P. falciparum* ortholog while 8 other genes were transcriptionally active and are found specifically in *P. yoelii*. Interestingly, the

mean YM:17x1.1 ratio for the PY02429 gene is 0, suggesting that this gene is absent or extremely polymorphic in YM.

Lastly, additional genes coding for hypothetical proteins not found in the previous comparisons to 17x1.1 were also selected. The YM:YA ratio comparison of hypothetical proteins containing transmembrane domains (Table 5.3c) yielded 6 genes (mean YM:YA < 0.5) that are more polymorphic in YM. Of these, two (PY04938 and PY05096) were transcriptionally active in the blood stages and are orthologous to *P. falciparum* while 3 others were transcriptionally active and are found specifically in *P. yoelii*. Similarly, 32 hypothetical proteins that were more polymorphic in YA (mean YM:YA > 2) and containing at least one transmembrane domain were discovered, with 14 of them being transcriptionally active in the blood stages and are orthologous to *P. falciparum*. Of these 32 genes, 8 were transcriptionally active and are found specifically in *P. yoelii*.

Based on this short survey, a total of 133 polymorphic hypothetical proteins containing transmembrane domains have been identified. Genes that showed some form of transcriptional regulation in the blood stage could function in host-parasite interactions while the presence or absence of a corresponding ortholog in *P. falciparum* suggest that this gene is either potentially common to other species or specific to *P. yoelii*.

Similar to the previous sections, these polymorphic genes were then further filtered to finally yield 70 genes that contain both transmembrane and signal peptide motifs (Table 5.4). The identification of these genes can pave the way to identify novel candidate genes responsible for either immune evasion or host-parasite interactions. Transcriptionally active genes that have a *P. falciparum* ortholog are perhaps more functionally important than the other variant *P. yoelii* specific hypothetical genes as

common orthologous genes imply some form of evolutionarily conserved function and are therefore good candidates as novel targets for therapeutics.

Py Acc. No.	Mean YM:YA	Predicted GO function	Pf ortholog	Peak Py transcription	PEXEL	Comment
PY00178	0.194108369	-	PFL0790w	R	-	
PY00311	2.552301779	-	-	-	-	<100nt
PY00440	0.24526414	-	-	T	-	
PY00575	0.225694457	-	-	-	-	
PY00596	0.390510258	-	PF07_0070	R	-	
PY00677	2.108088817	-	PFL1415w	T	-	
PY01073	2.656131632	-	MAL8P1.130	T	-	
PY01141	2.136184479	-	-	-	-	
PY01176	0.075542889	-	-	-	-	
PY01236	0.187569472	-	-	-	-	
PY01602	0.115745159	-	-	T	NRILADA	
PY01729	2.485597093	-	-	R	-	
PY02043	2.063701085	-	-	-	-	<100nt
PY02299	0.492596749	-	-	R	-	
PY02392	2.197184933	-	-	-	-	
PY02399	2.528146419	-	-	-	-	
PY02731	2.006309963	-	-	-	-	<100nt
PY03322	0.267079504	-	-	S	-	
PY03369	0.44293253	-	PF11_0073	R	-	
PY03405	0.243933825	-	-	S	NRILGDV	
PY03430	0.182946738	-	-	T	-	
PY03657	0.093601043	-	-	R	-	
PY03959	2.052166214	-	PF14_0094	T	-	
PY04022	2.713740119	-	-	R	-	
PY04317	0.140477275	-	-	R	-	
PY04398	0.494097902	-	-	-	-	
PY04420	0.260083489	-	-	-	-	
PY04579	0.442542375	-	-	T	-	
PY04866	0.335267466	-	PFD0555c	R	-	
PY04963	0.081994376	-	-	-	-	
PY05347	3.331712148	-	MAL13P1.49	T	-	
PY05454	0.035839501	-	-	-	DRILSDA	
PY05541	0.347113846	protein-S-isoprenylcysteine O-methyltransferase activity	PFL1780w	R	-	
PY06021	0.182617005	-	-	-	-	
PY06093	2.028908894	-	-	R	-	
PY06581	0.125703094	-	-	R	NRILGDV	
PY07087	0.407620455	-	-	-	-	
PY07167	0.330304671	-	-	-	-	<100nt
PY07279	0.114526325	-	-	-	-	
PY07504	0.049167083	-	MAL7P1.142	T	-	
PY07551	0.116545783	-	-	-	-	
PY07559	0.153392764	-	-	-	-	
PY07592	0.030410584	-	-	T	-	
PY07608	0.216201258	-	-	-	-	
PY07670	0.27977027	-	-	-	-	
PY07671	0.039966286	-	-	-	-	
PY07828	0.206175985	-	-	-	-	
PY07837	0.047655973	-	MAL7P1.142	-	-	
PY00143	2.814839169	-	MAL7P1.208	T	-	
PY00237	0.337200779	-	-	-	-	
PY00813	0.491453472	-	-	S	-	<100nt
PY01207	2.274640304	-	-	T	-	
PY01843	0.348016463	-	-	-	-	
PY02301	0.311836748	-	PF14_0344	-	-	
PY02356	0.084371335	-	-	-	-	
PY02382	2.376131252	-	-	T	-	
PY02533	0.205989895	-	-	R	-	
PY02630	2.643801024	-	-	-	-	<100nt
PY03085	3.356459116	-	-	-	-	<100nt
PY04137	0.353843883	-	PF14_0792	S	-	
PY04152	0.244394836	-	-	-	-	
PY04313	0.127906577	-	-	T	-	
PY04333	0.45371624	-	-	R	-	
PY05466	2.001986004	-	-	R	-	
PY06384	0.101965954	-	-	-	-	<100nt
PY06494	2.236093117	-	-	R	-	
PY07057	2.058102642	-	-	R	NRLLFET	
PY07278	0.287646983	-	-	R	-	
PY07458	0.290481596	-	-	-	-	
PY07676	0.397511905	-	-	R	-	

Table 5.4. List of polymorphic proteins containing both transmembrane and signal peptide motifs. The presence of PEXEL domains is indicated in the respective columns. Genes more variant in YM are represented by the a mean YM:YA ratio < 0.5 while genes more polymorphic in YA are represented by a Mean YM:YA ratio > 2.0.

5.4 Conclusions

While the genome of the reference strain of *P. yoelii* 17x1.1 has already been published (Carlton, J.M. *et al*, 2002), many experimental models still utilize related lines such as YA and YM to compare differences in virulence and as models for the evaluation of pathological features and immunological responses. By being classified into the same species, the YA and YM lines are understandably very similar to 17x1.1. However, clear differences in phenotype are immediately discernable as YA and 17x1.1 both typically lead to an acute infection that is usually cleared while YM is extremely virulent, often leading to massive parasitemias and death of the host animal due to anemia and other complications. As per the analysis in the previous chapter, the *P. yoelii* genome sequences were further checked for bacterial and host (rodent) DNA contamination and their presence indicated in the supplementary appendices.

It is clear from this simple survey that the genetic make-up of 17x1.1, YA and YM are significantly different. By virtue of the numbers of genes duplicated and have undergone changes in sequence, YM seems to be more genetically similar to 17x1.1 while YA seems to have diverged further.

The majority of these polymorphic genes are predicted to code for hypothetical proteins and since their function cannot be directly inferred from annotation or domain searches, further analysis was undertaken to select for novel parasite proteins with some confidence of being exported or exposed on the parasite surface that may mediate novel parasite-host interactions and hence exhibit high polymorphism in the face of strong immune selection pressure. While it is useful to search for the chromosomal location of these *P. yoelii* genes, the absence of this information in the public domain databases

makes it difficult to correlate highly polymorphic genes with a sub-telomeric chromosomal location. Therefore, the presence of transmembrane domains and signal peptide motifs were chosen to further enhance the confidence that these candidate genes are not false positives and are likely to be in recognized by the host.

Therefore, 2 additional levels of confidence were added to tables 5.1, 5.2 and 5.3. Genes coding for proteins with high confidence of being exposed to the host immune surveillance would have predicted transmembrane domains while genes that can be classified as highly probable of being externalized contain both transmembrane and signal peptide motifs.

While it would be useful in future to sequence a subset of these genes and compare the frequency of non-synonymous substitutions, currently available data comparing the ratio of non-synonymous (dN) to synonymous (dS) nucleotide substitution ratios between orthologous genes of *P. berghei* and *P. chabaudi* (Hall, N. *et al*, 2005) can be used to further increase the confidence of these shortlisted genes. Genes with high dN/dS ratios suggest that they are targeted by the host immune system and there hence they are under pressure to diversify. When the dN/dS ratios were compared to the respective *P. yoelii* orthologs, the coverage was firstly not complete with respect to the *P. yoelii* genomes and secondly, there was a big spread of dN/dS ratios without any significant trend. Thus, it is not possible to directly establish genes under immune selection using this method. Perhaps the dN/dS ratios were artificially skewed when sequences are compared between different species where differences can occur. However, the only gene in Table 5.4 to show a high dN/dS ratio of 1.0178 is a hypothetical protein (PY00143) that is orthologous to a rophtry associated membrane antigen.

It is perhaps worthwhile to examine the origins of these laboratory rodent malaria parasite lines so as to see if any potential links of lineage to gene polymorphisms can be inferred. The first isolate (17x) was isolated as a mild strain from *Thamnomys rutilans* in 1965 (Landau, I. and Chabaud, A.G., 1965) in La Maboque Field station in the Central African Republic. This initial stock was maintained in the Museum National d'Histoire Naturelle in Paris and then delivered to New York University (NYU) where Dr. Meir Yoeli obtained a virulent line called YM after thawing a vial of the original isolate (Yoelii, M. and Hargreaves, B.J., 1975). He then sent this line to Edinburgh in 1977 where Dr. David Walliker subsequently made clones from this line (Knowles, G., and Walliker, D., 1980) and named this clone YM after Dr. Meir Yoeli. An avirulent clone was also obtained and named YA. Hence, both YA and YM clones are derived from the original 17x isolate.

Another uncloned 17x line was passed from Dr. Irene Landau to Dr. Richard Beaudoin of the Navy Medical Research Institute in Bethesda, Maryland in 1986 (Charoenvit, Y. *et al*, 1987). Dr. Walter Weiss then further cloned this line in 1990 and named it 17x1.1 (Weiss, W.R. *et al*, 1989) and this clone was used as a reference clone in the sequencing of *P. yoelii* (Carlton, J.M. *et al*, 2002).

It is therefore evident that the original 17x isolate was made up of a mixed infection where both virulent and avirulent lines coexist and are somehow able to modulate the infection to a mild one. This is evident where single cloned parasites from this original stablate exhibited either virulent or avirulent phenotypes.

Data from comparative genome hybridization experiments of YA or YM against 17x1.1 suggests that there are differences in the repertoire of genes that have expanded in

the individual parasite lines. More importantly, a significant number of genes are polymorphic and variant as compared to the reference line, thus the initial observed phenomenon in the initial isolation of YM as reported in 1975 cannot be possibly due to a ‘spontaneous transformation’ event. It is more probable that the initial laboratory stock contains a mixed population of parasites and that several rounds of passage and deep-freeze has selected for this line as its highly invasive phenotype would surely outgrow and out compete with the less virulent line as they are continuously passaged. Since YA and YM are significantly different from each other and to 17x1.1, the spontaneous transformation theory does not hold and genomics data presented here further supports the notion of a mixed population of different lines from the initial isolate in 1965.

The data presented here also corroborates with previous comparative genetic analysis especially of the *ama-1* gene where it has been shown that the 5’ nucleotide sequences of this gene is identical in YM and 17x1.1 but more divergent in YA (Pattaradilokrat, S. *et al*, 2008). Another finding also confirmed that YM and YA are genetically distinct at the *ama-1* locus (Kappe, S.H. *et al*, 1996). In addition, amplified fragment length polymorphism (AFLP) analysis suggests that the genotypes of YM and YA differ from 17x1.1 by 6 and 147 markers respectively, thus also suggesting that the cloned *P. yoelii* stocks are represented by the YA genotype and a 17x genotype that includes 17x1.1, 17XL and YM (Pattaradilokrat, S. *et al*, 2008). The data presented in this chapter provides a significantly wider coverage with genes in YA and YM that differ from 17x1.1 by 637 and 198 respectively. While the data agrees that YA is divergent from 17x1.1, it suggests that YM is more divergent than 17x1.1 but less so than YA.

Although the genomic survey has already provided some clues on certain markers that may contribute to virulence, it is possible that there are more factors involved. The identification of hypothetical proteins that contain both transmembrane and signal peptide motifs and are transcriptionally active raises the possibility that the product of these genes are exposed to immune surveillance and are involved in novel parasite-host interactions. Current work from the lab using a few genes of interest has confirmed the presence of polymorphisms and antibodies against these proteins would be helpful in elucidating cellular localization. It would be interesting for further investigations to elucidate the actual functions of these genes and they might perhaps serve as targets for therapeutics in the future.

Differential transcriptional profile of genes between the virulent YM and the avirulent YA strain of *Plasmodium yoelii* with focus on the SERA proteases

6.1 Abstract

The rodent malaria parasite *Plasmodium yoelii* is often utilized as an *in vivo* model of pathogenesis, immune mechanisms and vaccine development. It relies on the comparison of a virulent strain such as YM to an avirulent strain such as YA. While genomic differences between these two lines have been characterized in the previous section, these differences cannot fully explain their differences in invasion phenotype. In this study, global differences in gene transcription were screened via microarray. Potential virulence markers were validated via real-time RT-PCR to confirm that such genes were transcribed more abundantly in YM than in YA. Another comparison with *P. yoelii* lines exhibiting varying levels of virulence further confirmed a selection of these genes as potential virulence markers based on the association of a higher abundance of these transcripts in the virulent lines. Two SERA proteases, three regulatory proteins, an anion transporter and an invasion-related protein were so far confirmed. Genetic knockouts of the individual SERA proteases were performed; with one member (Δ PySERA1) being more reticulocyte restricted while the other SERA knockout (Δ PySERA2) did not exhibit any significant reduction in virulence. Real-time RT-PCR confirmed the knockout phenotype while analysis of the other four SERA genes did not show any significant compensation at the transcript level. The results suggest that different SERAs may either have redundant functions or be more specifically involved in the proteolytic activation of merozoite surface proteins.

6.2 Introduction

6.2.1 The use of microarray as valuable tools for identification of differentially transcribed genes

The use of microarrays to analyze and compare differential transcription of genes has been well established as potential genes of interest can thus be quickly identified. Improvements in the oligonucleotide probe design algorithms have increased the coverage of target genes that can be hybridized with good signal to noise ratio. In addition, genes with very similar sequences such as members of multigene families can be specifically targeted with high confidence (Hu, G. *et al*, 2007). Competitive hybridization experiments where differential binding ratios of cDNA from a pathogenic line against a less pathogenic line are routinely performed to identify candidate genes involved in pathogenicity. Typically, these are genes that show higher levels of transcription in virulent species/lines as compared to their avirulent counterparts (Pannucci, J., *et al*, 2004). This has been clearly shown by studies that correlate known virulence genes in *Streptococci* to the morbidity of an infection and also identified other putative genes such as *scpA* that codes for a peptidase that inactivates the C5a complement protein (Sitkiewicz, I. and Musser, J.M., 2006). In the field of malaria, a recent survey of the *var* genes provided evidence that different subgroups of these genes showed differential expression, with group B *var* genes implicated in mild to severe symptoms while group C transcripts were associated with asymptomatic malaria (Kaestli, M. *et al*, 2006). Hence, the use of microarray can quickly facilitate the survey of thousands of genes and the association of differential gene expression to virulence can be applicable for gene discovery of putative virulence markers.

6.2.2 The role of malarial proteases in invasion and egress

One of the most critical processes in the multi-step erythrocyte invasion pathway is the active movement of the merozoite into the new host cell. The merozoite has to firstly bind to specific receptors on the surface of the erythrocyte and this signals the start of the active driving force to propel the parasite into the new host cell. Before invasion can happen, the merozoite has to disengage this receptor-ligand interaction and this process was shown in *P. falciparum* to be mediated by a subtilisin-like protease known as PfSUB2 that is released from the micronemes and then cleaves the ectodomains of AMA1 and MSP1 (O'Donnell, R.A. and Blackman, M.J., 2005; Harris, P.K. *et al*, 2005), thereby releasing the interactions between the parasite and the host surface to facilitate invasion. More recently, the rhomboid proteases in *P. falciparum* were able to proteolytically cleave other diverse parasite adhesive ligands including TRAP, CTRP, MTRAP, PFF0800c, EBA-175, BAEBL, JESEBL, MAEBL, AMA1, Rh1, Rh2a, Rh2b, and Rh4 at their transmembrane domain (Baker, R.P. *et al*, 2006). This group of proteases might well serve as potential targets for chemotherapy as they seem to be universal 'sheddas' in both the mammalian and insect life cycle stages of the parasite.

Another main function of malarial proteases lies in the process of egress where mature late-stage schizonts rupture to release the daughter merozoites. In *P. falciparum*, the proteolytically processed form of SERA5 mediates this mechanism. Processing of SERA5 is dependent on the both cysteine protease dipeptidyl peptidase 3 (DPAP3) and the subtilisin-family serine protease (PfSUB1) as key regulators of downstream protease effectors responsible for egress (Arastu-Kapur, S. *et al*, 2008). However, antibodies directed against the 47kDa amino-terminal domain of the SERA-

5 protein seemed to inhibit erythrocyte invasion (Fox, B.A. *et al*, 1997) and hence points to an additional role in mediating invasion.

6.2.3 The serine repeat antigen (SERA) multigene family of the malaria parasites

The *P. falciparum* SERA multigene family consists of 9 members and the first identified member, now known as SERA5, is expressed as a 126kDa precursor protein that is localized in the parasitophorous vacuole (Delplace, P. *et al*, 1987). Prior to schizont rupture and release of the daughter merozoites, the SERA precursor is processed into a 50kDa fragment and a 73kDa fragment that is made up of a 47 and 18kDa fragments joined by disulfide linkages (Debrabant, A. *et al*, 1992). The 47kDa fragment is thought to be a potential vaccine candidate as antibodies directed against this fragment conferred protection against malaria (Sugiyama, T., *et al*, 1996; Fox, B.A. *et al*, 1997; Pang, X.L. *et al*, 1999) while the development of specific IgG antibodies in patients have been associated with protection against severe malaria (Okech, B.A. *et al*, 2001; Okech, B. *et al*, 2006). The representative rodent parasite species, namely *P. berghei*, *P. chabaudi* and *P. yoelii* only possess five SERAs.

The genomic organization of these SERA genes (Figure 6.1) are conserved amongst the malaria parasite species, with the entire repertoire of these genes located in tandem on a single chromosome and flanked at both ends by two conserved hypothetical proteins (Arisue, N. *et al*, 2007). Due to chromosomal rearrangements, there is a single SERA translocated to another chromosomal position as seen in *P. falciparum* and potentially in *P. gallinaceum*. Of note, *P. vivax* seems to have undergone an expansion of the SERA multigene family; with a total of 12 members. Also, the possibility for the other less completely annotated and sequenced species

such as the rodent and avian parasites to possess more than the known repertoire of genes cannot be discounted without full closure of the sequence gaps.

Figure 6.1. Gene ordering of the SERA multigene family in eight representative *Plasmodium* species. Genes that are located on the same chromosome are joined in solid horizontal lines and these genes are predominantly located on the same chromosome with the exception of a ninth SERA located on chromosome 9 of *P. falciparum* and another potential SERA located on another chromosome in *P. gallinaceum*. Dashed vertical lines indicate close phylogenetic relationship of specific SERA genes across different species. Groups I, II, III and IV represent monophyletic groupings of positionally conserved SERA genes, with groups I to III containing a cysteine residue at the catalytic domain, while group IV genes has a serine in place of the cysteine. (Adapted from Arisue, N. *et al*, 2007)

The SERAs are highly conserved amongst the malaria parasite species, especially those that contain a cysteine at the active center of the protease domain as depicted by those classified in groups I to III (Figure 6.1). On the other hand, Group IV genes have the serine residue at the active center domain and are more varied, thus suggesting that groups I to III SERAs perform evolutionarily conserved functions as compared to the group IV SERAs. All the SERAs contain a characteristic 4 exon structure (Figure 6.2) and all 9 SERAs in *P. falciparum* were transcribed in the mature blood stage. The centrally located SERA5 and SERA6 genes cannot be disrupted, suggesting a conserved role of these 2 genes in mediating essential functions in the blood stages (Miller, S.K. *et al*, 2002; McCoubrie, J.E. *et al*, 2007). For the rodent parasite species, it was established in *P. berghei* that all the SERAs were transcribed in the blood stages except PbSERA5 that was only transcribed in the mosquito stages and that PbSERA3 was processed in a manner that is similar to the processed fragments of PfSERA5 (Schmidt-Christensen, A. *et al*, 2008).

Figure 6.2. Gene structure of a typical SERA gene. The SERA genes contain 4 exons and the location of the catalytic domain is clearly marked out in dotted lines. (Adapted from Miller, S.K. *et al*, 2002)

Perhaps the most widely studied member is PfSERA5 which was shown to be processed at an initial site near the N-terminal of the papain-like domain and then followed by a second cleavage at the C-terminal of this domain (Li, J. *et al*, 2002a; Li, J. *et al*, 2002b). A model for SERA processing was deduced when a subtilisin-like protease PfSUB-1 was found to be responsible for the processing of PfSERA4, 5 and 6 (Yeoh, S. *et al*, 2007). From this study, the proposed consensus sequence was deemed to first consist of a hydrophobic residue followed by a random amino acid, a small amino acid, a polar residue followed by four acidic residues. The remaining 56kDa central SERA region that contains the active catalytic domain is then processed by an unknown protease to yield the 50kDa residue of which its function is currently not known (Figure 6.3).

Immunofluorescence studies showed that both PfSERA4 and PfSERA5 co-localize with merozoite surface protein 1 (MSP1) in late stage schizonts, with a staining pattern that covers the entire parasitophorous vacuole but excludes the merozoite (Miller, S.K. *et al*, 2002). This suggests that the SERA might possibly be involved in proteolytically processing merozoite surface proteins (MSP1) prior to schizont rupture. Therefore, such a potential role is worth investigating. For example, prior to schizont rupture (late schizont stage), the huge MSP1 precursor protein is proteolytically cleaved into the 83kDa (N-terminus), 30 and 38 kDa (central region) and 42 kDa (C-terminus) fragments (Holder, A.A. *et al*, 1987). The fact that the SERAs co-localize with MSP1 in late stage schizonts certainly seem to suggest a role for the SERAs in processing MSP1 and potentially other proteins that are located on the surface of the merozoites.

Figure 6.3. A schematic model of PfSERA5 processing in the role of egress from late blood stage parasites. SERA5 is first detected in the parasitophorous vacuole as a 126kDa precursor and is then proteolytically cleaved by the regulated discharge of PfSUB1 to release the 56kDa-activated fragment that contains the papain-like domain (shown as a hatched region) and is responsible for the proteolytic events in host cell egress. The remaining 47 and 19kDa fragments were then joined via a disulphide bridge and associates with the surface of the merozoite. An unknown protease then processes the 56kDa fragment to a 50kDa fragment. (Adapted from Blackman, M.J., 2008)

6.3. Results and discussion

6.3.1 Transcriptomic differences between the avirulent YA and virulent YM lines of *P. yoelii*

In this experimental layout, the avirulent YA line was compared to the virulent YM line of the rodent malaria parasite *P. yoelii*. Total RNA from broad stages of rings, trophozoites and schizonts were obtained from a typical mixed asynchronous infection in Balb/c mice. Basically, reverse transcribed labeled cDNA from identically staged parasites from YM were compared to YA, i.e. hybridization ratios of YM/YA rings; YM/YA trophozoites; and YM/YA schizonts were obtained in duplicates. As with previous experiments, a background subtraction of twice the background intensity plus twice the standard deviation of the background intensity was used to filter off non-specific hybridizations. Only normalized stage-specific ratios of *P. yoelii* specific probes were considered for analysis and a simple heatmap depicting the spread of differentially expressed genes is shown in Figure 6.4. A total of 1518 genes were differentially expressed and the entire gene list can be found in Appendix J. In this dataset, a total of 830 hypothetical proteins were differentially transcribed in the asexual blood stages and since the function of these putative proteins have not been established and confirmed, it is difficult to speculate on their possible involvement in virulence. The next largest group consists of 374 PIRs, of which only 87 (23.26%) are preferentially expressed higher in YA than in YM, leaving the majority of the PIRs being more abundantly transcribed in YM than in YA. Since the main function of this particular *P. yoelii* multigene family lies in antigenic variation when they are expressed on the surface of the infected erythrocyte (Cunningham, D.A. *et al*, 2005), any possible role in virulence is much debatable.

Figure 6.4. Distribution of genes found to be differentially expressed in YM as compared to YA. Normalized median signal intensity ratios of YM/YA staged parasites in duplicate experiments were used for the comparison. The quality and reproducibility of the hybridizations were evident, with a Pearson correlation of 0.953.

On the other hand, there are 732 genes more highly transcribed in YA than YM. While most of these genes are hypothetical (63%), the PIRs do not dominate this list (12%). Genes that are implicated in invasion include rhoptry proteins (PY01526, PY01674, PY02635, PY04318 and PY06381). The p235 rhoptry protein E5 (PY06381) is expected as shown by the previous chapter of being absent in YM. This subset of genes also include five erythrocyte membrane protein PFEMP3 genes (PY01183, PY01667, PY03698, PY04922 and PY06052) but none of these genes were orthologous to any *P. falciparum* erythrocyte membrane protein and were also devoid of any significant domain hits.

There are also a number of genes that were transcribed only in the early ring stages as seen in the profile of a subset of genes without any signal in the later erythrocytic developmental stages (Figure 6.4; Appendix K). A total of 294 genes were only transcribed and mainly consist of 152 PIRs (52%) and 123 hypothetical genes (42%). Of interest, three genes annotated as CCAAT-box DNA binding protein subunit B (PY02242, PY06770 and PY00834) were potentially inaccurately described as no significant domains were present. Another DNA binding protein (PY00739) should function as a tRNA synthetase based on orthology and conserved domain searches while a predicted erythrocyte membrane protein PFEMP3 gene (PY04120) is also absent for any functional domains and is orthologous to a hypothetical gene in the other malaria parasite species.

A survey performed in a previous chapter established that a subset of these genes contains significant polymorphisms in both the YA and YM lines and hence hybridize to the probe with different efficiencies. As discussed previously, a caveat of the direct YA vs. YM comparative genome hybridization assumes that the differences do not arise from potential gene duplications. Together with the fact that the dynamic

range in microarray may not be necessarily as wide as in quantitative RT-PCR, normalizing the transcriptome comparison dataset may not be the most accurate way of mitigating such bias. Thus, a few genes of interest were selected on the basis that they have some potential in invasion, development and metabolic functions and are at least five-fold more abundantly transcribed in YM. These genes were selected from the remaining 314 genes that were not hypothetical proteins or PIRs and contain diverse functions such as proteases, invasion-related proteins, putative transcriptional factors / regulators and a transporter. While up regulated genes in a virulent strain are usually associated with putative virulence-associated genes, it is perhaps worthwhile to highlight that differences in gene expression is not the only factor. Studies on *P. falciparum* field isolates have shown that polymorphisms in the erythrocyte binding ligand Jesebl/EBA-181 led to the increase in the repertoire of erythrocyte receptors available for binding and thereby also potentially increasing the virulence of parasites (Mayer, D.C. *et al*, 2004). Recently, polymorphisms in PfEBA175 and PfRh1 were implicated in mediating disease severity and hence affecting virulence (Chokejindachai, W. and Conway, D.J., 2009). In this study, the selected up regulated genes will be targeted for confirmation via real-time RT-PCR using primers flanking the region targeted by the microarray probe. In order to account for sequence polymorphisms that may affect the efficiency of PCR, the genes of interest from each line were TOPO-cloned to firstly obtain evidence for sequence quality and as well as to provide a means to generate a standard curve for quantification in each parasite line.

6.3.2 Validation of selected genes of interest via real-time RT-PCR

A total of 10 genes that were more abundantly transcribed in YM than YA were chosen, together with 1 gene that was more abundantly expressed in YA than in YM for validation via real-time RT-PCR using individual line-specific (YM and YA) plasmid clones as standards of the same gene for normalization (Table 6.1). Their predicted functions were confirmed with domain searches in Pfam so as to have a certain confidence on their actual role. Two proteases (PY00291 & PY00292) were chosen, as they could both be potentially involved in proteolytic processing of invasion molecules and perhaps conferring an advantage to invasion efficiency simply due to a higher efficiency of generating active sites for invasion. Invasion-related proteins such as merozoite surface protein precursor (MSP8; PY06415) and a protein orthologous to MSP-7 (PY02147) were chosen as they are potential mediators of the initial attachment to the host cell prior to the formation of the junction leading to the irreversible commitment to host cell invasion. On the other hand, a PY235 member (PY06381) was shown to be absent in YM as it cannot be amplified via PCR. Regulatory proteins such as a chromosome segregating protein *smc1* (PY00653), a potential mediator in cyclic nucleotide biosynthesis and intracellular signaling (PY07180), and a potential chaperon mediating protein-protein interactions (PY02799) were also selected as there is some evidence that YM replicates more rapidly than an avirulent line (Fahey, J.R. and Spitalny, G.L., 1986). Two putative transcriptional factors (PY01120 & PY05489) were also screened as they contain nucleic acid binding domains and the general lack of transcription factors in the malaria parasites certainly warrants the investigation of these putative factors. Lastly, an anion transporter (PY02198) was also picked as the host cell environment differs, i.e. a reticulocyte that is preferred by the avirulent line is richer in terms of receptors and

hemoglobin; while the older normocytes which can support the growth of the virulent line, is less complex internal environment and also continuously loses receptors (Geminard, C. *et al*, 2002; Koury, M.J. *et al*, 2005).

The real-time RT-PCR data shows high congruence with the semi-quantitative microarray ratios, i.e. if a particular gene is more abundant in one than the other, it is also reflected though the amplitude may differ. However, real-time RT-PCR is considered more accurate than the microarray ratios that are semi-quantitative and limited in dynamic range. Hence these quantitative real-time RT-PCR values are taken to be the more accurate representation and also overcome sequence polymorphisms.

By genome comparison, the two proteases (PY00291 & PY00292) are considered identical while sequencing reveals a single base difference between YA and YM. Hybridization ratios of PY00292 peaked at 11-fold while the peak real-time RT-PCR value is more than 24,000-fold, revealing the limited dynamic range of microarrays, especially if transcript levels are high and signals are near saturation point. However, PY00291 had a lower peak real-time RT-PCR value of 3.80 as compared with a peak 11-fold difference obtained via microarray. This could be due to the low hybridization signal from YA that is near the lower limit and thus being normalized to a lower value and hence creating an illusion of a higher change in ratio. The other genes with highly similar gene sequence (PY07180 & PY02198) also confirmed the higher abundance of these transcripts in YM while discrepancies in the absolute values have been addressed above.

Four genes with highly biased YM/YA comparative genomic hybridization (PY06415, PY00653, PY01120, PY02799 & PY05489) were also selected for real-time RT-PCR confirmation as they represent interesting genes and the transcripts for

these genes were significantly higher in YM than YA (5-fold or more). The most significant difference lies in PY01120 as a peak 200-fold microarray ratio difference was mitigated to just less than 2-fold via real-time RT-PCR. This is to be expected as sequencing of this target fragment revealed significant polymorphisms at the region targeted by the probe, with only 12% similarity. Another three of these genes (PY06415, PY02799 & PY05489) were comparable in sequence, with >90% identity and therefore do not significantly mitigate the transcript ratios, with a peak real-time RT-PCR ratio of at least 3.8. In fact, for PY02799, the real-time RT-PCR results showed a much higher difference with a peak of 41-fold as compared to the microarray peak of almost 6-fold.

It was difficult to amplify the PY02147 fragment in YA and this suggests that the primer binding sites were extremely polymorphic. This gene was therefore excluded from real-time RT-PCR verification due to difficulties in generating a vector for a YA-specific gene quantization.

The final gene in this list is the p235 E5 rhoptry gene (PY06381). The YM/YA comparative genomic hybridization ratio of 0.09 suggests extreme polymorphism or even a deletion of this gene in YM. Via microarray, this gene is transcribed a thousand-fold higher in YA than YM. Attempts to amplify this gene in YM failed, thus suggesting that this gene is absent in YM. However, previous work on sequence diversity of the *P. yoelii* p235 rhoptry genes showed the presence of the E5 gene in YM (Narum, D.L. *et al*, 2001). Therefore, the inability to pull out this gene is potentially due to significant sequence polymorphisms.

Acc No.	Description	YM/YA transcript ratios			Real-time RT-PCR			YM/YA gDNA ratio	% match to oligo seq	
		Rings	Trophs	Schizonts	Rings	Trophs	Schizonts		YM	YA
PY00292	Papain family cysteine protease, putative	2.60	11.03	9.76	13099.40	24178.60	14045.92	0.93	100	98
PY00291	SERA-3	3.44	9.98	11.09	3.51	1.94	3.80	1.03	100	98
PY06415	merozoite surface protein precursor	50.05	46.33	36.88	3.55	2.55	6.20	5.60	100	95
PY00653	chromosome segregation protein smc1	37.67	132.02	226.13	3.68	2.33	3.52	8.98	100	85
PY07180	Dictyostelium discoideum SgcA	5.84	6.74	7.02	6.62	3.33	9.18	1.10	100	97
PY01120	Arabidopsis thaliana BHRAMA ortholog-related	6.35	11.81	200.44	-3.02	-1.88	1.88	8.96	100	12
PY02799	unnamed protein product	5.36	4.85	5.90	15.73	17.91	41.34	2.55	100	98
PY05489	histidine kinase DhKG	9.41	28.58	19.48	1.80	1.55	3.82	5.12	100	97
PY02198	arsenite transport subunit A	3.24	5.68	6.73	25.85	41.05	79.49	1.06	100	98
PY02147	Meloidogyne incognita COL-1 related	90.26	199.70	212.61	YA clone unavailable			1.52	100	NA
PY06381	p235 rhopty protein E5	-17.67	-474.30	-1045.70	exclusive in YA			0.09	NA	100

Table 6.1. A tabulated list showing the selected genes of interest. Their respective normalized microarray data, real-time RT-PCR values and relative polymorphisms according to the hybridization efficiencies (YM/YA gDNA ratio) and actual sequencing data (% match to oligo seq) are also shown. The mean of triplicate real-time RT-PCR values were taken and were significant at $p \leq 0.05$.

Where the gene sequences do not differ too much, the semi-quantitative microarray ratios do provide a good indication of differentially transcribed genes in related species lines. Differential genomic hybridization ratios effectively predict polymorphic gene sequences that can affect comparisons on differential gene expression and should be validated via real-time RT-PCR.

6.3.3 Real-time RT-PCR screen of the selected genes of interest to the different genetic lines of *P. yoelii* exhibiting different virulence phenotypes

Following confirmation of the relative levels of transcripts for the selected genes of interest in both YM and YA, the next question is to confirm their role in virulence across different parasite lines. The progeny of the genetic crosses between then virulent YM and avirulent AC lines (Walliker, D. *et al*, 1976) were used as these lines exhibit differences in virulence and in other phenotypic characteristics that have seemingly been inherited in a Mendelian manner. While the same passage numbers of the parasite lines were used to standardize parasite ‘age’, the three broad erythrocytic developmental stages were obtained from a non-uniform Histodenz gradient purification. A previous study on the impact of PY235 on virulence (Iyer, J.K. *et al*, 2007) has also established the correlation between the amounts of PY235 to the degree of virulence of these parasite lines (Figure 6.5). Hence, an identical approach was used in this study. Genes were assayed via real-time RT-PCR and those that showed significant increase in transcript levels across the virulent lines as compared to the avirulent lines would point to a universal role of this particular gene as a marker for virulence (Figure 6). In this screen, the TOPO-cloned YM gene fragment was used as a means to quantify actual transcript copy number of the corresponding genes in the other lines. Statistical significance was measured using a T-test for the gene in

each virulent line to compare with the highest level of transcription in an avirulent line.

Figure 6.5. Diagrammatic representation of the genetic cross between the virulent YM line and the avirulent AC line of *P. yoelii*. Six progeny clones (604, 605, 606, 607, 611 & 612) were used in the survey and their virulence was previously confirmed by Iyer, J.K. and co-workers (2007). In this case, virulence was determined based on host cell selectivity, with virulent lines invading erythrocytes of all ages while avirulent lines being restricted mainly to reticulocytes and intermediate lines being somewhat less virulent than the virulent lines.

For the PY00292 protease gene (Figure 6.6a), all three virulent lines and an intermediate virulent line showed significantly higher levels of transcripts as compared to the avirulent and intermediate virulent lines with an overall p -value of 0.00313. As for the other protease (PY00291; Figure 6.6b), transcription levels were only significantly higher in the three virulent lines with an overall p -value of 0.0000239. Although the intermediate virulent line did not exhibit any significant increase in transcript level, the involvement of this gene in all three virulent lines still validates it as a prime candidate universal marker for virulence.

The chromosomal segregation protein *smc1* gene (PY00653, Figure 6.6c) showed significantly higher levels of transcripts in all three virulent and the intermediate virulent lines with an overall p -value of 0.005. While a central role of this class of proteins is for the maintenance of chromosome structure (Saitoh, N. *et al*, 1995) and DNA repair (Lehmann, R., *et al*, 1995), their involvement in recombination (Jessberger, R., *et al*, 1996) suggests a possible role in generating antigenic diversity, especially for the genes located in the sub-telomeric regions that undergo frequent ectopic recombination (Freitas-Junior, L. *et al*, 2002). On the other hand, the PyMSP-8 gene (PY06415; Figure 6.6d) cannot be considered to be a universal marker for virulence because this gene is not more abundantly transcribed in one of the virulent lines. The p235 E5 gene (PY06381, Figure 6.6e) was established to be extremely polymorphic in the previous section and while variable levels of transcription was picked up, the failure to detect this transcript in the other lines can be attributed to sequence polymorphisms and therefore its role in virulence cannot be accurately measured. What is more certain is that the putative ATP and nucleic acid binding helicase (PY01120; Figure 6.6f) is not correlated with the virulent phenotype as only the intermediate virulent line showed higher gene transcription. Likewise, the cyclic

nucleotide biosynthesis and intracellular signaling gene (PY07180; Figure 6.6g) is not significantly transcribed in higher levels in one of the virulent lines (604) and therefore does not correlate to be a universal virulence marker.

(A)

(B)

(C)

(D)

(E)

(F)

(G)

(H)

(I)

(J)

(K)

Figure 6.6. Real-time RT-PCR screen of the progeny of the genetic cross between the virulent YM and avirulent AC lines of *P. yoelii*. The relative contributions of the genes (A) PY00292; (B) PY00291; (C) PY00653; (D) PY06415; (E) Py06381; (F) PY01120; (G) PY07180; (H) PY02799; (I) PY05489; (J) PY02198; (K) PY02147; were evaluated across the 6 genetic cross progenies previously verified as avirulent (AC, 611), intermediate avirulent (606), intermediate virulent (612) and virulent (604, 605, 607). A dashed line across demarcates the peak number of transcripts contributed by the avirulent lines with respect to the 18SrRNA control. (Legend: A = avirulent; V = virulent; IV = intermediate virulent; IA = intermediate strains; * = p -value < 0.05; ** = p -value < 0.01)

A putative chaperon involved in protein-protein interactions (PY02799; Figure 6.6h) showed higher levels of transcription in all the virulent and intermediate virulent lines with an overall p -value of 0.00565, thus validating its potential role as a universal virulence marker. Similarly, the histidine kinase gene (PY05489; Figure 6.6i) is transcribed more abundantly in the three virulent lines with an overall p -value of 0.00936, also inferring its role as a universal virulence marker. In the same way, the transcription levels of the arsenite transport subunit A gene (PY02198; Figure 6.6j) is higher in all the virulent and intermediate virulent lines, with an overall p -value of 0.000475, thus strongly suggesting a potential role as a universal virulence factor. Finally, the MSP-7 ortholog (PY02147; Figure 6.6k) is transcribed more abundantly in all the virulent and intermediate virulent lines, with an overall p -value of 0.0154, thus also strongly suggestive of its role as a universal virulence factor.

Seven genes have thus been identified as potential universal virulence markers and the two proteases were chosen for further characterization as they could be potentially be involved in processing invasion-related proteins and hence be associated with a less reticulocyte restricted phenotype.

6.3.4 Genomic location of the re-annotated *P. yoelii* SERA genes

In the previous section on the genomic differences between the *P. yoelii* virulent YM line and the avirulent line YA line, two proteases (PY00291 & PY00292) were identified as potential virulence markers as they are both found to be transcribed more abundantly in the virulent lines than in the avirulent lines. The PY00291 protease is currently annotated as SERA3 while PY00292 is named as a putative papain family cysteine protease. The relationship between these 2 genes was made clearer by a re-annotation project whereby putative SERA genes were re-annotated and phylogenetically ordered (Figure 6.1; Arisue, N. *et al*, 2007). Based on this nomenclature and gene order system, only 5 SERAs were found in the genome of the rodent malaria parasites. While the rodent SERA3, 4 and 5 are phylogenetically related to SERA6, 7 and 8 of *P. falciparum* respectively, the rodent parasite SERA1 and 2 contain a serine substitution in the active center of the cysteine protease domain and these SERAs are only closely related amongst the rodent parasite species.

Based on this finding, the SERA gene organization was proposed to follow this order, with PySERA1 to 5 associated with accession numbers PY00291, PY00292, PY00293, PY02062 and PY02063 respectively (Figure 6.7).

With the establishment of the genomic location, the biological function of these 2 SERA genes of interest, i.e. PY00291 and PY00292 can thus be investigated using biological tools such as protein tagging to elucidate localization and targeted gene disruption to ascertain its biological function.

Figure 6.7. Genomic organization of the SERA genes in *P. yoelii*. The five known SERAs are flanked by 2 conserved hypothetical proteins. The respective *P. yoelii* contig name is shown above the GenBank reference number while the associated gene accession numbers for each SERA gene are shown below their respective re-annotated SERA member gene. The two SERA genes of interest are boxed and correspond to SERA1 and SERA2. (Adapted from Arisue, N. *et al*, 2007)

6.3.5 Construction of GFP tagged PySERA1 and 2 proteins

The aims for a functional tagging of the SERA proteins are to firstly demonstrate that these genes can be targeted and that the availability of fluorescent parasites in aiding recovery of transfected parasites and also for tracking protein localization within the parasitized erythrocyte.

A bacterial-parasite shuttle vector named B3DhRap2/3 EGFP (Figure 6.8a) was available for use thanks to a kind donation by Blandine Franke-Fayard and Andrew Waters from the University of Leiden, the Netherlands. This is a targeting construct whereby a unique restriction site within the gene-of-interest (in this case Rap2/3) is cut by a specific restriction enzyme and thus generates a linearized vector that undergoes a single recombination event such that the vector is introduced into the genome (Figure 6.8b). The gene-of-interest is to be cloned via the unique *BamHI* and *NcoI* restriction sites to be in-frame with the eGFP marker so that the fluorescence signal is under control of the endogenous promoter. The unique cut site for SERA1 and 2 are *HincII* and *BsaBI* respectively.

Primer pairs (291Bam & 291Nco; 292Bam & 292Nco, Appendix 6) were used to generate eGFP tagged PySERA1 and PySERA2 respectively. The eGFP tag was designed to be as close as possible to the C-terminal of the full-length protein, with the position of the unique restriction site being most pertinent criteria. In both constructs, the active center of the cysteine protease domain was left untouched (Figure 6.9). Successful transfectants were typically detected 7 to 10 days post transfection as the wild-type parasites are killed off by the anti-folate drug pyrimethamine while recombinant parasites were resistant due to complementation by the human dihydrofolate reductase (hDHFR) gene conferred by a successful integration event into the chromosome.

Figure 6.8. Schematic depicting a single cross-over gene targeting construct strategy.

(A) In this example, the gene-of-interest (Rap2/3) is to be introduced in-frame to the eGFP sequence via the *Bam*HI and *Nco*I restriction sites. (B) A unique restriction site within the gene-of-interest in the vector is cut to linearize the vector and the entire vector can be integrated into the genome sequence. Successful transfectants can be screened via PCR to confirm the correct 5' and 3' integration events using primers (depicted as arrows) designed specifically to the vector and flanking genome sequences.

Figure 6.9. Schematic representation of the relative positions of the cysteine protease domain and the site of eGFP fusion with respect to the organization of the full-length SERA proteins. The C-terminal end of the truncated endogenous protein is determined by the antisense-strand primer as this primer contains an *NcoI* restriction site for cloning into the N-terminal end of the eGFP protein of the B3DhRap2/3 EGFP vector. (A) For SERA1 (PY00201), fusion of the eGFP occurred just beyond the active center of the cysteine protease domain. (B) For SERA2, (PY00292), eGFP fusion was located at the C-terminal end of the full-length protein. (Adapted from PlasmoDB gene query at <http://www.plasmodb.org>)

6.3.6 Screening for transfected parasites expressing GFP tagged PySERA1 and PySERA2 proteins

Perhaps the easiest way to screen for positive integrants is via PCR analysis of total genomic DNA. Host Balb/c mice were sacrificed when the parasitemia reached ~5% and whole blood was harvested for genomic DNA extraction and storage in liquid nitrogen. Although successfully transfected parasites are expressing eGFP and are fluorescent, PCR is a more sensitive method as there are usually break-through parasites that are wild-type escapees as the drug dosage through the drinking water may be variable. In addition, a small population of the DNA used for transfection may contain uncut vector and thus have conferred resistance when they are maintained episomally. Hence, PCR primers 291intF & EGFP_R (Appendix 6) were used to screen for 5' (1.8kb) and primers 291intR & Rap2/3 3'intF (Appendix 6) were used to screen for 3' (1.7kb) integration events in PySERA1 respectively (Figure 6.10a). Similarly, primers 292intF & EGFP_R (Appendix 6) were used to screen for 5' (1.8kb) and primers 292intR & Rap2/3_3'intF (Appendix 6) were used to screen for 3' (2.0kb) integration events in PySERA2 respectively (Figure 6.10b).

PCR positive samples were then subjected to FACs sorting of fluorescent schizont stage parasites and these sorted parasites were then re-introduced into Balb/c mice under drug pressure so as to expand the cloned out parasites and establish frozen stocks for further use.

Figure 6.10. Agarose gel electrophoresis profile of the PCR screen of eGFP tagged SERA genes in transfected parasites shown with the 1kb marker labeled as lane 'M'. (A) Lanes 1 – 3 represent the 5'-integration PCR screen of (1) wild-type, (2) vector control and (3) transfected parasite DNAs for PY00291. Transfected parasites are faintly positive for a 1.8kb target band while the wild-type control was negative and only primer dimers were present in the vector control. Lanes 4 -6 represent the 3'-integration PCR screen of (4) wild-type, (5) vector control and (6) transfected parasite DNAs for PY00291. Only the transfected parasites were PCR positive, showing a 1.7kb band. (B) PY00292 PCR screen for 5'- integration showed a positive band in transfected parasites (1.8kb; lane 1) but only primer dimers in wild-type parasites. (lane 3). Similarly, PCR screen for 3'- integration showed a positive band in transfected parasites (lane 2) and there was no product in wild-type parasites. (2.0kb; lane 4).

Western blots of eGFP tagged SERA1 and SERA2 clones using mice anti-GFP on *in vitro* cultured late schizonts of these clones also showed the precursor fusion proteins of ~150kDa for both clones (Figure 6.11). Since the construction of PySERA1 practically removed the C-terminal distal to the central protease domain, the resultant total molecular weight should consist of the remaining N'terminal (47kDa), central domain (55kDa) and eGFP (26.9kDa) for a total of 128.9kDa as shown (Figure 11a). The first cleavage site yields a predicted 82kDa eGFP fusion fragment while cleavage at the second site releases the 27kDa eGFP tag (Figure 6.11). For PySERA2, the position of the eGFP tag is such that this practically does not remove much of the endogenous protein-coding region (123.7kDa). Thus, the expected molecular weight of this fusion protein should lie around 150.6kDa (Figure 6.11a). The first cleavage site yields a predicted 104kDa eGFP fusion fragment while cleavage at the second site releases the shorter 49kDa eGFP fusion fragment (Figure 6.11). A 75kDa band was detected in both PySERAs while a 37kDa associated band in PySERA1 could also be attributed to unknown processing sites.

The intermediate bands represent processed forms of the PySERAs and appear similar to the blood stage Western blot profiles of a similar parasite *P. berghei* where the full length protein and the intermediate fragments have been shown using antibodies specific to the N terminal region and C terminal TAP-tag (Schmidt-Christensen, A. *et al*, 2008).

In summary, the full-length precursor PySERA1 and 2 proteins are processed to the expected intermediate forms as shown by an indirect investigation using anti-eGFP monoclonal antibodies against the eGFP-tagged fusion proteins. The results thus support the identity of these proteins as SERA proteases, thus validating the proposed gene annotation by Arisue and co-workers (2007).

Figure 6.11. Molecular analysis of GFP-tagged PySERA proteins. (A) Predicted band sizes for anti-GFP detection with the two processing sites marked by arrows and the eGFP tag is shown as a box. (B) Western blot analysis of *P. yoelii* transgenic clones expressing eGFP-tagged SERA proteins. Parasites are matured *in vitro* into late, segmented schizonts and are lysed in the Laemli loading buffer, ran on a 10% gel and probed with mouse anti-eGFP monoclonal antibodies. Lane M represents the protein molecular weight marker while lanes 2 and 4 represent wild-type parasite extracts. In lane 1, the PySERA1_eGFP clones show the precursor fusion protein (129kDa), intermediate processed fragments (82kDa) and the cleaved off eGFP (27kDa). PySERA2_eGFP extracts were loaded in lane 3 and also showed the precursor fusion protein (151kDa). The processed C-terminal SERA segment that is fused to eGFP can be detected as a ~50kDa band.

6.3.7 Localization of PySERA1 and PySERA2

The availability of cloned-out pure lines of transgenic *P. yoelii* clones expressing eGFP-tagged SERA proteins is certainly useful for investigating the subcellular location of SERA proteins in *P. yoelii*. Although these proteins are tagged at the C-terminal end and cleavage of both the N- and C-terminal portions is necessary for activation of the SERA proteins, visualization of the C-terminal end could still provide useful information on their subcellular location as there are no other targeting or signal sequences in the central active cysteine protease domain of the activated protein. Thus, live fluorescence microscopy was used to access their cellular localization.

Both PySERA1 (Figure 6.12) and PySERA2 (Figure 6.13) are expressed in the late schizont stages as earlier developmental stages are not fluorescent and only *in vitro* matured schizonts with distinct segmented merozoites show significant eGFP signal in the central region of the mature schizont. The use of DAPI as a counter-stain is useful to establish the location of the daughter merozoites within the parasitized erythrocyte. From these views, it can be clearly seen that PySERA1 is localized to a central region within the schizont and seems to surround the merozoites. An identical fluorescent pattern was also shown by Miller, S.K. and co-workers (2002) when they discovered the co-localization of PfSERA4 with PfSERA5 and MSP-1. However, PySERA2 seems to be more localized to the centre of the schizont and there is a faint associated with the merozoites. This could be due to visualization of an early stage schizont as the diffuse pattern of PYSERA1 is seen in late stage schizonts containing fully segmented merozoites.

Figure 6.12. A representative live fluorescence microscopy image of a PySERA1-eGFP tagged parasite clone. (A) DAPI DNA staining reveals two schizonts next to each other. (B) eGFP signal is clearly seen in 1 of the schizonts. (C) A merged image of image (A) and image (B) showing both merozoite and eGFP signals within the schizont. (D) A phase contrast view of the same field showing the morphology of the parasitized erythrocytes, with the more mature schizont marked out by an arrow. Note that the less mature schizont on the left does not express the fusion protein yet, correlating with the synthesis of these proteins prior to schizont rupture. A thick white line represents a scale bar of 10 μ M. This cropped view represents 2 infected erythrocytes out of a total of 11 erythrocytes in this field of view and is representative for a parasitemia of around 30%, with schizonts being the predominant erythrocytic developmental stage. A total of ten random fields were observed with similar results.

Figure 6.13. A representative live fluorescence microscopy image of a PySERA2-eGFP tagged parasite clone. (A) DAPI DNA staining illustrating the relative location of the daughter merozoites within a schizont. (B) eGFP signal is clearly seen in the same schizont. (C) A merged image of image (A) and image (B) showing both merozoite and eGFP signals within the schizont. (D) A phase contrast view of the same field showing the morphology of the parasitized erythrocyte, with the schizont marked out by an arrow. A thick white line represents a scale bar of 10 μ M. This cropped view represents 1 infected erythrocyte out of a total of 2 erythrocytes in this field of view and is representative for a parasitemia of around 30%, with schizonts being the predominant erythrocytic developmental stage. A total of ten random fields were observed with similar results.

The live imaging also managed to capture an elusive PySERA2-eGFP schizont that has just ruptured and in this example (Figure 6.14), the eGFP signal seems to be localized around the newly released merozoites and therefore suggests that the C-terminal fragment associates specifically with the merozoite surface. The PySERA2-eGFP parasite was constructed with an almost intact C-terminal region and it is likely that this C-terminal-eGFP fusion fragment is still able to associate with the N-terminal region and then this complex is able to still bind to the merozoite surface (Figure 6.3).

Therefore, although the initial live image of PySERA2 (Figure 6.13) did not suggest any sort of direct interaction with the merozoites, it is possible that the inactive nascent full length SERA protein is retained within the schizont and is only activated and released once it is processed and just prior to egress.

Figure 6.14. A representative live fluorescence microscopy image of a ruptured PySERA2-eGFP schizont clone. (A) DAPI DNA staining illustrating the relative location of the merozoites. (B) eGFP signal is clearly seen within the localized area of interest. (C) A merged image of image (A) and image (B) showing that the C-terminal PySERA2 (eGFP) is associated with the merozoite surface. (D) A phase contrast view of the same field showing the morphology of the released merozoites. A thick white line represents a scale bar of 10 μ M. This represents an isolated case where a recently ruptured schizont was imaged.

Having already established that the ultimate location of the processed *P. yoelii* SERA1 and SERA2 proteins to be similarly located in close association with the daughter merozoites, it can be further investigated to ascertain if this is a general surface association, or if it is more specific to the apical end of the merozoites. By taking advantage of our laboratory's stock of PY235 antibodies, the mouse 1431 peptide polyclonal antibody was chosen as it cross-reacts with most members of this multigene family and clearly labels the rhoptries of the merozoites. The eGFP-tagged SERA proteins are further subjected to a rabbit anti-GFP monoclonal antibody in order to detect eGFP in acetone-fixed slide samples.

Hence, immunofluorescence microscopy clearly shows that both PySERA1 (Figure 6.15) and PySERA2 (Figure 6.16) are located in a central region in the schizont and surrounds the merozoites that are counter-stained blue by DAPI. Although confirmation of co-localization with the merozoite surface was not performed, these stains strongly imply that the SERAs also surround the daughter merozoites. However, there was no detectable co-localization with the rhoptries, suggesting that any interactions of the SERAs to the merozoites are only restricted to the general surface of the merozoites, thereby implying that any potential proteolytic processing of the merozoite proteins are restricted to the merozoite surface proteins.

Figure 6.15. A representative immunofluorescence microscopy stain of PySERA1 showing: (A) DAPI-stained daughter merozoite DNA; (B) FITC-labeled PySERA1-eGFP; (C) rhodamine-labeled PY235 denoting rhoptry organelles. (D) Merged images showing the clustering of PySERA1, especially around the merozoite bodies. Note that the free merozoites below the schizont also seem to be enveloped by PySERA1 and that these merozoites might also express variant members of PY235 that are not detected by the 1431 polyclonal antibody. (E) Phase contrast view. A white thick line represents a scale of 20 μ M. A total of ten random fields containing 2 to 3 schizonts were observed with similar results.

Figure 6.16. A representative immunofluorescence microscopy stain of PySERA2 showing: (A) DAPI-stained daughter merozoite DNA; (B) FITC-labeled PySERA2-eGFP; (C) rhodamine-labeled PY235 denoting rhoptry organelles. (D) Merged images showing the clustering of PySERA1, especially around the merozoite bodies. Note that a small proportion of the merozoites might express variant members of PY235 that are not detected by 1431 peptide antibody and thus not stained red. (E) Phase contrast view. A white thick line represents a scale of 20 μ M. A total of ten random fields containing 2 to 3 schizonts were observed with similar results.

6.3.8 Targeted gene disruption of PySERA1 and PySERA2

Since both the *P. yoelii* SERA1 and SERA2 genes can be targeted by the creation of fusion proteins with an eGFP tag, it is now possible to explore targeted gene disruption strategies to knockout the cysteine protease domain and hence generate transgenic parasites lacking the functional active catalytic site.

A single crossover integration strategy was chosen to generate parasites deficient in PySERA1, known as Δ PySERA1 clones. The same B3DhRap2/3 EGFP cloning strategy was used (Figure 6.8), with the Rap2/3 gene replaced by the gene-of-interest, in this case, only exon 1 of PySERA1 (amplified by primers 291e1F & 291e1R; Appendix 6) was retained leading to a truncation of the rest of the protein due to the presence of a stop codon after the eGFP sequence (Figure 6.17). In this strategy, the unique linearization site for this construct is *BsaBI*.

Figure 6.17. Schematic representation of the relative positions of the cysteine protease domain and the site of eGFP fusion for a PySERA1 knockout with respect to the organization of the full length PySERA1 protein. An arrow denoted the position where the PySERA1 gene is truncated with the introduction of the eGFP gene from the targeting vector, clearly removing the cysteine protease domain. (Adapted from PlasmoDB gene query at <http://www.plasmodb.org>)

Another targeting strategy was used to knockout PySERA2 due to the lack of a suitable unique restriction site for linearizing the insertional targeting vector. In this case, a double crossover replacement construct strategy was used for this gene. Basically, flanking regions of the gene or domain of interest are cloned into 2 sites on a replacement vector, in this case the b3D.DT^H.^D vector (Figure 6.18a) was used (a kind donation by Blandine Franke-Fayard and Andrew Waters from the University of Leiden, the Netherlands). The region upstream of the protease domain of the target of interest, hereby known as region I, was amplified via PCR (using primers 292F_KpnI and 292R_HindIII; Appendix 6) and cloned into the *KpnI* and *HindIII* sites in the 5' to 3' direction. Similarly, the region downstream of the target of interest (protease domain), hereby known as region II, was PCR amplified (using primers 292F_EcoRI and 292R_SacII, Appendix 6) and cloned in via the *EcoRI* and *SacII* sites in the 5' to 3' direction. The construct will then be linearized at both the *KpnI* and *SacII* restriction sites, after which two site-specific recombination events will cause the replacement of the genomic sequence with the replacement construct sequence, thus effectively disrupting the gene or domain of interest as the target sequence in the genome will be replaced by the vector backbone sequence (Figure 6.18b).

Both constructs were electroporated into *P. yoelii* YM late stage schizonts and selected for by pyrimethamine that was provided in the drinking water. Resistant parasites that survived the drug selection are typically harvested around 10 days post-infected of electroporated parasites.

Figure 6.18. Targeted gene disruption of PySERA2 using a double crossover strategy.

(A) The upstream region I is introduced into the b3D.DT^H.^D vector via the *KpnI* and *HindIII* restriction sites while the downstream region II is cloned in via the *EcoRI* and *SacII* restriction sites. Resistance to anti-folate drugs is provided by complementation with the *Toxoplasma spp.* dihydrofolate reductase *TgDHFR*. (B) Linearization at the *KpnI* and *SacII* restriction sites linearizes the vector and a double crossover event replaces the genomic sequence containing the cysteine protease domain (between positions b and c) with the vector backbone sequence. Single crossover events as shown (I & II) can occur but will not efficiently remove the cysteine protease domain (Adapted from de Koning-Ward, T.F. *et al*, 2000).

The Δ PySERA1 parasites were screened via PCR to identify the characteristic regions generated by this recombination event (Figure 6.8b). Basically, the 5' integration event was screened using PCR primers 291ex1_5'int and EGFP_R (Appendix 6) for the expected 1.5kb band while the 3'-integration event was screened using PCR primers 291ex1_3'int and Rap2/3_3' intF (Appendix 6) for the expected 1.7kb band (Figure 6.19). The agarose gel electrophoresis profile of total blood DNA agrees with the theoretical sizes for both 5' and 3' integration events and therefore the infected erythrocytes do contain transgenic parasites that are deficient in PySERA1.

Similarly, the transfected PySERA2 knockouts known as Δ PySERA2 were also screened to identify regions that have correctly integrated and replaced the endogenous wild-type genome sequence. PCR primers 292int5'F and B3D5'Rint (Appendix 6) were used to screen for integration of region I while primers 292int3'R and B3D5'Fint (Appendix 6) were used to screen for region II (Figure 6.20a). The agarose gel electrophoresis profile of total blood DNA also agrees with the theoretical sizes of 1.0kb for both region I and region II integration events and therefore the infected erythrocytes do contain transgenic parasites that are deficient in PySERA2 (Figure 6.20b).

The results indicate that both the single crossover strategy for the targeted removal of PySERA1 and the double crossover replacement strategy for the replacement of PySERA2 were successful in generating viable transgenic parasites deficient in either of these two proteases when cloned.

Figure 6.19. PCR screen of the PySERA1 knockout parasites. Lane M denotes the 1kb marker while lanes 1 and 3 represent the 5' integration screen for transfected parasites and wild-type parasites control respectively. Similarly, lanes 2 and 4 represent the 3' integration screen for transfected parasites and wild-type parasites control respectively. Only transfected parasites were PCR positive for both 5' and 3' integration PCR screens at 1.5kb and 1.7kb respectively. The spurious bands at the bottom of the gel are likely to be primer dimers.

Figure 6.20. Detection of the PySERA2 knockout parasites. (A) Diagram showing successful targeted replacement with the relative location of the regions I and II (shown in hatched boxes) and their boundaries demarcated by their respective restriction sites. The locations of the PCR primers for replacement screen are indicated by thin arrows. (B) Agarose gel electrophoresis profile of Δ PySERA2 parasites with the 1kb marker loaded in lane M. Region I screen was positive for transfected parasites (lane 3) but negative for wild-type parasites (lane 2) and vector control (lane 1). Region II screen was positive for transfected parasites (lane 6) but absent for wild-type parasites (lane 5) and vector control (lane 4).

6.3.9 Clonal isolation of transgenic *P. yoelii* parasites containing gene disruptions of PySERA1 and PySERA2

Both SERA deficient transfectant parasites could contain mixed populations of break-through resistant wild-type parasites and also parasites that could somehow be genetically modified only at one and not at both sites. Therefore, stabilates were re-introduced into Balb/c mice and then limiting dilutions of 1 parasite per mouse were introduced into recipient mice for expansion and isolation of pure clones.

Parasites isolated from recipient mice infected with Δ PySERA1 parasites at 1 parasite per mouse were then screened for 5' and 3' integration of the single crossover insertional vector. The same PCR primers as described in the previous section were used to screen for 5' and 3' integration events (Figure 6.21). Notably, not all the cloned out parasites contain both the 5' and 3' integration sites, though most of them are positive for both. A further PCR screen for the possibility of wild-type parasites was also performed using primers 291e1F and 291Nco (Appendix 6) such that wild-type parasites would give a 3.5kb band and integrants would be negative as the vector backbone contributes an extra 7.8kb block of sequence (Figure 6.22). Clearly, parasites from mice 3, 13, 15 and 17 are still of a heterogeneous population where the wild-type parasites are still present.

Figure 6.21. Agarose gel electrophoresis profiles of PCR screens for the presence of correctly inserted targeting vector in recipient mice for the pure clone isolation of Δ PySERA1 parasites. Lane M represents the 1kb marker lane while the lane numbers represent the respective mice total DNA subjected to (A) 5' integration and (B) 3' integration screens.

Figure 6.22. Agarose gel electrophoresis profiles of PCR screens for the presence of wild-type parasites in recipient mice for the pure clone isolation of Δ PySERA1 parasites. Lane M represents the 1kb marker lane while lane C represents the control wild-type parasite DNA showing the target 3.5kb band. Parasites from mice 3, 13, 15 and 17 contain the 3.5kb band and thus positive for wild-type parasites. The spurious lower molecular weight bands are likely to represent non-specific bands.

Lastly, the 8 remaining pure clones of Δ PySERA1 parasites were subjected to a Southern blot screen to further confirm the findings from the previous PCR screens. The probe was synthesized via PCR using primers 291e1F & 291e1R (Appendix 6) that were previously used to target the coding region of the gene prior to the protease domain (Figure 6.17). DNA was digested with *KpnI* and the expected wild-type profile will result in a 11.3kb band, while parasites containing the integrated vector will generate a 2.5kb band due to the introduction of an additional *KpnI* restriction site (Figure 6.23a). From the blot itself (Figure 6.23b), the wild-type parasite DNA gave an expected band of 11.3kb. Although the vector control was overloaded on the gel, the expected vector size of 7.3kb can be also be seen together with some non-specific digestion fragments. On hindsight, this vector control is unnecessary as the aim is to isolate pure clones of Δ PySERA1 parasites that do not contain any wild-type parasites. More importantly, the cloned out parasites' DNA were positive for pure Δ PySERA1 clones for clones 1, 4, 7, 8, 16 and possibly also clone 18. Since clone 1 gave the strongest signal, this was then chosen as the stock of *P. yoelii* Δ PySERA1.

It is not clear why the other lanes did not give any bands on the Southern blot although all lanes showed the typical smear associated with a successful restriction digest and adequate sample loading. One possible explanation would be that the DNA from these clones contained significantly more mouse DNA than parasite DNA, thus lowering the sensitivity of detection.

Figure 6.23. Southern blot strategy for the detection of transgenic parasites that are pure clones of Δ PySERA1. (A) The location of the *KpnI* restriction sites are shown with respect to the contig MALPY00082. The coding region is boxed while the vector targeting sequence is in hashed boxes. (B) Lane M represents the 1kb marker, lane 1 represents the vector control and lane 2 represents the wild-type parasite DNA. The other lanes are for probing the cloned out parasites.

In the case for detecting pure cloned out parasites isolated from recipient mice infected with Δ PySERA2 parasites at 1 parasite per mouse, PCR screening for the entire replacement integration event hinged on the detection of regions I and II as 1.0 kb target bands as illustrated in Figure 6.20a. Twelve recipient mice were successfully parasitized and their total blood was harvested and screened for the replacement event (Figure 6.24). Wild-type parasite DNA were negative for both screens while a positive control using the initial mixed population of transfected parasites yielded the expected 1.0kb bands for both region I and II PCR screens. Only parasites derived from recipient mice positive for both regions were considered successfully cloned out as a double crossover replacement strategy is also amenable to single crossovers, thus resulting in insertion of the targeting vector without excising the endogenous target DNA site (de Koning-Ward, T.F. *et al*, 2000). Thus, only clones 2, 4 and 9 consist of parasites that have undergone successful double crossover recombination to generate the PySERA2 knockout genotype.

These clones were further screened for the presence of wild-type parasites using primers 292int5'F and 292wtR (Appendix 6) that flank the endogenous coding region I such that wild-type parasite should give an expected size of 1.0kb while successfully transfected parasites would have the anti-sense primer region replaced and hence yield no product. It can be clearly seen that clones 2, 4 and 9 that contains the correct replacement construct are pure clones devoid of any wild-type parasites (Figure 6.25). Certain lanes did not give any PCR products for all the screens and may be due to some inherent contamination in the DNA prep. In conclusion, clone 4 was then arbitrarily chosen as the stock of *P. yoelii* Δ PySERA2.

Figure 6.24. Agarose gel electrophoresis profiles of PCR screens for the presence of correctly inserted replacement targeting vector in recipient mice for the pure clone isolation of Δ PySERA2 parasites. Lane M represents the 1kb marker while lanes T represent the initial mixed parasite population containing the successful transfectants while lane W represent wild-type parasites. Lanes 1 to 12 in (A) represent the recipient mice DNA screened for region I of PySERA2 knockout while those in (B) were screened for region II of the PySERA2 knockout.

Figure 6.25. Agarose gel electrophoresis profiles of PCR screens for the presence of wild-type parasites in recipient mice for the pure clone isolation of Δ PySERA2 parasites. Lanes M represent the 1kb marker while lane T represents the initial mixed parasite population containing the successful transfectants. Clearly, lanes 1 to 12 that represent clones from each recipient mouse were evidently devoid of wild-type parasites.

6.3.10 Phenotypic assessment of virulence of *P. yoelii* Δ PySERA1 clone

Firstly, virulence can be defined by the rate of invasion and hence the rate of increase of parasitemia. Secondly, the preference or non-preference of parasites to invading reticulocytes can help to limit the expansion of the parasite population as these host cells make up only 2% of the total circulating erythrocytes. Next, the maximum parasitemia reached in the course of the infection and the amount of time required to clear the infection can also be used to determine virulence. Finally, the survival or death of the host animal is also a key factor to consider.

Purity cloned Δ PySERA1 parasites were first introduced into donor Balb/c mice and then grown up to 1 to 4 % parasitemia. An aliquot of blood from the tail vein was drawn and put into complete RPMI for direct cell count with a hemocytometer. Further dilutions in complete RPMI were made to a concentration of 1000 infected erythrocytes per 100 μ l of media. Five recipient mice were then chosen for monitoring the progress of infection at an identical dose of 1000 parasitized erythrocytes per mouse, injected via the tail vein.

As a control, wild-type parasites were also grown up in the same manner and similarly, five recipient mice were infected with 1000 parasitized erythrocytes per mouse via the tail vein. A broad overview of the parasitemias of these 2 groups of mice (Figure 6.26) exhibit the classical two-step increase in parasitemia as parasites preferentially invades reticulocytes that constitute 2% of the total erythrocyte population before switching to invading normocytes (Bugener, W., 1985; Fahey, J.R. and Spitalny, G.L., 1984). Both populations of mice clearly showed a lower invasion potential during days 3 and 4 due to their preference to invading reticulocytes. This is shown by their average parasitemias on days 3 and 4 to be less than 10% for both YM and Δ PySERA1. On the following day (day 5), both groups of parasites switched to

invading normocytes and this is clearly shown by the rapid invasion of erythrocytes from an average of 6.4% to 51.5% for YM and from 4.3% to 31.7% in for Δ PySERA1 (Figure 6.27). Although this translates from a 8.05-fold increase in parasitemia in YM as compared to a 7.37-fold increase in Δ PySERA1, the parasitemias of the YM and Δ PySERA1 sample populations were significantly different based on a T-test score of 0.00194. While the t-test assumes that the samples are normally distributed, another test for significance such as the non-parametric Mann-Whitney test can be used, as this test does not assume that the samples are normally distributed. The Mann-Whitney p -value for day 5 is also significant at 0.006 and 0.0121 for a one- and two-tailed p value respectively. In addition, the Mann-Whitney one-tailed p value for day 4 is significant at 0.0475, suggesting that the median parasitemia of Δ PySERA1 infected mice were significantly lower than YM infected mice. Also, the rate of invasion can be performed by measuring the slope of the increase of parasitemia from day 4 to day 5. The slope for YM is 0.451 while the slope for Δ PySERA1 is 0.274 and therefore the rate of invasion of YM is 1.646 times higher than Δ PySERA1.

Figure 6.26. Parasitemia curves of the sample populations of recipient mice infected with standardized doses of Δ PySERA1 and control wild-type parasites. (A) Δ PySERA1 parasitemia curve where 1 mouse died on day 8 and 2 died on day 9 while the remaining mice cleared the infection by day 25 post-infection. (B) Wild-type YM parasitemia curve where 1 mouse died on day 7 and 2 died on day 8 while the remaining mice cleared the infection by day 33 post-infection. An asterisk denotes the death of 1 animal and marks the last parasitemia reading of the previous day.

Figure 6.27. Average parasitemias growth curves of wild-type YM and Δ PySERA1 (291ko) parasites inoculated into 5 mice per sample population. The difference in parasitemia loads in the pick-up phase is most significant on day 5 as mice infected with Δ PySERA1 parasites had significantly less parasitemia than wild-type YM parasites. A T-test comparison of the parasitemias from mice infected with YM to mice infected with the PySERA1 knock out parasites at day 5 yielded a p -value of 0.00194. P values for the Mann Whitney test was 0.0475 (one tailed) on day 4 while values for day 5 were 0.006 (one-tailed) and 0.0121 (two-tailed). When the parasite load is at a plateau from days 5 to 19, mice infected with Δ PySERA1 parasites had moderately less parasitemia than wild-type YM parasites. Error bars represent the standard error of the mean (SEM) of the surviving mice (n). For YM, $n=5$ from day 0 till day 6, $n=4$ on day 7 and $n=2$ from day 8 till the end of the experiment. For PySERA1 knockout mice, $n=5$ from day 0 till day 7, $n=4$ on day 8 and $n=2$ from day 9 till the end of the experiment.

To note, YM infections in our facility are not all lethal and this phenomenon could be due to different components of feed as it was shown that mice fed with a diet high in oxidants were able to survive infections with a lethal strain of *P. yoelii* (Taylor, D.W. *et al*, 1997). The next significant observation was the mortality of the infected mice (Figure 6.26). One YM-infected mouse died first on day 7 and subsequently 2 more followed on the next day. For Δ PySERA1-infected mice, the first fatality occurred on day 8 where 1 mouse died and was followed by 2 more mice on the next day. The parasitemia of the individual YM-infected mice steadily increased and the readings on the day before they succumbed were 84.9%, 81.6% and 75.1%, thus implying that these animals died from fulminating parasitemia and anemia. However, only 1 Δ PySERA1-infected mouse (mouse 3) died from fulminating parasitemia at 74.5% on the day before it died. More significantly, the other two Δ PySERA1-infected mice that succumbed on day 9 actually had parasite loads that dropped from a peak of 65.9% on day 6 to 51.9% on day 8 for mouse 2 while the parasite load for mouse 5 dropped from a peak of 79.3% on day 7 to 52.2% on day 8. Clearly, these mice were on the way to clearing the parasites and the cause of death could be due to additional complications. Infections with virulent strains of *P. yoelii* have resulted in pathophysiological changes such as intravascular sequestration of brain capillaries (Yoelii, M. and Hargreaves, B.J., 1974). More recently, RANTES, CCR 1, CCR 3 and CCR 5 were implicated in cerebral inflammation and cellular damage during infections with a lethal strain of *P. yoelii* (Sarfo, B.Y. *et al*, 2005). Therefore, mice 2 and 5 infected with Δ PySERA1 parasites probably succumbed to these complications.

After day 5, mice that had not succumbed to the infections were typically suffering from a chronic infection. By comparing the average parasitemias of the remaining mice from both groups from day 5 to 19 where the parasitemia remains in a

plateau stage, mice infected with Δ PySERA1 parasites had a lower average parasitemia than YM-infected mice. After which, Δ PySERA1-infected mice began to clear the infection and were parasite-free by day 25. This is in significant contrast to YM-infected mice where parasites were cleared a whole 8 days (i.e. 8 rounds of the erythrocytic cycle) after Δ PySERA1-infected mice survived and cleared the infection.

A sample Giemsa-stained blood film from both populations during the chronic infection stage (Figure 6.28) revealed massive reticulocytosis in mice from both populations as a host response to replenish erythrocytes. Both parasite populations show reticulocyte preference as YM parasites seem to have up to two or three parasites per reticulocyte while Δ PySERA1 parasitized mice can be observed to contain up to three to five parasites per reticulocyte. While statistical methods to establish reticulocyte selectivity have not been established for chronic infections where reticulocytes significantly outnumber normocytes, there is some suggestion that multiple-infected reticulocytes were more prevalent in Δ PySERA1-infected mice than YM-infected mice, thus supporting the lower parasitemias of Δ PySERA1-infected mice during the chronic infection stage.

Based on this observation, a statistical measure of host cell selectivity via the Selective Index (SI) calculation (Simpson, J.A. *et al*, 1999) whereby the ratio of the observed number of multiple-infected cells to the expected number of multiple-infected cells, can be used to differentiate parasites of varying virulence. This method was validated recently with the correlation of low SI to virulent parasites that can invade erythrocytes of all ages while a high SI is representative of parasites that can only invade a limited repertoire of erythrocytes and hence correlate with a less virulent phenotype (Iyer, J. *et al*, 2007).

Figure 6.28. Giemsa-stained slides of the typical blood film smears of mice suffering from a chronic infection. Reticulocytes contain residual nucleic acids and are thus absorb more of the stain, hence the entire reticulocyte is stained a darker shade of purple-blue. At chronic infections, the host animal undergoes an increased rate of reticulocytosis, with reticulocytes making up almost half of the total erythrocyte repertoire. (A) A typical blood smear of a Δ PySERA1-infected mouse; (B) A typical blood smear of a YM-infected mouse. Both populations switched back to a reticulocyte-restricted invasion phenotype while total parasite counts support YM as being more invasive and less host cell restricted than Δ PySERA1.

As this statistical test was recommended to be performed on blood smears from 5% to 20% parasitemia, SI was determined when parasitemia was between these levels. On the other hand, those with more than 20% parasitized erythrocytes will also create bias in the data, as there are practically no more free reticulocytes. Thus, mice that exhibit parasite loads from 5% to 10% were assayed for SI so as to minimize variations due to differences in parasitemias and availability of reticulocytes.

It is clear from the previous section that the initial multiplication rate in Δ PySERA1-infected mice is less efficient than YM parasites. This was supported by measuring SI (Figure 6.29), where for Δ PySERA1 it was reproducible and tightly clustered with a mean of 4.650 while those for YM ranged from 1.071 to 1.807 with a mean of 1.434. A T-test significantly shows that Δ PySERA1 parasites have a significantly stronger preference for a subset of cells as compared to the wild-type YM (p -value = 0.000341). This explains also the apparent reticulocyte preference for Δ PySERA1 parasites in the chronic infection stage and may also contribute to its early clearance from the host animal as parasites within multiple-infected erythrocytes usually do not survive and differentiate into schizonts. Since a stronger reticulocyte preference would translate to less parasitized cells in subsequent rounds of erythrocyte invasion, nascent reticulocytes will have a higher chance of maturing into normocytes, thus reducing the pool of suitable host cells and effectively hastening clearance of the parasites.

Figure 6.29. A scatter plot of the Selective Index (SI) values of both Δ PySERA1 and YM parasites calculated from infected mice with parasitemia loads of 5% to 10%. A line across the data points denotes the mean SI value and the corresponding values for Δ PySERA1 and YM are 4.650 and 1.434 respectively (p -value = 0.000341). Each dot represents the SI of an individual mouse. Only two Δ PySERA1 infected mice had comparable parasitemias for SI calculation and are represented as two close data points.

6.3.11 Phenotypic assessment of virulence of *P. yoelii* Δ PySERA2 clone

Cloned Δ PySERA2 parasites were first introduced into donor Balb/c mice and then grown up to 1 to 4 % parasitemia. An aliquot of blood from the tail vein was drawn and put into complete RPMI for direct cell count with a hemocytometer. Further dilutions in complete RPMI were made to a concentration of 1000 infected erythrocytes per 100 μ l of media. Five recipient mice were then chosen for monitoring the progress of infection at an identical dose of 1000 parasitized erythrocytes per mouse, injected via the tail vein.

There is not much difference between the invasion profile of Δ PySERA2 and YM parasites (Figure 6.30). However, day 4 to day 5 is also the time where Δ PySERA2 parasites switch from a reticulocyte restricted invasion profile to invading normocytes. On this day, the average parasitemia of Δ PySERA2-infected mice increased from 7.3% to 36.2% while the count for YM-infected mice increased more significantly from 6.4% to 51.1%. This represents a 28.9-fold increase in Δ PySERA2-infected mice as compared to a 44.7-fold increase in YM-infected mice. The day 5 parasitemias of the YM and Δ PySERA2 sample populations were also significantly different based on a T-test score of 0.0149. The Mann-Whitney *p*-value for day 5 is also significant at 0.0183 and 0.0366 for a one- and two-tailed *p* value respectively.

Δ PySERA2-infected mice 1 and 5 that died on day 7 had constantly increasing parasitemias and most likely died from anemia. However, mice 2 and 3 that died on days 8 and 12 respectively actually had parasitemias that were on a decreasing trend. In addition, the remaining mouse 4 that died on day 20 slipped into a chronic infection. Although all the Δ PySERA2-infected mice died by day 20, the average parasitemias from both groups from day 5 to 19 during the chronic infection stage,

mice infected with Δ PySERA2 parasites had marginally lower parasitemia than YM-infected mice, with mortality occurring after recovering from a peak parasitemia.

Figure 6.30. Parasitemia curves of the sample populations of recipient mice infected with standardized doses of Δ PySERA2 and control wild-type parasites. (A) Δ PySERA2 parasitemia curve where 2 mice died on day 7, another died on day 8 and another mouse on day 12. The last mouse expired on day 20. (B) A scatter plot combining the parasitemia curves of Δ PySERA2 and YM infected mice. An asterisk denotes the death of 1 animal and marks the last parasitemia reading of the previous day. Error bars represent the standard error of the mean (SEM) of the surviving mice (n). For YM, n=5 from day 0 till day 6, n=4 on day 7 and n=2 from day 8 till the end of the experiment. For PySERA2 knockout mice, n=5 from day 0 till day 6, n=3 on day 7 and n=2 from day 8 till day 11. Only one PySERA2 knockout infected mouse survived till day 20.

The Selective Index (SI) calculation for Δ PySERA2 was also performed on mice with parasitemias from 5% to 10% (Figure 6.31). Although the mice were matched as closely as possible in terms of age and weight, the variation in SI is quite huge and although the average SI is 2.531 as compared to YM's 1.434, the difference is not statistically significant (p -value = 0.114).

Figure 6.31. A scatter plot of the Selective Index (SI) values of both Δ PySERA2 and YM parasites calculated from infected mice with parasitemia loads of 5% to 10%. A line across the data points denotes the mean SI value and the corresponding values for Δ PySERA1 and YM are 2.531 and 1.434 respectively (p -value = 0.114). Each dot represents the SI of an individual mouse.

While the relative transcription of *sera2* (PY00292) is higher in the virulent phenotypes of *P. yoelii*, the lack of a phenotypic change in the PySERA2 knockout clone was surprising. Although all PySERA2 knockout infected mice died earlier than YM infected mice, other factors such as a slight lag in invasiveness from day 4 to day 5 post-infection as well as a drop in parasitemia after an acute phase suggest that the PySERA2 knockout phenotype is neither more nor less virulent than YM. Perhaps the evidence for a role in virulence for PySERA2 is not strongly apparent from this data but a reasonable explanation might be due to a certain degree of redundancy where function can be compensated by the other SERAs.

6.3.12 Real-time RT-PCR analysis of the transcriptional profile of the SERA genes in both Δ PySERA1 and Δ PySERA2 parasite lines

The accession numbers of the five *P. yoelii* SERA genes have been identified in the previous section and it would be informative if there were some kind of compensatory mechanism whereby another SERA gene is up-regulated when one is disrupted (McCoulbrie, J.E. *et al*, 2007). Hence a quick screen of the entire known repertoire of *P. yoelii* SERA genes in the knockout lines could shed some light if this is indeed happening for this animal model species.

Total RNA from late schizont stage parasites was used as the template and while the knockout lines were confirmed to be deficient for their respective knockout transcripts. On the other hand, the other SERA genes did not exhibit any significant changes in transcriptional activity (Figure 6.32) and hence do not seem to be compensating for the lack of the disrupted gene in any way. Interestingly, the PbSERA5 gene was transcribed exclusively in the mosquito stage (Schmidt-Christensen, A. *et al*, 2008) but the RT-PCR data here clearly shows that the *P. yoelii* SERA5 gene is also transcribed in the erythrocyte stage. This might reflect some redundancy of this gene in *P. yoelii* such that it can potentially function in the blood stage.

Current sequence data only supports the rodent parasite species with having five SERA genes. This is significantly less than the numbers found in the more complete genomes of *P. falciparum* and *P. vivax* that number 9 and 12 respectively. Since the 2 PySERA knockouts are of the group with a serine substitution in the active cysteine protease domain. Since the SERAs in this group are much more varied in other malaria parasite species (Arisue, N. *et al*, 2007), the possibility of other undiscovered SERAs in *P. yoelii* cannot be discounted due to the poor genome

coverage of the rodent malaria parasite species. In addition, the fact that PbSERA5 is transcribed specifically in the mosquito stages (Schmidt-Christensen, A. *et al*, 2008) while PySERA5 (Figure 6.32) is transcribed in the *P. yoelii* schizonts seems to suggest that they are function differently or this gene is also required in the blood stages for *P. yoelii*. This difference could be species-specific or point to casual phylogenetic associations in the incomplete genomes of both species where the entire repertoire of the members of this multigene family have not been fully discovered and the current ‘best-hit fit’ in genome mining simply creates associations with the current genome sequence (Arisue, N. *et al*, 2007).

Figure 6.32. Real-time RT-PCR screen of wild type, Δ PySERA1 and Δ PySERA2 parasite clones for all five known SERA genes. The knockout lines were confirmed to be deficient for their respective knockout transcripts. On the other hand, the other SERA genes did not exhibit any significant change in transcriptional activity. Based on the re-annotation work by Arisue, N. and co-workers (2007), the *P. yoelii* SERAs from 1 to 5 corresponds to the following accession numbers: PY00291, PY00292, PY00293, PY02062 and PY02063 respectively.

6.4 Conclusions

The comparison of transcriptional activity of genes clearly isolated a small number of genes found to be transcribed significantly higher in YM than in YA, thus suggesting that these genes could be virulence markers. The absolute difference in transcriptional activity of these selected genes were assayed via real-time RT-PCR. Furthermore, real-time RT-PCR comparison of absolute transcript levels in the genetic crosses supported the role of eight genes as ‘universal’ markers of virulence, of which two were SERA proteases.

Based on the homology comparison work to identify SERA proteins in the malaria parasite species (Arisue, N. *et al*, 2007), five putative SERA proteins in *P. yoelii* were identified. With reference to their genomic location, their corresponding accession numbers were identified and two of these genes, PySERA1 and PySERA2 were chosen for further characterization as they were linked to virulence. These two genes were successfully tagged with eGFP and their predominant centralized intracellular location within a fully segmented schizont strongly suggests localization within the parasitophorous vacuole, similar to the *P. falciparum* SERA proteins. Western blots also gave the characteristic proteolytic processing patterns, thereby confirming their function as SERA proteins.

Genetic knockout clones deficient for the protease domain were also generated, with Δ PySERA1 parasites showing some preference towards reticulocytes as evident from a higher SI score and a moderately lower parasitemia load during the chronic infection stage where the reticulocyte count is significantly increased. Most significantly, mice that were chronically infected with Δ PySERA1 parasites cleared the infection a full eight days earlier than wild-type YM-infected mice. However, Δ PySERA2 parasites did not show any significant reticulocyte restriction. Although

mice that were chronically infected with Δ PySERA2 parasites had a marginally lower parasite burden and even suggested some form of parasite clearance, they all succumbed to the infection. Therefore, the Δ PySERA1 parasites seem to exhibit an intermediate virulent phenotype while Δ PySERA2 parasites did not show any significant change in virulence.

Real-time RT-PCR analysis of all the five PySERA genes did not reveal any transcriptional up regulation of the other remaining members, thus suggesting a redundancy in function amongst the members of this multigene family. Another possible reason for the lack of compensation by the other members could be due to the non-deleterious nature of the knockout phenotypes and thus there is no critical need to complement the function of the deleted gene. A strong possibility of other as yet undiscovered SERA genes in the incomplete genome cannot be excluded when compared to the huge repertoire of the members of this multigene family that have arisen through gene duplication events in one locus in the more completely annotated genomes of the human parasites *P. falciparum* and *P. vivax*.

SERA knockouts in the human malaria parasite *P. falciparum* have been generated without any significant phenotypic effects and only PfSERA5 and PfSERA6 cannot be disrupted (Miller, S.K. *et al*, 2002; McCoubrie, J.E. *et al*, 2007). Investigations using cultured blood are typically devoid of reticulocytes as they age rapidly in culture. Also, the 48 hour cycle of the human blood stage cycle means that any existing pool of existing reticulocytes would have matured into normocytes. The use of an *in vivo* system here accurately replicates the physiological and immunological pressures that the parasite would typically encounter in a natural host. It is clear that both the presence of reticulocytes at the beginning of the infection and the compensatory reticulocytosis generated by the host animal during the chronic

stage of the infection are significant to the outcome and are missing in culture systems. The observed modulation of infectivity and increase in reticulocyte specificity of the knockout parasites are obvious in Δ PySERA1 but not in Δ PySERA2 parasites. These observations point to some role in modulating virulence and that the SERAs may function co-operatively and perhaps with some of the members being more specific in their target while others being more relaxed in substrate specificity. A summary table of the *Plasmodia* SERA localization and knockout phenotype studies is shown in Table 6.2. Since the SERA proteins are potentially associated with the merozoite surface (Miller, S.K. *et al*, 2002) and that the cleaved off 47kDa and 18kDa fragments associate with the merozoite surface upon rupture of the schizont, it is possible to speculate that these SERA proteins could somehow aid in the proteolytically processing of proteins found on the merozoite surface. The presence of processed forms of merozoite surface 1 protein just prior to schizont rupture (Holder, A.A. *et al*, 1987) points to a proteolytic activation event. It would be interesting to establish whether the SERAs are indeed responsible and also to identify other potential targets for this class of proteases that can function beyond host cell egress. Also, we are attempting to generate PySERA1 and PySERA2 double knockout parasites for further investigations into the redundancy and invasive potential of these proteases.

SERA Localization	References
PfSERA3, 4, 5, 6 and 9 localize to the lumen of the parasitophorous vacuole	Delplace, P. <i>et al</i> , 1987 Knapp, P. <i>et al</i> , 1989 Knapp, P. <i>et al</i> , 1991 Aoki, S. <i>et al</i> , 2002 Miller, S.K. <i>et al</i> , 2002
PfSERA4 and PfSERA5 co-localize with MSP-1 in late stage schizonts	Miller, S.K. <i>et al</i> , 2002
PbSERA3 in hepatic schizonts localize mainly in the parasite cytoplasm	Schmidt-Christiansen, A. <i>et al</i> , 2008
PbSERA1 in hepatic schizonts localize mainly in the parasitophorous vacuole	Schmidt-Christiansen, A. <i>et al</i> , 2008
PySERA1 and PySERA2 potentially localize in the PV lumen	This study

SERA knockouts	References
PfSERA1, 2, 3, 4, 7, 8, 9 knockout lines successfully generated without significant phenotypic change	Miller, S.K. <i>et al</i> , 2002 McCoubrie, J.E. <i>et al</i> , 2007
PfSERA4 knockout lines seem to up regulate transcription of <i>Pfsera5</i>	McCoubrie, J.E. <i>et al</i> , 2007
PySERA1 and PySERA2 knockout lines successfully generated without compensatory increase in transcription of the other PySERAs	This study
PySERA1 knockout clone preferentially invades reticulocytes, inflicts a moderately lower parasitemia load on the host and chronic infections are cleared significantly earlier than wild-type parasites	This study

Table 6.2. Summary table of the SERA knockout phenotypes and localization studies in the malaria parasites. Previous work done on *P. falciparum* blood stage and *P. berghei* liver stage parasites are shown together with the data described in this study.

Conclusions

7.1. Comparative genome analysis of the three rodent malaria parasite species

Plasmodium berghei*, *Plasmodium chabaudi* and *Plasmodium yoelii

The rodent malaria parasite species, namely, *P. berghei*, *P. chabaudi* and *P. yoelii* are common laboratory parasite species used for the *in vivo* model for investigations related to pathogenesis, chemotherapy, host immune mechanisms and vaccine trials. Although the genome of the virulent human malaria parasite *P. falciparum* has been sequenced to a high degree of confidence and resolution, the rodent parasite genomes are still fraught with gaps and are poorly annotated. This is especially problematic when a gene of interest in the human malaria parasite cannot be matched to an ortholog in the rodent parasites, thereby presenting a stumbling block to further studies using the animal model systems.

By using our in-house pan-rodent oligonucleotide array chip, comparative genome hybridizations produced interesting results. Basically, the array probes were designed to annotated protein coding genes in as many rodent parasite species as possible, i.e. depending on sequence availability, a probe can hybridize to an orthologous gene in all three, two or only one species. Therefore, the presence of a signal for a species to which a probe has not been designed for suggests the presence of a missing gene. Existence for a selection of these ‘newly’ discovered genes was also confirmed by PCR. Based on this information, missing genes were appended to a synteny map using the most established genome of *P. falciparum* as an index. The addition of non-syntenic orthologs also filled additional missing genes. In conjunction, a bioinformatics approach via a tBLASTn search of the rodent genomes using the amino acid sequences of *P. falciparum* also detected the presence of

unannotated orthologs. These steps considerably increased the number of ‘core’ genes common to all the four species. The use of non-syntenic orthologs is immensely useful to finalize the presence or absence of genes that have been displaced via chromosomal rearrangements. Bioinformatics search using tBLASTn basically queries and pulls out genome sequences that contain potential coding information for genes with similar functions. The above-mentioned search complements the direct hybridization data as we cannot design the oligonucleotide probes to a proteing coding gene that is not annotated and thus cannot foresee potential sequence polymorphisms of a potential corresponding ortholog. While the above strategy significantly improves the linkage of orthologs between the four species, there still remain groups of *P. falciparum* genes lacking at least an ortholog from one of the rodent parasite species. These ‘missing’ genes were screened via PCR and the results suggest that a *P. falciparum* gene that has at least one rodent parasite ortholog could potentially also be orthologous to all three rodent parasite species.

A rodent specific orthology map was also constructed using the more complete and annotated *P. yoelii* as the index. Genes that are orthologous to *P. falciparum* were first removed from the list and then hybridization data together with a tBLASTn search in the other two species using the amino acid sequence of *P. yoelii* as a query was done. While most genes are common amongst the three rodent parasite species, other clusters of genes appear to specific only to *P. yoelii* or a combination of these three species. Where differences do occur, they typically consist of hypothetical proteins and genes duplicated from the core set. Since the majority of these genes are either redundant copies or have no known function, we can conclude that the rodent parasite species are indeed genetically similar and perhaps regulation of gene expression is the key to define species-specific phenotypic differences.

7.2. Genomic differences between the virulent YM strain and the avirulent YA strain of the rodent malaria parasite *Plasmodium yoelii*

Following the analysis of the three rodent parasite species, a more thorough comparison was done between the virulent YM line and the avirulent YA and 17x1.1 lines of *P. yoelii*. Since the oligonucleotide probes on the array were designed based on the reference 17x1.1 line, genomic DNA from the YM and YA lines were analyzed for performance on the chip. Competitive hybridizations with either of these lines against 17x1.1 yielded interesting data, with some genes showing higher ratios and other showing lower ratios, thus suggesting gene duplications and sequence polymorphisms respectively. Of note, the YM line contains 198 genes polymorphic to 17x1.1 while the YA line contains 637 polymorphic genes with respect to 17x1.1. While the PIRs are expected to be variant due to their role in antigenic variation, annotated genes such as AMA-1, merozoite surface proteins as well as members of the PY235 multigene family are also expected as these invasion-related proteins are constantly under immune pressure. In addition, the presence of a significant number of polymorphic hypothetical proteins is indeed puzzling. A subset of these hypothetical proteins were subsequently found to contain at least one transmembrane domain thus strongly implying that these proteins are also exposed to immune surveillance and may function as novel ligands or mediate some form of antigenic variation. Some of these genes are transcribed in the blood stages and also contain orthologs in *P. falciparum*, thus adding confidence to their postulated function and differentiates universal or rodent parasite specific genes.

The sheer difference in the number of polymorphic genes suggests that YA is more divergent than YM. Since these lines were obtained from an initial field isolate as a mild strain, we propose that this source contained a mixed population of parasites

with different virulence phenotypes and that virulence was somehow modulated in such a mixed population. The discovery of the YM line due to a 'spontaneous conversion' of an avirulent line to a virulent line is not supported by the genomics evidence presented here but rather suggests that the isolated lines have evolved independently of each other.

7.3. Differential transcriptional profile of genes between the virulent YM and the avirulent YA strain of *Plasmodium yoelii*

Having already established the genomic differences between the YM and YA lines of *P. yoelii*, more information on novel virulence markers can be elucidated by comparing the global transcriptional activity of genes in the erythrocytic developmental stage of the parasite. A selection of genes deemed to be transcribed more abundantly in YM than YA were validated via real-time RT-PCR. These genes were typically transcribed at least five-fold higher in YM than YA. The generation of line-specific quantification standards for real-time RT-PCR using common primers flanking the oligonucleotide-binding region enabled the elucidation of the extent of sequence polymorphisms.

These genes were then further assayed via real-time RT-PCR on the genetic cross between the virulent YM and avirulent AC lines. The progeny of these two parental lines were of different virulence phenotypes. Seven genes were found to be consistently more abundantly transcribed in the virulent lines and comprise of: two SERA proteases (PY00291 & PY00292), a chromosomal segregation protein *smc1* gene (PY00653), a putative chaperon involved in protein-protein interactions (PY02799), a histidine kinase (PY05489), an arsenite transport subunit A (PY02198) and an MSP-7 like gene (PY02147). These genes, by virtue of being more abundantly

transcribed in the virulent lines of *P. yoelii*, can thus be considered to be linked in mediating virulence.

7.4. Characterization of the SERA1 and SERA2 genes of *Plasmodium yoelii*

The SERA genes are highly conserved amongst the malaria parasite species and are involved in egress of daughter merozoites. Processed SERA fragments are associated with the merozoite surface and the 47kDa fragment is a potential vaccine candidate. Based on the transcriptomic analysis between the virulent and avirulent lines of *P. yoelii*, two SERA genes namely, PySERA1 (PY00291) and PySERA2 (PY00292) were identified as universal markers of virulence, suggesting an alternative role of these proteases in mediating virulence.

Both SERAs were eGFP-tagged and seemed to localize in a central region of a segmented late stage schizont and surrounds the merozoites but do not co-localization with the rhoptries of the daughter merozoites.

Functional gene knockouts of both PySERA1 and PySERA2 were constructed to evaluate their *in vivo* function and to ascertain their role in virulence. In order to measure virulence, a repertoire of quantitative characteristics must first be defined. These include: (i) rate of invasion, (ii) reticulocyte preference measured by Selective Index (SI), (iii) maximum parasitemia and time taken to clear the infection and (iv) the survival or death of the host animal.

PySERA1 knockout parasites were less invasive than the wild-type YM line when they both switched to invading normocytes after an early infection stage where reticulocytes were still available. The knockout line showed a significantly lower rate of invasion based on parasitemia counts in infected animals while the rate of invasion is 1.65 times less than YM. Reticulocyte preference as measured by SI was

significantly higher in the knockout as compared to YM. While the peak parasitemia levels were not that different, host mice actually showed a decrease in parasitemia levels following this peak before a few succumbed to the infection. Surviving host mice that were chronically parasitized had moderately less parasitemia than chronically infected YM mice. Most significantly, the two surviving PySERA1 knockout infected mice cleared the infection eight days earlier than wild-type YM infected mice. Since a stronger reticulocyte preference would translate to less parasitized cells in subsequent rounds of erythrocyte invasion, nascent reticulocytes will have a higher chance of maturing into normocytes, thus reducing the pool of suitable host cells and effectively hastening clearance of the parasites. Both populations of PySERA1 knockout and YM infected mice had only two animals remaining at the end of the experiment and although there was no change in the number of surviving mice at the end, the mice infected with PySERA1 knockout parasites always succumbed a day later than the population infected with wild-type YM parasites. Therefore, the mortality of mice infected with PySERA1 knockout parasites were significantly delayed by a day and their parasitemia loads were on the decline, suggesting that fulminating parasitemia was not the direct cause of death in the host animal. Therefore, with the satisfaction of the four criteria of virulence, PySERA1 knockout parasites are moderately less virulent than the wild-type YM line.

As for an identical study of PySERA2 knockout parasites, there is a significantly lower invasion rate when the parasites switched to invading normocytes, however the differences in SI was not significant and this knockout line failed to satisfy the other criteria.

Real-time RT-PCR analysis of these knockout lines confirmed the absence of their respective knockout transcripts and did not show any compensation via

transcriptional up regulation of the other members of this multi-gene family. The results suggest that while PySERA1 might be more specific in modulating virulence, others such as PySERA2 may function co-operatively and share some redundancy in function with the other SERAs.

References

- Adams, J.H., Blair, P.L., Kaneko, O., Peterson, D.S. (2001). An expanding ebl family of *Plasmodium falciparum*. Trends Parasitol. 17: 297 - 299.
- Adams, J.H., Sim, B.K., Dolan, S.A., Fang, X., Kaslow, D.C., Miller, L.H. (1992). A family of erythrocyte binding proteins of malaria parasites. Proc Natl Acad Sci USA. 89: 1085 - 7089.
- Aidoo, M., Terlouw, D.J., Kolczak, M.S., McElroy, P.D., ter Kuile, F.O., Kariuki, S., Nahlen, B.L., Lal, A.A., Udhayakumar, V. (2002). Protective effects of the sickle cell gene against malaria morbidity and mortality. Lancet. 359: 1311-1312.
- Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., Ménard, R. (2006). Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. Nat Med. 12: 220 – 224.
- Anstey, N.M., Weinberg, J.B., Hassanali, M.Y., Mwaikambo, E.D., Manyenga, D., Misukonis, M.A., Arnelle, D.R., Hollis, D., McDonald, M.I., Granger, D.L. (1996). Nitric oxide in Tanzanian children with malaria: inverse relationship between malaria severity and nitric oxide production/nitric oxide synthase type 2 expression. J Exp Med. 184: 557 - 567.
- Aoki, S., Li, J., Itagaki, S., Okech, B.A., Egwang, T.G., Matsuoka, H., Palacpac, N.M., Mitamura, T., Horii, T. (2002). Serine repeat antigen (SERA5) is predominantly expressed among the SERA multigene family of *Plasmodium falciparum*, and the acquired antibody titers correlate with serum inhibition of the parasite growth. J Biol Chem. 277: 47533 – 47540.
- Arastu-Kapur, S., Ponder, E.L., Fonovic, U.P., Yeoh, S., Yuan, F., Fonovic, M., Grainger, M., Phillips, C.I., Powers, J.C., Bogyo, M. (2008). Identification of proteases that regulate erythrocyte rupture by the malaria parasite *Plasmodium falciparum*. Nat Chem Biol. 4: 203 - 213.
- Arisue, N., Hirai, M., Arai, M., Matsuoka, H., Horii, T. (2007). Phylogeny and evolution of the SERA multigene family in the Genus *Plasmodium*. J Mol Evol. 65: 82 - 91.
- Avarind, L., Iyer, L.M., Wellems, T.E., Miller, L.H. (2003). *Plasmodium* biology: genomic gleanings. Cell. 115: 771 – 785.
- Babon, J.J., Morgan, W.D., Kelly, G., Eccleston, J.F., Feeney, J., Holder, A.A. (2007). Structural studies on *Plasmodium vivax* merozoite surface protein-1. Mol Biochem Parasitol. 153: 31-40.
- Baker, R.P., Wijetilaka, R., Urban, S. (2006). Two *Plasmodium* rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. PLoS Pathog. 2: e113.
- Balaji, S., Babu, M.M., Iyer, L.M., Aravind, L. (2005). Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Res. 33: 3944 – 4006.

References

- Barnwell, J.W., Nichols, M.E., Rubinstein, P. (1989). In vitro evaluation of the role of the Duffy blood group in erythrocyte invasion by *Plasmodium vivax*. J Exp Med. 169: 1795 - 1802.
- Barrett, J., Brophy, P.M., Hamilton, J.V. (2005). Analysing proteomic data. Int. J. Parasitol. 35: 543-553.
- Barrett, J., Jefferies, J.R., Brophy, P.M. (2000). Parasite proteomics. Parasitol. Today. 16: 400-403.
- Baum, J., Maier, A.G., Good, R.T., Simpson, K.M., Cowman, A.F. (2005). Invasion by *P. falciparum* merozoites suggests a hierarchy of molecular interactions. PLoS Pathog. 1: e37.
- Beaute-Lafitte, A., Altemayer-Caillard, V., Gonnet-Gonzales, F., Ramiaramanana, L., Chabaud, A.G., Landau, I. Int J Parasitol. 24: 981 – 986.
- Benovoy, D., Kwan, T., Majewski, J. (2008). Effect of polymorphisms within probe-target sequences on oligonucleotide microarray experiments. Nucleic Acids Res. 36: 4417 - 4413.
- Biron, D.G., Moura, H., Marche, L., Hughes, A.L., Thomas, F. (2005). Towards a new conceptual approach to 'parasitoproteomics'. Trends. Parasitol. 21: 162-168.
- Blackman, M.J. (2008). Malarial proteases and host cell egress: an 'emerging' cascade. Cell Microbiol. 10: 1925 - 1934.
- Blackman, M.J., Scott-Finnigan, T.J., Shai, S., Holder, A.A. (1994). Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. J Exp Med. 180: 389-393.
- Blackman, M.J., Whittle, H., Holder, A.A. (1991). Processing of the *Plasmodium falciparum* major merozoite surface protein-1: identification of a 33-kilodalton secondary processing product which is shed prior to erythrocyte invasion. Mol Biochem Parasitol. 49: 35-44.
- Blanc L, De Gassart A, Geminard C, Bette-Bobillo P, Vidal M. (2005). Exosome release by reticulocytes--an integral part of the red blood cell differentiation system. Blood Cells Mol Dis. 35: 21-26.
- Blythe, J.E., Suretheran, T., Preiser, P. (2003). STEVOR-a multidunctional protein? Mol Biochem Parasitol. 134: 11-15.
- Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., DeRisi, J.L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biol. 1: 85-100.

References

- Bozdech, Z., Mok, S., Hu, G., Imwong, M., Jaidee, A., Russell, B., Ginsburg, H., Nosten, F., Day, N.P., White, N.J., Carlton, J.M., Preiser, P.R. (2008). The transcriptome of *Plasmodium vivax* reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proc Natl Acad Sci U.S.A.* 105: 16290 – 16295.
- Braks, J.A., Mair, G.R., Franke-Fayard, B., Janse, C.J., Waters, A.P. (2008). A conserved U-rich RNA region implicated in regulation of translation in *Plasmodium* female gametes. *Nucleic Acids Res.* 36: 1176 – 1186.
- Browne, G.J., Proud, C.G. (2002). Regulation of peptide-chain elongation in mammalian cells. *Eur. J. Biochem.* 269: 5360-5368.
- Bunger, W. (1985). Virulence of *Plasmodium yoelii* line YM is caused by altered preference for parasitized and unparasitized immature and mature erythrocytes. *Trop Med Parasitol.* 36: 43-45.
- Carlberg, U., Nilsson, A., Nygard, O. (1990). Functional properties of phosphorylated elongation factor 2. *Eur. J. Biochem.* 191: 639-645.
- 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., Fraser-Liggett, C.M. (2008). Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455: 757-763.
- Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Perte, 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., Carucci, D.J. (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*. 419: 512-519.
- Carlton, J.M., Escalante, A.A., Neafsey, D., Volkman, S.K. (2008b). Comparative evolutionary genomics of human malaria parasites. *Trends Parasitol.* 24: 545-550.
- Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voûte, P.A., Heisterkamp, S., van Kampen, A., Versteeg, R. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*. 291: 1289 - 1292.
- Carr-Schmid, A., Valente, L., Loik, V.I., Williams, T., Starita, L.M., Kinzy, T.G. (1999). Mutations in elongation factor 1b, a guanine nucleotide exchange factor, enhance translational fidelity. *Mol. Cell. Biol.* 19: 5257-5266.

References

- Carsiotis, M., Stocker, B.A., Weinstein, D.L., O'Brien, A.D. (1989). A *Salmonella typhimurium* virulence gene linked to flg. Infect Immun. 57: 3276 - 3280.
- Carter, R., Walliker, D. (1975). New observations on the malaria parasites of rodents of the Central African Republic - *Plasmodium vinckei petteri* subsp. nov. and *Plasmodium chabaudi* Landau, 1965. Ann Trop Med Parasitol. 69:187-96.
- Carter, R. and Diggs, C.L. (1977). In: 'Parasitic Protozoa', vol.3 pp.359-465. Academic Press, New York.
- CDC Malaria at <http://www.cdc.gov/malaria/disease.htm>
- Chervitz, S.A., Aravind, L., Sherlock, G., Ball, C.A., Koonin, E.V., Dwight, S.S., Harris, M.A., Dolinski, K., Mohr, S., Smith, T., Weng, S., Cherry, J.M., Botstein, D. (1998). Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science. 282: 2022 – 2028.
- Charoenvit, Y., Leef, M.F., Yuan, L.F., Sedegah, M., Beaudoin, R.L. (1987). Characterization of *Plasmodium yoelii* monoclonal antibodies directed against stage-specific sporozoite antigens. Infect Immun. 55: 604 – 608.
- Chin, W., Contacos, P.G., Coatney, G.R., Kimball, H.R. (1965). A naturally acquired quotidian-type malaria in man transferable to monkeys. Science. 149: 865.
- Chokejindachai, W., Conway, D.J. (2009). Case-control approach to identify *Plasmodium falciparum* polymorphisms associated with severe malaria. PLoS ONE. 4: e5454.
- Cholera, R., Brittain, N.J., Gillrie, M.R., Lopera-Mesa, T.M., Diakité, S.A., Arie, T., Krause, M.A., Guindo, A., Tubman, A., Fujioka, H., Diallo, D.A., Doumbo, O.K., Ho, M., Wellems, T.E., Fairhurst, R.M. (2008). Impaired cytoadherence of *Plasmodium falciparum*-infected erythrocytes containing sickle hemoglobin. Proc Natl Acad Sci U.S.A. 105: 991-996.
- Cohen, B.A., Mitra, R.D., Hughes, J.D., Church, G.M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. Nat Genet. 26: 183 - 186.
- Coulson, R.M., Hall, N., Ouzounis, C.A. (2004). Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. Genome Res. 14: 1548-1554.
- Cowman, A.F., Crabb, B.S. (2006). Invasion of red blood cells by malarial parasites. Cell. 124: 755 - 766.
- Cristiano, F.A., Perez, M.A., Nicholls, R.S., Guerra, A.P. (2008). Polymorphism in the *Plasmodium vivax* msp 3: gene in field samples from Tierralta, Colombia. Mem Inst Oswaldo Cruz. 2008. 103:493 - 496.

References

- Cunningham, D.A., Jarra, W., Koernig, S., Fonager, J., Fernandez-Reyes, D., Blythe, J.E., Waller, C., Preiser, P.R., Langhorne, J. (2005). Host immunity modulates transcriptional changes in a multigene family (yir) of rodent malaria. *Mol Micro.* 58: 636-647.
- Curley, G.P., O'Donovan, S.M., McNally, J., Mullally, M., O'Hara, H., Troy, A., O'Callaghan, S.A., Dalton, J.P. (1994). Aminopeptidases from *Plasmodium falciparum*, *Plasmodium chabaudi chabaudi* and *Plasmodium berghei*. *J. Eukaryot. Microbiol.* 41: 119-123.
- Daly, T.M., Long, C.A. (1993). A recombinant 15-kilodalton carboxyl-terminal fragment of *Plasmodium yoelii yoelii* 17XL merozoite surface protein1 induces a protective immune response in mice. *Infect Immun.* 61: 2462 - 2467.
- Daran-Lapujade, P., Daran, J. M., Kotter, P., Petit, T., Piper, M. D. & Pronk, J. T. (2003). Comparative genotyping of the *Saccharomyces cerevisiae* laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays. *FEMS Yeast Res* 4: 259–269.
- de Koning-Ward, T.F., Janse, C.J., Waters, A.P. (2000). The development of genetic tools for dissecting the biology of malaria parasites. *Annu Rev Microbiol.* 54: 157 - 185.
- de Silva, E.K., Gehrke, A.R., Olszewski, K., Leon, I., Chahal, J.S., Bulyk, M.L., Llinas, M. (2008). Specific DNA-binding by Apicomplexan AP2 transcription factors. *Proc Natl Acad Sci USA.* 105: 8393 – 8398.
- Debrabant, A., Maes, P., Delplace, P., Dubremetz, J.F., Tartar, A., Camus, D. (1992). Intramolecular mapping of *Plasmodium falciparum* P126 proteolytic fragments by N-terminal amino acid sequencing. *Mol Biochem Parasitol.* 53: 89 - 95.
- del Portillo, H.A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C.P., Schneider, N.K., Villalobos, J.M., Rajandream, M.A., Harris, D., Pereira da Silva, L.H., Barrell, B., Lanzer, M. (2001). A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature.* 410: 839 - 842.
- Delahaye, N.F., Coltel, N., Puthier, D., Barbier, M., Benech, P., Joly, F., Iraqi, F.A., Grau, G.E., Nguyen, C., Rihet, P. (2007). Gene expression analysis reveals early changes in several molecular pathways in cerebral malaria-susceptible mice versus cerebral malaria-resistant mice. *BMC Genomics.* 8: 452.
- Delplace, P., Fortier, B., Tronchin, G., Dubremetz, J.F., Vernes, A. (1987). Localization, biosynthesis, processing and isolation of a major 126 kDa antigen of the parasitophorous vacuole of *Plasmodium falciparum*. *Mol Biochem Parasitol.* 23: 193 - 201.
- DeRisi, J.L., Iyer, V.R., Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science.* 278: 680-686.

References

- Despommier, D., Gwadz, R., Hotez, P., Knirsch, C. (2005). In: 'Parasitic Diseases', fifth edition. pp. 50 - 68. Apple Trees Productions, LLC.
- Dieye, A., Rogier, C., Trape, J.F., Sarthou, J.L., Druilhe, P. (1997). HLA class I-associated resistance to severe malaria: a parasitological re-assessment. *Parasitol Today*. 13: 48 – 49.
- Dluzewski, A.R., Ling, I.T., Hopkins, J.M., Grainger, M., Margos, G., Mitchell, G.H., Holder, A.A., Bannister, L.H. (2008). Formation of the food vacuole in *Plasmodium falciparum*: a potential role for the 19 kDa fragment of merozoite surface protein 1 (MSP1(19)). *PLoS ONE*. 3: e3085.
- Dong, Y., Glasner, J. D., Blattner, F. R. & Triplett, E. W. (2001). Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte *Klebsiella pneumoniae* 342, by microarray hybridization with *Escherichia coli* K-12 open reading frames. *Appl Environ Microbiol* 67: 1911–1921.
- Duraisingh, M.T., Triglia, T., Ralph, S.A., Rayner, J.C., Barnwell, J.W., McFadden, G.I., Cowman, A.F. (2003). Phenotypic variation of *Plasmodium falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO J*. 22: 1047 - 1057.
- Dvorak, J.A., Miller, L.H., Whitehouse, W.C., Shiroishi, T. (1975). Invasion of erythrocytes by malaria merozoites. *Science*. 187: 748-750.
- Ecker, A., Bushell, E.S., Tewari, R., Sinden, R.E. (2008). Reverse genetics screen identifies six proteins important for malaria development in the mosquito. *Mol Microbiol*. 70: 209 – 220.
- Escalante, A.A., Grebert, H.M., Chaiyaroj, S.C., Magris, M., Biswas, S., Nahlen, B.L., Lal, A.A. (2001). Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. X. Asembo Bay Cohort Project. *Mol Biochem Parasitol*. 113: 279 - 287.
- Fahey, J.R., Spitalny, G.L. (1984). Virulent and nonvirulent forms of *Plasmodium yoelii* are not restricted to growth within a single erythrocyte type. *Infect Immun*. 44: 151 - 156.
- Fahey, J.R., Spitalny, G.L. (1986). Rapid parasite multiplication rate, rather than immunosuppression, causes the death of mice infected with virulent *Plasmodium yoelii*. *Infect Immun*. 55: 490 - 493.
- Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., Witney, A.A., Wolters, D., Wu, Y., Gardner, M.J., Holder, A.A., Sinden, R.E., Yates, J.R., Carucci, D.J. (2002). A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*. 419: 520-526.

References

- Fonager, J., Cunningham, D., Jarra, W., Koernig, S., Henneman, A.A., Langhorne, J., Preiser, P. (2007). Transcription and alternative splicing in the yir multigene family of the malaria parasite *Plasmodium y. yoelii*: Identification of motifs suggesting epigenetic and post-transcriptional control of RNA expression. *Mol Biochem Parasitol.* 156: 1-11.
- Fox, B.A., Xing-Li, P., Suzue, K., Horii, T., Bzik, D.J. (1997). *Plasmodium falciparum*: an epitope within a highly conserved region of the 47-kDa amino-terminal domain of the serine repeat antigen is a target of parasite-inhibitory antibodies. *Exp Parasitol.* 85: 121 - 134.
- Franke-Fayard, B., Trueman, H., Ramesar, J., Mendoza, J., van der Keur, M., van der Linden, R., Sinden, R.E., Waters, A.P., Janse, C.J. (2004). A *Plasmodium berghei* reference line that constitutively expresses GFP at a high level throughout the complete life cycle. *Mol. Biochem. Parasitol.* 137: 23-33.
- Fraser, T.S., Kappe, S.H., Narum, D.L., VanBuskirk, K.M., Adams, J.H. (2001). Erythrocyte-binding activity of *Plasmodium yoelii* apical membrane antigen-1 expressed on the surface of transfected COS-7 cells. *Mol Biochem Parasitol.* 117: 49-59.
- Freeman, R.R., Trejdosiewicz, A.J., Cross, G.A. (1980). Protective monoclonal antibodies recognising stage-specific merozoite antigens of a rodent malaria parasite. *Nature* 284: 366-368.
- Galinski, M.R. and Barnwell, J.W. (1996). *Plasmodium vivax*: merozoites, invasion of reticulocytes and considerations for malaria vaccine development. *Parasitol. Today.* 12: 20-29.
- Gardiner, D.L., McCarthy, J.S., Trenholme, K.R. (2004). Malaria in the post-genomics era: light at the end of the tunnel or just another train? *Postgrad. Med. J.* 81: 505-509.
- 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., Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature.* 419: 498-511.
- Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Smith, H.O., Fraser, C.M., Venter, J.C., Hoffman, S.L. (1999). The malaria genome sequencing project: complete sequence of *Plasmodium falciparum* chromosome 2. *Parassitologia.* 41: 69 - 75.

References

- Gaur, D., Mayer, D.C., Miller, L.H. (2004). Parasite ligand-host receptor interactions during invasion of erythrocytes by *Plasmodium* merozoites. *Int. J. Parasitol.* 34: 1413-1429.
- Gaur, D., Furuya, T., Mu, J., Jiang, L.B., Su, X.Z., Miller, L.H. (2006). Upregulation of expression of the reticulocyte homnology gene 4 in *Plasmodium falciparum* clone Dd2 is associated with a switch in erythrocyte invasion pathway. *Mol Biochem Parasitol.* 145: 205 - 215.
- Gautret, P., Deharo, E., Tahar, R., Chabaud, A.G., Landau, I. (1995). The adjustment of the schizogonic cycle of *P. chabaudi chabaudi* in the blood to the circadian rhythm of the host. *Parasite.* 2: 69 - 74.
- Gavigan, C.S., Dalton, J.P., Bell, A. (2001). The role of aminopeptidases in haemoglobin degradation in *Plasmodium falciparum*-infected erythrocytes. *Mol. Biochem. Parasitol.* 117: 37-48.
- Geminard, C., de Gassart, A., Vidal, M. (2002). Reticulocyte maturation: mitoptosis and exosome release. *Biocell.* 26: 205-215.
- Gerold, P., Schofield, L., Blackman, M.J., Holder, A.A., Schwarz, R.T. (1996). Structural analysis of the glycosyl-phosphatidylinositol membrane anchor of the merozoite surface proteins-1 and -2 of *Plasmodium falciparum*. *Mol Biochem Parasitol.* 75: 131-143.
- Gilson, P.R., O'Donnell, R.A., Nebl, T., Sanders, P.R., Wickham, M.E., McElwain, T.F., de Koning-Ward, T.F., Crabb, B.S. (2008). MSP1(19) miniproteins can serve as targets for invasion inhibitory antibodies in *Plasmodium falciparum* provided they contain the correct domains for cell surface trafficking. *Mol Microbiol.* 68: 124-138.
- Good, M.F. (2001). Towards a blood-stage vaccine for malaria: Are we following all the leads? *Nat. Rev. Immunol.* 1: 117-125.
- Gunter, T.E., Gunter, K.K., Sheu, S.S., Gavin, C.E. (1994). Mitochondrial calcium transport: physiological and pathological relevance. *Am. J. Physiol.* 267: C313-339.
- Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., Florens, L., Janssen, C.S., Pain, A., Christophides, G.K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M.A., Ormond, D., Doggett, J., Trueman, H.E., Mendoza, J., Bidwell, S.L., Rajandream, M.A., Carucci, D.J., Yates, J.R. 3rd, Kafatos, F.C., Janse, C.J., Barrell, B., Turner, C.M., Waters, A.P., Sinden, R.E. (2005). A comprehensive study of the *Plasmodium* life cycle by genomic, transcriptome, and proteomic analyses. *Science.* 307: 82-86.
- Harris, P.K., Yeoh, S., Dluzewski, A.R., O'Donnell, R.A., Withers-Martinez, C., Hackett, F., Bannister, L.H., Mitchell, G.H., Blackman, M.J. (2005). Molecular identification of a malaria merozoite surface sheddase. *PloS Pathig.* 1: 241 - 251.

References

- Hernandez-Rivas, R., Mattei, D., Sterkers, Y., Peterson, D.S., Wellems, T.E., Scherf, A. (1997). Expressed var genes are found in *Plasmodium falciparum* subtelometic regions. *Mol Cell Biol.* 17: 604 - 611.
- Hill, A.V. (1992). Malaria resistance genes: a natural selection. *Trans R Soc Trop Med Hyg.* 86: 225-232.
- Hill, A.V., Elvin, J., Willis, A.C., Aidoo, M., Allsopp, C.E., Gotch, F.M., Gao, X.M., Takiguchi, M., Greenwood, B.M., Townsend, A.R., McMichael, A.J., Whittle, H.C. (1992). Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature.* 360: 434-439.
- Hobbs, M.R., Udhayakumar, V., Levesque, M.C., Booth, J., Roberts, J.M., Tkachuk, A.N., Pole, A., Coon, H., Kariuki, S., Nahlen, B.L., Mwaikambo, E.D., Lal, A.L., Granger, D.L., Anstey, N.M., Weinberg, J.B. (2002). A new NOS2 promoter polymorphism associated with increased nitric oxide production and protection from severe malaria in Tanzanian and Kenyan children. *Lancet.* 360: 1468-1475.
- Hodder, A.N., Crewther, P.E., Anders, R.F. (2001). Specificity of the protective antibody response to apical membrane antigen 1. *Infect Immun.* 69: 3286-3294.
- Hoffman, S.L., Bancroft, W.H., Gottlieb, M., James, S.L., Burroughs, E.C., Stephenson, J.R., Morgan, M.J. (1997). Funding for malaria genome sequencing. *Nature.* 387: 647.
- Holder, A.A., Freeman, R.R. (1981). Immunization against blood-stage rodent malaria using purified parasite antigens. *Nature.* 294: 361 - 364.
- Holder, A.A., Freeman, R.R. (1984). Protective antigens of rodent and human blood stage malaria. *Philos Trans R Soc Lond B Biol Sci.* 171 - 177.
- Holder, A.A., Sandhu, J.S., Hillman, Y., Davey, L.S., Nicholls, S.C., Cooper, H., Lockyer, M.J. (1987). Processing of the precursor to the major merozoite surface antigens of *Plasmodium falciparum*. *Parasitology.* 94: 199-208.
- Horrocks, P., Dechering, K., Lanzer, M. (1998). Control of gene expression in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 95: 171-181.
- Hu, G., Llinás, M., Li, J., Preiser, P.R., Bozdech, Z. (2007). Selection of long oligonucleotides for gene expression microarrays using weighted rank-sum strategy. *BMC Bioinformatics.* 8: 350.
- Hubbard, M.J., McHugh, N.J. (1996). Mitochondrial ATP synthase F1-beta-subunit is a calcium-binding protein. *FEBS Lett.* 391: 323-329.

References

- Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephanian, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., Linsley, P.S. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol.* 19: 342-347.
- Iyer, J.K., Amaladoss, A., Ganesan, S., Preiser, P.R. (2007). Variable expression of the 235 kDa rhoptry protein of *Plasmodium yoelii* mediate host cell adaptation and immune evasion. *Mol Microbiol.* 65: 333 - 346.
- Iyer, J.K., Gruner, A.C., Renia, L., Snounou, G., Preiser, P.R. (2007a). Invasion of host cells by malaria parasites: a tale of two protein families. *Mol Micro.* 65: 231 - 249.
- Janssen, C.S., Barrett, M.P., Turner, C.M., Phillips, R.S. (2002). A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proc Biol Sci.* 269: 431-436.
- Janssen, C.S., Phillips, R.S., Turner, C.M., Barrett, M.P. (2004). *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Res.* 32: 5712-5720.
- Jessberger, R., Riwar, B., Baehtold, H., Akhmedov, A. (1996). SMC proteins constitute two subunits of the mammalian recombination complex RC-1. *EMBO J.* 15: 4061 – 4068.
- Jewett, T.J., Sibley, L.D. (2003). Aldolase forms a bridge between cell surface adhesions and the actin cytoskeleton in apicomplexan parasites. *Mol. Cell.* 11: 885-894.
- Jiang, H., Yi, M., Mu, J., Zhang, L., Ivens, A., Klimczak, L.J., Huyen, Y., Stephens, R.M., Su, X.Z. (2008). Detection of genome-wide polymorphisms in the AT-rich *Plasmodium falciparum* genome using a high-density microarray. *BMC Genomics.* 9: 398
- Kaestli, M., Cockburn, I.A., Cortés, A., Baea, K., Rowe, J.A., Beck, H.P. (2006). Virulence of malaria is associated with differential expression of *Plasmodium falciparum* var gene subgroups in a case-control study. *J Infect Dis.* 193: 1567 - 1574.
- Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., Madore, S.J. (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res.* 28:4552-4557.
- Kappe, S.H., Adams, J.H. (1996). Sequence analysis of the apical membrane antigen-1 genes (*ama-1*) of *Plasmodium yoelii yoelii* and *Plasmodium berghei*. *Mol Biochem Parasitol.* 78: 279 – 283.

References

- Khan, S.M., Franke-Fayard, B., Mair, G.R., Lasonder, E., Janse, C.J., Mann, M., Waters, A.P. (2005). Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell*. 121: 675-687.
- Khan, S.M., Jarra, W., Preiser, P.R. (2001). The 235 kDa rhoptry protein of *Plasmodium (yoelii) yoelii*: function at the junction. *Mol Biochem Parasitol*. 117: 1 - 10.
- Killick-Kendrick, R., (1974). Parasitic protozoa of the blood of rodents: a revision of *Plasmodium berghei*. *Parasitol*. 69: 225 – 237.
- Killick-Kendrick, R. (1978). In: ‘Rodent malaria’ pp. 1 – 52. Academic Press, London.
- Knapp, B., Hundt, E., Nau, U., Kupper, H.A. (1989). Molecular cloning, genomic structure and localization in blood stage antigen of *Plasmodium falciparum* characterized by a serine stretch. *Mol Biochem Parasitol*. 32: 73 – 83.
- Knapp, B., Nau, U., Hundt, E., Kupper, H.A. (1991). A new blood stage antigen of *Plasmodium falciparum* highly homologous to the serine-stretch protein SERP. *Mol Biochem Parasitol*. 44: 1 – 13.
- Knowles, G., Walliker, D. (1980). Variable expression of virulence in the rodent malaria parasite *Plasmodium yoelii yoelii*. *Parasitol*. 81: 211-219.
- Knowles, R., Das Gupta, B.M. (1932). A study of monkey-malaria and its experimental transmission to man. *Ind Med Gaz*. 67: 301–320.
- Kooij, T.W., Carlton, J.M., Bidwell, S.L., Hall, N., Ramesar, J., Janse, C.J., Waters, A.P. (2005). A *Plasmodium* whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. *PloS Pathogen*. 1:e44.
- Kooij, T.W., Janse, C.J., Waters, A.P. (2006). *Plasmodium* post-genomics: better the bug you know? *Nat Rev Microbiol*. 4: 344-357.
- Koury, M.J., Koury, S.T., Kopsombut, P., Bondurant, M.C. (2005). In vitro maturation of nascent reticulocytes to erythrocytes. *Blood*. 105: 2168-2174.
- Kun, J.F., Mordmuller, B., Perkins, D.J., May, J., Mercereau-Puijalon, O., Alpers, M., Weinberg, J.B., Kremsner, P.G. (2001). Nitric oxide synthase 2 Lambarene (G-954C), increased nitric oxide production, and protection against malaria. *J Infect Dis*. 184: 330-336.
- Kyes, S., Horrocks, P., Newbold, C. (2001). Antigenic variation at the infected red cell surface in malaria. *Annu Rev Microbiol*. 55: 673-707.
- Landau, I., (1965). Description de *Plasmodium chabaudi* n. sp. Parasite de rongeurs africains. *C R Acad Sci Paris*. 260: 3758 – 3761.

References

- Landau, I. and Boulard, Y. (1978). In: Rodent Malaria. Life cycles and morphology. Pp 53-84. Academic Press, New York.
- Landau, I., Chabaud, A.G. (1965). [Natural infection by 2 plasmodia of the rodent *Thamnomys rutilans* in the Central African Republic.] C R Acad Sci Hebd Seances Acad Sci D. 261: 230 - 232.
- Landau, I., Gautret, P. (1998). In: 'Malaria: Parasite biology, pathogenesis, and protection' pp. 401 – 417. ASM Press, Washington D.C.
- Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., Mann, M. (2002). Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature. 419: 537-542.
- Le Roch, K.G., Johnson, J.R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S.F., Williamson, K.C., Holder, A.A., Carucci, D.J., Yates, J.R. 3rd, Winzeler, E.A. (2004). Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. Genome Res. 14: 2308-2318.
- Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., Winzeler, E.A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. Science. 301: 1503-1508.
- Lee, J.C., Gimm, J.A., Lo, A.J., Koury, M.J., Krauss, S.W., Mohandas, N., Chasis, J.A. (2004). Mechanism of protein sorting during erythroblast enucleation: role of cytoskeletal connectivity. Blood. 103: 1912-1919.
- Lehmann, A.R., Walicka, M., Griffiths, D.J., Murray, J.M., Watts, F.Z., McCready, S., Carr, A.M. (1995). The *rad18* gene of *Schizosaccharomyces pombe* defines a new subgroup of the SMC superfamily involved in DNA repair. Mol Cell Biol. 15: 7067 – 7080.
- Lercher, M.J., Urrutia, A.O., Hurst, L.D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet. 31: 180 - 183.
- Lew, V.L., Macdonald, L., Ginsburg, H., Krugliak, M., Tiffert, T. (2004). Excess haemoglobin digestion by malaria parasites: a strategy to prevent premature host cell lysis. Blood Cells Mol. Dis. 32: 353-359.
- Leyhausen, G., Schuster, D.K., Vaith, P., Zahn, R.K., Umezawa, H., Falke, D., Muller, W.E. (1983). Identification and properties of the cell membrane bound leucine aminopeptidase interacting with the potential immunostimulant and chemotherapeutic agent bestatin. Biochem. Pharmacol. 32: 1051-1057.
- Leyva, J.A., Bianchet, M.A., Amzel, L.M. (2003). Understanding ATP synthesis: structure and mechanism of the F1-ATPase (Review). Mol. Membr. Biol. 20: 27-33.

References

- Li, J., Gutell, R.R., Damberger, S.H., Wirtz, R.A., Kissinger, J.C., Rogers, M.J., Sattabongkot, J., McCutchan, T.F. (1997). Regulation and trafficking of three distinct 18 S ribosomal RNAs during development of the malaria parasite. *J. Mol. Biol.* 269: 203-213.
- Li, J., Matsuoka, H., Mitamura, T., Horii, T., (2002a). Characterization of proteases involved in the processing of *Plasmodium falciparum* serine repeat antigen (SERA). *Mol Biochem Parasitol.* 120: 177 - 186.
- Li, J., Mitamura, T., Fox, B.A., Bzik, D.J., Horii, T., (2002b). Differential localization of processed fragments of *Plasmodium falciparum* serine repeat antigen and further processing of its N-terminal 47 kDa fragment. *Parasitol Int.* 51: 343 - 352.
- Lilley, K.S., Razzaq, A., Dupree, P. (2001). Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr. Opin. Chem. Biol.* 6: 46-50.
- Lowther, W.T., Matthews, B.W. (2002). Metalloaminopeptidases: common functional themes in disparate structural surroundings. *Chem. Rev.* 102: 4581-4608.
- Mair, G.R., Braks, J.A., Garver, L.S., Wiegant, J.C., Hall, N., Dirks, R.W., Khan, S.M., Dimopoulos, G., Janse, C.J., Waters, A.P. (2006). Translational repression is essential for *Plasmodium* sexual development and mediated by a DDX6-type RNA helicase. *Science.* 313: 667 – 669.
- Mamillapalli, A., Sunil, S., Diwan, S.S., Sharma, S.K., Tyagi, P.K., Adak, T., Joshi, H., Malhotra, P. (2007). Polymorphism and epitope sharing between the alleles of merozoite surface protein-1 of *Plasmodium falciparum* among Indian isolates. *Malar J.* 6: 95.
- Mayer, D.C., Mu, J.B., Feng, X., Su, X.Z., Miller, L.H. (2002). Polymorphism in a *Plasmodium falciparum* erythrocyte-binding ligand changes its receptor specificity. *J. Exp Med.* 196: 1532 - 1538.
- Mayer, D.C., Mu, J.B., Kaneko, O., Duan, J., Su, X.Z., Miller, L.H. (2004). Polymorphism in the *Plasmodium falciparum* erythrocyte-binding ligand JESEBL/EBA-181 alters its receptor specificity. *Proc Natl Acad Sci USA.* 101: 2518 – 2523.
- McCoubrie, J.E., Miller, S.K., Sargeant, T., Good, R.T., Hodder, A.N., Speed, T.P., de Koning-Ward, T.F., Crabb, B.S. (2007). Evidence for a common role for the serine-type *Plasmodium falciparum* serine repeat antigen proteases: Implications for vaccine and drug design. *Infect Immun.* 75: 5565 - 5574.
- McCutchan, T.F., Li, J., McConkey, G.A., Rogers, M.J., Waters, A.P. (1995). The cytoplasmic ribosomal RNAs of *Plasmodium spp.* *Parasitol Today.* 11: 134-138.
- Mideo, N., Barclay, V.C., Chan, B.H., Savill, N.J., Read, A.F., Day, T. (2008). Understanding and predicting strain-specific patterns of pathogenesis in the rodent malaria *Plasmodium chabaudi*. *Am Nat.* 172: 214 - 238.

References

- Miklos, G.L., Maleszka, R. (2001). Protein functions and biological contexts. *Proteomics*. 1: 169-178.
- Militello, K.T., Dodge, M., Bethke, L., Wirth, D.F. (2004). Identification of regulatory elements in the *Plasmodium falciparum* genome. *Mol. Biochem. Parasitol.* 134: 75-88.
- Miller, L.H., Aikawa, M., Johnson, J.G., Shiroishi, T. (1979). Interaction between cytochalasin B-treated malarial parasites and erythrocytes. Attachment and junction formation. *J Exp Med.* 149: 172 - 184.
- Miller, L.H., Mason, S.J., Clyde, D.F., McGinniss, M.H. (1976). The resistance factor to *Plasmodium vivax* infection in blacks. The Duffy-blood-group genotype, FyFy. *N. Engl J Med.* 295: 302 - 304.
- Miller, L.H., McGinniss, M.H., Holland, P.V., Sigmon, P. (1978). The Duffy blood group phenotype in American blacks infected with *Plasmodium vivax* in Vietnam. *Am J Trop Med Hyg.* 27: 1069 - 1072.
- Miller, S.K., Good, R.T., Drew, D.R., Delorenzi, M., Sanders, P.R., Hodder, A.N., Speed, T.P., Cowman, A.F., de Koning-Ward, T.F., Crabb, B.S. (2002). A subset of *Plasmodium falciparum* SERA genes are expressed and appear to play an important role in the erythrocytic cycle. *J. Biol. Chem.* 277: 47524-47532.
- Mitchell, G.H., Thomas, A.W., Margos, G., Dluzewski, A.R., Bannister, L.H. (2004). Apical membrane antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites to host red blood cells. *Infect Immun.* 72: 154-158.
- Narum, D.L., Green, J.L., Ogun, S.A., Holder, A.A. (2001). Sequence diversity and antigenic polymorphism in the *Plasmodium yoelii* p235 high molecular mass rhoptry proteins and their genes. *Mol Biochem Parasitol.* 112: 193 - 200.
- Narum, D.L., Thomas, A.W. (1994). Differential localization of full-length and processed forms of PF83/AMA-1 an apical membrane antigen of *Plasmodium falciparum* merozoites. *Mol Biochem Parasitol.* 67: 59-68.
- Nirmalan, N., Sims, P.F., Hyde, J.E. (2004). Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition. *Mol. Microbiol.* 52: 1187-1199.
- NobelPrize.org at http://nobelprize.org/nobel_prizes/medicine/laureates/1907/laveran-lecture.html and <http://nobelprize.org/medicine/laureates/1902/ross-bio.html>
- O'Donnell, R.A., Blackman, M.J. (2005). The role of malaria merozoite proteases in red blood cell invasion. *Curr Opin Microbiol.* 8: 422 - 427.
- O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250: 4407-4421.

References

- Oakley, M.S., McCutchan, T.F., Anantharaman, V., Ward, J.M., Faucette, L., Erexson, C., Mahajan, B., Zheng, H., Majam, V., Aravind, L., Kumar, S. (2008). Host biomarkers and biological pathways that are associated with the expression of experimental cerebral malaria in mice. *Infect Immun.* 76: 4518 - 4529.
- Ogun, S.A., Holder, A.A. (1996). A high molecular mass *Plasmodium yoelii* rhoptry protein binds to erythrocytes. *Mol Biochem Parasitol.* 76: 321-324.
- Ogun, S.A., Scott-Finnigan, T.J., Narum, D.L., Holder, A.A. (2000). *Plasmodium yoelii*: effects of red blood cell modification and antibodies on the binding characteristics of the 235-kDa rhoptry protein. *Expt Parasitol.* 95: 187-195.
- Oka, A., Aikawa, M., Freeman, R.R., Holder, A.A., Fine, E. (1984). Ultrastructural localization of protective antigens of *Plasmodium yoelii* merozoites by the use of monoclonal antibodies and ultrathin cryomicrotomy. *Am J Trop Med Hyg.* 33: 342-346.
- Okech B.A., Nalunkuma, A., Okello, D., Pang, X.L., Suzue, K., Li, J., Horii, T., Egwang, T.G. (2001). Natural human immunoglobulin G subclass responses to *Plasmodium falciparum* serine repeat antigen in Uganda. *Am J Trop Med Hyg.* 65: 912 – 917.
- Okech, B., Mujuzi, G., Ogawal, A., Shirai, H., Horii, T., Egwang, T.G. (2006). High titres of IgG antibodies against *Plasmodium falciparum* serine repeat antigen 5 (SERA5) are associated with protection against severe malaria in Ugandan children. *Am J Trop Med Hyg.* 74: 191 - 197.
- Okoyeh, J.N., Pillai, C.R., Chitnis, C.E. (1999). *Plasmodium falciparum* field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of glycophorin A. *Infect Immun.* 67: 5784 - 5791.
- Omi, K., Ohashi, J., Patarapotikul, J., Hananantachai, H., Naka, I., Looareesuwan, S., Tokunaga, K. (2003). CD36 polymorphism is associated with protection from cerebral malaria. *Am J Hum Genet.* 72: 364-374.
- Ong, C., Ooi, C.H., Wang, D., Chong, H., Ng, K.C., Rodrigues, F., Lee, M.A., Tan, P. (2004). Patterns of large-scale genomic variation in virulent and avirulent *Burkholderia* species. *Genome Res.* 14: 2295-2307.
- Pain, A., Urban, B.C., Kai, O., Casals-Pascual, C., Shafi, J., Marsh, K., Roberts, D.J. (2001). A non-sense mutation in Cd36 gene is protection from severe malaria. *Lancet.* 357: 1502-1503.
- Pang, X.L., Mitamura, T., Horii, T. (1999). Antibodies reactive with the N-terminal domain of *Plasmodium falciparum* serine repeat antigen inhibit cell proliferation by agglutinating merozoites and schizonts. *Infect Immun.* 67: 1821 - 1827.
- Pannucci J, Cai H, Pardington PE, Williams E, Okinaka RT, Kuske CR, Cary RB. (2004). Virulence signature: microarray-based approaches to discovery and analysis. *Biosens Bioelectron.* 20: 706 - 718.

References

- Pattaradilokrat, S., Cheesman, S.J., Carter, R. (2008). Congenicity and genetic polymorphism in cloned lines derived from a single isolate of a rodent malaria parasite. *Mol Biochem Parasitol.* 157: 244 – 247.
- Pedersen, P.L., Amzel, L.M. (1993). ATP synthases. Structure, reaction centre, mechanism, and regulation of one of nature's most unique machines. *J.Biol. Chem.* 268: 9937-9940.
- Perkins, S.L., Sarkar, I.N., Carter, R. (2007). The phylogeny of rodent malaria parasites: Simultaneous analysis across three genomes. *Infect. Genet. Evol.* 7: 74-83.
- Pinder, J., Fowler, R., Bannister, L., Dluzewski, A., Mitchell, G.H. (2000). Motile systems in malaria merozoites: how is the red blood cell invaded? *Parasitol. Today.* 16: 240-245.
- Pinder, J.C., Fowler, R.E., Dluzewski, A.R., Bannister, L.H., Lavin, F.M., Mitchell, G.H., Wilson, R.J., Gratzer, W.B. (1998). Actomyosin motor in the merozoite of the malaria parasite, *Plasmodium falciparum*: implications for red cell invasion. *J. Cell Sci.* 111: 1831-1839.
- Preiser P, Kaviratne M, Khan S, Bannister L, Jarra W. (2000). The apical organelles of malaria merozoites: host cell selection, invasion, host immunity and immune evasion. *Microbes. Infect.* 2: 1461-1477.
- Preiser, P.R., Jarra, W., Capiod, T., Snounou, G. (1999). A rhoptry-protein-associated mechanism of clonal phenotypic variation in rodent malaria. *Nature.* 398: 618-622.
- Rayner, J.C., Galinski, M.R., Ingravallo, P., Barnwell, J.W. (2000). Two *Plasmodium falciparum* genes express merozoite proteins that are related to *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins involved in host cell selection and invasion. *Proc Natl Acad Sci USA.* 97: 9648 - 9653.
- Rayner, J.C., Vargas-Serrato, E., Huber, C.S., Galinski, M.R., Barnwell, J.W. (2001). A *Plasmodium falciparum* homolog of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway. *J Exp Med.* 194: 1571 - 1582.
- Ryazanov, A., Rudkin, B.B., Spirin, A.S. (1991). Regulation of protein synthesis at the elongation stage. New insights into the control of gene expression in eukaryotes. *FEBS Lett.* 285: 170-175.
- Saitoh, N., Goldberg, I., Earnshaw, W.C. (1995). The SMC proteins and the coming of age of the chromosome scaffold hypothesis. *Bioessays.* 17: 759 – 766.
- Sanni, L.A., Fonseca, L.F., Langhorne, J. (2002). In: *Malaria Methods and Protocols. Mouse models for erythrocyte-stage malaria.* Pp. 57-76. Humana Press.

References

- Sarfo, B.Y., Armah, H.B., Irune, I., Adjei, A.A., Olver, C.S., Singh, S., Lillard, J.W. Jr., Stiles, J.K. (2005). *Plasmodium yoelii* 17XL infection up-regulates RANTES, CCR1, CCR3 and CCR5 expression, and induces ultrastructural changes in the cerebellum. *Malar J.* 4: 63.
- Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., Cowman, A.F. (2006). Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* 7: R12.
- Saul, A., Prescott, N., Smith, F., Cheng, Q., Walliker, D. (1997). Evidence of cross-contamination among laboratory lines of *Plasmodium berghei*. *Mol Biochem Parasitol.* 84: 143 – 147.
- Schmidt-Christensen, A., Sturm, A., Horstmann, S., Heussler, V.T. (2008). Expression and processing of *Plasmodium berghei* SERA3 during liver stages. *Cell Microbiol.* 10: 1723 - 1734.
- Schmitz, S., Grainger, M., Howell, S., Calder, L.J., Gaeb, M., Pinder, J.C., Holder, A.A., Veigel, C. (2005). Malaria parasite actin filaments are very short. *J. Mol. Biol.* 349: 113-125.
- Shortt, H.E., Fairley, N.H., Covell, G., Shute, P.G., Garnham, P.C.C. (1951). The pre-erythrocytic stage of *Plasmodium falciparum*. *Trans. R. Soc. Trop. Med. Hyg.* 44: 405-419.
- Shortt, H.E., Garnham, P.C.C. (1948). Demonstration of a persisting exo-erythrocytic cycle in *Plasmodium cynomolgi* and its bearing on the production of relapses. *BMJ*: 1225-1228.
- Simpson, J.A., Silamut, K., Chotivanich, K., Pukrittayakamee, S., White, N.J. (1999). Red cell selectivity in malaria: a study of multiple-infected erythrocytes. *Trans R Soc Trop Med Hyg.* 93: 165 - 168.
- Singh, B., Sung, L.K., Matusop, A., Radhakrishnan, A., Shamsul, S.S., Cox-Singh, J., Thomas, A., Conway, D.J. (2004). A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet.* 363: 1017-1024.
- Singh, S., Kennedy, M.C., Long, C.A., Saul, A.J., Miller, L.H., Stowers, A.W. (2003). Biochemical and immunological characterization of bacterially expressed and refolded *Plasmodium falciparum* 42-kilodalton C-terminal merozoite surface protein 1. *Infect Immun.* 71: 6766 - 6774.
- Sinha, K.A., Keen, J.K., Ogun, S.A., Holder, A.A. (1996). Comparison of two members of a multigene family coding for high-molecular mass rhoptry proteins of *Plasmodium yoelii*. *Mol Biochem Parasitol.* 76: 329-332.
- Sitkiewicz, I., Musser, J.M. (2006). Expression microarray and mouse virulence analysis of four conserved two-component gene regulatory systems in Group A *Streptococcus*. *Infect Immun.* 74: 1339 - 1351.

References

- Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y., Hay, S.I. (2005). The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*. 434: 214-217.
- Stephens, R., Ndungu, F.M., Langhorne, J. (2009). Germinal centre and marginal zone B cells expand quickly in a second *Plasmodium chabaudi* malaria infection producing mature plasma cells. *Parasite Immunol.* 31: 20 - 31.
- Stubbs, J., Simpson, K.M., Triglia, T., Plouffe, D., Tonkin, C.J., Duraisingh, M.T., Maier, A.G., Winzeler, E.A., Cowman, A.F. (2005). Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. *Science*. 309: 1384 - 1387.
- Sueblinvong, T., Chutmongkonkul, M., Seesod, N., Thaithong, S. (1996). The variation of Pfl55/RESA gene in field isolates of *Plasmodium falciparum* from Thailand. *J Med Assoc Thai*. 79: S40 - 48.
- Sugiyama, T., Suzue, K., Okamoto, M., Inselburg, J., Tai, K., Horii, T. (1996). Production of recombinant SERA proteins of *Plasmodium falciparum* in *Escherichia coli* by using synthetic genes. *Vaccine*. 14: 1069 - 1076.
- Swardson-Olver, C.J., Dawson, T.C., Burnett, R.C., Peiper, S.C., Maeda, N., Avery, A.C. (2002). *Plasmodium yoelii* uses the murine Duffy antigen receptor for chemokines as a receptor for normocyte invasion and an alternative receptor for reticulocyte invasion. *Blood*. 99: 2677 - 2684.
- Tamasauskas, D., Powell, V., Saksela, K., Yazdanbakhsh, K. (2001). A homologous naturally occurring mutation in Duffy and CCR5 leading to reduced receptor expression. *Blood*. 97: 3651-3654.
- Taylor, A. (1993). Aminopeptidases: towards a mechanism of action. *Trends. Biochem. Sci.* 18: 167-172.
- Taylor, D.W., Levander, O.A., Krishna, V.R., Evans, C.B., Morris, V.C., Barta, J.R. (1997). Vitamin E-deficient diets enriched with fish oil suppress lethal *Plasmodium yoelii* infections in athymic and *scid/bg* mice. *Infect Immun*. 65: 197 – 202.
- Taylor, H.M., Grainger, M., Holder, A.A. (2002). Variation in the expression of a *Plasmodium falciparum* protein family implicated in erythrocyte invasion. *Infect Immun*. 70: 5779 - 5789.
- Topley, E., Bruce-Chwatt, L.J., Dorrell, J. (1970) Haematological study of a rodent malaria model. *J Trop Med Hyg*. 73: 1 - 8.
- Triglia, T., Duraisingh, M.T., Good, R.T., Cowman, A.F. (2005). Reticulocyte-binding protein homolog 1 is required for sialic acid-dependent invasion into human erythrocytes by *Plasmodium falciparum*. *Mol Microbiol*. 55: 162 - 174.

References

- Triglia, T., Healer, J., Caruana, S.R., Hodder, A.N., Anders, R.F., Crabb, B.S., Cowman, A.F. (2000). Apical membrane antigen 1 plays a central role in erythrocyte invasion by *Plasmodium* species. *Mol Microbiol.* 38: 706-718.
- Triglia, T., Thompson, J., Caruana, S.R., Delorenzi, M., Speed, T., Cowman, A.F. (2001). Identification of proteins from *Plasmodium falciparum* that are homologous to reticulocyte binding proteins in *Plasmodium vivax*. *Infect Immun.* 69: 1084 - 1092.
- Turowski, D., Wysocka, J., Butkiewicz, A.M. (1996). Peripheral blood reticulocytes and their reference range values for percentage, absolute count, and immature fraction in children, measured with flow cytometry. *Folia Histochem Cytobiol.* 38: 31-36.
- Uyemura, S.A., Luo, S., Moreno, S.N., Docampo, R. (2000). Oxidative phosphorylation, Ca(2+) transport, and fatty acid-induced uncoupling in malaria parasites mitochondria. *J. Biol. Chem.* 275: 9709-9715.
- van Dijk, M.R., Waters, A.P., Janse, C.J. (1995). Stable transfection of malaria parasite blood stages. *Science.* 268: 1358-1362.
- VanBuskirk, K.M., Sevova, E., Adams, J.H. (2004). Conserved residues in the *Plasmodium vivax* Duffy-binding ligand domain are critical for erythrocyte receptor recognition. *Proc. Natl. Acad. Sci. U.S.A.* 101: 15754-15759.
- Vaughan A, Chiu SY, Ramasamy G, Li L, Gardner MJ, Tarun AS, Kappe SH, Peng X. (2008). Assessment and improvement of the *Plasmodium yoelii yoelii* genome annotation through comparative analysis. *Bioinformatics* 24:i383-i389.
- Vincke, I.H., Lips, M. (1948). Un nouveau *Plasmodium* d'un rongeur sauvage du Congo, *Plasmodium berghei* n. sp. *Ann Soc belge Med Trop.* 28: 197 – 204.
- Waller, R.F., Keeling, P.J., Donald, R.G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A.F., Besra, G.S., Roos, D.S., McFadden, G.I. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc Natl Acad Sci U.S.A.* 95: 12352-12357.
- Walliker, D., Sanderson, A., Yoelii, M., Hargreaves, B.J. (1976). A genetic investigation of virulence in a rodent malaria parasite. *Parasitology.* 72: 183-194.
- Waters, A.P. (2002). Orthology between the genomes of *Plasmodium falciparum* and rodent malaria parasites: possible practical applications. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 357: 55-63.
- Waters, A.P., Syin, C., McCutchan, T.F. (1989). Developmental regulation of stage-specific ribosome populations in *Plasmodium*. *Nature.* 342: 438-440.
- Waters, A.P., Thomas, A.W., Deans, J.A., Mitchell, G.H., Hudson, D.E., Miller, L.H., McCutchan, T.F., Cohen, S. (1990). A merozoite receptor protein from *Plasmodium knowlesi* is highly conserved and distributed throughout *Plasmodium*. *J Biol Chem.* 265: 17974-17979.

References

- Weinbaum, F.I., Evans, C.B., Tigelaar, R.E. (1976). An in vitro assay for T cell immunity to malaria in mice. *J Immunol.* 116: 1280 – 1283.
- Wu, Y., Kirkman, L.A., Wellems, T.E. (1996). Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *PNAS.* 93: 1130-1134.
- Weiss, W.R., Good, M.F., Hollingdale, M.R., Miller, L.H., Berzofsky, J.A. (1989). Genetic control of immunity to *Plasmodium yoelii* sporozoites. *J Immunol.* 143: 4263 – 4266.
- Yamauchi, M., Coppi, A., Snounou, G., Sinnis, P. (2007). *Plasmodium* sporozoites trickle out of the injection site. *Cell Microbiol.* 9: 1215 – 1222.
- Yazdanbakhsh, K., Rios, M., Storry, J.R., Kosower, N., Parasol, N., Chaudhuri, A., Reid, M.E. (2000). Molecular mechanisms that lead to reduced expression of Duffy antigens. *Transfusion.* 40: 310-320.
- Yeoh, S., O'Donnel, R.A., Koussis, K., Dluzewski, A.R., Ansell, K.H., Osbourne, S.A., Hackett, F., Withers-Martinez, C., Mitchell, G.H., Bannister, L.H., Bryans, J.S., Kettleborough, C.A., Blackman, M.J. (2007). Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. *Cell* 131: 1072-1083.
- Yoelii, M., Hargreaves, B., Carter, R., Walliker, D. (1975). Sudden increase in virulence in a strain of *Plasmodium berghei yoelii*. *Ann Trop Med Parasitol.* 69: 173-178.
- Yoelii, M., Hargreaves, B.J. (1974). Brain capillary blockage produced by a virulent strain of rodent malaria. *Science.* 184: 572 - 573.
- Zhang, G.Q., Guan, Y.Y., Sheng, H.H., Zheng, B., Wu, S., Xiao, H.S., Tang, L.H. (2008). Multiplex PCR and oligonucleotide microarray for detection of single-nucleotide polymorphisms associated with *Plasmodium falciparum* drug resistance. *J Clin Microbiol.* 46: 2167 - 2174.

Appendix 1: Buffers and reagents for rodent parasites

Histodenz Working Solution (27.6%w/v)

27.6g of Histodenz (Sigma-Aldridge, USA) was dissolved into 60ml of buffered medium (5mM Tris-Cl pH7.5; 3mM KCl; 0.3mM EDTA.2Na (Sigma-Aldridge, USA)). The solution was then topped up to 100ml with the same buffered medium to obtain a density of 1.15g/ml at 20°C. The working solution was then autoclaved at 120°C for 20min and stored at 4°C.

50% Histodenz Gradient

A 10ml solution of 50%(v/v) Histodenz solution was freshly prepared by mixing 5ml of Histodenz working solution with 5ml of RPMI 1640 (Invitrogen, USA)

Discontinuous Histodenz Gradient

A total of 12ml of 4 individual 3ml step gradients of 80%, 70%, 60% and 50% (v/v) of Histodenz stock solution to RPMI 1640 (Invitrogen, USA) were carefully layered respectively on top of each other in a 15ml centrifuge tube (BD Biosciences, USA). Each centrifuge tube is sufficient to separate up to 2ml of parasitized blood that was resuspended back to its original volume in fetal calf serum (Invitrogen, USA).

Kreb's-Henseleit Buffer

9.55g of the powdered buffer was dissolved in 900ml of Milli-Q water (Millipore, USA). 2.1g sodium bicarbonate (Sigma-Aldridge, USA) was added and the pH was equilibrated. Additional Milli-Q water was then added to make up till 1000ml. The buffer was then sterilized immediately by filtration using a membrane with a porosity of 0.22 microns and then stored at 4°C.

Parasite freezing medium

Glycerol (Fluka, USA)	400ml
1.4M sodium lactate (Sigma-Aldridge, USA) (157g/L)	100ml
0.5M potassium chloride (Sigma-Aldridge, USA) (37.5g/L)	10ml
ddH ₂ O	450ml

The above were combined and the pH was adjusted to 7.4 by adding sodium dihydrogen phosphate (Fluka, USA).

Saline

Sodium chloride (Sigma-Aldridge, USA)	0.9g
Glucose (Sigma-Aldridge, USA)	0.2g
ddH ₂ O	to 100ml.

The buffer was then sterilized immediately by filtration using a membrane with a porosity of 0.22 microns and then stored in 10ml aliquots at 4°C.

Appendix 2: Buffers and reagents for proteomics

10X SDS electrophoresis running buffer

Tris 250mM, glycine 1.92M, SDS 1%

PlusOne™ Tris (Amersham Biosciences, UK)	60.5g
PlusOne™ Glycine (Amersham Biosciences, UK)	288g
SDS (Fluka, USA)	20g
Milli-Q water (Millipore, USA)	to 2000ml

This 10X buffer was diluted with Milli-Q water to the desired concentration.

10% ammonium persulfate

Ammonium persulfate (Sigma-Aldridge, USA)	0.1g
Milli-Q water (Millipore, USA)	to 1ml

The solution was freshly prepared and used immediately.

4x Tris-Cl pH8.8

PlusOne™ Tris (Amersham Biosciences, UK)	91g
Milli-Q water (Millipore, USA)	400ml

The solution was dissolved with a magnetic stirrer platform (IKA Werke, Germany) and the pH was adjusted to 8.8 with concentrated HCl (Sigma-Aldridge, USA). The solution was then topped up to 500ml with Milli-Q water and then filtered through a 0.45µm filter (Millipore, USA) before 2g of SDS (Fluka, USA) was dissolved into it. This solution was then stored at 4°C until use.

4x Tris-Cl pH6.8

PlusOne™ Tris (Amersham Biosciences, UK)	6.05g
Milli-Q water (Millipore, USA)	40ml

The solution was dissolved with a magnetic stirrer platform (IKA Werke, Germany) and the pH was adjusted to 6.8 with concentrated HCl (Sigma-Aldridge, USA). The solution was then topped up to 100ml with Milli-Q water and then filtered through a 0.45µm filter (Millipore, USA) before 0.4g of SDS (Fluka, USA) was dissolved into it. This solution was then stored at 4°C until use.

10% polyacrylamide gel

30% Acrylamide/Bis-acrylamide (29:1) (Biorad, USA)	1.67ml
4x Tris-Cl pH8.8 (Appendix 2)	1.25ml
10% ammonium persulfate (Appendix 2)	17µl
TEMED (Sigma-Aldridge, USA)	3.3µl
Milli-Q water (Millipore, USA)	2.08ml

The above mixture is enough for 1 mini gel (Bio-Rad, USA). After the plates were filled to the desired height, 1ml of isopropanol (Merck, USA) was pipetted as a layer onto each gel.

Stacking gel

30% Acrylamide/Bis-acrylamide (29:1) (Biorad, USA)	0.325ml
4x Tris-Cl pH6.8 (Appendix 2)	0.625ml
10% ammonium persulfate (Appendix 2)	12.5µl
TEMED (Sigma-Aldridge, USA)	2.5µl
Milli-Q water (Millipore, USA)	1.525ml

The above mixture is enough for 1 mini gel (Bio-Rad, USA).

2x SDS sample buffer

4x Tris-Cl pH6.8 (Appendix 2)	2.5ml
Glycerol (Sigma-Aldridge, USA)	2ml
SDS (Fluka, USA)	0.4g
DTT (dithiothreitol) (GE Healthcare, USA)	0.31g
Bromophenol Blue (USB, USA)	Few grains
ddH ₂ O	to 10ml

The mixture was stored in 1ml aliquots at -20°C and used as required.

Appendix 3: General buffers

Phosphate buffered saline (PBS)

The following were dissolved in 800ml distilled water:

NaCl (Sigma-Aldridge, USA)	8g
KCl (Sigma-Aldridge, USA)	0.2g
Na ₂ HPO ₄ (Sigma-Aldridge, USA)	1.44g
KH ₂ PO ₄ (Sigma-Aldridge, USA)	0.24g

The pH was adjusted to 7.4 with 1M HCl (Merck, USA) and the volume adjusted to 1L with additional distilled water. The buffer was then sterilized by autoclaving at 120°C for 20min.

1M NaOH

NaOH (Fluka, USA)	4g
Milli-Q water (Millipore, USA)	100ml

0.5M Sodium bicarbonate, pH9.0

Sodium bicarbonate (Sigma-Aldridge, USA)	42g
Milli-Q water (Millipore, USA)	to 1000ml

The pH was adjusted to 9.0 with 1M HCl (Merck, USA) and sterilized by autoclaving at 120°C for 20min.

20 X SSC

The following were dissolved in 800ml distilled water:

Sodium chloride (Sigma-Aldridge, USA)	175.3g
Sodium citrate (Sigma-Aldridge, USA)	88.2g

The pH was adjusted to 7.4 with 1M HCl (Merck, USA) and the volume adjusted to 1L with additional distilled water. The buffer was then sterilized by autoclaving at 120°C for 20min. Suitable concentrations were obtained by diluting with autoclaved distilled water.

1M HEPES buffer, pH7.5

HEPES (Sigma-Aldridge, USA)	23.83g
ddH ₂ O	80ml.

The pH was adjusted to 7.5 with potassium hydroxide (Merck, USA) and the volume adjusted to 100ml with additional distilled water. The buffer was then sterilized by autoclaving at 120°C for 20min.

LB medium

The following were dissolved in 800ml distilled water:

Bacto™-tryptone (Oxoid, UK)	10g
Bacto™-yeast extract (Oxoid, UK)	5g
NaCl (Sigma-Aldridge, USA)	5g

The pH was adjusted to 7.4 with sodium hydroxide (Sigma-Aldridge, USA) and the volume adjusted to 1L with additional distilled water. The medium was then sterilized by autoclaving at 120°C for 20min.

LB Ampicillin plates

Add 15g agar (Oxoid, UK) to 1 litre of LB medium, dissolve completely and equilibrate to pH 7.0. Autoclave the mixture and allow it to cool to 50°C before adding ampicillin (Gibco, USA) to a final concentration of 100µg/ml. Pour 30ml of the medium into 85mm petri dishes (BD-Bioscience, USA) and let the agar solidify. The set agar can be stored at 4°C up to a month before use.

Appendix 4: Microarray Buffers

Wash Solution

Ethanol (Merck, USA)	2400ml
Sodium hydroxide (Sigma-Aldridge, USA)	400g
Milli-Q water (Millipore, USA)	1600ml

Poly-l-lysine Solution

Milli-Q water (Millipore, USA)	2850ml
1 X PBS (Appendix 3)	350ml
Poly-l-lysine (Sigma-Aldridge, USA)	300ml

The above constituents were measured and mixed in a plastic measuring cylinder and then filtered through a 0.2µm filter (Millipore, USA) into a clean plastic bottle before use.

1M Sodium Borate, pH8.0

In 800ml of Milli-Q water, dissolved 61.84g of boric acid (Sigma-Aldridge, USA) and the pH adjusted with sodium hydroxide pellets (Sigma-Aldridge, USA) before topping up to 1L with Milli-Q water.

Blocking Solution

1-methyl-2-pyrrolidinone (Sigma-Aldridge, USA)	335ml
Succinic anhydride (Sigma-Aldridge, USA)	5.5g

Once the succinic anhydride was fully dissolved, immediately add 15ml of 1M sodium borate (Appendix 4) to the stirring mixture and use immediately once they are uniformly mixed.

30x aa-dUTP dNTPs

dATP (100mM, GE Healthcare, USA)	30µl
dCTP (100mM, GE Healthcare, USA)	15µl
dGTP (100mM, GE Healthcare, USA)	15µl
dTTP (100mM, GE Healthcare, USA)	15µl
5-aminoallyl-dUTP (100mM, Biotium, USA)	15µl
Molecular biology grade water (Eppendorf, Germany)	10µl

The mixture was aliquoted and can be diluted to a suitable concentration with molecular biology grade water and stored at -20°C.

Appendix 5: Southern Blot Buffers

Depurination Solution

12N concentrated HCl (Merck, USA)	12ml
ddH ₂ O	998ml

Concentrated HCl was added slowly to water and then mixed by several inversions.

Denaturation Solution

Sodium hydroxide (Sigma-Aldridge, USA)	20g
Sodium chloride (Sigma-Aldridge, USA)	87.66g
ddH ₂ O	to 1L

Neutralizing Solution

Tris-base (Sigma-Aldridge, USA)	121.12g
Sodium chloride (Sigma-Aldridge, USA)	87.66
ddH ₂ O	800ml

The solution was equilibrated to pH 8.0 by adding concentrated HCl (Merck, USA) and topped up with ddH₂O to a volume of 1L.

Maleic Acid Buffer (10x)

Maleic acid (Sigma-Aldridge, USA)	116.1g
Sodium chloride (Sigma-Aldridge, USA)	87.66g
ddH ₂ O	800ml

The pH was adjusted to 7.5 by adding sodium hydroxide pellets (Sigma-Aldridge, USA) and the volume topped up to 1L with ddH₂O.

Washing Solution I

In 800ml of ddH₂O, dissolve the following:

Disodium hydrogen phosphate (Sigma-Aldridge, USA)	2.84g
EDTA (Bio-Rad, USA)	292.25mg
SDS (Bio-Rad, USA)	10g

The pH was equilibrated to 7.5 with concentrated HCl (Merck, USA) and the buffer topped up with ddH₂O and filtered through a 0.2µm filter (Millipore, USA) into a clean plastic bottle before use.

Detection Buffer

In 800ml of ddH₂O, dissolve the following:

Tris-base (Sigma-Aldridge, USA)	12.11g
Sodium chloride (Sigma-Aldridge, USA)	5.84g

The pH was equilibrated to 9.5 with concentrated HCl (Merck, USA) and the buffer topped up with ddH₂O and filtered through a 0.2µm filter (Millipore, USA) into a clean plastic bottle before use.

Appendix 6: Primer sequences

Primer Name	Sequence (5' – 3')
PY00632_F	AATGTGAATCCCACGAGAGAA
PY00632_R	CGAAACGTCTCATGTTTAGAATAT
Py03414_F	TGAACAATCAAGAAGTAAAATTGG
Py03414_R	TGGATCAAAACTTTCAAAAATAGAA
PY04600_F	AAAAATTCAAACATGAAAGCAGAC
PY04600_R	CTTTTGTCAGCTCGTTTGTCAC
PY04485_F	CAAACCTGTTATTGCGTCTTATG
PY04485_R	AACTTGCATGATGGATTTGTG
PY02086_F	TTTACAGGTACGAATGATTAATTTCA
PY02086_R	AAATTTTGTCTTTCATCACTTCAA
PY04869_F	GGAAAAGAAAAGTAAATGGAAAGG
PY04869_R	TTGTTCACTACCTACCTTATGTAACC
PY06972_F	GCCATCCCAGTAGCAGAATTAC
PY06972_R	GAGTCCTTTTTGCCATCATCAG
PY05482_F	GAGTTCAAAGAAAATTTAATATCAGC
PY05482_R	AGTTAAGTATTTTAAAGCAGGA

Table 9.1. PCR primers for verification of newly annotated genes.

Primer Name	Sequence (5' – 3')
PFI0535w_F	CCCAAACCTCACAGCGATGAC
PFI0535w_R	AATAAGAGCGCCTTCCAATG
PF14_0473_F	TTGGTTTGACATCTATCACTTCG
PF14_0473_R	TTCTTGACGCGATTAAAAAGG
MAL13P1.345_F	CACCGATCGAATGATAATGG
MAL13P1.345_R	ACACAGACTTTCATAGGATATTTTCAG
PFL0595c_F	AATTCCGCATCAAAGTGTGG
PFL0595c_R	AAAGTTCGTGGGTATTTGTTCC
PFF1480w_(3'F)	GCATCAATCTGTTATTGGCTTG
PFF1480w_(3'R)	CAGTTCCTCCAGATAAAAGCAAAG
PF13_0131_F	AATAAATTAGTTTGTGTTTGTCCATTATAG
PF13_0131_R	AATTGTCTGGTGAATTTTCATCC
PFC0095c_F	CAAATAGGCTGTATAAGAAGGTGTG
PFC0095c_R	TGGCTCATAATAATTTGCCTTAG
PFL2450C_F	TCTGCCACAAACAAACATGG
PFL2450C_R	TAAACGTTTCCAGCATGGTTC
MAL8P1.310_F	CAGCCACAAGCCAGTTATCC
MAL8P1.310_R	CGCAGATGTCCAAAGACAAG
PFB0645c_F	GTCTTAATCTATTATTTGGTAGCATTC
PFB0645c_R	ACTGCAATAAGTGTATTTGATAAAGAG
PFF1480w_(5'F)	ACCCTCTTCTTTATATATACATTTTTTG
PFF1480w_(5'R)	CAAAATATGCATCAATCTGTTATTG
PF13_0278_F	TAAATATTCATAAATCATCTGCTTCTC
PF13_0278_R	ATTTGCTAAGGACATATAAAATGATG

Table 9.2. PCR primers for verification of unique sets of genes.

Primer Name	Sequence (5' – 3')
291F	AATGTTACAAATGTGCATTG
291R	AATGTTACAAATGTGCATTG
292F	GAACCAATACATCAGATCAAC
292R	GTTGACCAAGTACGTTTGG
653F	AAAGATCTACTAAACGAATGTG
653R	TTAATTTTGTTAATAACTTATC
1120F	TGGACGTCAA AATTGTAATG
1120R	CCTTTTTTTGTTCTCTTTTC
2147F	AAATGAAAGGGAAAATCG
2147R	ACCTGATCCACCTGAACC
2198F	CTCGATTATTTTAGCTCGAC
2198R	GGTTTATTTGTAAACTTTTGG
2799F	CTATAAAGGCTTTTACACGC
2799R	AATAATTTTGTCTGAGATAGCTC
5489F	AATAATACATCTTCATTTATGAGC
5489R	TTCATGTTCACTTATTATACATGC
6381F	GCTAATAAAACAAATGGGC
6381R	TTTTTCACAGAAGCATCTTC
6415F	ATCTTAAAGCAATGAAGAGTG
6415R	CTCAGGAATAATCATTGTAGC
7180F	GAAAAGTAACAATTATATTGTATCTG
7180R	AATTATTGTAAGTATTCTCATCAAC

Table 9.3. Real-time RT-PCR primers for validating differentially transcribed genes.

Primer Name	Sequence (5' – 3')
291Bam	TCCAGGATCCAAGAATGTCCCAAAGCTATGG
291Nco	CATTCCATGGTGGGAATCTGCTGCTTGTGC
291intF	TTTGATGGAGTTATAGATTTACCTTTACC
291intR	GCAGTTTGACATTCACTCCATTTCG
291e1F	AGCTGGATCCCCTAGCACGGATGATCAAAAGG
291e1R	AGCTCCATGGTTGGTTCGGTTTCTGAAGATGG
291ex1_3'int	TTCTTTTAAACATTATGTACTTGAATAGATG
291ex1_5'int	CAAATAACCAATACAAAAGGGGAAACTGG
292Bam	TCCAGGATCCCAGTTAGTGGTTTAGGTGTATTAG
292Nco	CATTCCATGGTTCTGCTGACATGGTTGTAG
292intF	TTGTCAATAGAAAGGGTTTGAAATTACC
292intR	CCATCTTGACTTGATCCTCCTCCAC
292F_KpnI	AGCTGGTACCTCAGCCAAACAAGGTTCTGGTTCTG
292R_HindIII	AGCTAAGCTTCATAATTCACATTTACCCATTCCTGG
292F_EcoRI	AGCTGAATTCACACTACAAAGTGGAGGAGGATCAAG
292R_SacII	AGCTCCGCGGAGTGGATCTCTGCTCATAGCTCTTG
292int5'F	ATGCTGGGGAATGACATAATAAATTGTTCGC
292int3'R	TTATACAGCACAAAAGTTACATTCATTATC
292wtR	CTTTAGTTAATAGTGAATGAACTTGGATTG
B3D3'Fint	TGTATGTTGTGTGCATATTTTTTTTTTGTG
B3D5'Rint	CATACACAAACATACAAAAATAAACACC
eGFF_R	TTCAGCTCGATGCGGTTCCACC
Rap2/3_3'intF	TGAGAAGAATGATTAAAGATCCCG

Table 9.4. PCR primers for SERA gene targeting.

Primer Name	Sequence (5' – 3')
291F	AATGTTACAAATGTGCATTG
291R	AATGTTACAAATGTGCATTG
292RTf	AAATGTGACAAAATAGCTACCAAATG
292RTTr	CATCACTTTGTCCAAGTCTAATAC
293RTf	TTGATAACAATGGATGATTTTGATG
293RTTr	TTCATTACCTAATTCAGCATTTTCC
2062RTf	AATAACGAAATTGACGATGAAAATG
2062RTTr	AAAGTGCTGAGCTTCTATAATGACC
2063RTf	TACGAAATAGTTGGGGTTCTAGATG
2063RTTr	ATTTGGTGGGTCTACTATCTTAGGG

Table 9.5. Real-time RT-PCR primers for PySERA gene

Appendix A: Identification of syntenic orthologs by microarray

PtGene	PtGene	PcGene	PvGene	Pf Product	Pf function
PFA0260C	PB001039.01.0	PC000215.05.0	PB001039.01.0	hypothetical protein	hypothetical
PFA0270C	PB001363.02.0	PB001363.02.0 + PY04768	PY04768	hypothetical protein	hypothetical
PFA0335W	PB000519.01.0	PB000519.01.0 + PY01029	PY01029	P. falciparum GTP binding protein RAB5	GTP-binding
PFA0345W	PB001391.02.0	PC000332.05.0	PB001391.02.0	centrin, putative	ca-binding
PFA0425C	PY00632	PC000290.05.0	PY00632	hypothetical protein	hypothetical
PFA0430C	PB000915.00.0	PB000915.00.0 + PY04869	PY04869	hypothetical protein	hypothetical
PFA0585W	PY01445	PY01445	PY01445	hypothetical protein	hypothetical
PFB0120W	-	PY02025	PY02025	early transcribed membrane protein, putative	early transcribed membrane protein,
PFB0330C	PB000352.01.0	PY02062 + PY00294	PY02062 & PY00294	cysteine protease, putative	protease
PFB0345C	PB000108.03.0	PB000108.03.0 + PY00291	PY00291	cysteine protease, putative	protease
PFB0505C	PB000108.00.0	PC001097.02.0	PB000108.00.0 + PC001097.02.0	beta-ketoacyl-acyl carrier protein synthase III precursor, putative	biosynthesis
PFB0620W	PB001507.02.0	PY03935	PY03935	hypothetical protein	hypothetical
PFB0725C	PB001189.00.0	PC000454.04.0	PB102042.00.0 + PB001189.00.0	hypothetical protein	hypothetical
PFB0805C	PB000257.02.0	-	PB000257.02.0	clathrin coat assembly protein, putative	transport
PFB0826C	PB001103.03.0	PB001103.03.0 + PY04711	PY04711	hypothetical protein	hypothetical
PFC0090W	PY06010	PY06010	PY06010	hypothetical protein, conserved	hypothetical
PFC0125W	PB300885.00.0 & PB000591.00.0	PB300885.00.0 + PB000591.00.0 + PY03961	PY03961	ABC transporter, putative	transport
PFC0360W	PB000835.01.0	PC000094.03.0	PB000835.01.0	hypothetical protein, conserved	hypothetical
PFC0441C	PC001356.02.0	PC001356.02.0	PY04581	hypothetical protein	hypothetical
PFC0525C	PB000288.00.0	PB000288.00.0 + PY01719	PY01719	glycogen synthase kinase, putative	kinase
PFC0600W	PC000010.05.0 + PY02285	PC000010.05.0	PY02285	hypothetical protein	hypothetical
PFC0615W	PB000077.03.0	PB000077.03.0 + PY07280	PY07280	hypothetical protein	hypothetical
PFC0635C	PB000857.02.0	PB000857.02.0 + PY04859	PY04859	translation initiation factor E4, putative	protein synthesis
PFC0645W	PB100427.00.0	PC105981.00.0	PC105981.00.0	hypothetical protein	hypothetical
PFC0670C	PB000496.00.0	PY05833	PY05833	hypothetical protein	hypothetical
PFC0740C	PY05143	PC000473.02.0	PY05143	binding protein, putative	protein-binding
PFC0770C	PY02733	PC001291.02.0	PY02733	kinesin-related protein, putative	motor
PFC0795W	PB108208.00.0	PY02651	PY02651	hypothetical protein	hypothetical
PFC0895W	PB000070.00.0	PB000070.00.0 + PY00599	PY00599	hypothetical protein	hypothetical
PFC1050W	PB000532.03.0	PY00358	PY00358	hypothetical protein	hypothetical
PFD0150W	PC000637.03.0 + PY02441	PC000637.03.0	PY02441	hypothetical protein	hypothetical
PFD0170C	PB001000.02.0	PY02446	PY02446	hypothetical protein	hypothetical
PFD0175C	PY02446-exon1	PY02446-exon1	PY02446-exon1	hypothetical protein	hypothetical
PFD0310W	-	PY05101	PY05101	sexual stage-specific protein precursor	sexual-stage
PFD0345C	PB000504.01.0	PB000504.01.0 + PY05694	PY05694	hypothetical protein	hypothetical
PFD0367W	PB001254.00.0	PB001254.00.0 + PY07641	PY07641	hypothetical protein	hypothetical
PFD0375W	PB001255.00.0	PY03396	PY03396	hypothetical protein	hypothetical
PFD0520C	PB000737.02.0	PB000737.02.0 + PY	PY02308	hypothetical protein	hypothetical
PFD0580C	PB300356.00.0	PC001211.02.0	PC001211.02.0	hypothetical protein	hypothetical
PFD0665C	PB000557.01.0	PB000557.01.0 + PY00544	PY00544	26S proteasome aaa-ATPase subunit Rpt3, putative	protease
PFD0675W	PB000464.01.0	PB000464.01.0 + PY00545	PY00545	ribosomal protein r10, putative	ribosomal
PFD0715C	PB000478.00.0	PB000478.00.0 + PY02201	PY02201	hypothetical protein, conserved	hypothetical
PFD0935C	PB000437.00.0	PC000783.01.0	PC000783.01.0	hypothetical protein	hypothetical
PFD0945C	-	PY05934	PY05934	hypothetical protein	hypothetical
PFD1025W	PY02467	PY02467	PY02467	erythrocyte membrane protein 1 (PEMP1)-like pseudogene	adhesion
PFD1066W	PY07678	PY07678	PY07678	hypothetical protein	hypothetical
PFD1045W	PY06448	PC000158.02.0	PY06448	hypothetical protein	hypothetical
PFE0085C	-	PY03919	PY03919	hypothetical protein	hypothetical
PFE0105C	PY04373	PY04373	PY04373	hypothetical protein	hypothetical
PFE0130C	PB000152.00.0	PC000791.04.0	PB000152.00.0	hypothetical protein	hypothetical
PFE0320W	PB105749.00.0	PB105749.00.0	PY03156	hypothetical protein	hypothetical
PFE0355C	PB000701.02.0	PB000701.02.0	PB000701.02.0	serine protease belonging to subtilisin family, putative	protease
PFE0815W	PB000111.00.0	PB000111.00.0 + PY05227	PY05227	hypothetical protein	hypothetical
PFE0540C	PB001081.03.0	PB001081.03.0 + PY04894 + PY00914	PY04894 & PY00914	hypothetical protein	hypothetical
PFE0600C	PC102462.00.0 + PY07249	PC102462.00.0	PY07249	hypothetical protein	hypothetical
PFE0615W	PC106709.00.0	PC106709.00.0	PY03183	hypothetical protein	hypothetical
PFE0640W	PY04627	-	PY04627	hypothetical protein	hypothetical
PFE1025C	PB301336.00.0	PY04488	PY04488	hypothetical protein	hypothetical
PFE1030C	PY07753	PY07753	PY07753	phosphomethylpyrimidine kinase, putative	kinase
PFE1040C	PB000732.03.0	PC000296.02.0	PC000296.02.0	hypothetical protein	hypothetical
PFE1075C	PY00872	PC000231.00.0	PY00872	hypothetical protein	hypothetical
PFE1100W	PB405740.00.0	PY06034	PY06034	hypothetical protein	hypothetical
PFE1205C	-	PY00157	PY00157	hypothetical protein	hypothetical
PFE1235C	PY07395	PC000218.04.0	PY07395	hypothetical protein	hypothetical
PFE1265W	PY02110	PC000780.01.0	PY02110	G-protein coupled receptor, putative	receptor
PFE1505W	PB000448.01.0	PY01811	PY01811	hypothetical protein	hypothetical
PFE1510C	PB000447.01.0	PB000447.01.0 + PY01812	PY01812	phosphoenolpyruvate/phosphate translocator precursor, putative	transport
PFE1540W	PB001036.00.0	PC000192.00.0	PB001036.00.0	hypothetical protein	hypothetical
PFE1550W	PY07379	PY07379	PY07379	hypothetical protein	hypothetical
PFE1560C	PY02968	PC000699.03.0	PY02968	hypothetical protein	hypothetical
PFF0130C	PB000865.01.0	PB000865.01.0 + PY00219	PY00219	hypothetical protein	hypothetical
PFF0195C	PB300118.00.0	PB300118.00.0	PY04445	hypothetical protein	hypothetical
PFF0240C	PB101509.00.0	PY03727	PY03727	hypothetical protein	hypothetical
PFF0245W	PB000858.00.0	PC000783.04.0	PB000858.00.0 + PC000783.04.0	50S ribosomal subunit L24, putative	ribosomal
PFF0510W	PY00496	PY00496	PY00496	histone H3, putative	DNA-binding
PFF0515C	PB000307.03.0	PB000307.03.0 + PY00497	PY00497	hypothetical protein, conserved	hypothetical
PFF0565C	PB300967.00.0	PB300967.00.0 + PY02914	PY02914	hypothetical protein, conserved	hypothetical
PFF0570C	PB301072.00.0	PB301072.00.0	PY02915	hypothetical protein	hypothetical
PFF0600W	PY03349	PY03349	PY03349	hypothetical protein	hypothetical
PFF0630C	PB000550.00.0	PB000550.00.0 + PY03097	PY03097	hypothetical protein, conserved	hypothetical
PFF0635W	PB107557.00.0	PB107557.00.0 + PY04882	PY04882	hypothetical protein	hypothetical
PFF0740C	PB000278.00.0	PC000183.03.0	PB000278.00.0	hypothetical protein	hypothetical
PFF07160W	PB000600.00.0	PB406000.00.0 + PY07541	PY07541	RNA and export factor binding protein, putative	RNA-binding
PFF0865W	PY05073	PY05073	PY05073	histone h3	DNA-binding
PFF0885W	PB000967.01.0	PB000967.01.0 + PY03375	PY03375	60S ribosomal protein L27a, putative	ribosomal
PFF0925W	PB000280.01.0	PB000280.01.0	PY04987	hypothetical protein	hypothetical
PFF1030W	PY04474	PC300145.00.0	PY04474	hypothetical protein, conserved	hypothetical
PFF1120C	PB001282.02.0	PC000386.01.0	PB001282.02.0 + PC000386.01.0	hypothetical protein	hypothetical
PFF1255W	PB000769.01.0	PY02046	PY02046	hypothetical protein	hypothetical
PFF1290C	PB000532.01.0	PB000532.01.0 + PY07531	PY07531	hypothetical protein	hypothetical
PFF1385C	PB300148.00.0	PB300148.00.0 + PY06369	PY06369	hypothetical protein	hypothetical
PFF1405C	PB000128.00.0 & PB100240.00.0	PC000071.01.0	PC000071.01.0	hypothetical protein	hypothetical
PFF07_0005	PB108278.00.0	PY02009	PY02009	lysophospholipases-like protein, putative	hydrolase
MAL7P1_13	PB107249.00.0 & PB301339.00.0	PC10670.00.0	PC10670.00.0	hypothetical protein	hypothetical
MAL7P1_17	PB105908.00.0	PY01949	PY01949	hypothetical protein	hypothetical
MAL7P1_320	PB105908.00.0	PB105908.00.0 + PY02795	PY02795	Ribosomal protein, L37e, putative	ribosomal
MAL7P1_36	PY01895	PC100836.00.0	PY01895	hypothetical protein	hypothetical
PFF07_0052	PB000890.00.0	PB000890.00.0 + PY06753	PY06753	hypothetical protein	hypothetical
PFF07_0060	PY01299	PC000719.04.0	PY01299	hypothetical protein, conserved	hypothetical
MAL7P1_120	PB000095.00.0	PC000382.01.0	PB000095.00.0	hypothetical protein	hypothetical
MAL7P1_154	PB000068.03.0	PY01310	PY01310	hypothetical protein	hypothetical
MAL7P1_155	PB000589.00.0	PC000568.02.0	PB000589.00.0 + PC000568.02.0	hypothetical protein	hypothetical
MAL7P1_161	PB106943.00.0 & PB106943.00.0	PB000994.00.0 + PY01469	PY01469	dynein light chain, putative	motor
PFB08_0136	PB106943.00.0	PB106943.00.0 + PB106943.00.0	PB106943.00.0 + PB106943.00.0	hypothetical protein	hypothetical
MAL8P1_158	PB106945.00.0-PB106944.00.0	PB106945.00.0-PB106944.00.0	PB106945.00.0-PB106944.00.0	hypothetical protein	hypothetical
MAL8P1_145	PB001044.01.0	PC001250.02.0	PB001044.01.0	hypothetical protein, conserved	hypothetical
PFB08_0122	PY00383 + PY05524	PC107744.00.0 & PC000312.04.0 & PC107742.00.0	PY00383 & PY05524 (exon1)	hypothetical protein	hypothetical
PFB08_0121	PB000654.02.0	PB000654.02.0 + PY00382	PY00382	peptidyl-prolyl cis-trans isomerase precursor	protein folding
MAL8P1_200	PB402059.00.0	PB402059.00.0 + PY05005	PY05005	hypothetical protein	hypothetical
PFB08_0099	PB000715.03.0	PB000715.03.0 + PY01656	PY01656	hypothetical protein, conserved	hypothetical
PFB08_0083	PB000460.03.0	PY03189	PY03189	hypothetical protein	hypothetical
MAL8P1_94	PY00458	PY00458	PY00458	hypothetical protein	hypothetical
MAL8P1_88	PB000867.03.0	PB000867.03.0 + PY05341	PY05341	hypothetical protein	hypothetical
MAL8P1_75	PB000873.00.0	PC000482.01.0	PB000873.00.0	hypothetical protein	hypothetical
MAL8P1_66	PB402680.00.0	PB402680.00.0 + PY06542	PY06542	hypothetical protein	hypothetical
MAL8P1_61	PB301041.00.0	PB301041.00.0	PY03073	hypothetical protein	hypothetical
PFB08_0041	PB001150.02.0	PC000338.04.0	PB001150.02.0 + PC000338.04.0	ribosome biogenesis protein nep1 homologue, putative	ribosomal
PFB08_0024	PY01887	PY01887	PY01887	hypothetical protein	hypothetical
MAL8P1_25	PY01888	PY01888	PY01888	hypothetical protein	hypothetical
MAL8P1_20	PC000181.00.0	PC000181.00.0	PY03586	hypothetical protein	hypothetical
PFB08_0015	PB000365.00.0	PB000365.00.0 + PY06194	PY06194	hypothetical protein	hypothetical
PFF0145W	PB000108.02.0	PB000108.02.0 + PY01822	PY01822	hypothetical protein	hypothetical
PFF0150C	PB000787.01.0	PB000787.01.0 + PY01823	PY01823	hypothetical protein	hypothetical
PFF0180W	PB000857.00.0	PY01155	PY01155	alpha tubulin	protein polymerization

Appendix A: Identification of syntenic orthologs by microarray

PfGene	PbGene	PcGene	PvGene	Pf Product	Pf function
PF0335w	PY05144	PC001118.02.0	PY03114	hypothetical protein	hypothetical
PF0370c	PB000321.02.0	PB000321.02.0 + PY04117	PY04117	subunit of proteasome activator complex, putative	proteasome
PF0375w	PC301277.00.0 + PY01737	PC301277.00.0	PY01737	ribosomal protein L35, putative	ribosomal
PF0440w	PB000599.00.0	PB000599.00.0 + PY03134	PY03134	hypothetical protein	hypothetical
PF0465c	PB000313.02.0	PB000313.02.0 + PY03129	PY03129	hypothetical protein	hypothetical
PF0560c	PB100146.00.1	PB100146.00.1	PY02769	hypothetical protein	hypothetical
PF0585c	PB000983.01.0	PB000983.01.0 + PY03424	PY03424	hypothetical protein	hypothetical
PF0700c	PB000080.00.0	PY04160	PY04160	Met-10 - like protein, putative	biosynthesis
PF0715w	PB000461.03.0	PB000461.03.0 + PY06377	PY06377	hypothetical protein, conserved	hypothetical
PF0740c	PB000414.02.0	PC000396.00.0	PB000414.02.0 + PC000396.00.0	ubiquitin conjugating enzyme, putative	ubiquitin-like
PF0790w	PC000085.02.0	PC000085.02.0	PY00221	thioredoxin, putative	hydrolase
PF0823w	PY07484	PY07484	PY07484	hypothetical protein, conserved	hypothetical
PF0910w	PB301025.00.0	PB301025.00.0 + PY07206	PY07206	DNA helicase, putative	RNA-binding
PF0915w	PY06799 + PY06800	PC107037.00.0	PY06799 + PY06800	hypothetical protein	hypothetical
PF0995w	PB000836.00.0	PB000836.00.0	PB000836.00.0	hypothetical protein	hypothetical
PF1035w	PC000013.04.0	PC000013.04.0	PY03605	hypothetical protein	hypothetical
PF11135c	PY02414	PY02414	PY02414	hypothetical protein	hypothetical
PF1200w	PB000420.00.0	PB000420.00.0	PY05197	hypothetical protein	hypothetical
PF1250w	PB000908.02.0	PB000908.02.0	PB000908.02.0	thioredoxin, putative	redox
PF1463w	-	PY04278	PY04278	hypothetical protein	hypothetical
PF1490c	PB001091.02.0	PB001091.02.0 + PY03175	PY03175	large cyclophilin-like protein	isomerase
PF1555w	PY03210	PC106873.00.0 & PC106872.00.0	PY03210	hypothetical protein	hypothetical
PF10_0033	PY02919	PY02919	PY02919	hypothetical protein	hypothetical
PF10_0035	PB000173.01.0	PB000173.01.0 + PY00940	PY00940	hypothetical protein	hypothetical
PF10_0042	PB000361.03.0	PC001089.02.0	PC001089.02.0	hypothetical protein	hypothetical
PF10_0065	PB000470.02.0	PC301244.00.0	PB000470.02.0	hypothetical protein	hypothetical
PF10_0095	PB000647.03.0	PY02531	PY02531	hypothetical protein	hypothetical
PF10_0098	PB000012.01.0	PB000012.01.0 + PY04010	PY04010	hypothetical protein	hypothetical
PF10_0100	PC000364.03.0 + PY06182	PC000364.03.0	PY06182	hypothetical protein	hypothetical
PF10_0103	PB001076.01.0	PB001076.01.0 + PY02091	PY02091	eukaryotic translation initiation factor 2, beta, putative	protein synthesis
PF10_0106	PY02088	PC000395.00.0	PY02088	hypothetical protein	hypothetical
PF10_0107	PB301233.00.0	PB301233.00.0 + PY02086	PY02086	hypothetical protein	hypothetical
PF10_0108	PB000114.00.0	PY02715	PY02715	hypothetical protein	hypothetical
PF10_0118	PB000672.00.0	PB000672.00.0 + PY03641	PY03641	hypothetical protein	hypothetical
PF10_0119	PY03476	PC001202.02.0	PY03476	hypothetical protein	hypothetical
PF10_0121	PB001303.00.0	PC001328.02.0	PB001303.00.0	hypoxanthine phosphoribosyltransferase	biosynthesis
PF10_0141	PB001067.02.0	PC05139	PB001067.02.0 + PY05139	cdk1, putative	kinase
PF10_0147	PY07439	-	PY07439	FAD synthetase, putative	biosynthesis
PF10_0148	PY07440	-	PY07440	hypothetical protein	hypothetical
PF10_0157	PY01540	PC101811.00.0	PY01540	hypothetical protein	hypothetical
PF10_0170	PY02639	PC000204.05.0	PY02639	hypothetical protein	hypothetical
PF10_0172	PY02641	PY02641	PY02641	hypothetical protein	hypothetical
PF10_0207	PB000228.02.0	PC000381.00.0	PB000228.02.0	hypothetical protein	hypothetical
PF10_0253	PB000356.01.0	PB000356.01.0 + PY03822	PY03822	hypothetical protein	hypothetical
PF10_0255	PB000066.00.1	PB000066.00.1 + PY03820	PY03820	hypothetical protein	hypothetical
PF10_0260	PB000688.02.0	PB000688.02.0	PB000688.02.0	hypothetical protein	hypothetical
PF10_0308	PB000440.01.0	PB000440.01.0 + PY05983	PY05983	hypothetical protein	hypothetical
PF10_0315	PB000353.01.0	PC000591.00.0	PB000353.01.0	hypothetical protein	hypothetical
PF10_0359	PB001110.00.01.0	PY01011	PY01011	hypothetical protein	hypothetical
PF10_0370	PB000485.00.0	PB000485.00.0 + PY04075	PY04075	hypothetical protein	hypothetical
PF10_0371	PB000878.01.0	PB000878.01.0	PY03525	hypothetical protein	hypothetical
PF10_0415	-	PY04876	PY04876	hypothetical protein, conserved	hypothetical
PF10_0374	-	PY02057	PY02057	PI 11-1 protein	peptidase
PF11_0043	PB000918.01.0	PB000918.01.0 + PY06972	PY06972	60S acidic ribosomal protein p1, putative	ribosomal
PF11_0045	PB000621.01.0	PC000340.02.0	PB000621.01.0 + PC000340.02.0	hypothetical protein	hypothetical
PF11_0062	PB300781.00.0	PC302253.00.0	PB300781.00.0	histone H2B	DNA-binding
PF11_0090	PY07549	PC300579.00.0	PY07549	hypothetical protein	hypothetical
PF11_0096	PB001001.01.0	PB001001.01.0 + PY05048	PY05048	casein kinase II, alpha subunit, putative	kinase
PF11_0106	PB000275.02.0	PB000275.02.0 + PY06787	PY06787	60S Ribosomal protein L36, putative	ribosomal
PF11_0110	PB405942.00.0	PY03703	PY03703	hypothetical protein	hypothetical
PF11_0119	PB000991.03.0	PY02704	PY02704	hypothetical protein	hypothetical
PF11_0130	PY04594	PY04594	PY04594	hypothetical protein	hypothetical
PF11_0134	PB001144.01.0	PB001144.01.0 + PY05482	PY05482	hypothetical protein	hypothetical
PF11_0152	PY00726	PY00726	PY00726	hypothetical protein	hypothetical
PF11_0155	PY00723	-	PY00723	hypothetical protein	hypothetical
PF11_0159	PB000984.00.0	PB000984.00.0	PB000984.00.0	hypothetical protein	hypothetical
PF11_0185	PC000723.03.0 + PY05215	PC000723.03.0	PY05215	hypothetical protein	hypothetical
PF11_0193	PY00910	PC104903.00.0	PY00910	hypothetical protein	hypothetical
PF11_0195	PB000134.01.0	PC000119.02.0	PB000134.01.0 + PC000119.02.0	hypothetical protein	hypothetical
PF11_0199	PY04676	PC000855.02.0	PY04676	hypothetical protein	hypothetical
PF11_0207	PB000140.01.0	PC000291.00.0	PB000140.01.0	hypothetical protein	hypothetical
PF11_0209	PY00688	PY00688	PY00688	hypothetical protein	hypothetical
PF11_0218	PB000157.00.0	PC000688.00.0	PB000157.00.0	hypothetical protein	hypothetical
PF11_0224	PY04421	PY04421	PY04421	circumsporozoite-related antigen	invasion
PF11_0225	PB000693.01.0	PB000693.01.0 + PY06050	PY06050	ABC transporter	ABC transporter
PF11_0230	PB000252.03.0	PB000252.03.0	PB000252.03.0	hypothetical protein	hypothetical
PF11_0232	PB000262.00.0	PC301868.00.0	PB000262.00.0	hypothetical protein	hypothetical
PF11_0236	PB000025.00.0	PY03228	PY03228	hypothetical protein	hypothetical
PF11_0241	PB000932.02.0	PC000604.04.0	PB000932.02.0	hypothetical protein	hypothetical
PF11_0244	PB000994.01.0	PC000047.05.0	PC000047.05.0	hypothetical protein	hypothetical
PF11_0255	PB000295.00.0	PC000207.04.0	PB000295.00.0	hypothetical protein	hypothetical
PF11_0281	PB000833.03.0	PC000167.02.0	PB000833.03.0	hypothetical protein	hypothetical
PF11_0296	PB000971.03.0	PC000402.00.0	PB000971.03.0	hypothetical protein	hypothetical
PF11_0299	PY08470	-	PY08470	hypothetical protein	hypothetical
PF11_0300	PB000516.00.0	PB000516.00.0	PB000516.00.0	hypothetical protein	hypothetical
PF11_0309	PB000732.00.0	PB000732.00.0 + PY05171	PY05171	hypothetical protein	hypothetical
PF11_0320	PY01167	-	PY01167	hypothetical protein	hypothetical
PF11_0325	PB000034.01.0	PB000034.01.0	PB000034.01.0	hypothetical protein	hypothetical
PF11_0332	PY07197	PC000261.01.0	PY07197	hypothetical protein	hypothetical
PF11_0334	PB000335.00.0	PC000335.00.0	PY05954	hypothetical protein	hypothetical
PF11_0337	PB300861.00.0	PY05952	PY05952	60S ribosomal protein L2, putative	ribosomal
PF11_0338	PB300862.00.0	PY05950	PY05950	Aquaglyceroporin	transport
PF11_0343	PB000822.01.0	PB000822.01.0	PB000822.01.0	hypothetical protein, conserved	hypothetical
PF11_0354	PY01923	-	PY01923	hypothetical protein	hypothetical
PF11_0356	PB000426.02.0	PC000460.01.0	PB000426.02.0 + PC000460.01.0	hypothetical protein	hypothetical
PF11_0360	PB000427.02.0	PB000427.02.0 + PY00472	PY00472	hypothetical protein	hypothetical
PF11_0370	PB001092.02.0	PC000427.01.0	PB001092.02.0	hypothetical protein	hypothetical
PF11_0387	PC000614.02.0	PC000614.02.0	PC000614.02.0	hypothetical protein	hypothetical
PF11_0388	PB000421.01.0	PC000613.02.0	PB000421.01.0	hypothetical protein	hypothetical
PF11_0390	PB300319.00.0	PY05022	PY05022	reticulocyte binding protein, putative	invasion
PF11_0398	PB000269.00.0	PC000208.05.0	PB000269.00.0	hypothetical protein	hypothetical
PF11_0406	PY02069	PY02069	PY02069	hypothetical protein	hypothetical
PF11_0412	PB000656.03.0	PC301156.00.0	PB000656.03.0	Vacuolar ATP synthase subunit F, putative	proton transport
PF11_0417	PY04791-PY04790	PY04791-PY04790	PY04791-PY04790	hypothetical protein	hypothetical
PF11_0427	PY00944	-	PY00944	dolichyl-phosphate b-D-mannosyltransferase, putative	biosynthesis
PF11_0428	PY03414	PC000657.02.0	PY03414	hypothetical protein, conserved	hypothetical
PF11_0434	PB000035.03.0	PY05499	PY05499	hypothetical protein	hypothetical
PF11_0436	PB000335.03.0	PB000335.03.0	PB000335.03.0	hypothetical protein	hypothetical
PF11_0454	PB000999.00.0	PC00108.02.0	PB000999.00.0	Ribosomal protein, 40S subunit, putative	ribosomal
PF11_0455	PY06955	PY06955	PY06955	hypothetical protein	hypothetical
PF11_0457	PB000664.03.0	PC000199.02.0	PB000664.03.0 + PC000199.02.0	hypothetical protein	hypothetical
PF11_0462	PY00720	PY00720	PY00720	hypothetical protein	hypothetical
PF11_0465	PB000159.03.0	PC001110.02.0	PB000159.03.0	dynamitin-like protein	GTP-binding
PF11_0470	PY02596-PY02597	PY02596-PY02597	PY02596-PY02597	hypothetical protein	hypothetical
PF11_0477	PY05259	PC001178.02.0	PY05259	CCAAT-box DNA binding protein subunit B	DNA-binding
PF11_0479	PB000656.02.0	PC301222.00.0	PC301222.00.0	hypothetical protein	hypothetical
PF11_0484	PB000802.02.0	PC300068.00.0	PB000802.02.0 + PC300068.00.0	hypothetical protein	hypothetical
PF11_0488	PB000797.02.0	PC000272.02.0	PB000797.02.0 + PC000272.02.0	hypothetical protein	hypothetical
PF11_0504	-	PY02244	PY02244	hypothetical protein	hypothetical
PF0095c	PB300929.00.0	PY04196	PY04196	hypothetical protein	hypothetical
PF0120c	PY01525	PC000004.00.0	PY01525	cyclophilin, putative	protein folding
PF0155c	-	PY05284	PY05284	hypothetical protein	hypothetical
PF0175c	PB000302.02.0	PC301176.00.0	PB000302.02.0 + PC301176.00.0	hypothetical protein, conserved	hypothetical
PF0190w	PB000336.03.0	PC000554.00.0	PB000336.03.0	ubiquitin-conjugating enzyme e2, putative	ubiquitin-like
PF0225c	PB000698.03.0	PC000672.01.0	PB000698.03.0 + PC000672.01.0	hypothetical protein	hypothetical
PF0300c	PC001060.02.0	PC001060.02.0	PC001060.02.0	phosphoesterase, putative	biosynthesis

Appendix A: Identification of syntenic orthologs by microarray

PfGene	PbGene	PcGene	PvGene	Pf Product	Pf function
PFL0340w	PY02726	PY02726	PY02726	hypothetical protein	hypothetical
PFL0350c	PB300899.00.0	-	PB300899.00.0	hypothetical protein	hypothetical
PFL0355c	PB000082.02.0	PC000440.02.0	PB000082.02.0	hypothetical protein	hypothetical
PFL0415w	PB301127.00.0	PB301127.00.0	PB301127.00.0	acyl carrier protein, mitochondrial precursor, putative	co-factor binding
PFL0470w	PB000830.01.0	PC000180.04.0	PB000830.01.0	hypothetical protein	hypothetical
PFL0485w	PY01828	PY01828	PY01828 (exon 3-5)	hypothetical protein	hypothetical
PFL0500w	PY03485	PC000359.03.0	PY03485	50S ribosomal protein L1, putative	ribosomal
PFL0515w	PB000770.00.0	PB000770.00.0 + PY07010	PY07010	hypothetical protein	hypothetical
PFL0560c	PB000046.02.0	PC000251.04.0	PC000251.04.0	minichromosome maintenance protein, putative	DNA-binding
PFL0585w	PY03971	PY03971	PY03971	PfPUB Plasmodium falciparum polyubiquitin	ubiquitin-like
PFL0615w	PB001413.02.0	PC000380.00.0	PB001413.02.0	hypothetical protein	hypothetical
PFL0750w	-	PY02275	PY02275	hypothetical protein	hypothetical
PFL0765w	PB001093.01.0	PC300772.00.0	PB001093.01.0 + PC300772.00.0	hypothetical protein	hypothetical
PFL0780w	PB001193.02.0	PB001193.02.0 + PY05585	PY05585	glycerol-3-phosphate dehydrogenase, putative	catabolism
PFL0860c	PB001085.02.0	PY04730	PY04730	hypothetical protein, conserved	hypothetical
PFL0875w	PB001084.02.0	PB001084.02.0 + PY04840	PY04840	hypothetical protein	hypothetical
PFL0895c	PB000406.00.0	PC000456.04.0	PB000406.00.0	hypothetical protein	hypothetical
PFL0910c	PC000648.00.0	PC000648.00.0	PC000648.00.0	hypothetical protein	hypothetical
PFL0925w	PB000621.03.0	PC000698.03.0	PB000621.03.0	Formin 2, putative	actin-binding
PFL0975w	PB001263.02.0	PC000706.04.0	PC000706.04.0	hypothetical protein	hypothetical
PFL0990w	PB000050.01.0	PC301490.00.0	PB000050.01.0 + PC301490.00.0	hypothetical protein	hypothetical
PFL1065c	PB000042.01.0	PB000042.01.0 + PY00581	PY00581	hypothetical protein	hypothetical
PFL1085w	PY03447	PY03447	PY03447	hypothetical protein	hypothetical
PFL1100w	PY03449	PC301101.00.0	PY03449	hypothetical protein	hypothetical
PFL1105c	PB300267.00.0	PB300267.00.0 + PY03450	PY03450	hypothetical protein	hypothetical
PFL1125w	PB001045.01.0	PC001079.02.0	PB001045.01.0 + PC001079.02.0	phospholipid-transporting ATPase, putative	transport
PFL1185c	PB000820.01.0	PC300668.00.0	PB000820.01.0	cytochrome c heme lyase, putative	biosynthesis
PFL1190c	PB000275.03.0	PC000685.01.0	PC000685.01.0	hypothetical protein	hypothetical
PFL1280w	PB300370.00.0	PC000841.02.0	PB300370.00.0	hypothetical protein	hypothetical
PFL1295w	PB300054.00.0	PY06998	PY06998	hypothetical protein	hypothetical
PFL1330c	PB106171.00.0	PC000056.03.0	PB106171.00.0	hypothetical protein	hypothetical
PFL1340c	PB000588.03.0	PB000588.03.0 + PY00537	PY00537	hypothetical protein	hypothetical
PFL1345c	PB000295.03.0	PC300268.00.0	PB000295.03.0	hypothetical protein, conserved	hypothetical
PFL1350w	PB001394.02.0	PB001394.02.0 + PY04000	PY04000	RNA pseudouridylate synthase, putative	RNA-binding
PFL1370w	PB001392.02.0	PC000368.02.0	PB001392.02.0 + PC000368.02.0	NIMA-related protein kinase (Pfkex-1)	kinase
PFL1380w	PB001089.01.0	PC000498.00.0	PB001089.01.0 + PC000498.00.0	hypothetical protein	hypothetical
PFL1410c	-	PY05035	PY05035	hypothetical protein	hypothetical
PFL1430c	PB000347.03.0	PC000466.01.0	PB000347.03.0 + PC000466.01.0	hypothetical protein	hypothetical
PFL1435c	PB000458.01.0	PC302159.00.0	PC302159.00.0	myosin d	motor activity
PFL1440c	PB000160.02.0	PY02515	PY02515	hypothetical protein	hypothetical
PFL1480w	PB001149.01.0	PC000536.02.0	PB001149.01.0	hypothetical protein	hypothetical
PFL1490w	PB000310.01.0	PY01572	PY01572	hypothetical protein	hypothetical
PFL1495w	PB000566.00.0	PC000443.03.0	PB000566.00.0	hypothetical protein	hypothetical
PFL1505c	PB001104.00.0	PB001104.00.0 + PY05255	PY05255	hypothetical protein	hypothetical
PFL1540c	PB001599.02.0	PC300155.00.0	PB001599.02.0	phenylalanyl-tRNA synthetase alpha chain, putative	biosynthesis
PFL1565c	PC000122.00.0 + PY04962	PC000122.00.0	PY04962	hypothetical protein	hypothetical
PFL1570w	PY04702	PY04702	PY04702	hypothetical protein	hypothetical
PFL1575w	PY04703	PC000259.04.0	PY04703	hypothetical protein	hypothetical
PFL1585c	PC000548.00.0 + PY04705	PC000548.00.0	PY04705	hypothetical protein	catabolism
PFL1620w	PB000094.01.0	PC000730.00.0	PB000094.01.0	asparagine/aspartate rich protein, putative	catabolism
PFL1670c	PB001238.02.0	PC001037.02.0	PB001238.02.0 + PC001037.02.0	hypothetical protein	hypothetical
PFL1685w	PB001294.00.0	PC000017.02.0	PB001294.00.0	hypothetical protein, conserved	hypothetical
PFL1695c	PY00509	PY00509	PY00509	hypothetical protein	hypothetical
PFL1740w	PB000820.00.0	PC000921.01.0	PB000820.00.0	hypothetical protein	hypothetical
PFL1750c	PB301293.00.0	PB301293.00.0 + PY04611	PY04611	hypothetical protein	hypothetical
PFL1830w	PB000920.02.0	PB000920.02.0 + PY06022	PY06022	ubiquitin-like protein, putative	ubiquitin-like
PFL1880w	PB001116.03.0	PC000662.04.0	PB001116.03.0	long-chain-fatty-acid-CoA ligase, putative	biosynthesis
PFL1910c	PB300903.00.0	PC300193.00.0	PB300903.00.0 + PC300193.00.0	hypothetical protein	hypothetical
PFL1940w	PB001039.02.0	PY02220	PY02220	3-hydroxyisobutyryl-coenzyme A hydrolase, putative	hydrolase
PFL2025w	PY01489	PY01489	PY01489	hypothetical protein	hypothetical
PFL2035c	PB106974.00.0	PC108254.00.0	PB106974.00.0	hypothetical protein	hypothetical
PFL2040w	PB000304.03.0	PB000304.03.0	PY01486	hypothetical protein	hypothetical
PFL2080c	PY01277	-	PY01277	hypothetical protein	hypothetical
PFL2130w	PB301277.00.0	PC301883.00.0	PC301883.00.0	hypothetical protein	hypothetical
PFL2135c	PB000587.02.0	PC001139.02.0	PB000587.02.0 + PC001139.02.0	hypothetical protein	hypothetical
PFL2170c	-	PC000684.02.0	PC000684.02.0	hypothetical protein	hypothetical
PFL2175w	PC000397.01.0	PC000397.01.0	PC000397.01.0	ubiquitin-conjugating enzyme e2, putative	ubiquitin-like
PFL2215w	PB000323.01.0	PB000323.01.0 + PY02240	PY02240	actin	structural
PFL2280w	PB000589.03.0	PC000214.03.0	PB000589.03.0	cyclin g-associated kinase, putative	kinase
PFL2290w	PB000590.03.0	PC000458.02.0	PB000590.03.0	preprocathepsin c precursor, putative	protease
PFL2320w	PB000245.02.0	PC000594.03.0	PB000245.02.0 + PC000594.03.0	hypothetical protein	hypothetical
PFL2350c	PB300475.00.0	PY02898	PY02898	hypothetical protein	hypothetical
PFL2360w	PY02900	-	PY02900	hypothetical protein	hypothetical
PFL2380c	PB000486.00.0	PC000594.02.0	PC000594.02.0	hypothetical protein	hypothetical
PFL2395c	PB000758.00.0	PC000337.00.0	PB000758.00.0	dimethyladenosine transferase, putative	biosynthesis
PFL2410w	PB000013.01.0	PC000869.02.0	PB000013.01.0	hypothetical protein	hypothetical
PFL2420w	PB300567.00.0	PC001038.02.0	PC001038.02.0	hypothetical protein	hypothetical
PFL2425w	PB000561.03.0	PC000725.04.0	PB000561.03.0	adaptor-related protein complex 3, sigma 2 subunit, putative	transport
PFL2470c	PY01335	PC000763.04.0	PY01335	hypothetical protein	hypothetical
PFL2490c	PC300131.00.0 + PY01704	PC300131.00.0	PY01704	hypothetical protein	hypothetical
PFL2520w	PB103906.00.0	PB103906.00.0 + PY01185	PY01185	reticulocyte-binding protein, putative	invasion
MAL13P1_60	PB000282.01.0	PY04764	PY04764	erythrocyte binding antigen 140	invasion
MAL13P1_20	PB001007.00.0	PC000410.03.0	PC000410.03.0	hypothetical protein	hypothetical
PF13_0023	PB000758.01.0	PC000833.03.0	PB000758.01.0 + PC000833.03.0	DNA-directed RNA polymerase 2, putative	DNA-binding
PF13_0026	PB000974.00.0	PB000974.00.0 + PY02188	PY02188	hypothetical protein	hypothetical
MAL13P1_26	PB000972.00.0	PB000972.00.0 + PY02189	PY02189	hypothetical protein	hypothetical
MAL13P1_28	PB000286.03.0	PB000286.03.0	PB000286.03.0	hypothetical protein	hypothetical
MAL13P1_29	PB106857.00.0	PB106857.00.0 + PY04953	PY04953	hypothetical protein	hypothetical
MAL13P1_31	PB000684.03.0	PB000684.03.0 + PY02588	PY02588	hypothetical protein	hypothetical
PF13_0033	PC000237.01.0 + PY02590	PC000237.01.0	PY02590	26S proteasome regulatory subunit, putative	protease
PF13_0037	PB000880.02.0	PB000880.02.0 + PY00037	PY00037	DEAD box helicase, putative	RNA-binding
PF13_0039	PY00034	PY00034	PY00034	hypothetical protein	hypothetical
MAL13P1_40	PB000719.00.0	PY04905	PY04905	hypothetical protein	hypothetical
MAL13P1_43	PY01749	PY01749	PY01749	hypothetical protein	hypothetical
PF13_0054	PB000940.03.0	PB000940.03.0 + PY03108	PY03108	hypothetical protein	hypothetical
MAL13P1_51	PY02796	PY02796	PY02796	Rab5B protein	GTP-binding
PF13_0067	PB000199.03.0	PC000047.01.0	PB000199.03.0	hypothetical protein	hypothetical
MAL13P1_57	PB001064.02.0	PB001064.02.0 + PY02335	PY02335	hypothetical protein	hypothetical
PF13_0072	PB000679.00.0	PC106856.00.0	PB000679.00.0	hypothetical protein	hypothetical
MAL13P1_64	PB000080.01.0	PC300321.00.0	PB000080.01.0	ubiquitin-like protein nedd8 homologue, putative	ubiquitin-like
PF13_0079	PB000440.02.0	PY02835	PY02835	hypothetical protein	hypothetical
MAL13P1_69	PB001008.01.0	PB001008.01.0 + PY02834	PY02834	septum formation protein MAF homologue, putative	inhibitor of septum formation
MAL13P1_71	PY01883	PY01883	PY01883	hypothetical protein	hypothetical
MAL13P1_77	PB000729.03.0	PC000614.01.0	PB000729.03.0	hypothetical protein	hypothetical
MAL13P1_80	PY04587	PY04587	PY04587	hypothetical protein	hypothetical
PF13_0089	PB000565.02.0	PC000666.03.0	PB000565.02.0 + PC000666.03.0	hypothetical protein	hypothetical
PF13_0090	PB000767.00.0	PB000767.00.0	PY02241	ADP-ribosylation factor, putative	GTP-binding
MAL13P1_86	PY02436	PY02436	PY02436	cholinephosphate cytidyltransferase	biosynthesis
MAL13P1_88	PY03408	PY03408	PY03408	hypothetical protein	hypothetical
PF13_0098	PB001375.02.0	PB001375.02.0 + PY00193	PY00193	hypothetical protein	hypothetical
PF13_0104	PY03274	PY03274	PY03274	hypothetical protein	hypothetical
MAL13P1_105	PB000997.03.0	PB000997.03.0 + PY01941	PY01941	hypothetical protein	hypothetical
PF13_0117	PY03871	PY03871	PY03871	hypothetical protein, conserved	hypothetical
MAL13P1_112	PY07138	-	PY07138	hypothetical protein, conserved	hypothetical
MAL13P1_113	PY07140	PY07140	PY07140	hypothetical protein	hypothetical
PF13_0121	PB000860.02.0	PC000907.01.0	PB000860.02.0 + PC000907.01.0	dihydrolipoamide succinyltransferase, putative	biosynthesis
MAL13P1_118	PB000520.03.0	PB000520.03.0 + PY03988	PY03988	cAMP-specific 3',5'-cyclic phosphodiesterase 4D, putative	hydrolase
PF13_0125	PB000367.00.0	PC000036.00.0	PB000367.00.0 + PC000036.00.0	hypothetical protein	hypothetical
MAL13P1_128	PY01584	PY01584	PY01584	hypothetical protein	hypothetical
PF13_0137	PB000665.00.0	PC001093.02.0	PB000665.00.0	hypothetical protein	hypothetical
PF13_0138	PY02716	PC000431.01.0	PY02716	hypothetical protein, conserved	hypothetical

Appendix A: Identification of syntenic orthologs by microarray

PfGene	PbGene	PcGene	PvGene	Pf Product	Pf function
MAL13P1.145		PC000819.01.0	PC000819.01.0	hypothetical protein	hypothetical
MAL13P1.153	PB000362.02.0	PB000362.02.0	PY00262	hypothetical protein	hypothetical
PF13.0159	PY06728	-	PY06728	predicted nucleotidyltransferase, putative	biosynthesis
PF13.0158	PB000277.00.0	PB000277.00.0 + PY06727	PY06727	hypothetical protein	hypothetical
MAL13P1.151	PB000780.00.0	PC302386.00.0	PB000780.00.0	hypothetical protein	hypothetical
PF13.0155	PB001453.02.0	PC301179.00.0	PB001453.02.0	hypothetical protein	hypothetical
MAL13P1.150	PB001120.00.0	PC000231.05.0	PB001120.00.0 + PC000231.05.0	hypothetical protein	hypothetical
MAL13P1.157	PB001450.02.0	PC000086.04.0	PC000086.04.0	hypothetical protein	hypothetical
MAL13P1.160	PB000970.02.0	PB000970.02.0 + PY04180	PY04180	unknown	unknown
PF13.0165	PY02209	PY02209	PY02209	hypothetical protein	hypothetical
PF13.0166	PY02207	PY02207	PY02207	protein kinase, putative	kinase
PF13.0167	PY00481	PY00481	PY00481	hypothetical protein	hypothetical
MAL13P1.158	PB001246.00.0	PC001216.02.0	PC001216.02.0	hypothetical protein	hypothetical
PF13.0170	PY02178	PY02178	PY02178	glutaminyl-tRNA synthetase, putative	protein synthesis
PF13.0171	PY07409	PY07409	PY07409	60S ribosomal protein L23, putative	ribosomal
PF13.0172	PY02568	PY02568	PY02568	hypothetical protein	hypothetical
PF13.0173	PY02567	PY02567	PY02567	hypothetical protein	hypothetical
PF13.0175	PY05726	PY05726	PY05726	hypothetical protein, conserved	hypothetical
PF13.0176	PY05725	PY05725	PY05725	apurinic/aprimidinic endonuclease Apn1	endonuclease
PF13.0177	PY05724	PY05724	PY05724	ATP-dependent RNA helicase, putative	RNA-binding
PF13.0178	PY01848	PY01848	PY01848	translation initiation factor 6, putative	protein synthesis
PF13.0179	PY01849	PY01849	PY01849	isoleucine-tRNA ligase, putative	protein synthesis
PF13.0180	PY01850	PY01850	PY01850	chaperonin, putative	protein folding
PF13.0189	PB0003161.02.0	PC302461.00.0	PB0003161.02.0	hypothetical protein	hypothetical
MAL13P1.176	PB104017.00.0	PB104017.00.0 + PY03046	PC3046	unknown	unknown
PF13.0202	PY01507	PY01507	PY01507	hypothetical protein	hypothetical
MAL13P1.185	-	PY04971	PY04971	CDK-related protein kinase 6	kinase
MAL13P1.186	-	PY04970	PY04970	1-deoxy-D-xylulose 5-phosphate synthase	biosynthesis
PF13.0210	PB000357.02.0	PC000956.01.0	PB000357.02.0 + PC000956.01.0	hypothetical protein	hypothetical
PF13.0211	PY02109	PC001360.02.0	PY02109	calcium-dependent protein kinase	kinase
MAL13P1.202	PY00593	PY00593	PY00593	hypothetical protein	hypothetical
PF13.0217	PB300284.00.0	PB300284.00.0	PB300284.00.0	hypothetical protein	hypothetical
PF13.0218	PB000651.03.0	PC000139.05.0	PB000651.03.0	ABC transporter, putative	transport
PF13.0221	PY04558	PC000264.02.0	PY04558	hypothetical protein	hypothetical
PF13.0228	PB000942.03.0	PC000692.04.0	PB000942.03.0	40S ribosomal subunit protein S6, putative	ribosomal
PF13.0230	PB000438.02.0	PC301424.00.0	PB000438.02.0 + PC301424.00.0	hypothetical protein	hypothetical
MAL13P1.221	PY06210	PY06210	PY06210	aspartate carbamoyltransferase	biosynthesis
MAL13P1.223	PY00585	PY00585	PY00585 (exon1-2)	hypothetical protein	hypothetical
MAL13P1.231	PY02510	PY02510	PY02510	Pfsec61	secretion
PF13.0249	PB000150.00.0	PC000734.03.0	PC000734.03.0	hypothetical protein	hypothetical
MAL13P1.242	PB001275.00.0	PB001275.00.0 + PY02626	PY02626	hypothetical protein	hypothetical
MAL13P1.246	PB000755.03.0	PC000695.02.0	PB000755.03.0 + PC000695.02.0	hypothetical protein	hypothetical
PF13.0256	PB300425.00.0	PB300425.00.0 + PY02889	PY02889	hypothetical protein	hypothetical
MAL13P1.248	PY02890	PC000053.01.0	PY02890	nucleoside diphosphate hydrolase	hydrolase
PF13.0263	PY05605	PC000750.02.0	PY05605	hypothetical protein	hypothetical
PF13.0265	PB000345.03.0	PC000748.02.0	PB000345.03.0	hypothetical protein	hypothetical
MAL13P1.255	PB301033.00.0	PY06977	PY06977	hypothetical protein	hypothetical
PF13.0266	PY06978	PC000456.00.0	PY06978	hypothetical protein	hypothetical
MAL13P1.257	PB000122.01.0	PC301297.00.0	PB000122.01.0	hypothetical protein, conserved	hypothetical
MAL13P1.258	PB001126.01.0	PC000386.00.0	PB001126.01.0	hypothetical protein	hypothetical
MAL13P1.261	PC000221.03.0	PC000221.03.0	PC000221.03.0	hypothetical protein	hypothetical
PF13.0274	PB001638.02.0	PC000569.01.0	PB001638.02.0 + PC000569.01.0	hypothetical protein, conserved	hypothetical
PF13.0280	PB001048.00.0	PC000489.02.0	PB001048.00.0	ER lumen protein retaining receptor	protein retention in ER
MAL13P1.274	PY04697	PY04697	PY04697	serine/threonine protein phosphatase pPp5	phosphatase
PF13.0297	PC000763.02.0	PC000763.02.0	PC000763.02.0	hypothetical protein, conserved	hypothetical
MAL13P1.286	-	PY01567	PY01567	hypothetical protein	hypothetical
MAL13P1.287	PY01181	PY01181	PY01181	hypothetical protein	hypothetical
MAL13P1.290	PB000691.03.0	PC301822.00.0	PC301822.00.0	high mobility group protein 4, putative	DNA-binding
MAL13P1.289	PB0000661.03.0	PC000066.01.0	PC000066.01.0	mitotic control protein disc3 homologue, putative	cell-cycle
MAL13P1.297	PB000284.03.0	PB000284.03.0	PB000284.03.0	ADP-ribosylation factor, putative	signaling
PF13.0319	PY00950	PY00950	PY00950	hypothetical protein	hypothetical
MAL13P1.301	PY00951-PY00952	PY00951-PY00952	PY00951-PY00952	guanylyl cyclase	biosynthesis
PF13.0321	PC300159.00.0 + PY01090	PY02725	PY02725	hypothetical protein	hypothetical
MAL13P1.307	PY01759	PC300159.00.0	PY01759	hypothetical protein, conserved	hypothetical
MAL13P1.308	PY01758	-	PY01758	proliferating cell nuclear antigen	DNA-binding
PF13.0328	PY06706	PY06706	PY06706	hypothetical protein	hypothetical
PF13.0329	PY06706	PY06706	PY06706	hypothetical protein	hypothetical
PF13.0330	PB000877.01.0	PC000398.03.0	PB000877.01.0 + PC000398.03.0	ATP-dependent DNA helicase, putative	DNA-binding
MAL13P1.313	PY00918	PY00918	PY00918	hypothetical protein	hypothetical
MAL13P1.315	PB000357.03.0	PB000357.03.0	PB000357.03.0	hypothetical protein	hypothetical
PF13.0334	PB000811.03.0	PB000811.03.0 + PY05803	PY05803	polynucleotide kinase, putative	kinase
MAL13P1.318	PB000644.02.0	PB000644.02.0	PB000644.02.0	hypothetical protein	hypothetical
MAL13P1.320	PB000081.00.0	PC000477.01.0	PB000081.00.0 + PC000477.01.0	hypothetical protein	hypothetical
PF13.0342	-	PC000649.02.0	PC000649.02.0	hypothetical protein	hypothetical
PF13.0347	PY03338	PY03338	PY03338	hypothetical protein	hypothetical
MAL13P1.331	PY05033	PY05033	PY05033	hypothetical protein, conserved	hypothetical
PF13.0353	PY03083	PC000358.03.0	PY03083	NADH-cytochrome b5 reductase, putative	electron transport
MAL13P1.342	PB300696.00.0	PB300696.00.0 + PY03079	PY03079	hypothetical protein	hypothetical
MAL13P1.348	PB000170.00.0	PC000729.00.0	PB000170.00.0	hypothetical protein	hypothetical
PF13.0360	PC000007.05.0	PC000007.05.0	PC000007.05.0	hypothetical protein	hypothetical
MAL13P1.349	PB000738.03.0	PB000738.03.0 + PY06440	PY06440	hypothetical protein	hypothetical
MAL13P1.352	PB000733.00.0	PC300446.00.0	PC300446.00.0	hypothetical protein	hypothetical
PF14.0021	PB001105.03.0	PB001105.03.0	PB001105.03.0	hypothetical protein	hypothetical
PF14.0032	PB000021.02.0	PB000021.02.0 + PY05274	PY05274	hypothetical protein	hypothetical
PF14.0054	PY06983	PC000642.00.0	PY06983	hypothetical protein, conserved	hypothetical
PF14.0056	PB001563.02.0	PC000409.00.0	PB001563.02.0	hypothetical protein	hypothetical
PF14.0062	PB000977.01.0	PC000432.02.0	PB000977.01.0	hypothetical protein	hypothetical
PF14.0068	PB000893.01.0	PC000051.00.0	PB000893.01.0 + PC000051.00.0	fibrillarin, putative	RNA-binding
PF14.0069	PY01070	PC000053.00.0	PY01070	hypothetical protein	hypothetical
PF14.0072	PY05265	PY05265	PY05265	hypothetical protein, conserved	hypothetical
PF14.0073	PY05266-PY05267	PY05266-PY05267	PY05266-PY05267	hypothetical protein	hypothetical
PF14.0081	PB000359.00.0	PB000359.00.0 + PY06657	PY06657	DNA repair helicase, putative	RNA-binding
PF14.0085	PB000852.00.0	PB000852.00.0 + PY01375	PY01375	hypothetical protein	hypothetical
PF14.0089	PY03352	PY03352	PY03352	hypothetical protein	hypothetical
PF14.0094	PB000902.03.0	PC000246.03.0	PB000902.03.0	hypothetical protein	hypothetical
PF14.0106	PB000077.00.0	PB000077.00.0 + PY04030	PY04030	hypothetical protein	hypothetical
PF14.0110	PB000753.00.0	PB000753.00.0 + PY00165	PY00165	hypothetical protein	hypothetical
PF14.0133	PB000224.00.0	PC000302.05.0	PB000224.00.0	ATP-dependent transporter, putative	transport
PF14.0135	-	PY01402	PY01402	hypothetical protein	hypothetical
PF14.0139	PB000548.02.0	PC000162.04.0	PB000548.02.0 + PC000162.04.0	hypothetical protein	hypothetical
PF14.0142	PB000102.01.0	PB000102.01.0 + PY07156	PY07156	serine/threonine protein phosphatase, putative	phosphatase
PF14.0143	PB001658.02.0	PC001012.02.0	PB001658.02.0	hypothetical protein	hypothetical
PF14.0145	PB000468.01.0	PC000088.01.0	PB000088.01.0	hypothetical protein	hypothetical
PF14.0152	PY00348	-	PY00348	hypothetical protein	hypothetical
PF14.0160	PB000564.02.0	PC000375.02.0	PB000564.02.0 + PC000375.02.0	hypothetical protein	hypothetical
PF14.0170	PB000517.00.0	PC302651.00.0	PB000517.00.0 + PC302651.00.0	hypothetical protein	hypothetical
PF14.0171	PB300976.00.0	PB300976.00.0 + PY03056	PY03056	hypothetical protein	hypothetical
PF14.0185	PB001045.03.0	PC000921.02.0	PB001045.03.0 + PC000921.02.0	ATP-dependent RNA helicase, putative	RNA-binding
PF14.0188	PY05087	PC000112.04.0	PY05087	hypothetical protein	hypothetical
PF14.0190	PY01208	PC001000.02.0	PY01208	hypothetical protein	hypothetical
PF14.0195	PY05718	PC000467.02.0	PY05718	hypothetical protein	hypothetical
PF14.0204	PB000263.03.0	PC000465.02.0	PB000263.03.0 + PC000465.02.0	hypothetical protein	hypothetical
PF14.0219	PY00020	PY00020	PY00020 (part)	hypothetical protein	hypothetical
PF14.0226	-	PY05674	PY05674	hypothetical protein	hypothetical
PF14.0234	PB000432.03.0	PC000158.04.0	PB000432.03.0 + PC000158.04.0	DNA-directed DNA polymerase, putative	DNA-binding
PF14.0237	PB001004.03.0	PC000043.02.0	PB001004.03.0 + PC000043.02.0	hypothetical protein	hypothetical
PF14.0240	PB001001.03.0	PB001001.03.0 + PY05142	PY05142	ribosomal protein L21e, putative	ribosomal
PF14.0245	PB000139.02.0	PC000013.02.0	PB000139.02.0	hypothetical protein	hypothetical
PF14.0250	PB000864.01.0	PC300698.00.0	PB000864.01.0	hypothetical protein	hypothetical
PF14.0251	PY07136	PY07136	PY07136	hypothetical protein	hypothetical

Appendix A: Identification of syntenic orthologs by microarray

PfGene	PbGene	PcGene	PvGene	Pf Product	Pf function
PF14_0257	PY06840	-	PV06840	hypothetical protein, conserved	hypothetical
PF14_0260	PB301387.00.0	PB301387.00.0 + PY05760	PY05760	hypothetical protein	hypothetical
PF14_0272	PB000018.03.0	PC000688.04.0	PB000018.03.0 + PC000688.04.0	hypothetical protein	hypothetical
PF14_0273	PB001655.02.0	PC000686.04.0	PB001655.02.0	hypothetical protein	hypothetical
PF14_0290	PB000717.01.0	PC000343.00.0	PB000717.01.0	hypothetical protein	hypothetical
PF14_0318	PB000076.02.0	PC302404.00.0	PB000076.02.0	hypothetical protein	hypothetical
PF14_0319	PB301174.00.0	PC000725.01.0	PC000725.01.0	hypothetical protein	hypothetical
PF14_0320	PB000288.01.0	-	PB000288.01.0	hypothetical protein	hypothetical
PF14_0336	PY05583	PY05583	PY05583	hypothetical protein	hypothetical
PF14_0340	PY00617	PY00617	PY00617	hypothetical protein	hypothetical
PF14_0343	PY00620	PY00620	PY00620	hypothetical protein	hypothetical
PF14_0349	PB000566.01.0	PC000777.04.0	PB000566.01.0	histidine triad protein, putative	nucleotide-binding
PF14_0351	PB001268.00.0	PC302463.00.0	PB001268.00.0 + PC302463.00.0	hypothetical protein	hypothetical
PF14_0353	PY04927	PC000456.03.0	PY04927	hypothetical protein	hypothetical
PF14_0366	PB000487.03.0	PC000540.00.0	PB000487.03.0 + PC000540.00.0	small subunit DNA primase	DNA-binding
PF14_0377	PB001163.00.0	PB001163.00.0 + PY06882	PY06882	vesicle-associated membrane protein, putative	transport
PF14_0385	PB000139.00.0	PB000139.00.0	PY02582	hypothetical protein	hypothetical
PF14_0389	PB000736.00.0	PC000165.05.0	PB000736.00.0 + PC000165.05.0	hypothetical protein	hypothetical
PF14_0390	PY05640	PY05640	PY05640	hypothetical protein	hypothetical
PF14_0391	PB001103.02.0	PC000766.01.0	PB001103.02.0 + PC000766.01.0	ribosomal protein L1, putative	ribosomal
PF14_0392	PB001104.02.0	PC000765.01.0	PB001104.02.0 + PC000765.01.0	Ser/Thr protein kinase, putative	kinase
PF14_0397	PB0000214.03.0	PC302474.00.0	PB0000214.03.0	hypothetical protein, conserved	hypothetical
PF14_0399	PB001029.00.0	PB001029.00.0 + PY00881	PY00881	ADP-ribosylation-like factor, putative	GTP-binding
PF14_0403	PB000231.02.0	PB000231.02.0 + PY02453	PY02453	protein prenyltransferase alpha subunit, putative	biosynthesis
PF14_0404	PB000385.03.0	PY04986	PY04986	hypothetical protein	hypothetical
PF14_0405	PY05024	PY05024	PY05024	hypothetical protein	hypothetical
PF14_0406	PY05023	-	PY05023	hypothetical protein	hypothetical
PF14_0420	PB300889.00.0	PB300889.00.0 + PY01677	PY01677	hypothetical protein	hypothetical
PF14_0429	PB000381.00.0	PC000197.02.0	PB000381.00.0 + PC000197.02.0	RNA helicase, putative	RNA-binding
PF14_0430	PB000410.02.0	PC000195.02.0	PB000410.02.0 + PC000195.02.0	hypothetical protein	hypothetical
PF14_0431	PB000409.02.0	PC000196.02.0	PB000409.02.0 + PC000196.02.0	serine/threonine kinase-1	kinase
PF14_0434	PB001121.01.0	PC000349.02.0	PB001121.01.0 + PC000349.02.0	hypothetical protein	hypothetical
PF14_0435	PB000829.02.0	PC000293.04.0	PB000829.02.0	hypothetical protein	hypothetical
PF14_0436	PB000433.03.0	PB000433.03.0 + PY01902	PY01902	helicase, truncated, putative	RNA-binding
PF14_0441	PB000132.02.0	PY07062	PY07062	pyruvate dehydrogenase E1 beta subunit, putative	catabolism
PF14_0444	PB000652.00.0	PC000112.02.0	PC000112.02.0	hypothetical protein	hypothetical
PF14_0448	PB001359.02.0	PC000201.02.0	PB001359.02.0 + PC000201.02.0	ribosomal protein S2, putative	ribosomal
PF14_0449	PY06705	PY06705	PY06705	hypothetical protein	hypothetical
PF14_0452	PB000326.02.0	PC000200.00.0	PB000326.02.0	hypothetical protein	hypothetical
PF14_0457	-	PY01105	PY01105	hypothetical protein	hypothetical
PF14_0458	PB000074.01.0	PC001031.02.0	PB000074.01.0 + PC001031.02.0	hypothetical protein	hypothetical
PF14_0464	PY05111	PC001160.02.0	PY05111	hypothetical protein	hypothetical
PF14_0465	PB300099.00.0	PY05110	PY05110	hypothetical protein	hypothetical
PF14_0466	-	PY05109	PY05109	hypothetical protein	hypothetical
PF14_0468	PB000928.00.0	PC000671.00.0	PB000928.00.0 + PC000671.00.0	hypothetical protein	hypothetical
PF14_0471	PB000902.02.0	PC000242.00.0	PB000902.02.0	hypothetical protein	hypothetical
PF14_0474	PY07340	PY07340	PY07340	hypothetical protein	hypothetical
PF14_0475	PY07341	-	PY07341	hypothetical protein	hypothetical
PF14_0485	PB000475.01.0	PY01992	PY01992	hypothetical protein	hypothetical
PF14_0487	PB0000801.00.0	PB0000801.00.0 + PY05354	PY05354	hypothetical protein	hypothetical
PF14_0488	PB000050.02.0	PB000050.02.0 + PY01627	PY01627	hypothetical protein	hypothetical
PF14_0498	PB000620.00.0	PB000620.00.0 + PY06810	PY06810	hypothetical protein	hypothetical
PF14_0499	PB000596.00.0	PB000596.00.0 + PY05178	PY05178	hypothetical protein	hypothetical
PF14_0516	PB000239.02.0	PB000239.02.0	PB000239.02.0	serine/threonine-protein kinase	kinase
PF14_0536	PB000326.00.0	PC000000.04.0	PC000000.04.0	hypothetical protein	hypothetical
PF14_0537	PB000327.00.0	PC300237.00.0	PB000327.00.0	hypothetical protein	hypothetical
PF14_0542	PB001193.00.0	-	PB001193.00.0	hypothetical protein, conserved	hypothetical
PF14_0555	PY00107	PC000527.01.0	PY00107	hypothetical protein	hypothetical
PF14_0559	PB000674.03.0	PB000674.03.0 + PY05450	PY05450	hypothetical protein	hypothetical
PF14_0561	PB000473.01.0	PB000473.01.0 + PY04686	PY04686	hypothetical protein	hypothetical
PF14_0563	PB000472.01.0	PC000054.02.0	PB000472.01.0	DEAD-box RNA helicase, putative	RNA-binding
PF14_0577	PY01753 & PY01754	PY01753 & PY01754	PY01753 & PY01754	hypothetical protein	hypothetical
PF14_0582	-	PY05163	PY05163	hypothetical protein	hypothetical
PF14_0583	PB000810.00.0	PB000810.00.0	PB000810.00.0	hypothetical protein	hypothetical
PF14_0584	PB000754.01.0	PB000754.01.0 + PY04143	PY04143	ribosomal protein S4, putative	ribosomal
PF14_0588	PY00515	PY00515	PY00515	hypothetical protein	hypothetical
PF14_0600	PC000217.05.0 + PY03282	PC000217.05.0	PY03282	hypothetical protein	hypothetical
PF14_0607	PB000015.03.0	PC000384.01.0	PC000384.01.0	hypothetical protein	hypothetical
PF14_0610	PY04485	PC300054.00.0	PY04485	hypothetical protein	hypothetical
PF14_0613	PB001604.02.0	PC000708.03.0	PB001604.02.0 + PC000708.03.0	hypothetical protein	hypothetical
PF14_0617	PB001609.02.0	PC000030.00.0	PB001609.02.0 + PC000030.00.0	hypothetical protein	hypothetical
PF14_0624	PY02005	PY02005	PY02005	hypothetical protein	hypothetical
PF14_0634	PY06249	PC000469.04.0	PY06249	hypothetical protein	hypothetical
PF14_0643	PB000131.00.0	PC301306.00.0	PB000131.00.0	hypothetical protein, conserved	hypothetical
PF14_0650	PB000305.01.0	PY02060	PY02060	hypothetical protein	hypothetical
PF14_0659	PB000694.00.0	PY03644	PY03644	hypothetical protein	hypothetical
PF14_0661	PB300379.00.0	PB300379.00.0 + PY03646	PY03646	hypothetical protein, conserved	hypothetical
PF14_0666	PY01697	PY01697	PY01697	hypothetical protein	hypothetical
PF14_0668	PB000086.02.0	PC000078.00.0	PB000086.02.0	hypothetical protein	hypothetical
PF14_0670	PY06653	-	PY06653	hypothetical protein	hypothetical
PF14_0671	PB000062.02.0	PC300793.00.0	PB000062.02.0	hypothetical protein	hypothetical
PF14_0676	PB000621.00.0	PB000621.00.0 + PY06665	PY06665	20S proteasome beta 4 subunit, putative	protease
PF14_0684	PB108045.00.0-PB108046.00.0	PB108045.00.0-PB108046.00.0	PB108045.00.0-PB108046.00.0	hypothetical protein	hypothetical
PF14_0685	PB108044.00.0	PB108044.00.0	PB108044.00.0	hypothetical protein	hypothetical
PF14_0686	PB000580.03.0	PC000862.00.0	PB000580.03.0	hypothetical protein	hypothetical
PF14_0690	PB000660.03.0	PC000537.01.0	PB000660.03.0	histone deacetylase, putative	DNA-binding
PF14_0691	PB000661.03.0	PC000548.02.0	PB000661.03.0	hypothetical protein	hypothetical
PF14_0698	PY04292	PY04292	PY04292	hypothetical protein	hypothetical
PF14_0699	PB000183.02.0	PC000199.03.0	PB000183.02.0	hypothetical protein, conserved	hypothetical
PF14_0706	PY05084	PY05084	PY05084	hypothetical protein	hypothetical
PF14_0707	PB000387.00.0	PC000733.03.0	PB000387.00.0 + PC000733.03.0	hypothetical protein	hypothetical
PF14_0722	PB001174.01.0	PC000328.01.0	PB001174.01.0 + PC000328.01.0	hypothetical protein	hypothetical

The shaded region represents missing syntenic rodent parasite genes. When a corresponding missing gene is found via microarray hybridization, the accession number of the oligo(s) that cross-hybridized were filled in. Putative functions of these genes were referenced to *P. falciparum* as it is the most complete and correctly annotated malaria parasite species.

Pf-Pc-Py (genes common to *P. falciparum*, *P. chabaudi* and *P. yoelii*)

PfGene	Genomic Location	Position	PbGene	PcGene	PyGene	Pf Product	Orthologs	Domain	Domain description	Pb Contig	Pc Contig	Py Contig
PfL0400w	MAL12: 359,990 - 360,494 (+)	ST	-	-	-	50S ribosomal protein L29, putative	(MALPY01248) + (chab0685h03.q1k)	SSF46561	Ribosomal protein L29 (L29p)	N	Y	Y
PfF1_0132	MAL13: 972,048 - 973,237 (-)	C	-	PC000321.03.0	-	60S ribosomal protein L23a, putative	PY04600	PF03939	ribosomal L23 protein	Y	Y	Y
PfF0465c	MAL6: 399,809 - 401,620 (-)	C	-	PC000861.01.0	-	cardiolipin synthetase, putative	-	PF00614	Phospholipase D Active site motif	Y	Y	Y
PfI0985c	MAL9: 825,323 - 827,130 (-)	C	-	PC101641.00.0	PY05250	chaperone protein	-	PF00226	DnaJ domain: heat shock protein binding	Y	Y	Y
PfB0120w *	MAL2: 127,994 - 128,314 (+)	ST	-	(PY00205)	PY00205	early transcribed membrane protein, putative	-	TIGR01495	Early transcribed membrane protein (ETAMP)	N	Y	Y
PfI1_0257	MAL11: 970104 - 971315 (-)	C	-	PC000511.00.0	PY00818	ethanolamine kinase, putative	-	PF01633	Choline/ethanolamine kinase	Y	Y	Y
PfI1_0298	MAL11: 1,113,994 - 1,115,475 (-)	C	-	-	-	GP18p transamidase	PY03470 + (chab0403h02.p1k)	-	-	Y	N	Y
PfF0590c	MAL6: 505,155 - 511,125 (-)	C	-	PC000855.03.0	PY06519	homologue of human HSPC025	-	-	-	Y	Y	Y
MAL13P1.125	MAL13: 951,356 - 952,680 (+)	C	-	-	PY00521	hypothetical protein	(chab0222H11.p2kA7g)	-	-	Y	Y	Y
MAL13P1.211	MAL13: 1,645,097 - 1,646,697 (+)	C	-	-	-	hypothetical protein	(PC_RP0116) + (MALPY00877)	-	-	Y	Y	Y
MAL13P1.264	MAL13: 2077095 - 2079483 (+)	C	-	PC000280.05.0	PY04526	hypothetical protein	-	PF00400	WD domain, G-beta repeat: protein-protein interactions	Y	Y	Y
MAL13P1.339	MAL13: 2,679,915 - 2,681,301 (+)	T	-	-	PY00079	hypothetical protein	PC301876.00.0	-	-	Y	Y	Y
MAL13P1.345	MAL13: 2,714,187 - 2,715,711 (-)	T	-	-	PY04218	hypothetical protein	(chab0479d11.q1k)	SSF57889	cysteine-rich domain superfamily	Y	Y	Y
MAL7P1.152	MAL7: 1,222,817 - 1,225,780 (+)	T	-	PC000002.03.0	PY01289	hypothetical protein	-	-	-	Y	Y	Y
MAL8P1.155	MAL8: 125,998 - 127,290 (+)	T	-	-	PY02607	hypothetical protein	(Pch0619c08.q1c)	-	-	Y	Y	Y
MAL8P1.75b	#N/A	#N/A	-	PC000315.00.0	PY04363	hypothetical protein	-	PF01926	GTPase of unknown function	Y	Y	Y
MAL8P1.86	MAL8: 710,666 - 711,721 (-)	C	-	PC000582.01.0	-	hypothetical protein	(MALPY01723)	-	-	Y	Y	Y
Pf07_0013	MAL7: 184,742 - 187,204 (-)	ST	-	PC107002.00.0	-	hypothetical protein	(MALPY00637)	-	-	Y	Y	Y
Pf07_0022	MAL7: 311,466 - 316,064 (+)	C	-	PC000421.02.0	-	hypothetical protein	(MALPY00278)	-	-	N	Y	Y
Pf07_0039	MAL7: 484,101 - 484,724 (+)	C	-	PC000715.04.0	PY02007	hypothetical protein	-	SSF50249	Nucleic acid-binding proteins superfamily	Y	Y	Y
Pf07_0041	MAL7: 515,435 - 516,814 (+)	C	-	PC300540.00.0	PY04692	hypothetical protein	-	-	-	Y	Y	Y
Pf07_0109	MAL7: 1,198,679 - 1,199,191 (-)	ST	-	PC000757.02.0	PY02101	hypothetical protein	-	-	-	Y	Y	Y
Pf08_0061	MAL8: 778,386 - 780,776 (-)	C	-	PC000240.00.0	PY04364	hypothetical protein	-	-	-	Y	Y	Y
Pf10_0074	MAL10: 299,280 - 299,609 (+)	ST	-	PC000518.03.0	PY00006	hypothetical protein	-	-	-	Y	Y	Y
Pf10_0093	MAL10: 387,027 - 388,438 (+)	C	-	PC000114.02.0	PY02533	hypothetical protein	-	SSF81606	Protein serine/threonine phosphatase 2C, catalytic domain superfamily	Y	Y	Y
Pf10_0185	MAL10: 773,584 - 774,489 (+)	C	-	PC001014.02.0	-	hypothetical protein	(MALPY01214)	-	-	Y	Y	Y
Pf10_0250	MAL10: 1,069,222 - 1,073,322 (-)	C	-	PC000757.04.0	PY03388 & PY03389	hypothetical protein	-	-	-	Y	Y	Y
Pf10_0258	MAL10: 1,101,816 - 1,103,156 (+)	C	-	PC000604.00.0	-	hypothetical protein	(MALPY02322)	-	-	Y	Y	Y
Pf10_0265	MAL10: 1,126,521 - 1,131,956 (+)	C	-	PC001286.02.0	PY06061	hypothetical protein	-	SSF81995	beta-sandwich fold of Sec23/24 superfamily	Y	Y	Y
Pf11_0075	MAL11: 267,547 - 267,897 (+)	ST	-	PC301990.00.0	-	hypothetical protein	(MALPY01885)	-	-	Y	Y	Y
Pf11_0135	MAL11: 488,480 - 489,045 (-)	C	-	-	-	hypothetical protein	(MALPY01751) + (chab1000h01.q1k)	-	-	Y	Y	Y
Pf11_0179	MAL11: 661,999 - 662,385 (+)	C	-	PC000136.02.0	PY06289	hypothetical protein	-	-	-	Y	Y	Y
Pf11_0238	MAL11: 868,387 - 869,151 (-)	C	-	-	PY06138	hypothetical protein	(chab0438a06.p1k)	-	-	Y	N	Y
Pf11_0285	MAL11: 1,065,783 - 1,066,567 (+)	C	-	-	-	hypothetical protein	(MALPY01905) + (PC_PH4525)	-	-	Y	Y	Y
Pf11_0322	MAL11: 1,199,251 - 1,201,803 (+)	C	-	PC300184.00.0	PY06085	hypothetical protein	-	-	-	Y	Y	Y
Pf11_0323	MAL11: 1,201,920 - 1,203,617 (-)	C	-	PC000031.04.0	PY06084	hypothetical protein	-	-	-	Y	Y	Y
Pf11_0324	MAL11: 1,204,295 - 1,209,585 (-)	C	-	PC000188.01.0	PY06083	hypothetical protein	-	-	-	N	Y	Y
Pf11_0328	MAL11: 1,234,059 - 1,236,308 (-)	C	-	PC000245.00.0	-	hypothetical protein	PY06628	SSF55729	acyl-coA N-acyltransferase (Nat)	Y	Y	Y
Pf11_0434 *	MAL11: 1,692,894 - 1,693,538 (-)	ST	-	(PY05499)	PY05499	hypothetical protein	-	-	-	Y	Y	Y
Pf11_0456	MAL11: 1,762,253 - 1,766,647 (+)	ST	-	PC000349.05.0	PY00296	hypothetical protein	-	SSF51156	Insect cysteine-rich antifreeze protein superfamily	Y	Y	Y
Pf11_0504 *	MAL11: 1,941,677 - 1,942,585 (+)	T	-	(PY02244)	PY02244	hypothetical protein	-	SSF51225	Fibre shaft of virus attachment proteins	N	N	Y
Pf11_0526	MAL11: 477,484 - 478,320 (+)	C	-	PC000254.03.0	-	hypothetical protein	(MALPY01751)	-	-	N	Y	Y
Pf13_0131	MAL13: 970,479 - 971,537 (+)	C	-	PC000708.04.0	-	hypothetical protein	PY04599	PF00583	acetyltransferase	Y	Y	Y
Pf13_0136	MAL13: 989,669 - 990,799 (+)	C	-	PC000210.02.0	PY06639	hypothetical protein	-	-	-	Y	Y	Y
Pf13_0255	MAL13: 1,949,635 - 1,951,183 (+)	C	-	PC300960.00.0	-	hypothetical protein	(MALPY01608)	-	-	Y	Y	Y
Pf13_0278	MAL13: 2,136,061 - 2,139,555 (+)	C	-	PC000860.02.0	-	hypothetical protein	PY04659	PS50102	eukaryotic RNA recognition motif	N	Y	Y
Pf13_0321 *	MAL13: 2,422,466 - 2,424,109 (-)	ST	-	(PY02725)	PY02725	hypothetical protein	-	-	-	Y	Y	Y
Pf13_0337	MAL13: 2,547,577 - 2,548,260 (-)	T	-	-	PY00840	hypothetical protein	(chab0831g07.p1k)	-	-	Y	N	Y
Pf13_0342 *	MAL13: 2,605,033 - 2,606,079 (+)	T	-	PC000669.02.0	(PC000669.02.0)	hypothetical protein	-	-	-	Y	Y	Y
Pf14_0046	MAL14: 170,000 - 170,893 (+)	T	-	-	-	hypothetical protein	(MALPY03022) + (chab0979c02.p1k)	-	-	Y	Y	Y
Pf14_0093	MAL14: 370,146 - 373,006 (+)	ST	-	-	PY03960	hypothetical protein	(chab0802H12.q1k)	-	-	Y	Y	Y
Pf14_0129	MAL14: 523,890 - 525,305 (-)	ST	-	-	-	hypothetical protein	(MALPY01822) + (chab0536e04.q1k)	SSF46689	Homeodomain-like superfamily: transcription factor	Y	Y	Y
Pf14_0138	MAL14: 553,639 - 555,772 (-)	ST	-	-	-	hypothetical protein	(MALPY01804) + (chab0763g10.p1k)	PF04683	Adhesion regulating molecule conserved region	Y	Y	Y
Pf14_0258	MAL14: 1,095,628 - 1,097,509 (+)	C	-	-	-	hypothetical protein	PC104734.00.0 + (MALPY02326)	-	-	N	Y	Y
Pf14_0259	MAL14: 1,097,980 - 1,099,893 (-)	C	-	PC000960.02.0	-	hypothetical protein	PY06720	-	-	N	Y	Y
Pf14_0306	MAL14: 1,280,830 - 1,281,245 (-)	C	-	PC001212.02.0	-	hypothetical protein	(MALPY02262)	-	-	Y	Y	Y
Pf14_0457 *	MAL14: 1,965,118 - 1,965,705 (+)	C	-	(PY01105)	PY01105	hypothetical protein	-	-	-	Y	Y	Y
Pf14_0466 *	MAL14: 2,001,026 - 2,002,049 (-)	C	-	(PY05109)	PY05109	hypothetical protein	-	PF01661	Macro domain; ADP-ribose binding	Y	Y	Y
Pf14_0582 *	MAL14: 2,482,711 - 2,485,022 (+)	C	-	(PY05163)	PY05163	hypothetical protein	PC400383.00.0	SSF55979	DNA clamp	Y	Y	Y
PfB0775w	MAL2: 701,244 - 702,893 (+)	C	-	PC000481.00.0	PY04634	hypothetical protein	-	SSF51225	Fibre shaft of virus attachment proteins	Y	Y	Y
PfB0835c	MAL2: 735,437 - 736,996 (-)	C	-	PC300318.00.0	PY03798	hypothetical protein	-	SSF54637	Thioesterase/thiol ester dehydrase-isomerase superfamily	Y	Y	Y
PfB0886c	MAL2: 773,333 - 773,752 (-)	ST	-	PC000497.02.0	PY04744	hypothetical protein	-	-	-	Y	Y	Y
PfB0910w	MAL2: 791,462 - 792,379 (+)	ST	-	-	PY04497	hypothetical protein	PC104786.00.0	-	-	N	Y	Y
PfC0650w	MAL3: 622,829 - 626,852 (+)	C	-	-	PY05835	hypothetical protein	(chab0444b01.q1k)	-	-	Y	Y	Y
PfC1040w	MAL3: 875,640 - 876,941 (+)	T	-	PC000072.00.0	PY00356	hypothetical protein	-	-	-	Y	Y	Y
PfD0325w	MAL4: 353,226 - 353,645 (+)	C	-	-	PY06037	hypothetical protein	(chab0569e06.p1k)	-	-	Y	Y	Y
PfD0690c	MAL4: 641,010 - 645,886 (-)	C	-	PC301863.00.0	PY00550	hypothetical protein	-	-	-	N	Y	Y
PfD0692c	MAL4: 654,372 - 655,371 (-)	C	-	-	-	hypothetical protein	(MALPY02876) + (chab0657f08.p1k)	PF00036	EF hand; Ca ion binding	N	N	Y
PfD0820w	MAL4: 749,130 - 750,072 (+)	C	-	PC000469.00.0	-	hypothetical protein	(MALPY01325)	-	-	Y	Y	Y
PfD0860w	MAL4: 790,602 - 793,411 (+)	C	-	PC300419.00.0	PY03318	hypothetical protein	-	-	-	Y	Y	Y
PfD0920w	MAL4: 856,773 - 859,222 (+)	C	-	PC300097.00.0	-	hypothetical protein	(MALPY01750)	-	-	N	Y	Y
PfD0925w	MAL4: 859,490 - 860,546 (+)	C	-	PC301913.00.0	PY05479	hypothetical protein	-	-	-	N	Y	Y
PfD0945c *	MAL4: 873,481 - 875,091 (-)	C	-	(PY05934)	PY05934	hypothetical protein	-	SSF90123	Multidrug resistance ABC transporter MsbA, N-terminal domain superfamily	Y	Y	Y
PfD1004c	MAL4: 992,282 - 993,323 (-)	ST	-	PC000028.04.0	PY02466	hypothetical protein	-	SSF55194	Ribosome recycling factor, RRF	N	Y	Y
PfD1085w	MAL4: 1,052,136 - 1,052,918 (+)	ST	-	PC000162.02.0	PY06450	hypothetical protein	-	-	-	Y	Y	Y

Pf-Pc-Py (genes common to *P. falciparum*, *P. chabaudi* and *P. yoelii*)

PfGene	Genomic Location	Position	PbGene	PcGene	PyGene	Pf Product	Orthologs	Domain	Domain description	Pb Contig	Pc Contig	Py Contig
PFE0085c *	MAL5: 86,820 - 89,201 (-)	T	-	(PY03919)	PY03919	hypothetical protein	-	PF00515	Tetratricopeptide repeat; protein-protein interaction	N	Y	Y
PFE0345c	MAL5: 292,696 - 293,474 (-)	C	-	PC000266.03.0	PY06226	hypothetical protein	-	-	-	Y	Y	Y
PFE0620c	MAL5: 539,729 - 542,591 (-)	C	-	PC000164.00.0	PY03181	hypothetical protein	-	-	-	Y	Y	Y
PFE0940c	MAL5: 783,653 - 784,088 (-)	C	-	-	PY03392	hypothetical protein	(Pch0737g02.p1c)	-	-	Y	Y	Y
PFE0985w	MAL5: 812,645 - 814,912 (+)	C	-	PC000934.02.0	PY00703	hypothetical protein	-	PF07967	C3HC zinc finger-like: zinc ion binding	Y	Y	Y
PFE1205c *	MAL5: 1,008,631 - 1,009,399 (-)	C	-	(PY00157)	PY00157	hypothetical protein	-	-	-	Y	Y	Y
PFE1425c	MAL5: 1,176,962 - 1,177,279 (-)	ST	-	-	PY00692	hypothetical protein	(chab0698b01.q1k) + (chab0698b01.q1k)	-	-	Y	Y	Y
PFF0215w	MAL6: 184,889 - 185,840 (+)	ST	-	-	PY05528	hypothetical protein	(chab0924d02.q1k)	-	-	Y	Y	Y
PFF0425w	MAL6: 356,225 - 356,272 (+)	C	-	-	PY02260	hypothetical protein	(PC_RP1951)	-	-	Y	Y	Y
PFF0990c	MAL6: 848,402 - 850,861 (-)	C	-	PC301391.00.0	PY04227	hypothetical protein	-	PDO91196	unknown function	Y	Y	Y
PFF1005w	MAL6: 857,723 - 858,607 (+)	C	-	-	PY04224	hypothetical protein	(chab0560d01.p1k)	-	-	Y	N	Y
PFF1220w	MAL6: 1,025,180 - 1,026,337 (+)	C	-	-	PY02329	hypothetical protein	(chab0478a02.p1k)	-	-	Y	Y	Y
PFI0765w	MAL9: 660,167 - 660,893 (+)	C	-	PC302035.00.0	PY07490	hypothetical protein	-	-	-	Y	Y	Y
PFI1055w	MAL9: 875,592 - 876,857 (+)	C	-	PC000831.00.0	-	hypothetical protein	(MALPY01448)	-	-	Y	Y	Y
PFI1440w	MAL9: 1,171,329 - 1,173,473 (+)	C	-	PC000775.01.0	PY05297	hypothetical protein	-	-	-	Y	Y	Y
PFI1463w	MAL9: 1,187,653 - 1,188,886 (+)	C	-	(PY04278)	PY04278	hypothetical protein	-	-	-	Y	Y	Y
PFI1675w	MAL9: 1,354,570 - 1,355,484 (+)	ST	-	PC001412.02.0	PY05237	hypothetical protein	-	-	-	Y	Y	Y
PFI1690c	MAL9: 1,364,245 - 1,365,988 (-)	ST	-	PC000803.04.0	-	hypothetical protein	(MALPY01654)	-	-	Y	Y	Y
PFL0135w	MAL12: 158,230 - 160,131 (+)	T	-	-	PY05185	hypothetical protein	PC108806.00.0	-	-	Y	Y	Y
PFL0435w	MAL12: 403,345 - 404,809 (+)	ST	-	PC301697.00.0	PY06577	hypothetical protein	-	SSF46565	Chaperone J-domain superfamily; heat shock protein binding	Y	Y	Y
PFL1000c	MAL12: 825,690 - 827,705 (-)	C	-	-	PY04906	hypothetical protein	(chab0632c08.q1k)	-	-	Y	N	Y
PFL1095c	MAL12: 920,781 - 922,407 (-)	C	-	PC000003.02.0	-	hypothetical protein	(MALPY00995)	-	-	Y	Y	Y
PFL1410c *	MAL12: 1,192,877 - 1,199,203 (-)	C	-	(PY05035)	PY05035	hypothetical protein	-	SSF90123	Multidrug resistance ABC transporter MsbA, N-terminal domain superfamily	N	Y	Y
PFL1535w	MAL12: 1,321,970 - 1,329,230 (+)	C	-	PC000200.05.0	PY06044	hypothetical protein	-	SSF82171	Dipeptidyl peptidase IV/CD26, N-terminal domain superfamily	Y	Y	Y
PFL2170c *	MAL12: 1,888,238 - 1,892,686 (-)	ST	-	PC000684.02.0	(PC000684.02.0)	hypothetical protein	-	-	-	Y	Y	Y
PFL2185w	MAL12: 1,897,254 - 1,899,826 (+)	ST	-	PC001315.02.0	-	hypothetical protein	PY06544	-	-	Y	Y	Y
PFL2435w	MAL12: 2,075,776 - 2,076,501 (+)	T	-	-	-	hypothetical protein	(MALPY02265) + (chab0598d12.q1k)	SSF50784	Transcription factor III (TFIIA), beta barrel domain	Y	Y	Y
MAL13P1.212	MAL13: 1,669,725 - 1,670,987 (-)	C	-	-	-	hypothetical protein, conserved	PY07307	-	-	Y	Y	Y
PFI0_O415 *	MAL10: 1,502,954 - 1,504,321 (-)	ST	-	(PY04876)	PY04876	hypothetical protein, conserved	-	-	-	Y	Y	Y
PFI1_0409	MAL11: 1,593,963 - 1,594,948 (+)	C	-	PC302418.00.0	-	hypothetical protein, conserved	(MALPY01728)	PF03665	uncharacterized protein family	Y	Y	Y
PFB0194w	MAL2: 192,818 - 194,199 (+)	C	-	PC001341.02.0	PY00894	hypothetical protein, conserved	-	-	-	Y	Y	Y
PFF0605c	MAL6: 522,244 - 522,660 (-)	C	-	-	PY03102	hypothetical protein, conserved	(chab0083h05.q1k)	-	-	Y	Y	Y
PFF0280w	MAL6: 236,394 - 237,248 (+)	ST	-	PC300302.00.0	PY00627	hypothetical protein, conserved	-	-	-	N	Y	Y
PFD0805w	MAL4: 733,248 - 734,111 (+)	C	-	PC000104.03.0	PY02863	prohibitin-like protein, putative	-	-	-	N	Y	Y
MAL8P1.310	MAL8: 1,289,259 - 1,289,658 (-)	T	-	-	-	putative senescence-associated protein	(MALPY02241) + (chab0089b03.p1k)	-	-	N	N	Y
PFI1_0260	MAL11: 981,944 - 982,765 (+)	C	-	PC000283.03.0	-	ribosomal protein L35, putative	PY03511	PF00831	ribosomal L29 protein	Y	Y	Y
PFD0990w	MAL4: 936,320 - 937,120 (+)	C	-	PC000317.01.0	PY06329	ribosome recycling factor, putative	-	PF01765	Ribosome recycling factor	N	Y	Y
PFF1425w	MAL6: 1,210,422 - 1,212,764 (+)	ST	-	PC301509.00.0	PY04169	RNA binding protein, putative	-	PF00076	RNA recognition motif	Y	Y	Y
PFD0310w *	MAL4: 335,693 - 336,166 (+)	C	-	(PY05101)	PY05101	sexual stage-specific protein precursor	-	-	-	Y	Y	Y
PFB0480w	MAL2: 436,050 - 436,994 (+)	C	-	PC300128.00.0	-	syntxin, putative	(MALPY01221)	PF00804	syntxin (exocytosis)	Y	Y	Y
PFO8_0055	MAL8: 856,751 - 857,838 (-)	C	-	PC000115.01.0	PY03839	u3 small nucleolar ribonucleoprotein protein	-	PF04427	Brix domain; ribosome biogenesis	Y	Y	Y

Total: 116

Known orthologs are appended into the orthologs column while a potential gene detected by tBLASTn is also appended this way and the contig name is shown in parentheses.

The last 3 columns represent the presence/absence of a corresponding RMP contig with respect to the *P. falciparum* gene.

The relative chromosomal position in *P. falciparum* is denoted by T: telomeric; ST: sub-telomeric and C: centromere (bounded by the sub-telomeric regions at each end of the chromosomal arms).

*: Corresponding gene discovered in via cross-hybridization with orthologous genes from another rodent malaria parasite species as shown in bold parentheses.

289

P. yoelii specific genes that are transcriptionally inactive

Py_ID	Pb_ID	Pc_ID	Pf_ID	Py_name	Py_location	Pb_tblastn	Pc_tblastn	Pf_tblastn	Pb_cgh	Pb_S	Pc_cgh	Pc_S	Py_cgh	Py_S
PY04963	-	-	-	hypothetical protein	MALPY01553: 2,191 - 3,065 (-)	-	-	-	-	-	-	-	-	-
PY04995	-	-	-	hypothetical protein	MALPY01564: 692 - 793 (+)	-	-	-	-	-	-	-	-	-
PY05011	-	-	-	hypothetical protein	MALPY01571: 2,199 - 2,442 (+)	-	-	-	-	-	-	-	-	-
PY05012	-	-	-	hypothetical protein	MALPY01571: 2,552 - 2,785 (+)	-	-	-	-	-	-	-	-	-
PY05173	-	-	-	hypothetical protein	MALPY01633: 2,462 - 2,553 (+)	-	-	-	-	-	-	-	-	-
PY05240	-	-	-	hypothetical protein	MALPY01655: 8,179 - 8,303 (+)	-	-	-	-	-	-	-	-	-
PY05288	-	-	-	hypothetical protein	MALPY01674: 507 - 608 (+)	-	-	-	-	-	-	-	-	-
PY05304	-	-	-	hypothetical protein	MALPY01678: 2,573 - 2,707 (+)	-	-	-	-	-	-	-	-	-
PY05308	-	-	-	hypothetical protein	MALPY01679: 3,178 - 3,288 (+)	-	-	-	-	-	-	-	-	-
PY05351	-	-	-	hypothetical protein	MALPY01692: 5,389 - 5,460 (-)	-	-	-	-	-	-	-	-	-
PY05407	-	-	-	hypothetical protein	MALPY01716: 7,080 - 7,247 (-)	-	-	-	-	-	-	-	-	-
PY05411	-	-	-	hypothetical protein	MALPY01717: 2,473 - 2,598 (-)	-	-	-	-	-	-	-	-	-
PY05440	-	-	-	hypothetical protein	MALPY01732: 6,249 - 6,764 (-)	-	-	-	-	-	-	-	-	-
PY05506	-	-	-	hypothetical protein	MALPY01763: 2,797 - 2,898 (-)	-	-	-	-	-	-	-	-	-
PY05546	-	-	-	hypothetical protein	MALPY01780: 209 - 466 (-)	-	-	-	-	-	-	-	-	-
PY05660	-	-	-	hypothetical protein	MALPY01829: 7,262 - 7,957 (-)	-	-	-	-	-	-	-	-	-
PY05672	-	-	-	hypothetical protein	MALPY01834: 2,558 - 2,627 (-)	-	-	-	-	-	-	-	-	-
PY05687	-	-	-	hypothetical protein	MALPY01843: 322 - 648 (-)	-	-	-	-	-	-	-	-	-
PY05700	-	-	-	hypothetical protein	MALPY01848: 5,118 - 5,213 (-)	-	-	-	-	-	-	-	-	-
PY05713	-	-	-	hypothetical protein	MALPY01854: 8,921 - 9,438 (-)	-	-	-	-	-	-	-	-	-
PY05800	-	-	-	hypothetical protein	MALPY01898: 1,658 - 2,186 (-)	-	-	-	-	-	-	-	-	-
PY05823	-	-	-	hypothetical protein	MALPY01908: 2,669 - 2,845 (+)	-	-	-	-	-	-	-	-	-
PY05851	-	-	-	hypothetical protein	MALPY01921: 5,779 - 6,451 (+)	-	-	-	-	-	-	-	-	-
PY05863	-	-	-	hypothetical protein	MALPY01926: 4,511 - 5,002 (-)	-	-	-	-	-	-	-	-	-
PY05872	-	-	-	hypothetical protein	MALPY01931: 2,231 - 2,315 (+)	-	-	-	-	-	-	-	-	-
PY05874	-	-	-	hypothetical protein	MALPY01932: 1,793 - 1,856 (+)	-	-	-	-	-	-	-	-	-
PY05896	-	-	-	hypothetical protein	MALPY01940: 5,445 - 5,654 (-)	-	-	-	-	-	-	-	-	-
PY05924	-	-	-	hypothetical protein	MALPY01952: 2,655 - 2,729 (+)	-	-	-	-	-	-	-	-	-
PY05944	-	-	-	hypothetical protein	MALPY01961: 5,803 - 6,201 (+)	-	-	-	-	-	-	-	-	-
PY05958	-	-	-	hypothetical protein	MALPY01967: 947 - 1,375 (+)	-	-	-	-	-	-	-	-	-
PY05989	-	-	-	hypothetical protein	MALPY01981: 1,513 - 2,146 (+)	-	-	-	-	-	-	-	-	-
PY06021	-	-	-	hypothetical protein	MALPY01998: 2,092 - 2,199 (+)	-	-	-	-	-	-	-	-	-
PY06028	-	-	-	hypothetical protein	MALPY02003: 1 - 84 (+)	-	-	-	-	-	-	-	-	-
PY06030	-	-	-	hypothetical protein	MALPY02003: 2,250 - 2,342 (-)	-	-	-	-	-	-	-	-	-
PY06058	-	-	-	hypothetical protein	MALPY02018: 3,454 - 3,664 (-)	-	-	-	-	-	-	-	-	-
PY06090	-	-	-	hypothetical protein	MALPY02031: 37 - 317 (-)	-	-	-	-	-	-	-	-	-
PY06096	-	-	-	hypothetical protein	MALPY02031: 2,547 - 2,671 (+)	-	-	-	-	-	-	-	-	-
PY06152	-	-	-	hypothetical protein	MALPY02058: 9,997 - 10,090 (-)	-	-	-	-	-	-	-	-	-
PY06167	-	-	-	hypothetical protein	MALPY02065: 2,492 - 2,579 (-)	-	-	-	-	-	-	-	-	-
PY06169	-	-	-	hypothetical protein	MALPY02066: 2,050 - 2,278 (+)	-	-	-	-	-	-	-	-	-
PY06174	-	-	-	hypothetical protein	MALPY02067: 5,941 - 6,012 (-)	-	-	-	-	-	-	-	-	-
PY06186	-	-	-	hypothetical protein	MALPY02071: 2,797 - 2,863 (-)	-	-	-	-	-	-	-	-	-
PY06187	-	-	-	hypothetical protein	MALPY02072: 446 - 565 (+)	-	-	-	-	-	-	-	-	-
PY06206	-	-	-	hypothetical protein	MALPY02081: 5,783 - 5,943 (+)	-	-	-	-	-	-	-	-	-
PY06216	-	-	-	hypothetical protein	MALPY02085: 2,434 - 2,534 (-)	-	-	-	-	-	-	-	-	-
PY06266	-	-	-	hypothetical protein	MALPY02103: 2,392 - 2,634 (+)	-	-	-	-	-	-	-	-	-
PY06280	-	-	-	hypothetical protein	MALPY02109: 2,937 - 3,029 (-)	-	-	-	-	-	-	-	-	-
PY06319	-	-	-	hypothetical protein	MALPY02126: 7,389 - 7,459 (+)	-	-	-	-	-	-	-	-	-
PY06336	-	-	-	hypothetical protein	MALPY02136: 3,405 - 3,518 (+)	-	-	-	-	-	-	-	-	-
PY06384	-	-	-	hypothetical protein	MALPY02157: 2,539 - 2,626 (-)	-	-	-	-	-	-	-	-	-
PY06515	-	-	-	hypothetical protein	MALPY02216: 2,911 - 3,252 (+)	-	-	-	-	-	-	-	-	-
PY06548	-	-	-	hypothetical protein	MALPY02231: 2,223 - 2,302 (-)	-	-	-	-	-	-	-	-	-
PY06593	-	-	-	hypothetical protein	MALPY02257: 1,612 - 1,678 (-)	-	-	-	-	-	-	-	-	-
PY06598	-	-	-	hypothetical protein	MALPY02261: 226 - 302 (+)	-	-	-	-	-	-	-	-	-
PY06616	-	-	-	hypothetical protein	MALPY02268: 2,284 - 2,379 (+)	-	-	-	-	-	-	-	-	-
PY06669	-	-	-	hypothetical protein	MALPY02297: 1,357 - 1,419 (-)	-	-	-	-	-	-	-	-	-
PY06680	-	-	-	hypothetical protein	MALPY02302: 2,615 - 2,692 (+)	-	-	-	-	-	-	-	-	-
PY06689	-	-	-	hypothetical protein	MALPY02307: 3,314 - 3,414 (-)	-	-	-	-	-	-	-	-	-
PY06708	-	-	-	hypothetical protein	MALPY02320: 489 - 1,151 (-)	-	-	-	-	-	-	-	-	-
PY06716	-	-	-	hypothetical protein	MALPY02324: 408 - 1,476 (-)	-	-	-	-	-	-	-	-	-
PY06731	-	-	-	hypothetical protein	MALPY02332: 557 - 772 (-)	-	-	-	-	-	-	-	-	-
PY06745	-	-	-	hypothetical protein	MALPY02337: 878 - 979 (+)	-	-	-	-	-	-	-	-	-
PY06747	-	-	-	hypothetical protein	MALPY02338: 322 - 389 (+)	-	-	-	-	-	-	-	-	-
PY06761	-	-	-	hypothetical protein	MALPY02347: 2,237 - 2,529 (+)	-	-	-	-	-	-	-	-	-
PY06783	-	-	-	hypothetical protein	MALPY02359: 2,708 - 2,941 (+)	-	-	-	-	-	-	-	-	-
PY06798	-	-	-	hypothetical protein	MALPY02369: 5,180 - 5,270 (-)	-	-	-	-	-	-	-	-	-
PY06833	-	-	-	hypothetical protein	MALPY02388: 2,494 - 4,255 (+)	-	-	-	-	-	-	-	-	-
PY06878	-	-	-	hypothetical protein	MALPY02413: 1,761 - 2,075 (-)	-	-	-	-	-	-	-	-	-
PY06879	-	-	-	hypothetical protein	MALPY02413: 3,547 - 3,663 (+)	-	-	-	-	-	-	-	-	-
PY06888	-	-	-	hypothetical protein	MALPY02419: 1,524 - 1,620 (-)	-	-	-	-	-	-	-	-	-
PY06892	-	-	-	hypothetical protein	MALPY02421: 3,144 - 3,377 (+)	-	-	-	-	-	-	-	-	-
PY06893	-	-	-	hypothetical protein	MALPY02421: 3,825 - 3,922 (+)	-	-	-	-	-	-	-	-	-
PY06940	-	-	-	hypothetical protein	MALPY02447: 2,321 - 2,417 (-)	-	-	-	-	-	-	-	-	-
PY06947	-	-	-	hypothetical protein	MALPY02451: 3,288 - 3,370 (-)	-	-	-	-	-	-	-	-	-
PY06987	-	-	-	hypothetical protein	MALPY02471: 740 - 994 (+)	-	-	-	-	-	-	-	-	-
PY07073	-	-	-	hypothetical protein	MALPY02525: 2,063 - 2,125 (-)	-	-	-	-	-	-	-	-	-
PY07074	-	-	-	hypothetical protein	MALPY02525: 2,357 - 2,509 (-)	-	-	-	-	-	-	-	-	-
PY07075	-	-	-	hypothetical protein	MALPY02525: 2,587 - 2,683 (+)	-	-	-	-	-	-	-	-	-
PY07081	-	-	-	hypothetical protein	MALPY02528: 3,102 - 3,768 (-)	-	-	-	-	-	-	-	-	-
PY07082	-	-	-	hypothetical protein	MALPY02528: 4,188 - 4,310 (-)	-	-	-	-	-	-	-	-	-
PY07087	-	-	-	hypothetical protein	MALPY02531: 1,874 - 2,137 (-)	-	-	-	-	-	-	-	-	-
PY07096	-	-	-	hypothetical protein	MALPY02535: 2,003 - 2,083 (-)	-	-	-	-	-	-	-	-	-
PY07126	-	-	-	hypothetical protein	MALPY02548: 1,954 - 2,053 (-)	-	-	-	-	-	-	-	-	-
PY07133	-	-	-	hypothetical protein	MALPY02552: 3,356 - 2,435 (-)	-	-	-	-	-	-	-	-	-
PY07135	-	-	-	hypothetical protein	MALPY02553: 2,099 - 2,171 (-)	-	-	-	-	-	-	-	-	-
PY07212	-	-	-	hypothetical protein	MALPY02600: 2,347 - 2,421 (+)	-	-	-	-	-	-	-	-	-
PY07213	-	-	-	hypothetical protein	MALPY02600: 2,515 - 2,577 (+)	-	-	-	-	-	-	-	-	-
PY07236	-	-	-	hypothetical protein	MALPY02613: 2,201 - 2,326 (-)	-	-	-	-	-	-	-	-	-
PY07241	-	-	-	hypothetical protein	MALPY02618: 877 - 975 (+)	-	-	-	-	-	-	-	-	-
PY07247	-	-	-	hypothetical protein	MALPY02619: 1,733 - 1,928 (+)	-	-	-	-	-	-	-	-	-
PY07348	-	-	-	hypothetical protein	MALPY02680: 2,120 - 2,199 (-)	-	-	-	-	-	-	-	-	-
PY07353	-	-	-	hypothetical protein	MALPY02684: 890 - 962 (+)	-	-	-	-	-	-	-	-	-
PY07360	-	-	-	hypothetical protein	MALPY02688: 2,008 - 2,103 (-)	-	-	-	-	-	-	-	-	-
PY07429	-	-	-	hypothetical protein	MALPY02732: 456 - 1,845 (-)	-	-	-	-	-	-	-	-	-
PY07437	-	-	-	hypothetical protein	MALPY02738: 1,846 - 1,923 (+)	-	-	-	-	-	-	-	-	-
PY07438	-	-	-	hypothetical protein	MALPY02738: 2,082 - 2,146 (-)	-	-	-	-	-	-	-	-	-
PY07456	-	-	-	hypothetical protein	MALPY02750: 3,070 - 3,848 (-)	-	-	-	-	-	-	-	-	-
PY07523	-	-	-	hypothetical protein	MALPY02795: 2,142 - 2,267 (-)	-	-	-	-	-	-	-	-	-
PY07534	-	-	-	hypothetical protein	MALPY02801: 2,046 - 2,128 (-)	-	-	-	-	-	-	-	-	-
PY07563	-	-	-	hypothetical protein	MALPY02815: 1 - 98 (+)	-	-	-	-	-	-	-	-	-
PY07565	-	-	-	hypothetical protein	MALPY02815: 453 - 530 (+)	-	-	-	-	-	-	-	-	-
PY07619	-	-	-	hypothetical protein	MALPY02838: 2,276 - 2,353 (+)	-	-	-	-	-	-	-	-	-
PY07620	-	-	-	hypothetical protein	MALPY02838: 2,989 - 3,073 (+)	-	-	-	-	-	-	-	-	-
PY07625	-	-	-	hypothetical protein	MALPY02842: 2,319 - 2,415 (-)	-	-	-	-	-	-	-	-	-
PY07830	-	-	-	hypothetical protein	MALPY02960: 219 - 332 (+)	-	-	-	-	-	-	-	-	-
PY02547	-	-	-	Predicted membrane protein	MALPY00699: 4,645 - 5,181 (-)	-	-	-	-	-	-	-	-	-

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

P. yoelii and *P. berghei* specific genes that are transcriptionally inactive

Py_ID	Pb_ID	Pc_ID	Pf_ID	Py_name	Py_location	Pb_tblastn	Pc_tblastn	Pf_tblastn	Pb_cgh	Pb_S	Pc_cgh	Pc_S	Py_cgh	Py_S
PY07811	-	-	-	Cellulose binding domain, putative	MALPY02945: 712 - 2,266 (+)	-	-	-	Y	-	-	-	Y	-
PY07816	-	-	-	high molecular weight glutenin subunit	MALPY02947: 673 - 2,549 (+)	-	-	-	Y	-	-	-	Y	-
PY00341	-	-	-	hypothetical protein	MALPY00098: 351 - 439 (+)	-	-	-	Y	-	-	-	Y	-
PY00393	-	-	-	hypothetical protein	MALPY00111: 1,908 - 1,988 (+)	-	-	-	Y	-	-	-	Y	-
PY02615	-	-	-	hypothetical protein	MALPY00721: 2,833 - 2,922 (-)	-	-	-	Y	-	-	-	Y	-
PY03364	PB104817.00.0	-	-	hypothetical protein	MALPY00970: 7,630 - 7,767 (+)	-	-	-	Y	-	-	-	Y	-
PY03387	-	-	-	hypothetical protein	MALPY00973: 5,665 - 5,766 (+)	-	-	-	Y	-	-	-	Y	-
PY03416	-	-	-	hypothetical protein	MALPY00983: 1,941 - 2,021 (+)	-	-	-	Y	-	-	-	Y	-
PY03743	-	-	-	hypothetical protein	MALPY01104: 2,700 - 2,763 (-)	-	-	-	Y	-	-	-	Y	-
PY03944	-	-	-	hypothetical protein	MALPY01196: 2,792 - 2,885 (-)	-	-	-	Y	-	-	-	Y	-
PY04268	-	-	-	hypothetical protein	MALPY01287: 6,170 - 6,242 (-)	-	-	-	Y	-	-	-	Y	-
PY04538	-	-	-	hypothetical protein	MALPY01392: 3,920 - 3,981 (-)	-	-	-	Y	-	-	-	Y	-
PY04908	-	-	-	hypothetical protein	MALPY01526: 111 - 212 (+)	-	-	-	Y	-	-	-	Y	-
PY05097	-	-	-	hypothetical protein	MALPY01603: 1,761 - 1,847 (+)	-	-	-	Y	-	-	-	Y	-
PY05291	-	-	-	hypothetical protein	MALPY01674: 2,116 - 2,189 (-)	-	-	-	Y	-	-	-	Y	-
PY05416	-	-	-	hypothetical protein	MALPY01719: 2,039 - 2,122 (-)	-	-	-	Y	-	-	-	Y	-
PY06047	-	-	-	hypothetical protein	MALPY02013: 1,704 - 1,868 (-)	-	-	-	Y	-	-	-	Y	-
PY06235	-	-	-	hypothetical protein	MALPY02091: 1,505 - 1,689 (-)	-	-	-	Y	-	-	-	Y	-
PY06268	-	-	-	hypothetical protein	MALPY02104: 2,558 - 2,649 (-)	-	-	-	Y	-	-	-	Y	-
PY06399	PB403123.00.0, PB102802.00.0	-	-	hypothetical protein	MALPY02167: 2,873 - 3,017 (+)	PB_RP0150	-	-	Y	-	-	-	Y	-
PY06429	-	-	-	hypothetical protein	MALPY02184: 2,263 - 2,588 (+)	-	-	-	Y	-	-	-	Y	-
PY06439	-	-	-	hypothetical protein	MALPY02188: 5,055 - 5,663 (+)	-	-	-	Y	-	-	-	Y	-
PY06917	-	-	-	hypothetical protein	MALPY02434: 1,404 - 1,469 (-)	-	-	-	Y	-	-	-	Y	-
PY06935	-	-	-	hypothetical protein	MALPY02444: 1,533 - 1,644 (-)	-	-	-	Y	-	-	-	Y	-
PY07287	-	-	-	hypothetical protein	MALPY02640: 2,807 - 2,945 (+)	PB_PH4755	-	-	Y	-	-	-	Y	-
PY07483	-	-	-	hypothetical protein	MALPY02767: 2,618 - 2,683 (-)	-	-	-	Y	-	-	-	Y	-
PY07735	-	-	-	hypothetical protein	MALPY02909: 272 - 1,075 (+)	-	-	-	Y	-	-	-	Y	-
PY07741	-	-	-	hypothetical protein	MALPY02913: 1,490 - 2,625 (-)	-	-	-	Y	-	-	-	Y	-
PY07744	-	-	-	hypothetical protein	MALPY02914: 25 - 1,040 (+)	-	-	-	Y	-	-	-	Y	-
PY07745	-	-	-	hypothetical protein	MALPY02914: 1,114 - 2,480 (+)	-	-	-	Y	-	-	-	Y	-
PY07746	-	-	-	hypothetical protein	MALPY02915: 48 - 455 (+)	-	-	-	Y	-	-	-	Y	-
PY07749	-	-	-	hypothetical protein	MALPY02916: 1,535 - 2,444 (+)	-	-	-	Y	-	-	-	Y	-
PY07757	-	-	-	hypothetical protein	MALPY02921: 197 - 1,485 (+)	-	-	-	Y	-	-	-	Y	-
PY07764	-	-	-	hypothetical protein	MALPY02924: 1,862 - 2,474 (+)	-	-	-	Y	-	-	-	Y	-
PY07771	-	-	-	hypothetical protein	MALPY02927: 93 - 332 (+)	-	-	-	Y	-	-	-	Y	-
PY07785	-	-	-	hypothetical protein	MALPY02934: 12 - 1,868 (-)	-	-	-	Y	-	-	-	Y	-
PY07791	-	-	-	hypothetical protein	MALPY02936: 1,510 - 2,471 (+)	-	-	-	Y	-	-	-	Y	-
PY07795	-	-	-	hypothetical protein	MALPY02937: 597 - 2,044 (+)	-	-	-	Y	-	-	-	Y	-
PY07796	-	-	-	hypothetical protein	MALPY02938: 121 - 1,494 (+)	-	-	-	Y	-	-	-	Y	-
PY07812	-	-	-	hypothetical protein	MALPY02946: 28 - 591 (+)	-	-	-	Y	-	-	-	Y	-
PY07820	-	-	-	hypothetical protein	MALPY02950: 1,693 - 2,109 (-)	-	-	-	Y	-	-	-	Y	-
PY07842	-	-	-	hypothetical protein	MALPY02967: 2,496 - 2,593 (+)	-	-	-	Y	-	-	-	Y	-
PY01604	-	-	-	nucleosome assembly protein-related	MALPY00434: 319 - 970 (+)	-	-	-	Y	-	-	-	Y	-
PY07797	-	-	-	Prenyltransferase and squalene oxidase repeat, putative	MALPY02938: 1,515 - 2,586 (+)	-	-	-	Y	-	-	-	Y	-
PY07738	-	-	-	unknown-related	MALPY02911: 909 - 2,094 (+)	-	-	-	Y	-	-	-	Y	-

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

P. yoelii and *P. chabaudi* specific genes that are transcriptionally active

Py ID	Pb ID	Pc ID	Pf ID	Py name	Py location	Pb tblastn	Pc tblastn	Pf tblastn	Pb_cgh	Pb_S	Pc_cgh	Pc_S	Py_cgh	Py_S
PY01635	-	-	-	40s ribosomal protein s14 (fragment). [common tobacco]-rel	MALPY00440: 5,927 - 6,082 (-)	-	-	-	-	-	Y	R	Y	R
PY07683	-	-	-	Adhesion lipoprotein, putative	MALPY02880: 32 - 1,617 (+)	-	-	-	-	-	Y	R	Y	R
PY07740	-	-	-	contains similarity to collagens	MALPY02913: 32 - 1,150 (+)	-	-	-	-	-	Y	T	Y	R
PY05444	-	-	-	homeobox-containing protein	MALPY01735: 162 - 1,109 (+)	-	PC_PH1847	-	-	-	-	-	Y	T
PY00012	-	-	-	hypothetical protein	MALPY00005: 415 - 1,151 (+)	-	chab0380d05.p1k	-	-	-	-	-	Y	T
PY00023	-	-	-	hypothetical protein	MALPY00005: 29,987 - 30,595 (-)	-	PC_PH4198	MAL14	-	-	Y	R	Y	R
PY00151	-	PC302050.00.0	-	hypothetical protein	MALPY00040: 14,493 - 14,572 (-)	-	-	-	-	-	-	-	Y	R
PY00317	-	-	-	hypothetical protein	MALPY00090: 27 - 954 (+)	-	chab0188b01.q1k	-	-	-	-	-	Y	R
PY00737	-	-	-	hypothetical protein	MALPY00197: 302 - 445 (+)	-	-	-	-	-	Y	R	Y	R
PY00936	-	-	-	hypothetical protein	MALPY00251: 10,391 - 10,543 (+)	-	-	-	-	-	Y	T	Y	T
PY01118	-	-	-	hypothetical protein	MALPY00297: 2,359 - 3,339 (+)	-	chab0462f06.q1k	MAL2	-	-	-	-	Y	R
PY01165	-	-	-	hypothetical protein	MALPY00307: 3,213 - 3,635 (+)	-	chab0221a01.p1k	-	-	-	-	-	Y	R
PY01613	-	-	-	hypothetical protein	MALPY00436: 6,218 - 6,574 (+)	-	PC_RP2175	-	-	-	-	-	Y	R
PY01673	-	-	-	hypothetical protein	MALPY00451: 9,268 - 9,684 (+)	-	-	-	-	-	Y	R	Y	S
PY01687	-	-	-	hypothetical protein	MALPY00455: 4,341 - 5,809 (+)	-	chab0317a02.p1k	-	-	-	-	-	Y	T
PY01690	-	-	-	hypothetical protein	MALPY00455: 13,729 - 14,463 (-)	-	Pch0069e02.q1c	-	-	-	-	-	Y	T
PY01751	-	-	-	hypothetical protein	MALPY00473: 1,500 - 1,928 (+)	-	chab0855c10.p1k	MAL14	-	-	-	-	Y	R
PY02408	-	-	-	hypothetical protein	MALPY00663: 1,787 - 2,234 (+)	-	-	-	-	-	Y	R	Y	T
PY02717	-	-	-	hypothetical protein	MALPY00752: 1,880 - 1,987 (+)	-	-	-	-	-	Y	R	Y	S
PY02892	-	PC300310.00.0	-	hypothetical protein	MALPY00814: 416 - 1,240 (-)	-	PC_RP2191	MAL5	-	-	-	-	Y	R
PY03040	-	-	-	hypothetical protein	MALPY00864: 10,746 - 11,106 (-)	-	chab0361b09.q1k	-	-	-	-	-	Y	R
PY03065	-	-	-	hypothetical protein	MALPY00871: 7,173 - 7,394 (+)	-	chab0373d12.q1k	-	-	-	-	-	Y	S
PY03182	-	PC106710.00.0	-	hypothetical protein	MALPY00906: 3,684 - 5,205 (-)	-	PC_PH4779	-	-	-	Y	R	Y	R
PY03192	-	-	-	hypothetical protein	MALPY00912: 4,481 - 5,104 (-)	-	Pch0069e02.q1c	-	-	-	-	-	Y	T
PY03322	-	-	-	hypothetical protein	MALPY00950: 3,939 - 4,611 (+)	-	PC_RP4480	-	-	-	-	-	Y	S
PY03324	-	-	-	hypothetical protein	MALPY00951: 261 - 336 (-)	-	-	-	-	-	Y	R	Y	T
PY03382	-	-	-	hypothetical protein	MALPY00972: 20,097 - 21,135 (-)	-	PC_RP5270	MAL6	-	-	-	-	Y	T
PY03471	-	-	-	hypothetical protein	MALPY01003: 3,366 - 3,719 (+)	-	chab0433b04.p1k	-	-	-	-	-	Y	R
PY03624	-	-	-	hypothetical protein	MALPY01062: 1,874 - 1,966 (+)	-	-	-	-	-	Y	R	Y	R
PY03634	-	-	-	hypothetical protein	MALPY01069: 221 - 643 (+)	-	chab0819c02.q1kw	-	-	-	-	-	Y	R
PY03648	-	-	-	hypothetical protein	MALPY01075: 1,118 - 1,565 (-)	-	chab0219b05.q1k	-	-	-	-	-	Y	T
PY03653	-	PC109216.00.0 PC101365.00.0 PC405542.00.0 PC107363.00.0	-	hypothetical protein	MALPY01076: 5,994 - 6,257 (+)	-	PC_RP5153	-	-	-	-	-	Y	R
PY03818	-	-	-	hypothetical protein	MALPY01132: 1,429 - 1,581 (+)	-	-	-	-	-	Y	T	Y	R
PY03937	-	-	-	hypothetical protein	MALPY01176: 23,632 - 24,577 (+)	-	chab0555c06.p1k	-	-	-	-	-	Y	S
PY03950	-	-	-	hypothetical protein	MALPY01182: 1,057 - 4,773 (+)	-	-	-	-	-	Y	T	Y	R
PY03997	-	PC000039.04.0	-	hypothetical protein	MALPY01201: 6,422 - 6,821 (+)	-	PC_RP4053	MAL5	-	-	-	-	Y	R
PY04020	-	-	-	hypothetical protein	MALPY01209: 573 - 656 (+)	-	-	-	-	-	Y	R	Y	R
PY04060	-	-	-	hypothetical protein	MALPY01221: 16,108 - 16,398 (+)	-	-	-	-	-	Y	R	Y	T
PY04110	-	-	-	hypothetical protein	MALPY01235: 65 - 487 (+)	-	-	-	-	-	Y	R	Y	R
PY04433	-	-	-	hypothetical protein	MALPY01350: 3,493 - 3,646 (+)	-	chab0412b11.q1k	-	-	-	-	-	Y	T
PY04585	-	-	-	hypothetical protein	MALPY01410: 5,361 - 5,585 (+)	-	-	-	-	-	Y	R	Y	S
PY04652	-	-	-	hypothetical protein	MALPY01430: 9,274 - 9,375 (+)	-	-	-	-	-	Y	R	Y	R
PY04915	-	-	-	hypothetical protein	MALPY01529: 119 - 2,025 (+)	-	PC_PH0150	MAL6	-	-	-	-	Y	R
PY04974	-	-	-	hypothetical protein	MALPY01556: 4,156 - 4,239 (+)	-	-	-	-	-	Y	R	Y	R
PY05224	-	-	-	hypothetical protein	MALPY01651: 2,247 - 2,375 (+)	-	-	-	-	-	Y	S	Y	T
PY05310	-	-	-	hypothetical protein	MALPY01680: 5,139 - 5,723 (-)	-	chab0671f11.q1k	-	-	-	-	-	Y	R
PY05393	-	-	-	hypothetical protein	MALPY01710: 2,120 - 3,100 (-)	-	-	-	-	-	Y	R	Y	R
PY05419	-	-	-	hypothetical protein	MALPY01721: 2,057 - 2,963 (+)	-	-	-	-	-	Y	R	Y	R
PY05446	-	-	-	hypothetical protein	MALPY01735: 4,621 - 5,023 (+)	-	chab0515e01.p1k	-	-	-	-	-	Y	T
PY05487	-	-	-	hypothetical protein	MALPY01754: 166 - 1,812 (-)	-	PC_RP3942	MAL13	-	-	Y	R	Y	S
PY05690	-	-	-	hypothetical protein	MALPY01844: 2,409 - 3,206 (+)	-	PC_RP2180	-	-	-	-	-	Y	T
PY05761	-	-	-	hypothetical protein	MALPY01877: 1,876 - 2,743 (-)	-	chab0596b12.p1k	MAL14	-	-	-	-	Y	T
PY05932	-	-	-	hypothetical protein	MALPY01957: 32 - 783 (+)	-	PC_RP0790	MAL4	-	-	-	-	Y	R
PY05933	-	-	-	hypothetical protein	MALPY01957: 896 - 1,868 (+)	-	PC_RP0790	MAL4	-	-	-	-	Y	T
PY06042	-	-	-	hypothetical protein	MALPY02010: 8,398 - 8,631 (+)	-	PC_RP5222	-	-	-	Y	R	Y	R
PY06342	-	PC107437.00.0	-	hypothetical protein	MALPY02138: 2,096 - 3,832 (-)	-	-	-	-	-	-	-	Y	R
PY06356	-	-	-	hypothetical protein	MALPY02143: 4,770 - 4,849 (-)	-	-	-	-	-	Y	R	Y	T
PY06506	-	-	-	hypothetical protein	MALPY02213: 623 - 1,235 (+)	-	chab0502f12.q1k	-	-	-	-	-	Y	R
PY06587	-	-	-	hypothetical protein	MALPY02255: 869 - 2,413 (-)	-	PC_RP2406	MAL6	-	-	-	-	Y	S
PY06633	-	-	-	hypothetical protein	MALPY02278: 442 - 909 (+)	-	-	-	-	-	Y	R	Y	R
PY06688	-	-	-	hypothetical protein	MALPY02307: 2,277 - 3,142 (+)	-	PC_RP2264	MAL9	-	-	-	-	Y	T
PY06749	-	-	-	hypothetical protein	MALPY02338: 1,806 - 1,971 (-)	-	-	-	-	-	Y	T	Y	R

PY06750	-	PC000275.01.0	-	hypothetical protein	MALPY02339: 36 - 855 (-)	-	PC_RP1313	-	-	-	Y	R	Y	S
PY06842	-	-	-	hypothetical protein	MALPY02392: 872 - 1,780 (+)	-	PC_RP4010	-	-	-	Y	R	Y	R
PY06932	-	PC000089.01.0	-	hypothetical protein	MALPY02442: 83 - 2,344 (+)	-	PC_RP1103	MAL5	-	-	Y	S	Y	T
PY07107	-	-	-	hypothetical protein	MALPY02541: 10,442 - 10,753 (+)	-	chab0758e03.q1k	-	-	-	-	-	Y	T
PY07230	-	PC108730.00.0	-	hypothetical protein	MALPY02612: 4,671 - 5,419 (+)	-	PC_RP4991	-	-	-	-	-	Y	S
PY07231	-	-	-	hypothetical protein	MALPY02612: 5,489 - 5,819 (-)	-	-	-	-	-	Y	T	Y	T
PY07244	-	-	-	hypothetical protein	MALPY02617: 2,784 - 3,908 (-)	-	PC_RP2212	-	-	-	-	-	Y	T
PY07300	-	-	-	hypothetical protein	MALPY02649: 237 - 1,192 (+)	-	PC_RP0421	MAL10	-	-	-	-	Y	T
PY07301	-	-	-	hypothetical protein	MALPY02649: 1,302 - 1,972 (+)	-	PC_RP0421	MAL10	-	-	Y	R	Y	T
PY07507	-	-	-	hypothetical protein	MALPY02784: 20 - 466 (+)	-	PC_RP0180	-	-	-	-	-	Y	S
PY07617	-	PC000690.04.0	-	hypothetical protein	MALPY02837: 126 - 1,548 (+)	-	PC_RP4824	-	-	-	Y	R	Y	T
PY07731	-	-	-	hypothetical protein	MALPY02907: 27 - 682 (+)	-	-	-	-	-	Y	R	Y	S
PY06196	-	-	-	latent nuclear antigen-related	MALPY02076: 2,694 - 3,113 (+)	-	PC_RP2521	-	-	-	-	-	Y	S
PY04177	-	PC400691.00.0	-	Streptococcus pyogenes AMV253, putative	MALPY01250: 11,282 - 13,018 (+)	-	PC_PH0640	MAL9	-	-	Y	R	Y	R
PY01689	-	-	-	unidentified reading frame	MALPY00455: 11,538 - 12,080 (-)	-	-	-	-	-	Y	R	Y	S

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Genes common to *P. yoelii*, *P. berghei* and *P. chabaudi* that are transcriptionally inactive

Py_ID	Pb_ID	Pc_ID	Pf_ID	Py_name	Py_location	Pb_tblastn	Pc_tblastn	Pf_tblastn	Pb_cgh	Pb_S	Pc_cgh	Pc_S	Py_cgh	Py_S
PY07736	-	-	-	Acyltransferase, putative	MALPY02910: 20 - 1,425 (-)	-	-	-	Y	-	Y	-	Y	-
PY07495	-	-	-	arabinogalactan protein	MALPY02775: 1 - 736 (+)	PB_RP2601	-	-	Y	-	Y	-	Y	-
PY07807	-	-	-	Bradyrhizobium japonicum ID89-related	MALPY02943: 1,038 - 2,816 (-)	-	-	-	Y	-	Y	-	Y	-
PY05845	PB000442.01.0, PB000174.00.0, PB000158.02.0, PB001595.02.0	PC000204.00.0, PC001201.02.0, PC000611.01.0, PC000499.04.0	PFI0550w, MAL7P1.92	cysteine repeat modular protein 2 PbCRM2-related	MALPY01919: 1 - 3,189 (+)	PB_RP2144	PC_RP0227	MAL7	Y	-	Y	-	Y	-
PY07668	PB405793.00.0	-	-	Endonuclease/Exonuclease/phosphatase family	MALPY02872: 119 - 856 (+)	PB_PH2489	PC_PH3354	-	Y	-	Y	-	Y	-
PY05598	PB107174.00.0	-	-	homeobox-containing protein	MALPY01801: 39 - 1,073 (+)	-	PC_RP2056	MAL4	Y	-	Y	-	Y	-
PY00059	-	-	-	hypothetical protein	MALPY00014: 2,023 - 3,453 (+)	-	-	-	Y	-	Y	-	Y	-
PY00282	-	PC100185.00.0, PC108532.00.0, PC406031.00.0, PC402603.00.0, PC402348.00.0, PC106059.00.0, PC402851.00.0	-	hypothetical protein	MALPY00080: 6,296 - 7,326 (+)	-	-	-	Y	-	Y	-	Y	-
PY00311	-	-	-	hypothetical protein	MALPY00089: 2,330 - 2,401 (-)	-	-	-	Y	-	Y	-	Y	-
PY00453	PB102091.00.0	-	-	hypothetical protein	MALPY00127: 9,449 - 9,769 (+)	PB_RP0914	PC_RP2517	-	Y	-	Y	-	Y	-
PY00820	-	-	-	hypothetical protein	MALPY00219: 11,739 - 11,906 (-)	-	-	-	Y	-	Y	-	Y	-
PY00966	-	-	-	hypothetical protein	MALPY00261: 2,881 - 3,919 (-)	PB_RP0576	PC_RP2319	MAL12	Y	-	Y	-	Y	-
PY01053	-	-	-	hypothetical protein	MALPY00278: 1,202 - 1,285 (-)	-	-	-	Y	-	Y	-	Y	-
PY01319	-	-	-	hypothetical protein	MALPY00347: 2,126 - 2,219 (-)	-	-	-	Y	-	Y	-	Y	-
PY01332	-	-	-	hypothetical protein	MALPY00351: 2,104 - 2,212 (+)	-	-	-	Y	-	Y	-	Y	-
PY01339	-	PC107725.00.0	-	hypothetical protein	MALPY00354: 11,522 - 11,647 (+)	-	-	-	Y	-	Y	-	Y	-
PY01342	-	-	-	hypothetical protein	MALPY00354: 20,981 - 22,056 (+)	-	-	-	Y	-	Y	-	Y	-
PY01593	-	-	-	hypothetical protein	MALPY00428: 7,908 - 8,202 (-)	-	-	-	Y	-	Y	-	Y	-
PY01642	-	-	-	hypothetical protein	MALPY00442: 4,234 - 4,644 (+)	PB_RP0216	PC_RP3181	MAL14	Y	-	Y	-	Y	-
PY01694	-	-	-	hypothetical protein	MALPY00457: 5,573 - 5,768 (+)	-	-	-	Y	-	Y	-	Y	-
PY01924	-	-	-	hypothetical protein	MALPY00527: 1,077 - 1,169 (+)	-	-	-	Y	-	Y	-	Y	-
PY01928	-	-	-	hypothetical protein	MALPY00527: 7,414 - 7,521 (+)	-	-	-	Y	-	Y	-	Y	-
PY02182	-	-	-	hypothetical protein	MALPY00600: 10,281 - 10,411 (+)	-	-	-	Y	-	Y	-	Y	-
PY02363	-	-	-	hypothetical protein	MALPY00647: 2,868 - 2,965 (-)	-	-	-	Y	-	Y	-	Y	-
PY02597	-	-	-	hypothetical protein	MALPY00714: 7,822 - 8,210 (+)	PB_RP3971	-	-	Y	-	Y	-	Y	-
PY02630	-	-	-	hypothetical protein	MALPY00726: 2,439 - 2,505 (+)	-	-	-	Y	-	Y	-	Y	-
PY02655	-	-	-	hypothetical protein	MALPY00734: 3,248 - 3,467 (+)	-	-	-	Y	-	Y	-	Y	-
PY02742	PB104976.00.0	PC109417.00.0, PC100306.00.0, PC100307.00.0	-	hypothetical protein	MALPY00762: 7,437 - 8,613 (+)	PB_RP2341	PC_RP0196	-	Y	-	Y	-	Y	-
PY03054	-	PC000521.00.0	-	hypothetical protein	MALPY00868: 10,339 - 11,426 (+)	PB_RP2012	PC_RP0601	MAL11	Y	-	Y	-	Y	-
PY03067	-	-	-	hypothetical protein	MALPY00872: 1,926 - 2,040 (+)	-	-	-	Y	-	Y	-	Y	-
PY03172	-	-	-	hypothetical protein	MALPY00901: 2,758 - 3,323 (+)	-	-	-	Y	-	Y	-	Y	-
PY03291	-	PC106813.00.0	-	hypothetical protein	MALPY00939: 423 - 1,213 (-)	-	-	-	Y	-	Y	-	Y	-
PY03340	-	PC108271.00.0, PC107487.00.0, PC101433.00.0, PC404087.00.0, PC403777.00.0, PC101818.00.0, PC107028.00.0, PC104414.00.0, PC109237.00.0, PC108229.00.0, PC102089.00.0, PC404185.00.0, PB105405.00.0, PB105350.00.0, PB104816.00.0	-	hypothetical protein	MALPY00957: 1,840 - 3,657 (+)	PB_RP2528	PC_RP1133	-	Y	-	Y	-	Y	-
PY03483	-	-	-	hypothetical protein	MALPY01006: 5,449 - 5,832 (-)	-	-	-	Y	-	Y	-	Y	-
PY03700	-	-	-	hypothetical protein	MALPY01088: 2,503 - 2,691 (+)	-	-	-	Y	-	Y	-	Y	-
PY04650	-	-	-	hypothetical protein	MALPY01430: 7,752 - 7,844 (+)	-	-	-	Y	-	Y	-	Y	-
PY04654	-	PC401259.00.0	-	hypothetical protein	MALPY01430: 9,964 - 10,447 (+)	PB_PH2244	PC_PH1156	MAL7	Y	-	Y	-	Y	-
PY04657	-	-	-	hypothetical protein	MALPY01430: 12,272 - 12,763 (+)	-	-	-	Y	-	Y	-	Y	-
PY04658	-	PC100999.00.0	-	hypothetical protein	MALPY01430: 13,416 - 13,806 (+)	-	-	-	Y	-	Y	-	Y	-
PY04815	-	-	-	hypothetical protein	MALPY01491: 2,500 - 2,604 (+)	-	-	-	Y	-	Y	-	Y	-

Py_ID	Pb_ID	Pc_ID	Pf_ID	Py_name	Py_location	Pb_blastn	Pc_blastn	Pf_blastn	Pb_cgh	Pb_S	Pc_cgh	Pc_S	Py_cgh	Py_S
PY05038	-	-	-	hypothetical protein	MALPY01582: 2,349 - 2,542 (-)	-	-	-	Y	-	Y	-	Y	-
PY05134	-	-	-	hypothetical protein	MALPY01617: 608 - 721 (+)	-	-	-	Y	-	Y	-	Y	-
PY05256	-	-	-	hypothetical protein	MALPY01664: 210 - 681 (+)	PB_RP2012	PC_RP2555	MAL11	Y	-	Y	-	Y	-
PY05277	-	-	-	hypothetical protein	MALPY01670: 2,939 - 3,029 (-)	-	-	-	Y	-	Y	-	Y	-
PY05302	-	-	-	hypothetical protein	MALPY01677: 2,230 - 2,322 (-)	-	-	-	Y	-	Y	-	Y	-
PY05327	-	-	-	hypothetical protein	MALPY01685: 10,513 - 10,749 (+)	-	-	-	Y	-	Y	-	Y	-
PY05329	-	-	-	hypothetical protein	MALPY01686: 1,074 - 1,778 (-)	-	-	-	Y	-	Y	-	Y	-
PY05376	-	-	-	hypothetical protein	MALPY01704: 2,561 - 2,851 (-)	-	-	-	Y	-	Y	-	Y	-
PY05477	-	-	-	hypothetical protein	MALPY01748: 11,905 - 12,975 (-)	PB_RP2068	-	MAL14	Y	-	Y	-	Y	-
PY05483	PB001073.00.0	-	-	hypothetical protein	MALPY01751: 14,837 - 15,913 (+)	PB_RP0861	-	MAL11	Y	-	Y	-	Y	-
PY05623	-	-	-	hypothetical protein	MALPY01810: 1,732 - 4,302 (+)	-	-	-	Y	-	Y	-	Y	-
PY05929	-	-	-	hypothetical protein	MALPY01954: 5,441 - 5,541 (-)	-	-	-	Y	-	Y	-	Y	-
PY06219	PB105442.00.0	-	-	hypothetical protein	MALPY02087: 1,256 - 1,537 (-)	-	-	-	Y	-	Y	-	Y	-
PY06363	PB402608.00.0, PB401466.00.0, PB406032.00.0, PB402902.00.0, PB402109.00.0	PC400472.00.0	-	hypothetical protein	MALPY02146: 168 - 2,051 (+)	PB_PH1825	PC_PH1433	-	Y	-	Y	-	Y	-
PY06364	-	-	-	hypothetical protein	MALPY02146: 2,060 - 2,362 (-)	PB_PH2059	PC_PH1433	-	Y	-	Y	-	Y	-
PY06433	-	-	-	hypothetical protein	MALPY02185: 8,709 - 9,203 (-)	-	-	-	Y	-	Y	-	Y	-
PY06563	-	-	-	hypothetical protein	MALPY02241: 3,783 - 3,981 (+)	-	-	-	Y	-	Y	-	Y	-
PY06564	-	-	-	hypothetical protein	MALPY02241: 3,987 - 4,079 (+)	-	-	-	Y	-	Y	-	Y	-
PY06629	-	-	-	hypothetical protein	MALPY02275: 222 - 438 (+)	-	-	-	Y	-	Y	-	Y	-
PY06938	-	-	-	hypothetical protein	MALPY02446: 2,086 - 2,284 (+)	-	-	-	Y	-	Y	-	Y	-
PY07234	-	-	-	hypothetical protein	MALPY02613: 1,688 - 1,879 (-)	-	-	-	Y	-	Y	-	Y	-
PY07279	-	-	-	hypothetical protein	MALPY02635: 2,406 - 2,905 (+)	-	-	-	Y	-	Y	-	Y	-
PY07299	PB100178.00.0	-	-	hypothetical protein	MALPY02648: 1,432 - 1,935 (-)	PB_PH2762	PC_PH2449	-	Y	-	Y	-	Y	-
PY07332	-	PC103610.00.0	-	hypothetical protein	MALPY02672: 1,685 - 2,224 (+)	-	-	-	Y	-	Y	-	Y	-
PY07366	PB402902.00.0, PB402109.00.0, PB401466.00.0, PB402608.00.0, PB406032.00.0	PC400472.00.0	-	hypothetical protein	MALPY02691: 108 - 1,541 (+)	PB_PH2638	PC_PH1433	-	Y	-	Y	-	Y	-
PY07377	PB401969.00.0	-	-	hypothetical protein	MALPY02696: 1,581 - 1,930 (+)	PB_PH2658	PC_PH1433	-	Y	-	Y	-	Y	-
PY07475	-	-	-	hypothetical protein	MALPY02763: 2,364 - 2,439 (-)	-	-	-	Y	-	Y	-	Y	-
PY07580	-	-	-	hypothetical protein	MALPY02820: 1,083 - 1,491 (+)	-	-	-	Y	-	Y	-	Y	-
PY07602	PB402902.00.0, PB402608.00.0, PB401466.00.0, PB402109.00.0, PB406032.00.0	PC400472.00.0	-	hypothetical protein	MALPY02829: 36 - 1,336 (+)	PB_PH2638	PC_PH1433	-	Y	-	Y	-	Y	-
PY07612	-	-	-	hypothetical protein	MALPY02834: 49 - 140 (+)	-	-	-	Y	-	Y	-	Y	-
PY07640	-	-	-	hypothetical protein	MALPY02851: 2,538 - 2,606 (+)	-	-	-	Y	-	Y	-	Y	-
PY07650	PB107308.00.0	-	-	hypothetical protein	MALPY02860: 1,780 - 2,127 (+)	PB_RP3196	PC_RP4247	MAL11	Y	-	Y	-	Y	-
PY07670	PB301146.00.0, PB300349.00.0, PB300555.00.0, PB300337.00.0	-	-	hypothetical protein	MALPY02873: 1 - 1,518 (-)	PB_PH1821	-	-	Y	-	Y	-	Y	-
PY07671	-	-	-	hypothetical protein	MALPY02873: 1,621 - 1,864 (-)	PB_PH1808	-	-	Y	-	Y	-	Y	-
PY07673	PB404704.00.0, PB402568.00.0, PB402012.00.0, PB405216.00.0	PC401560.00.0	-	hypothetical protein	MALPY02874: 333 - 839 (+)	PB_PH1541	PC_PH4430	-	Y	-	Y	-	Y	-
PY07686	-	-	-	hypothetical protein	MALPY02881: 806 - 2,235 (+)	-	-	-	Y	-	Y	-	Y	-
PY07691	-	-	-	hypothetical protein	MALPY02884: 5,316 - 6,725 (+)	-	-	-	Y	-	Y	-	Y	-
PY07696	-	-	-	hypothetical protein	MALPY02887: 25 - 393 (+)	-	-	-	Y	-	Y	-	Y	-
PY07706	-	-	-	hypothetical protein	MALPY02893: 3,122 - 3,269 (+)	-	-	-	Y	-	Y	-	Y	-
PY07730	-	-	-	hypothetical protein	MALPY02906: 2,177 - 2,871 (-)	-	-	-	Y	-	Y	-	Y	-
PY07760	-	-	-	hypothetical protein	MALPY02923: 28 - 774 (-)	-	-	-	Y	-	Y	-	Y	-
PY07762	-	-	-	hypothetical protein	MALPY02924: 31 - 592 (+)	-	-	-	Y	-	Y	-	Y	-
PY07763	-	-	-	hypothetical protein	MALPY02924: 1,087 - 1,848 (+)	-	-	-	Y	-	Y	-	Y	-
PY07766	-	-	-	hypothetical protein	MALPY02926: 148 - 774 (+)	-	-	-	Y	-	Y	-	Y	-
PY07768	-	-	-	hypothetical protein	MALPY02926: 2,023 - 2,688 (+)	-	-	-	Y	-	Y	-	Y	-
PY07769	-	-	-	hypothetical protein	MALPY02926: 2,785 - 3,006 (-)	-	-	-	Y	-	Y	-	Y	-
PY07773	-	-	-	hypothetical protein	MALPY02927: 1,378 - 2,050 (+)	-	-	-	Y	-	Y	-	Y	-
PY07776	-	-	-	hypothetical protein	MALPY02929: 1,495 - 3,140 (-)	-	-	-	Y	-	Y	-	Y	-
PY07783	-	-	-	hypothetical protein	MALPY02933: 213 - 887 (-)	-	-	-	Y	-	Y	-	Y	-
PY07799	-	-	-	hypothetical protein	MALPY02940: 398 - 3,650 (+)	-	-	-	Y	-	Y	-	Y	-

Py_ID	Pb_ID	Pc_ID	Pf_ID	Py_name	Py_location	Pb_tblastn	Pc_tblastn	Pf_tblastn	Pb_cgh	Pb_S	Pc_cgh	Pc_S	Py_cgh	Py_S
PY07802	-	-	-	hypothetical protein	MALPY02941: 2,508 - 2,705 (-)	-	-	-	Y	-	Y	-	Y	-
PY07808	-	-	-	hypothetical protein	MALPY02944: 35 - 874 (-)	-	-	-	Y	-	Y	-	Y	-
PY07809	-	-	-	hypothetical protein	MALPY02944: 884 - 1,204 (+)	-	-	-	Y	-	Y	-	Y	-
PY07817	-	-	-	hypothetical protein	MALPY02948: 575 - 2,808 (+)	-	-	-	Y	-	Y	-	Y	-
PY07828	-	-	-	hypothetical protein	MALPY02958: 1 - 2,007 (+)	PB_PH2715	PC_PH1729	-	Y	-	Y	-	Y	-
PY07698	-	-	-	nif-specific regulatory protein	MALPY02888: 1 - 2,890 (+)	-	-	-	Y	-	Y	-	Y	-
PY07288	-	-	-	RNase H, putative	MALPY02641: 1 - 2,928 (-)	-	-	-	Y	-	Y	-	Y	-
PY07613	-	-	-	titin-L1 fusion protein-related	MALPY02834: 269 - 1,581 (-)	PB_PH2204	PC_PH0450	-	Y	-	Y	-	Y	-
PY00129	PB000020.02.0	PC000997.02.0	-	TPR Domain, putative	MALPY00036: 47,470 - 51,333 (-)	PB_RP2018	PC_RP2643	MAL14	Y	-	Y	-	Y	-
PY03737	PB300508.00.0	PC000321.01.0	-	ubiquitin-conjugating enzyme-related	MALPY01101: 12,187 - 12,733 (-)	PB_PH2145	PC_RP1366	MAL5	Y	-	Y	-	Y	-
PY07374	-	-	-	Unknown protein	MALPY02696: 1 - 920 (+)	PB_PH2692	PC_PH1433	-	Y	-	Y	-	Y	-

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY04930	235 kDa rhostry protein	0.188726896	Y	-	-	nucleotide binding	PFL2520w, PF13_0198, MAL13P1.176	T	
PY00319	aconitate hydratase 1	0.1953125	-	-	-	-	PF13_0229	T	
PY01581	apical membrane antigen-1	0.263161465	Y	Y	-	-	PF11_0344	T	
PY01120	Arabidopsis thaliana BRAHMA ortholog-related	0.087581823	-	-	-	-	PFB0730w	S	
PY02350	asparagine-rich protein	0.247372473	-	-	-	-	PF13_0283	S	
PY03561	asparagine-rich protein	0.198816599	-	-	-	-	-	R	
PY02085	aspartyl protease-like	0.449359186	Y	Y	-	aspartic-type endopeptidase activity, pepsin A activity	PF13_0133	T	
PY05495	AT hook motif, putative	0.462582384	Y	-	-	ATP binding	PFC0325c	T	
PY05109	ATPase associated with chromosome architecture/replication	0.486252979	-	-	-	-	PF14_0466	R	
PY00253	CCAAT-box DNA binding protein subunit B	0.211340206	-	-	-	-	PF11315c	-	
PY00386	CCAAT-box DNA binding protein subunit B	0.436754044	Y	-	-	-	PF14_0154	S	
PY00689	CCAAT-box DNA binding protein subunit B	0.49816213	-	-	-	-	PF14_0271	-	
PY00912	CCAAT-box DNA binding protein subunit B	0.476601258	-	-	-	-	-	R	
PY02393	CCAAT-box DNA binding protein subunit B	0.484081796	-	-	-	-	PF11150w	R	
PY05072	CCAAT-box DNA binding protein subunit B	0.268359938	Y	-	-	-	PFF0870w	T	
PY05674	CCAAT-box DNA binding protein subunit B	0.375909186	-	-	-	-	-	T	
PY02289	CCAAT-box DNA binding protein subunit B, putative	0.284247001	-	-	-	-	PFL0530c	R	
PY00653	chromosome segregation protein smc1	0.070084697	-	-	-	-	PF11_0317	T	
PY02565	cir1 protein, putative	0.126632008	Y	-	-	-	-	-	
PY04787	cir1 protein, putative	0.066859165	Y	-	-	-	-	-	
PY07826	cir1 protein, putative	0.325172375	Y	-	-	-	-	R	
PY02757	cir2 protein	0.288997004	Y	-	-	-	-	-	
PY01337	Coronin	0.474945943	-	-	-	-	PFL2460w	-	
PY03618	coronin binding protein	0.261461516	-	-	-	-	PF14_0673	R	
PY02252	deoxyribose-phosphate aldolase	0.484817309	-	-	-	-	PF10_0210	S	
PY03839	Domain of unknown function, putative	0.058335432	-	-	-	-	PF08_0055	R	
PY04724	Drosophila melanogaster BcDNA.GH02833	0.32669579	-	-	-	-	MAL7P1.113	R	
PY07506	Drosophila melanogaster CG12781 gene product	0.251256941	Y	Y	-	-	-	S	
PY04549	Drosophila melanogaster CG15040 gene product	0.313961708	Y	-	-	-	-	T	
PY06583	Endonuclease/Exonuclease/phosphatase family	0.425161979	-	-	-	-	PFA0350w	R	
PY00637	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	0.479518116	-	-	-	-	PF08_0012	T	
PY06125	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	0.42223404	-	-	-	-	PF13_0056	R	
PY07447	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	0.479101151	-	-	-	-	-	T	
PY05489	histidine kinase DhkG	0.231107984	-	-	-	-	PF13_0237	T	
PY05189	homeobox-containing protein	0.49278955	-	-	-	-	PF10_0045	R	
PY05572	hydrolase, alpha/beta fold family, putative	0.372326162	-	-	-	-	PF14_0015, PF08_0022, PFC0065c	S	
PY00007	hypothetical protein	0.389209796	-	-	-	-	PF10_0075	R	
PY00178	hypothetical protein	0.194108369	Y	Y	-	-	PFL0790w	R	
PY00237	hypothetical protein	0.337200779	-	Y	-	-	-	-	
PY00307	hypothetical protein	0.377734382	Y	-	-	-	-	-	
PY00330	hypothetical protein	0.336667749	Y	-	-	-	-	-	
PY00388	hypothetical protein	0.463493841	-	-	-	-	PFE0415w	-	
PY00408	hypothetical protein	0.379948327	-	-	-	-	-	R	
PY00441	hypothetical protein	0.414209321	-	-	-	-	PFB0475c	T	
PY00492	hypothetical protein	0.477637014	-	-	-	-	-	R	<100nt
PY00575	hypothetical protein	0.225694457	Y	Y	-	-	-	-	
PY00576	hypothetical protein	0.381926774	Y	-	NRILGDV	-	-	-	
PY00596	hypothetical protein	0.390510258	Y	Y	-	-	PF07_0070	R	
PY00600	hypothetical protein	0.334071537	Y	-	-	-	PF10_0199	R	
PY00636	hypothetical protein	0.318472315	Y	-	-	-	-	T	
PY00647	hypothetical protein	0.375452813	-	-	-	-	-	R	
PY00694	hypothetical protein	0.134004638	-	-	-	-	PFE1435c	R	
PY00723	hypothetical protein	0.19765337	-	-	-	-	PF11_0155	R	
PY00813	hypothetical protein	0.491453472	-	Y	-	-	-	S	<100nt
PY00871	hypothetical protein	0.216720202	-	-	-	-	-	-	
PY00890	hypothetical protein	0.297404577	Y	-	-	metal ion transmembrane transporter activity	PF11_0210	-	
PY00936	hypothetical protein	0.308474054	-	-	-	-	-	T	
PY00953	hypothetical protein	0.17056961	-	-	-	-	PF10780w	R	

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY00958	hypothetical protein	0.43124066	-	-	-	-	-	-	<100nt
PY00980	hypothetical protein	0.144820946	-	-	-	-	-	-	
PY00991	hypothetical protein	0.441519194	-	-	-	-	-	S	<100nt
PY01003	hypothetical protein	0.245084998	-	-	-	-	-	T	
PY01006	hypothetical protein	0.367198783	Y	-	-	-	-	T	
PY01012	hypothetical protein	0.474758383	-	-	-	-	PF10_0358	R	
PY01019	hypothetical protein	0.273764669	-	-	-	-	-	-	
PY01057	hypothetical protein	0.333925098	-	-	-	-	-	-	
PY01137	hypothetical protein	0.331241324	-	-	-	-	PFL1030w	S	
PY01160	hypothetical protein	0.29207903	Y	-	-	-	-	R	
PY01176	hypothetical protein	0.075542889	Y	Y	-	-	-	-	
PY01219	hypothetical protein	0.49936102	-	-	-	-	-	T	
PY01236	hypothetical protein	0.187569472	Y	Y	-	-	-	-	
PY01262	hypothetical protein	0.242942337	-	-	-	-	-	T	<100nt
PY01280	hypothetical protein	0.388287033	-	-	-	-	PFE1410c	T	
PY01308	hypothetical protein	0.435633619	-	-	-	-	PF07_0119	R	
PY01326	hypothetical protein	0.111179162	-	-	-	-	-	-	
PY01401	hypothetical protein	0.432817399	-	-	-	-	-	S	
PY01405	hypothetical protein	0.457651482	-	-	-	-	-	-	
PY01416	hypothetical protein	0.412732238	-	-	-	-	-	-	
PY01443	hypothetical protein	0.283754324	-	-	-	-	-	R	
PY01507	hypothetical protein	0.367737899	Y	-	-	-	PF13_0202	S	
PY01588	hypothetical protein	0.290156951	-	-	-	-	-	R	
PY01602	hypothetical protein	0.115745159	Y	Y	NRILADA	-	-	T	
PY01650	hypothetical protein	0.09371254	Y	-	-	-	-	-	
PY01673	hypothetical protein	0.276612474	-	-	-	-	-	-	
PY01686	hypothetical protein	0.370226234	Y	-	-	-	-	-	
PY01747	hypothetical protein	0.148357261	-	-	-	-	-	-	
PY01772	hypothetical protein	0.377627484	-	-	-	-	-	T	
PY01809	hypothetical protein	0.363633976	Y	-	-	-	PFL2390c	R	
PY01838	hypothetical protein	0.391275634	-	-	-	-	-	-	
PY01843	hypothetical protein	0.348016463	-	Y	-	-	-	-	
PY01875	hypothetical protein	0.409281993	-	-	-	-	-	T	
PY01975	hypothetical protein	0.239344087	Y	-	-	-	-	-	
PY01978	hypothetical protein	0.114356127	Y	-	-	-	-	T	
PY01985	hypothetical protein	0.140164506	Y	-	-	-	-	-	
PY02031	hypothetical protein	0.377308493	-	-	-	-	PFF1160w	R	
PY02053	hypothetical protein	0.36828481	-	-	-	-	MAL7P1.123	S	
PY02078	hypothetical protein	0.344163752	Y	-	-	-	PF10_0290	T	
PY02103	hypothetical protein	0.233818876	-	-	-	-	PF13_0183	-	
PY02133	hypothetical protein	0.404094225	-	-	-	-	-	T	
PY02152	hypothetical protein	0.396349484	Y	-	-	-	MAL8P1.13, PF11_0172	-	
PY02184	hypothetical protein	0.053544365	Y	-	-	-	-	-	<100nt
PY02199	hypothetical protein	0.456421645	Y	-	-	binding	PFD0720w	S	
PY02259	hypothetical protein	0.418874886	-	-	-	-	-	-	
PY02301	hypothetical protein	0.311836748	-	Y	-	-	PF14_0344	-	
PY02313	hypothetical protein	0.201353928	-	-	-	-	-	-	
PY02356	hypothetical protein	0.084371335	-	Y	-	-	-	-	
PY02366	hypothetical protein	0.348915055	-	-	-	-	-	T	
PY02434	hypothetical protein	0.326097952	-	-	-	-	PF11_0388	S	
PY02465	hypothetical protein	0.434691785	-	-	-	-	PFD1030c	T	
PY02467	hypothetical protein	0.173762008	-	-	-	-	-	T	
PY02468	hypothetical protein	0.178571429	-	-	-	-	-	-	
PY02470	hypothetical protein	0.338286995	-	-	-	-	PFL1190c	-	
PY02489	hypothetical protein	0.316732827	-	-	-	-	PF11050c	T	
PY02546	hypothetical protein	0.390737896	-	-	-	-	-	T	
PY02628	hypothetical protein	0.258943043	-	-	-	-	-	-	<100nt
PY02729	hypothetical protein	0.22556391	-	-	-	-	-	-	
PY02789	hypothetical protein	0.433926333	-	-	-	-	PF11_0078	-	
PY02792	hypothetical protein	0.193602215	-	-	-	-	PFE0240c	T	
PY02895	hypothetical protein	0.059303563	-	-	-	-	PFE1060c	R	

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY02934	hypothetical protein	0.427348591	Y	-	-	-	PFE0260w	T	
PY02954	hypothetical protein	0.280202829	-	-	-	-	-	-	
PY03061	hypothetical protein	0.485024321	Y	-	-	-	-	-	
PY03084	hypothetical protein	0.175580358	-	-	-	-	PF10_0146	-	
PY03257	hypothetical protein	0.238409819	-	-	-	-	PFE0455w	T	
PY03260	hypothetical protein	0.144180628	-	-	-	-	-	S	
PY03297	hypothetical protein	0.484509899	Y	-	-	-	-	S	
PY03322	hypothetical protein	0.267079504	Y	Y	-	-	-	S	
PY03340	hypothetical protein	0.043442739	Y	-	-	-	-	-	
PY03347	hypothetical protein	0.427909537	-	-	-	-	-	-	
PY03352	hypothetical protein	0.42085738	-	-	-	-	-	-	
PY03369	hypothetical protein	0.44293253	Y	Y	-	-	PF14_0089	S	
PY03405	hypothetical protein	0.243933825	Y	Y	NRILGDV	-	PF11_0073	R	
PY03430	hypothetical protein	0.182946738	Y	Y	-	-	-	S	
PY03458	hypothetical protein	0.286092967	-	-	-	-	-	T	
PY03467	hypothetical protein	0.375080895	-	-	-	-	PFC0701w	R	
PY03474	hypothetical protein	0.218143661	-	-	-	-	PF11_0458	-	
PY03484	hypothetical protein	0.487177702	Y	-	-	protein binding	-	S	<100nt
PY03507	hypothetical protein	0.279273706	Y	-	-	-	PFL0505c	S	
PY03568	hypothetical protein	0.063969287	-	-	-	-	-	R	
PY03586	hypothetical protein	0.295582908	-	-	-	-	-	-	
PY03608	hypothetical protein	0.250295123	-	-	-	-	MAL8P1.20	T	
PY03621	hypothetical protein	0.169521224	-	-	-	-	-	-	<100nt
PY03624	hypothetical protein	0.041615592	-	-	-	-	-	-	<100nt
PY03644	hypothetical protein	0.414340232	-	-	-	-	PF14_0659	T	
PY03657	hypothetical protein	0.093601043	Y	Y	-	-	-	R	
PY03708	hypothetical protein	0.389063568	Y	-	-	-	-	R	
PY03759	hypothetical protein	0.40374265	-	-	-	-	-	R	
PY03836	hypothetical protein	0.369464084	-	-	-	-	-	R	
PY03848	hypothetical protein	0.086627988	Y	-	-	-	-	-	
PY03950	hypothetical protein	0.18083261	-	-	-	-	-	R	
PY03951	hypothetical protein	0.083472965	-	-	-	-	-	T	
PY03952	hypothetical protein	0.158979174	Y	-	-	-	-	-	
PY03968	hypothetical protein	0.271069025	Y	-	-	-	-	R	
PY03975	hypothetical protein	0.4857117043	-	-	-	-	-	-	<100nt
PY03993	hypothetical protein	0.392984054	Y	-	-	-	PFE0580w	T	
PY03997	hypothetical protein	0.063708694	-	-	-	-	-	R	
PY04109	hypothetical protein	0.447354193	Y	-	-	-	-	-	
PY04110	hypothetical protein	0.492631889	-	-	-	-	-	-	
PY04123	hypothetical protein	0.268393569	Y	-	-	-	-	-	
PY04137	hypothetical protein	0.353843883	-	Y	-	-	PF14_0792	S	
PY04152	hypothetical protein	0.244394836	-	Y	-	-	-	-	
PY04210	hypothetical protein	0.412658782	Y	-	-	-	-	R	
PY04227	hypothetical protein	0.243772055	-	-	-	-	PFF0990c	T	
PY04279	hypothetical protein	0.209309761	-	-	-	-	-	S	
PY04316	hypothetical protein	0.337571052	Y	-	-	-	-	R	
PY04333	hypothetical protein	0.45371624	-	Y	-	-	-	R	
PY04355	hypothetical protein	0.384750173	Y	-	-	-	PF11295c, PFB0465c	S	
PY04384	hypothetical protein	0.247781402	-	-	-	-	-	-	
PY04397	hypothetical protein	0.155943399	-	-	-	-	-	R	
PY04398	hypothetical protein	0.494097902	Y	Y	-	-	-	-	
PY04420	hypothetical protein	0.260083489	Y	Y	-	-	-	-	
PY04550	hypothetical protein	0.120040442	-	-	-	-	PFA0325w	T	
PY04555	hypothetical protein	0.358938547	-	-	-	-	-	-	
PY04683	hypothetical protein	0.221740654	Y	-	-	-	PF08_0090	-	
PY04734	hypothetical protein	0.418129803	Y	-	NRILGDV	-	-	-	
PY04827	hypothetical protein	0.269411547	-	-	-	-	-	-	
PY04866	hypothetical protein	0.335267466	Y	Y	-	-	PFD0555c	R	
PY05012	hypothetical protein	0.467778729	-	-	-	-	-	-	
PY05013	hypothetical protein	0.45506429	-	-	-	-	-	-	
PY05041	hypothetical protein	0.22811988	-	-	-	-	PFB0280w	R	

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Preptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY05097	hypothetical protein	0.285727503	-	-	-	-	-	-	<100nt
PY05134	hypothetical protein	0.458536601	-	-	-	-	-	-	
PY05167	hypothetical protein	0.235696053	-	-	-	-	PF08_0058	S	
PY05195	hypothetical protein	0.146980104	-	-	-	-	PF11190w	R	
PY05202	hypothetical protein	0.284188317	Y	-	-	-	PF10_0188	R	
PY05219	hypothetical protein	0.442072807	-	-	-	-	MAL13P1.333	R	
PY05240	hypothetical protein	0.445269976	Y	-	-	-	-	-	
PY05291	hypothetical protein	0.244446026	-	-	-	-	-	-	<100nt
PY05351	hypothetical protein	0.424426808	-	-	-	-	-	-	<100nt
PY05456	hypothetical protein	0.152806145	-	-	-	-	-	R	<100nt
PY05541	hypothetical protein	0.347113846	Y	Y	-	protein-S-isoprenylcysteine O-methyltransferase activity	PFL1780w	R	
PY05594	hypothetical protein	0.301665077	-	-	-	-	PF14_0047	-	
PY05635	hypothetical protein	0.195261981	-	-	-	-	-	R	
PY05687	hypothetical protein	0.358123953	-	-	-	-	-	-	
PY05720	hypothetical protein	0.248786638	-	-	-	-	-	R	<100nt
PY05775	hypothetical protein	0.270409747	-	-	-	-	PFA0305c	T	
PY05783	hypothetical protein	0.162515781	-	-	-	-	PF10_0251	T	
PY05795	hypothetical protein	0.332129191	-	-	-	-	PF14_0252	T	
PY05813	hypothetical protein	0.428254385	-	-	-	-	PF10_0243	S	
PY05917	hypothetical protein	0.481758931	-	-	-	-	-	S	
PY05998	hypothetical protein	0.317603646	-	-	-	-	-	T	
PY06021	hypothetical protein	0.182617005	Y	Y	-	-	-	-	
PY06092	hypothetical protein	0.458500365	-	-	-	-	-	R	
PY06103	hypothetical protein	0.429702572	Y	-	-	-	-	-	
PY06159	hypothetical protein	0.308970931	Y	-	-	-	-	R	
PY06193	hypothetical protein	0.128418307	-	-	-	-	PF08_0016	R	
PY06229	hypothetical protein	0.209823408	-	-	-	-	-	R	
PY06317	hypothetical protein	0.220874221	Y	-	-	-	-	T	
PY06353	hypothetical protein	0.194487923	-	-	-	-	-	S	
PY06384	hypothetical protein	0.101965954	-	Y	-	-	-	-	<100nt
PY06510	hypothetical protein	0.474388179	-	-	-	-	-	-	
PY06570	hypothetical protein	0.316881697	-	-	-	-	-	-	<100nt
PY06579	hypothetical protein	0.143761913	-	-	-	-	-	S	
PY06581	hypothetical protein	0.125703094	Y	Y	NRILGDV	-	-	R	
PY06594	hypothetical protein	0.411663148	-	-	-	-	-	R	<100nt
PY06617	hypothetical protein	0.26386137	Y	-	-	-	-	-	
PY06664	hypothetical protein	0.345995212	Y	-	-	-	PF10_0205	T	
PY06690	hypothetical protein	0.376215929	-	-	-	-	-	T	
PY06731	hypothetical protein	0.322557754	Y	-	-	-	-	-	
PY06781	hypothetical protein	0.294425087	-	-	-	-	PFL0445w	S	
PY06793	hypothetical protein	0.338886461	-	-	-	-	PF14_0200	S	
PY06815	hypothetical protein	0.185422174	-	-	-	-	-	R	
PY06850	hypothetical protein	0.071834938	-	-	-	-	-	T	
PY06879	hypothetical protein	0.499137579	-	-	-	-	-	-	
PY06881	hypothetical protein	0.204395126	Y	-	-	-	-	R	
PY06892	hypothetical protein	0.405363288	-	-	-	-	-	-	
PY06940	hypothetical protein	0.496500041	-	-	-	-	-	-	<100nt
PY06958	hypothetical protein	0.249197612	-	-	-	-	PF11_0269	-	
PY06989	hypothetical protein	0.366662886	-	-	-	-	-	R	<100nt
PY07087	hypothetical protein	0.407620455	Y	Y	-	-	-	-	
PY07091	hypothetical protein	0.12713344	-	-	-	-	-	-	
PY07109	hypothetical protein	0.14272402	-	-	-	-	-	T	
PY07133	hypothetical protein	0.222443457	-	-	-	-	-	-	<100nt
PY07167	hypothetical protein	0.330304671	Y	Y	-	-	-	-	<100nt
PY07170	hypothetical protein	0.417095754	-	-	-	-	-	R	
PY07188	hypothetical protein	0.261185246	-	-	-	-	-	-	
PY07213	hypothetical protein	0.372209149	-	-	-	-	-	-	<100nt
PY07217	hypothetical protein	0.241242804	Y	-	-	-	-	T	
PY07225	hypothetical protein	0.326744603	-	-	-	-	PF14_0680	R	
PY07231	hypothetical protein	0.114978059	Y	-	-	-	-	-	
PY07234	hypothetical protein	0.287119747	-	-	-	-	-	-	
PY07241	hypothetical protein	0.423781836	-	-	-	-	-	-	<100nt

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY07242	hypothetical protein	0.474070092	Y	-	-	-	-	-	
PY07282	hypothetical protein	0.21686747	-	-	-	-	-	-	
PY07286	hypothetical protein	0.285197993	-	-	-	-	-	-	
PY07362	hypothetical protein	0.479659066	-	-	-	-	MAL13P1.327	R	
PY07378	hypothetical protein	0.053104679	-	-	-	-	-	T	
PY07393	hypothetical protein	0.187888275	-	-	-	-	-	R	<100nt
PY07451	hypothetical protein	0.077285485	-	-	-	-	-	R	<100nt
PY07458	hypothetical protein	0.290481596	-	Y	-	-	-	-	
PY07478	hypothetical protein	0.204049062	-	-	-	-	-	-	
PY07534	hypothetical protein	0.298165698	-	-	-	-	-	-	<100nt
PY07548	hypothetical protein	0.224075834	-	-	-	-	-	-	<100nt
PY07559	hypothetical protein	0.153392764	Y	Y	-	-	-	-	
PY07562	hypothetical protein	0.287003482	Y	-	-	-	-	T	
PY07608	hypothetical protein	0.216201258	Y	Y	-	-	-	-	
PY07633	hypothetical protein	0.217667375	Y	-	-	-	-	-	
PY07651	hypothetical protein	0.359178774	-	-	-	-	PF11_0476	S	
PY05438	Leishmania major ppg3	0.066787768	Y	-	-	-	-	R	
PY03256	lipase precursor gehm. 6/101	0.362692079	-	-	-	-	PFE0460c	T	
PY07153	Maf-like protein, putative	0.456278359	-	-	-	-	PF10310w	T	
PY02239	mature-parasite-infected erythrocyte surface antigen	0.270383938	-	-	-	-	-	T	
PY06415	merozoite surface protein precursor	0.19675253	Y	-	-	-	PFE0120c	R	
PY02559	methionine aminopeptidase, type II, putative	0.22782859	-	-	-	-	PF14_0327	R	
PY05984	moaA / nifB / pqqE family, putative	0.343341646	-	-	-	-	PFE1240w	R	
PY05981	MORN repeat, putative	0.255719477	-	-	-	-	PF10_0306	S	
PY04548	para-aminobenzoic acid synthetase	0.251473414	-	-	-	-	PF11100w	R	
PY07597	Plasmodium falciparum CG6	0.349289606	-	-	-	-	PFF0340c, PFC0205c, PFO7_0036	S	
PY02647	Plasmodium vivax PV1H14195_P-related	0.091198117	-	-	-	-	-	R	
PY05888	Plasmodium vivax PV1H14220_P	0.288738078	Y	-	-	ubiquinol-cytochrome-c reductase activity	PFC0262c	T	
PY05715	PR7 protein, putative	0.459924857	-	-	-	-	PF14_0201	T	
PY03610	Predicted membrane protein	0.49033821	Y	Y	-	-	-	T	
PY04363	probable tRNA modification GTPase trme	0.372046488	Y	Y	-	GTP binding	-	-	
PY00927	prolyl-tRNA synthetase-related	0.262596963	-	Y	-	-	PF11240c	R	
PY02094	proteasome subunit alpha type 5	0.465741861	-	-	-	-	PFO7_0112	R	
PY02685	proteasome subunit beta type 1-related	0.475671281	-	-	-	-	PFE0915c	S	
PY00120	putative 26S proteasome subunits	0.230909626	-	-	-	-	PF14_0025	S	
PY00090	putative bir1 protein	0.479163145	Y	-	-	-	-	-	
PY00175	putative bir1 protein	0.100290947	Y	-	-	-	-	-	
PY00400	putative bir1 protein	0.050945048	Y	-	-	-	-	-	
PY01110	putative bir1 protein	0.075177735	Y	-	-	-	-	R	
PY01481	putative bir1 protein	0.171190761	Y	-	-	-	-	-	
PY02131	putative bir1 protein	0.30292418	Y	-	-	-	-	-	
PY02354	putative bir1 protein	0.096651584	Y	-	-	-	-	R	
PY02357	putative bir1 protein	0.136819763	Y	-	-	-	-	-	
PY02368	putative bir1 protein	0.083656708	Y	-	-	-	-	R	
PY02449	putative bir1 protein	0.083568742	Y	-	-	-	-	-	
PY02544	putative bir1 protein	0.201263534	Y	-	-	-	-	-	
PY02760	putative bir1 protein	0.023249735	Y	-	-	-	-	S	
PY02928	putative bir1 protein	0.221271644	Y	-	-	-	-	T	
PY03018	putative bir1 protein	0.302103993	Y	-	-	-	-	R	
PY03043	putative bir1 protein	0.154174576	Y	-	-	-	-	R	
PY03044	putative bir1 protein	0.33739059	Y	Y	-	-	-	R	
PY03332	putative bir1 protein	0.030877309	Y	-	-	-	-	-	
PY03359	putative bir1 protein	0.087185218	Y	-	-	-	-	-	
PY03400	putative bir1 protein	0.183147096	Y	-	-	-	-	-	
PY03436	putative bir1 protein	0.090050757	Y	-	-	-	-	-	
PY03695	putative bir1 protein	0.10287322	Y	-	-	-	-	-	
PY03809	putative bir1 protein	0.253682896	Y	-	-	-	-	R	
PY04321	putative bir1 protein	0.191748968	Y	-	-	-	-	-	
PY04758	putative bir1 protein	0.303055326	-	-	-	-	-	-	
PY04770	putative bir1 protein	0.428909976	Y	-	-	-	-	-	

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Preptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY05220	putative bir1 protein	0.04683923	Y	-	-	-	-	R	
PY05362	putative bir1 protein	0.020817667	Y	-	-	-	-	-	
PY05820	putative bir1 protein	0.212012911	Y	-	-	-	-	-	
PY06141	putative bir1 protein	0.23854544	Y	-	-	-	-	-	
PY06220	putative bir1 protein	0.069557799	Y	-	-	-	-	R	
PY06760	putative bir1 protein	0.415365791	Y	-	-	-	-	-	
PY06830	putative bir1 protein	0.041753258	Y	Y	-	-	-	-	
PY06988	putative bir1 protein	0.373932668	Y	-	-	-	-	-	
PY07012	putative bir1 protein	0.161143358	Y	-	-	-	-	-	
PY07169	putative bir1 protein	0.076297434	Y	-	-	-	-	-	
PY07187	putative bir1 protein	0.108583542	Y	-	-	-	-	-	
PY07226	putative bir1 protein	0.304091072	Y	-	-	-	-	-	
PY07246	putative bir1 protein	0.139513095	Y	-	-	-	-	T	
PY07252	putative bir1 protein	0.113062622	Y	-	-	-	-	-	
PY07304	putative bir1 protein	0.495576185	Y	-	-	-	-	-	
PY07356	putative bir1 protein	0.41495005	Y	-	-	-	-	-	
PY07516	putative bir1 protein	0.128965758	Y	-	-	-	-	-	
PY03455	putative dual-specificity protein phosphatase	0.391538877	-	-	-	-	PFC0380w	S	
PY02994	putative methionyl-tRNA synthetase	0.273142477	-	-	-	-	PF14_0401	-	
PY00969	putative ring finger protein	0.195822454	-	-	-	-	PFC0845c	S	
PY07229	putative small nuclear ribonucleoprotein polypeptide G	0.466625502	-	-	-	-	MAL8P1.48	R	
PY05710	Putative snoRNA binding domain, putative	0.323882405	-	-	-	-	PF10_0085	R	
PY05764	putative ubiquinone biosynthesis protein	0.308584816	-	-	-	-	PF11_0128	R	
PY00268	putative yir1 protein	0.178289532	Y	-	-	-	-	-	
PY00279	putative yir1 protein	0.056265486	Y	-	-	-	-	R	
PY00301	putative yir1 protein	0.428393814	Y	Y	-	-	-	-	
PY00635	putative yir1 protein	0.163379259	Y	-	-	-	-	-	
PY00989	putative yir1 protein	0.102880067	Y	-	-	-	-	-	
PY01174	putative yir1 protein	0.08181733	Y	-	-	-	-	R	
PY01417	putative yir1 protein	0.303776701	Y	-	-	-	-	T	
PY01466	putative yir1 protein	0.250057511	Y	-	-	-	-	-	
PY01594	putative yir1 protein	0.052764629	-	-	-	-	-	-	
PY01595	putative yir1 protein	0.155859615	Y	-	-	-	-	R	
PY01769	putative yir1 protein	0.110222777	Y	-	-	-	-	-	
PY02038	putative yir1 protein	0.197390659	Y	-	-	-	-	-	
PY02137	putative yir1 protein	0.435282231	Y	-	-	-	-	-	
PY02163	putative yir1 protein	0.330419235	Y	-	-	-	-	-	
PY02629	putative yir1 protein	0.427928733	Y	Y	-	-	-	-	
PY02827	putative yir1 protein	0.078468802	Y	-	-	-	-	-	
PY03029	putative yir1 protein	0.094167258	Y	-	-	-	-	-	
PY03341	putative yir1 protein	0.139797609	Y	-	-	-	-	R	
PY03355	putative yir1 protein	0.207117487	Y	-	-	-	-	-	
PY03428	putative yir1 protein	0.136536285	Y	-	-	-	-	-	
PY03579	putative yir1 protein	0.194402852	Y	-	-	-	-	-	
PY03976	putative yir1 protein	0.204481701	Y	-	-	-	-	-	
PY03983	putative yir1 protein	0.197980122	Y	-	-	-	-	-	
PY03985	putative yir1 protein	0.128506632	Y	-	-	-	-	R	
PY04151	putative yir1 protein	0.332847019	Y	-	-	-	-	-	
PY04312	putative yir1 protein	0.137957976	Y	-	-	-	-	R	
PY04334	putative yir1 protein	0.129396117	Y	-	-	-	-	-	
PY04396	putative yir1 protein	0.399659298	Y	-	-	-	-	-	
PY04414	putative yir1 protein	0.203621699	Y	-	-	-	-	-	
PY04425	putative yir1 protein	0.283200277	Y	-	-	-	-	-	
PY04515	putative yir1 protein	0.151888323	Y	-	-	-	-	-	
PY04601	putative yir1 protein	0.068926324	Y	-	-	-	-	-	
PY04735	putative yir1 protein	0.061588238	Y	-	-	-	-	-	
PY04805	putative yir1 protein	0.11420888	Y	-	-	-	-	R	
PY05010	putative yir1 protein	0.438707841	Y	-	-	-	-	-	
PY05243	putative yir1 protein	0.134907332	Y	-	-	-	-	-	
PY05244	putative yir1 protein	0.169154061	Y	-	-	-	-	R	
PY05301	putative yir1 protein	0.193547512	Y	-	-	-	-	-	
PY05307	putative yir1 protein	0.445824824	Y	-	-	-	-	-	

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY05363	putative yir1 protein	0.132389827	Y	-	-	-	-	T	
PY05758	putative yir1 protein	0.161092494	-	Y	-	-	-	-	
PY06113	putative yir1 protein	0.262834903	Y	-	-	-	-	-	
PY06153	putative yir1 protein	0.090156228	Y	-	-	-	-	-	
PY06155	putative yir1 protein	0.184634702	Y	-	-	-	-	-	
PY06188	putative yir1 protein	0.075464881	Y	-	-	-	-	R	
PY06231	putative yir1 protein	0.288413801	Y	-	-	-	-	-	
PY06508	putative yir1 protein	0.224628116	-	-	-	-	-	R	
PY06547	putative yir1 protein	0.056012921	Y	-	-	-	-	-	
PY06567	putative yir1 protein	0.140150504	Y	-	-	-	-	-	
PY06585	putative yir1 protein	0.185970445	Y	-	-	-	-	-	
PY06678	putative yir1 protein	0.390667226	Y	-	-	-	-	R	
PY06754	putative yir1 protein	0.300291545	Y	-	-	-	-	R	
PY06803	putative yir1 protein	0.156231934	Y	-	-	-	-	-	
PY06805	putative yir1 protein	0.134680409	Y	-	-	-	-	-	
PY06818	putative yir1 protein	0.363605413	Y	-	-	-	-	-	
PY06819	putative yir1 protein	0.174735291	Y	-	-	-	-	-	
PY06870	putative yir1 protein	0.157038744	Y	-	-	-	-	-	
PY06891	putative yir1 protein	0.191102719	Y	-	-	-	-	R	
PY06995	putative yir1 protein	0.109943644	Y	-	-	-	-	-	
PY07125	putative yir1 protein	0.075142909	Y	-	-	-	-	R	
PY07269	putative yir1 protein	0.372519908	Y	-	-	-	-	-	
PY07354	putative yir1 protein	0.138367239	Y	-	-	-	-	R	
PY07386	putative yir1 protein	0.228906477	Y	-	-	-	-	-	
PY07426	putative yir1 protein	0.079795227	Y	-	-	-	-	-	
PY07457	putative yir1 protein	0.176373065	Y	-	-	-	-	-	
PY07524	putative yir1 protein	0.045892165	Y	-	-	-	-	R	
PY07628	putative yir1 protein	0.148199064	Y	-	-	-	-	R	
PY00049	putative yir2 protein	0.13793183	Y	Y	-	-	-	-	
PY00236	putative yir2 protein	0.340013888	Y	-	-	-	-	-	
PY00350	putative yir2 protein	0.130588427	Y	-	-	-	-	R	
PY00578	putative yir2 protein	0.116399361	Y	-	-	-	-	-	
PY00612	putative yir2 protein	0.254929576	Y	-	-	-	-	R	
PY00642	putative yir2 protein	0.468456007	Y	-	-	-	-	-	
PY01161	putative yir2 protein	0.346837185	Y	-	-	-	-	-	
PY02107	putative yir2 protein	0.261464475	Y	-	-	-	-	R	
PY02924	putative yir2 protein	0.160002979	Y	-	-	-	-	-	
PY03041	putative yir2 protein	0.237927898	Y	-	-	-	-	R	
PY03194	putative yir2 protein	0.261684552	Y	-	-	-	-	R	
PY03331	putative yir2 protein	0.285079594	Y	-	-	-	-	-	
PY03542	putative yir2 protein	0.295171503	Y	Y	-	-	-	-	
PY03569	putative yir2 protein	0.138936934	Y	-	-	-	-	-	
PY03580	putative yir2 protein	0.031970846	Y	-	-	-	-	R	
PY03758	putative yir2 protein	0.405062573	Y	-	-	-	-	R	
PY03972	putative yir2 protein	0.13996169	Y	-	-	-	-	-	
PY04052	putative yir2 protein	0.154640045	Y	-	-	-	-	R	
PY04211	putative yir2 protein	0.078566555	Y	-	-	-	-	R	
PY04249	putative yir2 protein	0.084935689	-	-	-	-	-	R	
PY04461	putative yir2 protein	0.241956105	Y	-	-	-	-	R	
PY04667	putative yir2 protein	0.138567306	Y	-	-	-	-	T	
PY05021	putative yir2 protein	0.450742894	Y	-	-	-	-	R	
PY05410	putative yir2 protein	0.040776582	Y	-	-	-	-	R	
PY05442	putative yir2 protein	0.055366535	Y	-	-	-	-	R	
PY05455	putative yir2 protein	0.066119524	Y	-	-	-	-	R	
PY05656	putative yir2 protein	0.294522629	Y	-	-	-	-	-	
PY05792	putative yir2 protein	0.161542867	Y	-	-	-	-	-	
PY05942	putative yir2 protein	0.141511745	Y	-	-	-	-	R	
PY06279	putative yir2 protein	0.188722032	Y	-	-	-	-	R	
PY06340	putative yir2 protein	0.097323627	Y	-	-	-	-	R	
PY06821	putative yir2 protein	0.224881498	Y	-	-	-	-	-	
PY06880	putative yir2 protein	0.451510583	Y	-	-	-	-	-	
PY06934	putative yir2 protein	0.365357289	Y	-	-	-	-	-	

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY07036	putative yir2 protein	0.089401493	Y	-	-	-	-	R	
PY07072	putative yir2 protein	0.273743103	-	-	-	-	-	-	
PY07127	putative yir2 protein	0.214777363	Y	-	-	-	-	R	
PY07466	putative yir2 protein	0.085638092	Y	Y	-	-	-	T	
PY07499	putative yir2 protein	0.138351511	-	-	-	-	-	-	
PY00001	putative yir3 protein	0.489773916	Y	-	-	-	-	-	
PY00084	putative yir3 protein	0.33268789	Y	-	-	-	-	-	
PY00089	putative yir3 protein	0.0965364	Y	-	-	-	-	R	
PY00171	putative yir3 protein	0.145543268	Y	Y	-	-	-	-	
PY00213	putative yir3 protein	0.219002113	Y	-	-	-	-	-	
PY00305	putative yir3 protein	0.400631401	Y	-	-	-	-	-	
PY00392	putative yir3 protein	0.058319681	Y	-	-	-	-	-	
PY00485	putative yir3 protein	0.243258846	Y	-	-	-	-	-	
PY00486	putative yir3 protein	0.069533072	Y	Y	-	-	-	-	
PY00490	putative yir3 protein	0.323363722	Y	-	-	-	-	R	
PY00493	putative yir3 protein	0.300315086	Y	-	-	-	-	-	
PY00579	putative yir3 protein	0.15450226	-	-	-	-	-	-	
PY00580	putative yir3 protein	0.0615575	Y	-	-	-	-	R	
PY00842	putative yir3 protein	0.389964637	Y	-	-	-	-	R	
PY00923	putative yir3 protein	0.134770571	Y	-	-	-	-	-	
PY01079	putative yir3 protein	0.087291513	Y	Y	-	-	-	-	
PY01238	putative yir3 protein	0.101889169	Y	-	-	-	-	-	
PY01278	putative yir3 protein	0.332071599	Y	Y	-	-	-	-	
PY01318	putative yir3 protein	0.131356791	Y	-	-	-	-	R	
PY01331	putative yir3 protein	0.047212255	Y	-	-	-	-	-	
PY01720	putative yir3 protein	0.171236701	Y	-	-	-	-	-	
PY01810	putative yir3 protein	0.187615725	Y	-	-	-	-	R	
PY01842	putative yir3 protein	0.215183888	Y	-	-	-	-	R	
PY01872	putative yir3 protein	0.059532035	Y	-	-	-	-	R	
PY02011	putative yir3 protein	0.108094686	Y	-	-	-	-	R	
PY02036	putative yir3 protein	0.202336103	Y	-	-	-	-	-	
PY02831	putative yir3 protein	0.313909383	Y	-	-	-	-	-	
PY02972	putative yir3 protein	0.117870983	Y	-	-	-	-	-	
PY02997	putative yir3 protein	0.49166716	Y	Y	-	-	-	R	
PY03042	putative yir3 protein	0.038325237	Y	-	-	-	-	R	
PY03045	putative yir3 protein	0.105954452	Y	-	-	-	-	-	
PY03333	putative yir3 protein	0.163816817	Y	-	-	-	-	-	
PY03632	putative yir3 protein	0.25183523	Y	Y	-	-	-	-	
PY03636	putative yir3 protein	0.211018492	Y	-	-	-	-	R	
PY03760	putative yir3 protein	0.076945156	Y	-	-	-	-	-	
PY03849	putative yir3 protein	0.39848426	Y	-	-	-	-	-	
PY03986	putative yir3 protein	0.091322527	Y	-	-	-	-	-	
PY04006	putative yir3 protein	0.233252558	Y	Y	-	-	-	-	
PY04084	putative yir3 protein	0.11976208	Y	-	-	-	-	-	
PY04144	putative yir3 protein	0.203746445	Y	-	-	-	-	R	
PY04233	putative yir3 protein	0.294960861	Y	-	-	-	-	-	
PY04276	putative yir3 protein	0.125741128	Y	-	-	-	-	-	
PY04430	putative yir3 protein	0.187003294	Y	-	-	-	-	-	
PY04696	putative yir3 protein	0.23218575	Y	-	-	-	-	-	
PY04795	putative yir3 protein	0.285169253	Y	-	-	-	-	-	
PY04845	putative yir3 protein	0.139856718	Y	-	-	-	-	-	
PY04909	putative yir3 protein	0.460028883	Y	-	-	-	-	-	
PY04918	putative yir3 protein	0.353481383	Y	-	-	-	-	R	
PY04939	putative yir3 protein	0.298741838	Y	-	-	-	-	T	
PY05146	putative yir3 protein	0.36763723	Y	-	-	-	-	R	
PY05150	putative yir3 protein	0.137352853	Y	-	-	-	-	-	
PY05343	putative yir3 protein	0.171358634	Y	-	-	-	-	-	
PY05349	putative yir3 protein	0.063456188	Y	-	-	-	-	-	
PY05350	putative yir3 protein	0.269226938	Y	-	-	-	-	-	
PY05403	putative yir3 protein	0.240026596	-	-	-	-	-	-	
PY05428	putative yir3 protein	0.284802425	Y	Y	-	-	-	R	
PY05636	putative yir3 protein	0.144761624	Y	-	-	-	-	-	

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY05637	putative yir3 protein	0.107839894	Y	-	-	-	-	-	
PY05638	putative yir3 protein	0.228942908	Y	Y	-	-	-	-	
PY05673	putative yir3 protein	0.103271215	Y	-	-	-	-	S	
PY05708	putative yir3 protein	0.443517072	Y	-	-	-	-	R	
PY05709	putative yir3 protein	0.161810825	Y	-	-	-	-	-	
PY05719	putative yir3 protein	0.152956033	Y	Y	-	-	-	R	
PY05794	putative yir3 protein	0.150795743	Y	-	-	-	-	R	
PY05828	putative yir3 protein	0.208612966	Y	-	-	-	-	-	
PY05943	putative yir3 protein	0.311286038	Y	Y	-	-	-	-	
PY06126	putative yir3 protein	0.392313809	Y	-	-	-	-	-	
PY06131	putative yir3 protein	0.324565349	Y	Y	-	-	-	R	
PY06151	putative yir3 protein	0.239686772	Y	-	-	-	-	-	
PY06172	putative yir3 protein	0.148053999	Y	-	-	-	-	-	
PY06185	putative yir3 protein	0.105286299	Y	-	-	-	-	-	
PY06383	putative yir3 protein	0.372043773	Y	-	-	-	-	R	
PY06398	putative yir3 protein	0.252294547	Y	-	-	-	-	-	
PY06428	putative yir3 protein	0.098076082	Y	Y	-	-	-	-	
PY06569	putative yir3 protein	0.296058052	Y	-	-	-	-	-	
PY06619	putative yir3 protein	0.427923552	Y	-	-	-	-	-	
PY06666	putative yir3 protein	0.163752742	Y	-	-	-	-	R	
PY06667	putative yir3 protein	0.297040903	Y	-	-	-	-	-	
PY06674	putative yir3 protein	0.298491704	Y	-	-	-	-	-	
PY06679	putative yir3 protein	0.250072638	Y	-	-	-	-	-	
PY06710	putative yir3 protein	0.494243634	Y	-	-	-	-	-	
PY06743	putative yir3 protein	0.161239417	Y	-	-	-	-	R	
PY06794	putative yir3 protein	0.079449922	Y	-	-	-	-	T	
PY06795	putative yir3 protein	0.386740573	Y	-	-	-	-	R	
PY06861	putative yir3 protein	0.240898027	Y	Y	-	-	-	R	
PY06982	putative yir3 protein	0.251684034	Y	-	-	-	-	-	
PY06997	putative yir3 protein	0.325536749	Y	Y	-	-	-	-	
PY07028	putative yir3 protein	0.070898818	Y	-	-	-	-	-	
PY07042	putative yir3 protein	0.345156611	Y	-	-	-	-	-	
PY07044	putative yir3 protein	0.110144845	Y	-	-	-	-	-	
PY07099	putative yir3 protein	0.442955855	Y	-	-	-	-	-	
PY07163	putative yir3 protein	0.158146667	Y	-	-	-	-	-	
PY07171	putative yir3 protein	0.173174908	Y	-	-	-	-	-	
PY07198	putative yir3 protein	0.239843621	Y	-	-	-	-	-	
PY07233	putative yir3 protein	0.099541748	Y	-	-	-	-	-	
PY07261	putative yir3 protein	0.124285681	Y	Y	-	-	-	-	
PY07284	putative yir3 protein	0.306101001	Y	Y	-	-	-	-	
PY07315	putative yir3 protein	0.212639291	Y	Y	-	-	-	-	
PY07413	putative yir3 protein	0.312432635	Y	Y	-	-	-	-	
PY07501	putative yir3 protein	0.179898743	Y	-	-	-	-	R	
PY07510	putative yir3 protein	0.205800583	Y	-	-	-	-	T	
PY07518	putative yir3 protein	0.116362894	Y	-	-	-	-	R	
PY07527	putative yir3 protein	0.233286938	Y	-	-	-	-	R	
PY07533	putative yir3 protein	0.079552355	Y	-	-	-	-	-	
PY07546	putative yir3 protein	0.235887694	Y	-	-	-	-	-	
PY07635	putative yir3 protein	0.071969417	Y	-	-	-	-	-	
PY07666	putative yir3 protein	0.407380704	Y	-	-	-	-	-	
PY07843	putative yir3 protein	0.117241276	Y	-	-	-	-	-	
PY07849	putative yir3 protein	0.220878805	Y	-	-	-	-	-	
PY00303	putative yir4 protein	0.133988757	Y	-	-	-	-	-	
PY00494	putative yir4 protein	0.083409267	Y	-	-	-	-	R	
PY00574	putative yir4 protein	0.049505408	Y	-	-	-	-	-	
PY00577	putative yir4 protein	0.166919491	Y	-	-	-	-	-	
PY00610	putative yir4 protein	0.112930383	Y	-	-	-	-	R	
PY00661	putative yir4 protein	0.18119626	Y	Y	-	-	-	S	
PY00843	putative yir4 protein	0.300190912	Y	-	-	-	-	R	
PY00957	putative yir4 protein	0.098499062	Y	-	-	-	-	-	
PY01441	putative yir4 protein	0.060198243	Y	-	-	-	-	R	
PY01774	putative yir4 protein	0.130327563	Y	-	-	-	-	R	

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY01871	putative yir4 protein	0.204119335	Y	-	-	-	-	-	
PY01873	putative yir4 protein	0.224710031	Y	-	-	-	-	-	
PY02183	putative yir4 protein	0.356603368	Y	-	-	-	-	R	
PY02278	putative yir4 protein	0.296904981	Y	-	-	-	-	-	
PY02572	putative yir4 protein	0.056238441	Y	-	-	-	-	-	
PY02687	putative yir4 protein	0.120765742	Y	-	-	-	-	-	
PY02688	putative yir4 protein	0.218543415	Y	-	-	-	-	-	
PY02702	putative yir4 protein	0.13735513	Y	-	-	-	-	-	
PY02782	putative yir4 protein	0.41775422	Y	-	-	-	-	-	
PY02970	putative yir4 protein	0.153610717	Y	-	-	-	-	R	
PY02971	putative yir4 protein	0.074967846	Y	-	-	-	-	R	
PY02973	putative yir4 protein	0.073325218	Y	-	-	-	-	R	
PY02996	putative yir4 protein	0.20096102	Y	-	-	-	-	S	
PY03010	putative yir4 protein	0.086761138	Y	-	-	-	-	T	
PY03334	putative yir4 protein	0.215926249	Y	-	-	-	-	R	
PY03402	putative yir4 protein	0.297682278	Y	-	-	-	-	R	
PY03403	putative yir4 protein	0.172049809	Y	-	-	-	-	-	
PY03539	putative yir4 protein	0.441065676	Y	-	-	-	-	-	
PY03540	putative yir4 protein	0.222722191	Y	-	-	-	-	-	
PY03583	putative yir4 protein	0.17963365	Y	-	-	-	-	-	
PY03729	putative yir4 protein	0.393385204	Y	-	-	-	-	-	
PY03747	putative yir4 protein	0.342039977	Y	-	-	-	-	R	
PY03837	putative yir4 protein	0.319426059	Y	-	-	-	-	R	
PY04083	putative yir4 protein	0.359712529	Y	-	-	-	-	-	
PY04204	putative yir4 protein	0.134630629	Y	-	-	-	-	-	
PY04267	putative yir4 protein	0.07524868	Y	-	-	-	-	-	
PY04319	putative yir4 protein	0.125331365	Y	-	-	-	-	R	
PY04412	putative yir4 protein	0.429821551	Y	-	-	-	-	R	
PY04423	putative yir4 protein	0.126687689	Y	-	-	-	-	R	
PY04466	putative yir4 protein	0.418034682	Y	-	-	-	-	R	
PY04537	putative yir4 protein	0.28800109	Y	-	-	-	-	-	
PY04541	putative yir4 protein	0.311121006	Y	-	-	-	-	-	
PY04554	putative yir4 protein	0.294645104	Y	-	-	-	-	-	
PY04806	putative yir4 protein	0.301739166	Y	-	-	-	-	R	
PY04855	putative yir4 protein	0.417197452	Y	-	-	-	-	R	
PY04931	putative yir4 protein	0.103430483	Y	-	-	-	-	-	
PY04965	putative yir4 protein	0.283850021	-	-	-	-	-	R	
PY04979	putative yir4 protein	0.105158546	Y	-	-	-	-	-	
PY04988	putative yir4 protein	0.09919749	Y	-	-	-	-	-	
PY05044	putative yir4 protein	0.106849942	Y	-	-	-	-	-	
PY05242	putative yir4 protein	0.192150898	Y	-	-	-	-	-	
PY05289	putative yir4 protein	0.442844477	Y	-	-	-	-	-	
PY05624	putative yir4 protein	0.161710467	Y	-	-	-	-	R	
PY05675	putative yir4 protein	0.143158112	Y	-	-	-	-	-	
PY05796	putative yir4 protein	0.464524186	Y	Y	-	-	-	-	
PY05822	putative yir4 protein	0.138112842	Y	-	-	-	-	R	
PY05873	putative yir4 protein	0.095849211	Y	-	-	-	-	R	
PY05895	putative yir4 protein	0.469067349	Y	Y	-	-	-	-	
PY06337	putative yir4 protein	0.112790009	Y	-	-	-	-	-	
PY06499	putative yir4 protein	0.286389094	Y	Y	-	-	-	R	
PY06755	putative yir4 protein	0.38354928	Y	-	-	-	-	-	
PY06763	putative yir4 protein	0.079636964	Y	-	-	-	-	R	
PY06796	putative yir4 protein	0.059912723	Y	-	-	-	-	-	
PY06829	putative yir4 protein	0.375251878	Y	-	-	-	-	R	
PY06990	putative yir4 protein	0.302609235	Y	-	-	-	-	-	
PY07027	putative yir4 protein	0.079740222	Y	-	-	-	-	R	
PY07035	putative yir4 protein	0.155452155	Y	-	-	-	-	-	
PY07053	putative yir4 protein	0.09826147	Y	-	-	-	-	R	
PY07149	putative yir4 protein	0.082118956	Y	-	-	-	-	R	
PY07190	putative yir4 protein	0.053115768	Y	-	-	-	-	-	
PY07305	putative yir4 protein	0.381479957	Y	-	-	-	-	R	
PY07325	putative yir4 protein	0.080483663	Y	-	-	-	-	R	

Appendix D: Polymorphic genes in YA with respect to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY07342	putative yir4 protein	0.405138909	Y	-	-	-	-	R	
PY07406	putative yir4 protein	0.182232883	Y	-	-	-	-	-	
PY07450	putative yir4 protein	0.123708287	Y	-	-	-	-	-	
PY07522	putative yir4 protein	0.081112009	Y	-	-	-	-	-	
PY07528	putative yir4 protein	0.239192299	Y	-	-	-	-	-	
PY07659	putative yir4 protein	0.216258818	Y	-	-	-	-	-	
PY04496	replication factor C, 140 kDa subunit	0.222714397	-	-	-	-	PFB0895c	T	
PY01365	rhoptry protein	0.479942739	Y	Y	-	-	MAL13P1.176, PF13_0198, PFL2520w	S	
PY01016	RING-finger protein	0.389460556	-	Y	-	-	-	S	
PY02776	ring-infested erythrocyte surface antigen, putative	0.498045834	Y	-	-	-	-	S	
PY00973	RNA recognition motif, putative	0.33590281	-	-	-	-	PFC0865w	S	
PY06442	Saccharomyces cerevisiae ORF 2310-related	0.424261869	-	-	-	-	-	S	
PY00908	similar to S. cerevisiae SIK1	0.368173095	-	-	-	-	PF11_0191	R	
PY05606	Smt3 activating enzyme 1	0.126737614	-	-	-	-	PF13_0264	R	
PY03158	splicing factor, arginine/serine-rich 12	0.494152008	-	-	-	-	PFF0300w	S	
PY05218	synthetic antigen of P.falciparum, putative	0.331655899	-	-	-	-	MAL13P1.334	-	
PY03247	TERT gene	0.498504383	Y	-	-	-	PFF0965c	T	
PY00825	tRNA-guanine transglycosylase-related	0.499494638	-	-	-	-	PFL2030w, PFO7_0071	R	
PY03314	tRNA-pseudouridine synthase I	0.463062056	Y	-	-	tRNA binding, tRNA-pseudouridine synthase activity	PFE0815w	R	
PY04254	tyrosyl-tRNA synthetase 1-related	0.372598649	-	Y	-	-	PF11_0181	R	
PY02799	unnamed protein product	0.459272513	-	-	-	-	MAL13P1.52	R	
PY06180	unnamed protein product	0.322188998	Y	-	-	GTP binding	PF14_0292	R	
PY04969	unnamed protein product-related	0.240410837	-	-	-	-	PF13_0208	S	
PY02528	uroporphyrinogen decarboxylase, putative	0.35581	-	-	-	-	PFF0360w	T	
PY05083	Zinc finger, C2H2 type, putative	0.409701037	-	-	-	-	PF14_0707	S	

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Appendix E: Gene expansions in YA as compared to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya:17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY03219	Arabidopsis thaliana BRAHMA ortholog	2.508240182	-	-	-	-	PF14_0059	R	
PY00976	Capn7-related	2.16017316	Y	-	-	calcium-dependent cysteine-type endopeptidase acti	MAL13P1.310	S	
PY03609	chloroquine resistance marker protein	2.104782439	-	-	-	-	-	T	
PY07491	cir2 protein, putative	7.042284717	Y	-	-	-	-	-	
PY04439	DNA-directed RNA polymerase beta' chain	2.040816327	Y	-	-	DNA-directed RNA polymerase activity, DNA binding	PF13_0150, PFC0805w	S	
PY00023	hypothetical protein	2.058318692	-	-	-	-	-	R	
PY00473	hypothetical protein	2.369484223	-	-	-	-	PF13_0084	T	
PY00510	hypothetical protein	2.246575342	Y	-	-	-	-	T	
PY00658	hypothetical protein	2.050807067	-	-	-	-	-	-	
PY00758	hypothetical protein	2.142322097	-	-	-	-	PFC0820w	T	
PY00895	hypothetical protein	2.375586854	Y	-	-	-	-	T	
PY00931	hypothetical protein	2.040650407	-	-	-	-	PF14_0544	S	
PY00987	hypothetical protein	2.306362059	Y	-	-	-	-	-	
PY00990	hypothetical protein	2.15703491	-	-	-	-	-	-	
PY01078	hypothetical protein	3.168650554	-	-	-	-	-	R	
PY01097	hypothetical protein	3.282028578	-	-	-	-	-	-	<100nt
PY01125	hypothetical protein	2.401097643	-	-	-	-	PFB0255w	S	
PY01142	hypothetical protein	2.04960559	Y	-	-	-	-	T	
PY01237	hypothetical protein	5.256736431	-	-	-	-	-	S	
PY01688	hypothetical protein	2.381930216	-	-	-	-	-	S	
PY01767	hypothetical protein	2.325679348	Y	-	-	-	-	R	
PY01768	hypothetical protein	2.173263477	-	-	-	-	-	R	
PY01979	hypothetical protein	2.317191944	-	-	-	-	-	S	<100nt
PY02125	hypothetical protein	2.091671592	-	-	-	-	-	R	
PY02144	hypothetical protein	2.149789863	-	-	-	-	-	R	
PY02172	hypothetical protein	2.032258065	-	-	-	-	-	S	<100nt
PY02391	hypothetical protein	2.299299882	-	-	-	-	PF10705w	T	
PY02720	hypothetical protein	2.295951684	Y	-	-	-	MAL13P1.189	S	
PY03116	hypothetical protein	2.111111111	-	-	-	-	-	R	<100nt
PY03444	hypothetical protein	2.660459369	Y	-	-	-	-	R	
PY03806	hypothetical protein	2.103321409	-	-	-	-	-	-	
PY03887	hypothetical protein	22.81842457	-	-	-	-	MAL13P1.135	R	
PY04069	hypothetical protein	2.074650699	-	-	-	-	-	-	
PY04153	hypothetical protein	2.180135135	Y	-	-	-	-	R	
PY04271	hypothetical protein	7.118009321	-	-	-	-	-	T	
PY04426	hypothetical protein	2.261868948	-	-	-	-	-	-	
PY04579	hypothetical protein	2.850579592	Y	-	-	-	-	T	
PY04652	hypothetical protein	2.106908973	-	-	-	-	-	R	
PY04738	hypothetical protein	3.259581587	-	-	-	-	-	R	<100nt
PY04810	hypothetical protein	2.602725628	Y	-	-	-	-	-	
PY04963	hypothetical protein	2.79677219	Y	-	-	-	-	-	
PY05074	hypothetical protein	2.602150538	-	-	-	-	-	T	
PY05075	hypothetical protein	2.003679546	Y	-	-	-	-	-	
PY05506	hypothetical protein	2.56762941	Y	-	-	-	-	-	
PY05862	hypothetical protein	3.84431143	-	-	-	-	-	R	
PY05929	hypothetical protein	3.634748928	-	-	-	-	-	-	<100nt
PY06154	hypothetical protein	2.065613127	-	-	-	-	-	R	
PY06270	hypothetical protein	2.486171468	-	-	-	-	-	-	
PY06319	hypothetical protein	2.283842795	Y	-	-	-	-	-	<100nt
PY06477	hypothetical protein	2.009065156	Y	-	-	-	-	S	
PY06833	hypothetical protein	2.313868613	Y	-	-	-	-	-	
PY06991	hypothetical protein	2.163356559	-	-	-	-	-	-	<100nt
PY07018	hypothetical protein	2.170931525	Y	-	-	-	-	R	

Appendix E: Gene expansions in YA as compared to 17x1.1

Appendices

Py Acc. No.	Py Name	av ya:17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY07360	hypothetical protein	2.500589871	-	-	-	-	-	-	<100nt
PY07472	hypothetical protein	2.271932593	-	-	-	-	-	T	
PY07565	hypothetical protein	2.016129032	-	-	-	-	-	-	<100nt
PY07620	hypothetical protein	2.057170383	-	-	-	-	-	-	<100nt
PY07663	hypothetical protein	2.543838974	-	-	-	-	-	-	
PY07774	inorganic phosphate transporter	2.347985348	Y	-	-	inorganic phosphate transmembrane transporter acti	MAL13P1.206	-	
PY05959	putative bir1 protein	2.029863457	Y	-	-	-	-	R	
PY06495	putative bir1 protein	2.254231097	Y	-	-	-	-	-	
PY07264	putative bir1 protein	2.647786784	Y	-	-	-	-	R	
PY07474	putative bir1 protein	2.212389381	Y	-	-	-	-	-	
PY01162	putative yir1 protein	2.697913967	Y	-	-	-	-	R	
PY01454	putative yir1 protein	3.216646267	Y	-	-	-	-	-	
PY04505	putative yir1 protein	4.730761815	Y	-	-	-	-	-	
PY04785	putative yir1 protein	2.053665521	Y	-	-	-	-	-	
PY04923	putative yir1 protein	2.854929625	Y	-	-	-	-	-	
PY05698	putative yir1 protein	2.284629527	Y	-	-	-	-	-	
PY06271	putative yir1 protein	2.897875656	Y	-	-	-	-	R	
PY01865	putative yir2 protein	2.310224789	Y	-	-	-	-	R	
PY02562	putative yir2 protein	2.208992254	Y	-	-	-	-	-	
PY06536	putative yir2 protein	4.052790143	Y	-	-	-	-	R	
PY01788	putative yir3 protein	2.214692737	Y	-	-	-	-	-	
PY03091	putative yir3 protein	2.844448217	Y	-	-	-	-	-	
PY05923	putative yir3 protein	2.25113579	Y	-	-	-	-	R	
PY07618	putative yir3 protein	3.321325626	Y	-	-	-	-	-	
PY00814	putative yir4 protein	2.097619359	Y	-	-	-	-	-	
PY03089	putative yir4 protein	2.401338449	Y	-	-	-	-	R	
PY04419	putative yir4 protein	3.958873042	Y	-	-	-	-	-	
PY05507	putative yir4 protein	2.821257468	Y	-	-	-	-	-	
PY07080	putative yir4 protein	2.139945109	-	-	-	-	-	R	
PY04559	RNA lariat debranching enzyme	2.057695869	-	-	-	-	PF13_0222	T	
PY00947	RNA recognition motif, putative	2.113744076	-	-	-	-	PF13_0318	T	
PY04902	Uncharacterized ACR, COG2106, putative	2.083568316	-	-	-	-	PF14_0307	R	
PY07738	unknown-related	2.061262654	-	-	-	-	-	-	Bacterial

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Appendix F: Polymorphic genes in YM as compared to 17x1.1

Appendices

Py Acc. No.	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY03688	5702-7336, putative	0.064713065	-	-	-	-	PF14_0700	T	
PY00319	aconitate hydratase 1	0	-	-	-	-	PF13_0229	T	
PY04723	Arabidopsis thaliana At3g58560/F14P22_150-related	0	-	-	-	-	PFE0980c	T	
PY00253	CCAAT-box DNA binding protein subunit B	0	-	-	-	-	PFI1315c	-	
PY04786	cir1 protein, putative	0.361196551	Y	-	-	-	-	-	
PY04960	Drosophila melanogaster CG8797 gene product-related	0.47358604	Y	-	-	transferase activity	PFL1075w, PF14_0404, PFA0280w	T	
PY06985	Endonuclease/Exonuclease/phosphatase family	0.112786436	-	-	-	-	-	-	Rodent
PY07668	Endonuclease/Exonuclease/phosphatase family	0.037343672	-	-	-	-	-	-	Rodent
PY07680	hydroxyproline-rich glycoprotein DZ-HRGP-related	0.492169733	-	-	-	-	-	R	
PY00330	hypothetical protein	0.33618515	Y	-	-	-	-	-	
PY00439	hypothetical protein	0.076385457	Y	-	-	-	-	-	
PY00440	hypothetical protein	0.24526414	Y	Y	-	-	-	T	
PY00980	hypothetical protein	0.117667962	-	-	-	-	-	-	
PY01237	hypothetical protein	0.089606243	-	-	-	-	-	S	
PY01904	hypothetical protein	0.474621479	Y	-	-	-	-	R	
PY02429	hypothetical protein	0	Y	-	-	-	PFL0555c	R	
PY02468	hypothetical protein	0.232432432	-	-	-	-	-	-	
PY02729	hypothetical protein	0.245070423	-	-	-	-	-	-	
PY02985	hypothetical protein	0.192291066	Y	-	-	-	PFB0600c	R	
PY03172	hypothetical protein	0.369104198	-	-	-	-	-	-	
PY03644	hypothetical protein	0.380713342	-	-	-	-	PF14_0659	T	
PY03887	hypothetical protein	0	-	-	-	-	MAL13P1.135	R	
PY04271	hypothetical protein	0	-	-	-	-	-	T	
PY04313	hypothetical protein	0.127906577	-	Y	-	-	-	T	
PY04314	hypothetical protein	0.492882562	-	-	-	-	-	-	
PY04315	hypothetical protein	0.09164679	Y	-	-	-	-	R	
PY04316	hypothetical protein	0.417417141	Y	-	-	-	-	R	
PY04317	hypothetical protein	0.140477275	Y	Y	-	-	-	R	
PY04384	hypothetical protein	0.193297917	-	-	-	-	-	-	
PY04463	hypothetical protein	0.312920537	-	-	-	-	-	-	<100nt
PY04465	hypothetical protein	0.117121267	-	-	-	-	-	R	<100nt
PY04933	hypothetical protein	0.158826448	-	-	-	-	PFD0385w	T	
PY04963	hypothetical protein	0.081994376	Y	Y	-	-	-	-	
PY05408	hypothetical protein	0.42831017	Y	-	-	-	-	R	
PY05409	hypothetical protein	0.080375844	-	-	-	-	-	R	
PY05454	hypothetical protein	0.035839501	Y	Y	DRILSDA	-	-	-	
PY05687	hypothetical protein	0.288059427	-	-	-	-	-	-	
PY05720	hypothetical protein	0.220555973	-	-	-	-	-	R	<100nt
PY05956	hypothetical protein	0.140770838	-	-	-	-	-	T	
PY06336	hypothetical protein	0.428810448	-	-	-	-	-	-	
PY06363	hypothetical protein	0.058575878	-	-	-	-	-	R	
PY06364	hypothetical protein	0.045394962	-	-	-	-	-	-	
PY06365	hypothetical protein	0.076765545	-	-	-	-	-	R	<100nt
PY06366	hypothetical protein	0.048291993	-	-	-	-	-	R	<100nt
PY06904	hypothetical protein	0.42549929	-	-	-	-	PFB0161c	S	

Appendix F: Polymorphic genes in YM as compared to 17x1.1

Appendices

Py Acc. No.	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY07015	hypothetical protein	0.046988491	-	-	-	-	-	R	
PY07016	hypothetical protein	0.069253976	-	-	-	-	-	R	
PY07091	hypothetical protein	0.085987524	-	-	-	-	-	-	
PY07170	hypothetical protein	0.159573807	-	-	-	-	-	R	
PY07278	hypothetical protein	0.287646983	-	Y	-	-	-	R	
PY07279	hypothetical protein	0.114526325	Y	Y	-	-	-	-	
PY07282	hypothetical protein	0	-	-	-	-	-	-	
PY07366	hypothetical protein	0.044082639	-	-	-	-	-	-	
PY07367	hypothetical protein	0.06352404	-	-	-	-	-	R	
PY07376	hypothetical protein	0.031948084	-	-	-	-	-	-	
PY07377	hypothetical protein	0.094930281	-	-	-	-	-	-	
PY07503	hypothetical protein	0.039907387	-	-	-	-	-	T	
PY07504	hypothetical protein	0.049167083	Y	Y	-	-	MAL7P1.142	T	
PY07551	hypothetical protein	0.116545783	Y	Y	-	-	-	-	
PY07552	hypothetical protein	0.031715877	-	-	-	-	-	T	
PY07553	hypothetical protein	0.02592152	-	-	-	-	-	R	
PY07554	hypothetical protein	0.04684938	-	-	-	-	-	R	
PY07555	hypothetical protein	0.031169944	-	-	-	-	-	-	<100nt
PY07556	hypothetical protein	0.027486085	-	-	-	-	-	S	
PY07557	hypothetical protein	0.026843536	-	-	-	-	-	-	<100nt
PY07563	hypothetical protein	0.368566014	-	-	-	-	-	-	<100nt
PY07564	hypothetical protein	0.069142933	-	-	-	-	-	T	
PY07565	hypothetical protein	0.344056777	-	-	-	-	-	-	<100nt
PY07567	hypothetical protein	0.046217881	-	-	-	-	-	R	
PY07568	hypothetical protein	0.024889769	-	-	-	-	-	-	
PY07569	hypothetical protein	0.028083231	-	-	-	-	-	-	<100nt
PY07570	hypothetical protein	0.054712291	-	-	-	-	-	-	
PY07571	hypothetical protein	0.024832985	-	-	-	-	-	-	<100nt
PY07572	hypothetical protein	0.033336457	-	-	-	-	-	R	<100nt
PY07573	hypothetical protein	0.02652471	-	-	-	-	-	-	<100nt
PY07580	hypothetical protein	0.426504082	-	-	-	-	-	-	
PY07582	hypothetical protein	0.462840792	-	-	-	-	-	S	
PY07585	hypothetical protein	0.029136209	-	-	-	-	-	R	
PY07586	hypothetical protein	0.033499328	-	-	-	-	-	-	
PY07587	hypothetical protein	0.029513383	-	-	-	-	-	-	
PY07588	hypothetical protein	0.025651025	-	-	-	-	-	-	
PY07589	hypothetical protein	0.0664915	-	-	-	-	-	R	
PY07590	hypothetical protein	0.036022675	-	-	-	-	-	-	<100nt
PY07591	hypothetical protein	0.051176146	-	-	-	-	-	T	<100nt
PY07592	hypothetical protein	0.030410584	Y	Y	-	-	-	T	
PY07602	hypothetical protein	0.063359818	-	-	-	-	-	-	
PY07603	hypothetical protein	0.09412788	-	-	-	-	-	R	<100nt
PY07604	hypothetical protein	0.039310337	-	-	-	-	-	R	
PY07605	hypothetical protein	0.124560341	-	-	-	-	-	R	
PY07612	hypothetical protein	0.050752843	-	-	-	-	-	-	<100nt
PY07649	hypothetical protein	0.050537881	Y	-	-	-	-	T	
PY07663	hypothetical protein	0.185083738	-	-	-	-	-	-	

Appendix F: Polymorphic genes in YM as compared to 17x1.1

Appendices

Py Acc. No.	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY07664	hypothetical protein	0.224778616	-	-	-	-	-	-	
PY07665	hypothetical protein	0.0383869	Y	-	-	-	-	-	
PY07670	hypothetical protein	0.27977027	Y	Y	-	-	-	-	
PY07671	hypothetical protein	0.039966286	Y	Y	-	-	-	-	
PY07828	hypothetical protein	0.206175985	Y	Y	-	-	-	-	
PY07837	hypothetical protein	0.047655973	Y	Y	-	-	MAL7P1.142	-	
PY07845	hypothetical protein	0.187619143	Y	-	-	-	-	-	
PY07846	hypothetical protein	0.074460408	-	-	-	-	-	-	
PY07847	hypothetical protein	0.028156348	-	-	-	-	-	-	
PY07852	hypothetical protein	0.082999047	-	-	-	-	-	T	<100nt
PY07853	hypothetical protein	0.039250621	-	-	-	-	-	-	<100nt
PY07854	hypothetical protein	0.031073464	-	-	-	-	-	-	
PY07855	hypothetical protein	0.034934179	-	-	-	-	-	T	
PY07856	hypothetical protein	0.045718413	-	-	-	-	-	-	<100nt
PY07857	hypothetical protein	0.105996334	-	-	-	-	-	-	
PY07859	hypothetical protein	0.035933311	-	-	-	-	-	-	<100nt
PY07014	integrase-related	0.0743883	-	-	-	-	-	R	Rodent
PY06984	L1 transposable element	0.020343378	-	-	-	-	-	R	Rodent
PY01605	NAP-like protein	0.122709261	-	-	-	-	PF10930c	R	
PY06381	p235 rhopty protein E5	0.10404195	Y	Y	-	-	MAL13P1.176 , PFL2520w, PF13_0198	S	
PY07579	pORF1	0.07512946	-	-	-	-	-	-	Rodent
PY01269	protein phosphatase 2a, regulatory subunit b'	0.195266272	Y	-	-	phosphatase activator activit	PF14_0280	T	
PY00979	putative bir1 protein	0.233557017	Y	-	-	-	-	-	
PY01966	putative bir1 protein	0.049148323	Y	-	-	-	-	-	
PY03095	putative bir1 protein	0.162825534	Y	-	-	-	-	-	
PY03332	putative bir1 protein	0.026555808	Y	-	-	-	-	-	
PY04737	putative bir1 protein	0.129276564	Y	-	-	-	-	-	
PY04804	putative bir1 protein	0.064757986	Y	-	-	-	-	-	
PY05220	putative bir1 protein	0.025084188	Y	-	-	-	-	R	
PY07169	putative bir1 protein	0.033905904	Y	-	-	-	-	-	
PY00969	putative ring finger protein	0.205405405	-	-	-	-	PFC0845c	S	
PY00279	putative yir1 protein	0.456007223	Y	-	-	-	-	R	
PY01594	putative yir1 protein	0.062482829	-	-	-	-	-	-	
PY03029	putative yir1 protein	0.089507941	Y	-	-	-	-	-	
PY03825	putative yir1 protein	0.153011778	Y	Y	-	-	-	-	
PY04312	putative yir1 protein	0.069588924	Y	-	-	-	-	R	
PY04425	putative yir1 protein	0.456957218	Y	-	-	-	-	-	
PY04478	putative yir1 protein	0.13627153	Y	-	-	-	-	-	
PY04515	putative yir1 protein	0.146338935	Y	-	-	-	-	-	
PY06819	putative yir1 protein	0.126513833	Y	-	-	-	-	-	
PY07090	putative yir1 protein	0.130016235	Y	-	-	-	-	-	
PY07426	putative yir1 protein	0.05928924	Y	-	-	-	-	-	
PY07457	putative yir1 protein	0.134216711	Y	-	-	-	-	-	
PY00049	putative yir2 protein	0.089420889	Y	Y	-	-	-	-	

Appendix F: Polymorphic genes in YM as compared to 17x1.1

Appendices

Py Acc. No.	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY00578	putative yir2 protein	0.449701835	Y	-	-	-	-	-	
PY02562	putative yir2 protein	0.187809784	Y	-	-	-	-	-	
PY03094	putative yir2 protein	0.132309972	Y	-	-	-	-	-	
PY03194	putative yir2 protein	0.184609069	Y	-	-	-	-	R	
PY03331	putative yir2 protein	0.246156044	Y	-	-	-	-	-	
PY03966	putative yir2 protein	0.124111027	Y	-	-	-	-	R	
PY04211	putative yir2 protein	0.063747751	Y	-	-	-	-	R	
PY04249	putative yir2 protein	0.119299058	-	-	-	-	-	R	
PY04964	putative yir2 protein	0.085768547	Y	-	-	-	-	R	
PY05731	putative yir2 protein	0.041616138	Y	-	-	-	-	-	
PY06549	putative yir2 protein	0.045063139	Y	-	-	-	-	R	
PY00084	putative yir3 protein	0.448775056	Y	-	-	-	-	-	
PY00490	putative yir3 protein	0.455886777	Y	-	-	-	-	R	
PY00493	putative yir3 protein	0.4801899	Y	-	-	-	-	-	
PY00923	putative yir3 protein	0.088921779	Y	-	-	-	-	-	
PY01238	putative yir3 protein	0.116597447	Y	-	-	-	-	-	
PY01961	putative yir3 protein	0.311006104	Y	-	-	-	-	R	
PY01963	putative yir3 protein	0.400184495	Y	-	-	-	-	-	
PY02036	putative yir3 protein	0.141796484	Y	-	-	-	-	-	
PY02701	putative yir3 protein	0.301472771	Y	-	-	-	-	-	
PY03045	putative yir3 protein	0.079786658	Y	-	-	-	-	-	
PY03193	putative yir3 protein	0.320793488	Y	-	-	-	-	-	
PY03195	putative yir3 protein	0.129405921	Y	-	-	-	-	-	
PY03333	putative yir3 protein	0.156023118	Y	-	-	-	-	-	
PY03588	putative yir3 protein	0.386306651	Y	Y	-	-	-	R	
PY04464	putative yir3 protein	0.126234883	Y	-	-	-	-	-	
PY05617	putative yir3 protein	0.205228462	Y	Y	-	-	-	-	
PY05719	putative yir3 protein	0.182111904	Y	Y	-	-	-	R	
PY06398	putative yir3 protein	0.200934125	Y	-	-	-	-	-	
PY06920	putative yir3 protein	0.348965824	Y	-	-	-	-	-	
PY07263	putative yir3 protein	0.211032329	Y	-	-	-	-	-	
PY07500	putative yir3 protein	0.072178862	Y	-	-	-	-	R	
PY07527	putative yir3 protein	0.222608368	Y	-	-	-	-	R	
PY07594	putative yir3 protein	0.215680963	Y	-	-	-	-	-	
PY07632	putative yir3 protein	0.403170428	Y	-	-	-	-	-	
PY00177	putative yir4 protein	0.062755169	Y	-	-	-	-	R	
PY00924	putative yir4 protein	0.055710626	Y	Y	-	-	-	R	
PY00957	putative yir4 protein	0.15806279	Y	-	-	-	-	-	
PY01101	putative yir4 protein	0.258881476	-	-	-	-	-	-	
PY02278	putative yir4 protein	0.087012592	Y	-	-	-	-	-	
PY02298	putative yir4 protein	0.469295114	Y	-	-	-	-	-	
PY02572	putative yir4 protein	0.386881192	Y	-	-	-	-	-	
PY02873	putative yir4 protein	0.357150436	Y	-	-	-	-	-	
PY03315	putative yir4 protein	0.308827268	Y	-	-	-	-	R	
PY03330	putative yir4 protein	0.07444662	Y	Y	-	-	-	R	
PY03334	putative yir4 protein	0.147661983	Y	-	-	-	-	R	
PY03729	putative yir4 protein	0.28596794	Y	-	-	-	-	-	

Appendix F: Polymorphic genes in YM as compared to 17x1.1

Appendices

Py Acc. No.	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY04204	putative yir4 protein	0.113535866	Y	-	-	-	-	-	
PY04965	putative yir4 protein	0.173717679	-	-	-	-	-	R	
PY05242	putative yir4 protein	0.362005082	Y	-	-	-	-	-	
PY06337	putative yir4 protein	0.101805258	Y	-	-	-	-	-	
PY06410	putative yir4 protein	0.291788608	-	-	-	-	-	-	
PY06411	putative yir4 protein	0.163808827	Y	-	-	-	-	S	
PY06807	putative yir4 protein	0.06048728	Y	-	-	-	-	R	
PY07281	putative yir4 protein	0.245230129	Y	-	-	-	-	-	
PY07659	putative yir4 protein	0.163488788	Y	-	-	-	-	-	
PY07669	Reverse transcriptase, putative	0.033846216	-	-	-	-	-	-	Rodent
PY07288	RNase H, putative	0.064439518	-	-	-	-	-	-	Rodent
PY07375	similar to L1 retrotransposon, TF subfamily L1MD-TF	0.038113861	-	-	-	-	-	R	Rodent
PY07613	titin-L1 fusion protein-related	0.050938424	-	-	-	-	-	-	Rodent
PY07374	Unknown protein	0.064445998	-	-	-	-	-	-	Rodent
PY03098	Unknown-related	0.498266667	-	-	-	-	PFF0625w	R	

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Py Acc. No	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcript	Comment
PY00427	3'-nucleotidase/nuclease	2.122816423	-	Y	-	-	PF14_0117	R	
PY04665	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (ec 2.7.1.1)	2.086104503	Y	Y	-	kinase activity, ATP binding	PFE0150c	S	
PY01334	93 kDa protein	2.276995305	-	Y	-	-	-	R	
PY06147	93 kDa protein	2.022910345	-	Y	-	-	-	R	
PY04270	ABC transporter	2.102164808	Y	-	-	ATP binding	-	T	
PY05445	asparagine-rich protein	2.110414053	-	-	-	-	PFF0195c	T	
							PFB0815w, PFF0520w, PFC0420w, PF13_0211, PF07_0072		
PY03298	calmodulin-domain protein kinase	5.174899866	-	-	-	-	-	T	
PY07491	cir2 protein, putative	3.381586883	Y	Y	-	-	-	-	
PY07191	DNA mismatch repair protein msh2	2.617728548	-	-	-	-	-	T	
PY00165	Drosophila melanogaster CG15040 gene product	2.080482047	Y	-	-	-	PF14_0110	T	
PY02872	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	2.352954943	-	-	-	-	PF11_0421	T	
PY05846	Frost-related	2.466456119	-	Y	-	-	-	T	
PY00065	hypothetical protein	2.339156634	-	-	-	-	PFI0245c	T	
PY00235	hypothetical protein	2.057410763	Y	Y	-	-	-	-	
PY00359	hypothetical protein	2.073082964	-	-	-	-	PFC1055w	S	
PY00464	hypothetical protein	2.196426941	-	-	-	-	PFL0115w	S	
PY00473	hypothetical protein	2.156992843	-	-	-	-	PF13_0084	T	
PY00510	hypothetical protein	2.010443864	Y	-	-	-	-	T	
PY00658	hypothetical protein	2.039425265	-	-	-	-	-	-	
PY00750	hypothetical protein	2.061827957	-	-	-	-	-	-	<100nt
PY00752	hypothetical protein	2.014195548	-	-	-	-	-	S	
PY00807	hypothetical protein	2.7483066	-	-	-	-	-	R	
PY00937	hypothetical protein	2.503329749	-	-	-	-	-	S	
PY00990	hypothetical protein	3.416505788	-	-	-	-	-	-	
PY01073	hypothetical protein	2.008894536	Y	Y	-	-	MAL8P1.130	T	
PY01083	hypothetical protein	3.094224924	Y	-	-	-	-	-	
PY01097	hypothetical protein	3.7850237	-	-	-	-	-	-	<100nt
PY01221	hypothetical protein	2.216110628	-	-	-	-	-	T	<100nt
PY01289	hypothetical protein	2.27512283	-	-	-	-	MAL7P1.152	R	
PY01503	hypothetical protein	2.64109589	Y	-	-	-	MAL13P1.182	R	
PY01685	hypothetical protein	2.054636925	Y	Y	-	-	-	-	<100nt
PY01686	hypothetical protein	2.089936328	Y	-	-	-	-	-	
PY01688	hypothetical protein	2.421128377	-	-	-	-	-	S	
PY01729	hypothetical protein	2.04075311	Y	Y	-	-	-	R	
PY01789	hypothetical protein	2.123161765	Y	-	-	-	-	-	<100nt
PY01819	hypothetical protein	2.401887282	Y	-	-	-	-	T	
PY01979	hypothetical protein	4.145309576	-	-	-	-	-	S	<100nt
PY02367	hypothetical protein	2.015127389	-	-	-	-	-	-	
PY02479	hypothetical protein	2.320798617	-	-	-	-	-	-	
PY02518	hypothetical protein	2.116722506	-	Y	-	-	-	R	
PY02531	hypothetical protein	2.138148338	-	-	-	-	PF10_0095	S	
PY02662	hypothetical protein	2.260322723	Y	-	NRILADA	-	-	R	
PY02726	hypothetical protein	2.076318088	-	-	-	-	PFL0340w	R	
PY02744	hypothetical protein	2.473722102	-	-	-	-	-	T	
PY02822	hypothetical protein	2.145423181	-	-	-	-	-	T	<100nt
PY03103	hypothetical protein	2.038690476	Y	Y	-	-	PFC0575w	T	
PY03249	hypothetical protein	2.052504143	-	-	-	-	PFF0185c	T	
PY03414	hypothetical protein	2.042483492	-	-	-	-	PF11_0428	T	
PY03427	hypothetical protein	5.4	-	Y	-	-	PFI0565w	T	
PY03444	hypothetical protein	2.618385294	Y	Y	-	-	-	R	
PY03500	hypothetical protein	2.18612244	-	-	-	-	-	R	<100nt
PY03727	hypothetical protein	2.045710651	Y	-	-	-	PFF0240c	S	
PY03735	hypothetical protein	2.212568462	Y	Y	-	-	PFE1340w	R	
PY03757	hypothetical protein	2.363992023	Y	-	-	-	-	R	
PY03836	hypothetical protein	2.068463719	-	-	-	-	-	R	
PY04153	hypothetical protein	3.489075677	Y	-	-	-	-	R	
PY04522	hypothetical protein	2.671037825	-	Y	-	-	-	T	
PY04562	hypothetical protein	2.074502253	-	-	-	-	-	T	

Py Acc. No	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcript	Comment
PY04652	hypothetical protein	2.772453933	-	-	-	-	-	R	
PY04656	hypothetical protein	2.189163722	-	-	-	-	-	S	
PY04716	hypothetical protein	2.154035037	-	-	-	-	PFL0225c	R	
PY04810	hypothetical protein	2.388025334	Y	-	-	-	-	-	
PY04841	hypothetical protein	2.386001632	-	-	-	-	-	S	
PY04843	hypothetical protein	2.200598802	-	-	-	-	-	-	
PY04941	hypothetical protein	2.245738636	-	-	-	-	-	-	
PY05042	hypothetical protein	2.732144734	-	-	-	-	-	S	<100nt
PY05087	hypothetical protein	2.017731594	Y	-	-	-	PF14_0188	R	
PY05170	hypothetical protein	2.223765603	-	-	-	-	PF11_0308	R	
PY05347	hypothetical protein	2.655882993	Y	Y	-	-	MAL13P1.49	T	
PY05348	hypothetical protein	2.011344729	-	-	-	-	PF13_0055	R	
PY05465	hypothetical protein	2.314716362	-	-	-	-	-	R	
PY05466	hypothetical protein	2.007217881	-	Y	-	-	-	R	
PY05470	hypothetical protein	2.054545455	-	-	-	-	PF10_0336	R	
PY05610	hypothetical protein	2.001311344	Y	-	-	-	PF13_0035	R	
PY05632	hypothetical protein	2.033094765	-	-	-	-	-	R	
PY05723	hypothetical protein	2.108713999	Y	-	-	-	-	R	
PY05863	hypothetical protein	2.329630277	Y	-	-	-	-	-	
PY05939	hypothetical protein	2.107618688	Y	-	-	-	-	T	
PY06021	hypothetical protein	2.007096879	Y	Y	-	-	-	-	
PY06098	hypothetical protein	2.120686787	Y	-	-	-	-	R	
PY06289	hypothetical protein	2.391599261	Y	Y	-	-	PF11_0179	R	
PY06319	hypothetical protein	2.002475248	Y	Y	-	-	-	-	<100nt
PY06423	hypothetical protein	2.256983834	-	-	-	-	-	T	
PY06456	hypothetical protein	2.01056338	-	-	-	-	PFD1115c	R	
PY06563	hypothetical protein	2.040106078	-	-	-	-	-	-	
PY06616	hypothetical protein	2.2153787	-	-	-	-	-	-	<100nt
PY06991	hypothetical protein	2.265211662	-	-	-	-	-	-	<100nt
PY07034	hypothetical protein	2.119347095	Y	-	-	-	-	T	
PY07056	hypothetical protein	2.335064375	-	Y	-	-	-	R	
PY07360	hypothetical protein	2.275998411	-	-	-	-	-	-	<100nt
PY07530	hypothetical protein	2.028771269	-	-	-	-	PFF1285w	R	
PY07655	hypothetical protein	2.184626792	-	-	-	-	-	R	
PY07798	hypothetical protein	2.386333771	-	-	-	-	-	S	<100nt
PY01618	Krueppel-like protein	2.025319638	-	-	-	-	PFE0895c	-	
PY05090	LCCL domain-containing protein CCP2-related	2.234267839	-	Y	-	-	PF14_0532	S	
PY06687	Leucine Rich Repeat, putative	2.293399665	-	-	-	-	-	T	
PY06554	MAP kinase-related	2.079897867	-	-	-	-	MAL13P1.84	S	
PY02063	Papain family cysteine protease, putative	2.027355623	-	Y	-	-	PFB0325c	-	
PY05696	peroxisomal-coenzyme a synthetase	2.033587627	Y	-	-	thiamin pyrophosphate binding, magnesium ion binding, catalytic ac	PFF0945c	R	
PY05064	Plasmodium falciparum CG8	2.093659646	Y	-	-	-	PF07_0032	T	
PY05627	Plasmodium vivax PV1H14065_P-related	2.037642045	-	-	-	-	PFC0135c	R	
PY02951	protoporphyrinogen oxidase	2.212433123	Y	-	-	protoporphyrinogen oxidase activity	PF10_0275	R	
PY03399	putative bir1 protein	2.037688442	Y	-	-	-	-	-	
PY04770	putative bir1 protein	2.006128694	Y	-	-	-	-	-	
PY06020	putative bir1 protein	2.035955056	Y	-	-	-	-	-	
PY05061	putative transporter protein CG10	2.384078212	Y	-	-	-	MAL7P1.27	R	
PY01454	putative yir1 protein	2.098989487	Y	-	-	-	-	-	
PY05244	putative yir1 protein	2.324167026	Y	-	-	-	-	R	
PY06062	putative yir1 protein	2.015151515	Y	-	-	-	-	-	
PY06897	putative yir1 protein	2.015622369	Y	-	-	-	-	-	
PY07454	putative yir1 protein	2.180687848	Y	-	-	-	-	S	
PY02519	putative yir2 protein	2.158682635	Y	-	-	-	-	R	
PY05656	putative yir2 protein	3.041790678	Y	-	-	-	-	-	
PY06536	putative yir2 protein	2.02069602	Y	-	-	-	-	R	
PY07127	putative yir2 protein	2.172925148	Y	-	-	-	-	R	
PY00428	putative yir3 protein	2.225935102	Y	Y	-	-	-	-	
PY00644	putative yir3 protein	2.012019659	Y	-	-	-	-	-	
PY02364	putative yir3 protein	2.456296736	Y	-	-	-	-	-	
PY05437	putative yir3 protein	2.218181818	Y	-	-	-	-	R	
PY06126	putative yir3 protein	2.031042129	Y	-	-	-	-	-	

Appendix G: Gene expansions in YM as compared to 17x1.1

Appendices

Py Acc. No	Py Name	Mean YM: 17x	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcript	Comment
PY06383	putative yir3 protein	2.324709639	Y	-	-	-	-	R	
PY06649	putative yir3 protein	2.0221996	Y	-	-	-	-	-	
PY07028	putative yir3 protein	3.714297402	Y	-	-	-	-	-	
PY07042	putative yir3 protein	2.588790056	Y	-	-	-	-	-	
PY00310	putative yir4 protein	2.278947368	Y	-	-	-	-	-	
PY04998	putative yir4 protein	3.035128806	Y	-	-	-	-	-	
PY06834	putative yir4 protein	2.071529444	Y	-	-	-	-	R	
PY07053	putative yir4 protein	2.128603012	Y	-	-	-	-	R	
PY03645	Ser/Thr protein phosphatase, putative	2.374384763	-	Y	-	-	PF14_0660	T	
PY00295	SUMO-1 activating enzyme subunit 1	2.201612008	-	-	-	-	PF11_0457	R	
PY01554	transcriptional regulatory protein, Sir2 family	2.119470472	-	-	-	-	PF13_0152	R	
PY02959	transmission-blocking target antigen s230 precursor	2.088586341	-	-	-	-	-	R	
PY03107	vacuolar membrane protein pep3	2.29405544	-	-	-	-	PF13_0053	R	
PY00231	Yersinia pestis BoA-like protein	2.238319585	-	-	-	-	PFE0790c	S	

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY03219	Arabidopsis thaliana BRAHMA ortholog	0.479115721	-	-	-	-	PF14_0059	R	
PY03609	chloroquine resistance marker protein	0.489710073	-	-	-	-	-	T	
PY03616	chloroquine resistance marker protein	0.480762958	-	-	-	-	PF14_0675	T	
PY07491	cir2 protein, putative	0.48018321	Y	Y	-	-	-	-	
PY06985	Endonuclease/Exonuclease/phosphatase family	0.096791235	-	-	-	-	-	-	Rodent
PY07668	Endonuclease/Exonuclease/phosphatase family	0.027135519	-	-	-	-	-	-	Rodent
PY03626	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	0.410594635	-	-	-	-	MAL13P1.63	R	
PY00439	hypothetical protein	0.089110272	Y	-	-	-	-	-	
PY00440	hypothetical protein	0.34707397	Y	Y	-	-	-	T	
PY00443	hypothetical protein	0.484125791	-	-	-	-	-	R	
PY00553	hypothetical protein	0.45154321	-	-	-	-	-	-	<100nt
PY00931	hypothetical protein	0.432248034	-	-	-	-	PF14_0544	S	
PY01237	hypothetical protein	0.017045984	-	-	-	-	-	S	
PY01962	hypothetical protein	0.491139674	-	-	-	-	-	R	<100nt
PY02299	hypothetical protein	0.492596749	Y	Y	-	-	-	R	
PY02533	hypothetical protein	0.205989895	-	Y	-	-	-	R	
PY03172	hypothetical protein	0.250848058	-	-	-	-	-	-	
PY03805	hypothetical protein	0.331808798	-	-	-	-	-	R	
PY03806	hypothetical protein	0.404233439	-	-	-	-	-	-	
PY04313	hypothetical protein	0.101690358	-	Y	-	-	-	T	
PY04314	hypothetical protein	0.360588752	-	-	-	-	-	-	
PY04315	hypothetical protein	0.086767119	Y	-	-	-	-	R	
PY04317	hypothetical protein	0.125412057	Y	Y	-	-	-	R	
PY04463	hypothetical protein	0.21829561	-	-	-	-	-	-	<100nt
PY04465	hypothetical protein	0.09266184	-	-	-	-	-	R	<100nt
PY04579	hypothetical protein	0.442542375	Y	Y	-	-	-	T	
PY04738	hypothetical protein	0.219351506	-	-	-	-	-	R	<100nt
PY04933	hypothetical protein	0.106777992	-	-	-	-	PFD0385w	T	
PY04938	hypothetical protein	0.287073979	Y	-	-	-	PF11_0276	R	
PY04952	hypothetical protein	0.476008291	-	-	-	-	-	R	
PY04963	hypothetical protein	0.029317503	Y	Y	-	-	-	-	
PY04993	hypothetical protein	0.426245546	-	-	-	-	PF10_0240	R	
PY05074	hypothetical protein	0.405990984	-	-	-	-	-	T	
PY05096	hypothetical protein	0.440543927	Y	-	-	-	MAL7P1.17	T	
PY05187	hypothetical protein	0.470466919	-	-	-	-	-	R	
PY05408	hypothetical protein	0.247150824	Y	-	-	-	-	R	
PY05409	hypothetical protein	0.087248441	-	-	-	-	-	R	
PY05454	hypothetical protein	0.031749242	Y	Y	DRILSDA	-	-	-	
PY05506	hypothetical protein	0.454352046	Y	-	-	-	-	-	
PY05862	hypothetical protein	0.32244439	-	-	-	-	-	R	
PY05956	hypothetical protein	0.131817579	-	-	-	-	-	T	
PY06363	hypothetical protein	0.044966256	-	-	-	-	-	R	
PY06364	hypothetical protein	0.034345999	-	-	-	-	-	-	
PY06366	hypothetical protein	0.028319016	-	-	-	-	-	R	<100nt
PY07015	hypothetical protein	0.037646655	-	-	-	-	-	R	
PY07016	hypothetical protein	0.043952263	-	-	-	-	-	R	
PY07170	hypothetical protein	0.382583148	-	-	-	-	-	R	
PY07278	hypothetical protein	0.224080709	-	Y	-	-	-	R	
PY07279	hypothetical protein	0.063593663	Y	Y	-	-	-	-	
PY07366	hypothetical protein	0.038526015	-	-	-	-	-	-	
PY07367	hypothetical protein	0.047582536	-	-	-	-	-	R	
PY07376	hypothetical protein	0.025899364	-	-	-	-	-	-	
PY07377	hypothetical protein	0.071782424	-	-	-	-	-	-	
PY07503	hypothetical protein	0.028672538	-	-	-	-	-	T	
PY07504	hypothetical protein	0.038880232	Y	Y	-	-	MAL7P1.142	T	
PY07551	hypothetical protein	0.094984988	Y	Y	-	-	-	-	
PY07552	hypothetical protein	0.023540643	-	-	-	-	-	T	
PY07553	hypothetical protein	0.020146223	-	-	-	-	-	R	
PY07554	hypothetical protein	0.035472258	-	-	-	-	-	R	

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcript	Comment
PY07555	hypothetical protein	0.027236699	-	-	-	-	-	-	<100nt
PY07556	hypothetical protein	0.02642709	-	-	-	-	-	S	
PY07557	hypothetical protein	0.022256231	-	-	-	-	-	-	<100nt
PY07563	hypothetical protein	0.245724989	-	-	-	-	-	-	<100nt
PY07564	hypothetical protein	0.059420414	-	-	-	-	-	T	
PY07565	hypothetical protein	0.170652161	-	-	-	-	-	-	<100nt
PY07567	hypothetical protein	0.03800043	-	-	-	-	-	R	
PY07568	hypothetical protein	0.020233351	-	-	-	-	-	-	
PY07569	hypothetical protein	0.024021461	-	-	-	-	-	-	<100nt
PY07570	hypothetical protein	0.045013726	-	-	-	-	-	-	
PY07571	hypothetical protein	0.018388564	-	-	-	-	-	-	<100nt
PY07572	hypothetical protein	0.024246454	-	-	-	-	-	R	<100nt
PY07573	hypothetical protein	0.021901035	-	-	-	-	-	-	<100nt
PY07580	hypothetical protein	0.297728726	-	-	-	-	-	-	
PY07581	hypothetical protein	0.481707545	-	-	-	-	-	-	
PY07582	hypothetical protein	0.293241716	-	-	-	-	-	S	
PY07585	hypothetical protein	0.020504108	-	-	-	-	-	R	
PY07586	hypothetical protein	0.032135194	-	-	-	-	-	-	
PY07587	hypothetical protein	0.024389421	-	-	-	-	-	-	
PY07588	hypothetical protein	0.023599148	-	-	-	-	-	-	
PY07589	hypothetical protein	0.052198311	-	-	-	-	-	R	
PY07590	hypothetical protein	0.033673784	-	-	-	-	-	-	<100nt
PY07591	hypothetical protein	0.032357972	-	-	-	-	-	T	<100nt
PY07592	hypothetical protein	0.024709141	Y	Y	-	-	-	T	
PY07602	hypothetical protein	0.06368317	-	-	-	-	-	-	
PY07603	hypothetical protein	0.080945838	-	-	-	-	-	R	<100nt
PY07604	hypothetical protein	0.025718386	-	-	-	-	-	R	
PY07605	hypothetical protein	0.085830338	-	-	-	-	-	R	
PY07612	hypothetical protein	0.043053912	-	-	-	-	-	-	<100nt
PY07649	hypothetical protein	0.033569154	Y	-	-	-	-	T	
PY07663	hypothetical protein	0.072757647	-	-	-	-	-	-	
PY07664	hypothetical protein	0.149654418	-	-	-	-	-	-	
PY07665	hypothetical protein	0.031538607	Y	-	-	-	-	-	
PY07670	hypothetical protein	0.243113085	Y	Y	-	-	-	-	
PY07671	hypothetical protein	0.030064745	Y	Y	-	-	-	-	
PY07676	hypothetical protein	0.397511905	-	Y	-	-	-	R	
PY07677	hypothetical protein	0.360095946	-	-	-	-	-	-	
PY07699	hypothetical protein	0.497120139	-	-	-	-	-	R	Bacterial
PY07814	hypothetical protein	0.479394916	Y	-	-	-	-	T	Bacterial
PY07828	hypothetical protein	0.159475134	Y	Y	-	-	-	-	
PY07837	hypothetical protein	0.037818516	Y	Y	-	-	MAL7P1.142	-	
PY07845	hypothetical protein	0.164794926	Y	-	-	-	-	-	
PY07846	hypothetical protein	0.05563828	-	-	-	-	-	-	
PY07847	hypothetical protein	0.019911273	-	-	-	-	-	-	
PY07852	hypothetical protein	0.056675616	-	-	-	-	-	T	<100nt
PY07853	hypothetical protein	0.038772757	-	-	-	-	-	-	<100nt
PY07854	hypothetical protein	0.027095417	-	-	-	-	-	-	
PY07855	hypothetical protein	0.02388228	-	-	-	-	-	T	
PY07856	hypothetical protein	0.040557345	-	-	-	-	-	-	<100nt
PY07857	hypothetical protein	0.087618822	-	-	-	-	-	-	
PY07859	hypothetical protein	0.038842321	-	-	-	-	-	-	<100nt
PY07774	inorganic phosphate transporter	0.35585571	Y	-	-	inorganic phosphate transmembrane transporter activ	MAL13P1.206	-	Bacterial
PY07014	integrase-related	0.051063512	-	-	-	-	-	R	Rodent
PY06984	L1 transposable element	0.025856792	-	-	-	-	-	R	Rodent
PY01605	NAP-like protein	0.092732697	-	-	-	-	PF10930c	R	
							MAL13P1.176, PFL2520w, PF13_0198		
PY06381	p235 rhopty protein E5	0.090432246	Y	Y	-	-		S	
PY07579	pORF1	0.099980881	-	-	-	-	-	-	Rodent

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcript	Comment
PY00979	putative bir1 protein	0.279314825	Y	-	-	-	-	-	
PY01966	putative bir1 protein	0.077672065	Y	-	-	-	-	-	
PY03095	putative bir1 protein	0.188590192	Y	-	-	-	-	-	
PY04737	putative bir1 protein	0.106436488	Y	-	-	-	-	-	
PY04804	putative bir1 protein	0.074311754	Y	-	-	-	-	-	
PY06742	putative bir1 protein	0.472768308	Y	-	-	-	-	-	
PY07169	putative bir1 protein	0.444391144	Y	-	-	-	-	-	
PY01162	putative yir1 protein	0.435810878	Y	-	-	-	-	R	
PY03825	putative yir1 protein	0.16668851	Y	Y	-	-	-	-	
PY04478	putative yir1 protein	0.178290603	Y	-	-	-	-	-	
PY04505	putative yir1 protein	0.304164528	Y	-	-	-	-	-	
PY04923	putative yir1 protein	0.298830833	Y	Y	-	-	-	-	
PY07090	putative yir1 protein	0.180300965	Y	-	-	-	-	-	
PY02562	putative yir2 protein	0.085020572	Y	-	-	-	-	-	
PY03094	putative yir2 protein	0.112708028	Y	-	-	-	-	-	
PY03966	putative yir2 protein	0.117283616	Y	-	-	-	-	R	
PY04964	putative yir2 protein	0.142980254	Y	-	-	-	-	R	
PY05731	putative yir2 protein	0.032827899	Y	-	-	-	-	-	
PY06536	putative yir2 protein	0.498593796	Y	-	-	-	-	R	
PY06549	putative yir2 protein	0.040522317	Y	-	-	-	-	R	
PY01788	putative yir3 protein	0.470810418	Y	-	-	-	-	-	
PY01963	putative yir3 protein	0.478567399	Y	-	-	-	-	-	
PY02701	putative yir3 protein	0.349735706	Y	-	-	-	-	-	
PY03091	putative yir3 protein	0.376588882	Y	Y	-	-	-	-	
PY03193	putative yir3 protein	0.283482055	Y	-	-	-	-	-	
PY03195	putative yir3 protein	0.151133222	Y	-	-	-	-	-	
PY04464	putative yir3 protein	0.213303139	Y	-	-	-	-	-	
PY05617	putative yir3 protein	0.172655608	Y	Y	-	-	-	-	
PY05923	putative yir3 protein	0.496312558	Y	-	-	-	-	R	
PY06920	putative yir3 protein	0.243659486	Y	-	-	-	-	-	
PY07263	putative yir3 protein	0.140120701	Y	-	-	-	-	-	
PY07500	putative yir3 protein	0.129000839	Y	-	-	-	-	R	
PY07594	putative yir3 protein	0.158129864	Y	-	-	-	-	-	
PY07618	putative yir3 protein	0.38641232	Y	-	-	-	-	-	
PY07632	putative yir3 protein	0.398606235	Y	-	-	-	-	-	
PY00177	putative yir4 protein	0.077653224	Y	-	-	-	-	R	
PY00924	putative yir4 protein	0.098484664	Y	Y	-	-	-	R	
PY01101	putative yir4 protein	0.281715379	-	-	-	-	-	-	
PY02278	putative yir4 protein	0.293065451	Y	-	-	-	-	-	
PY02873	putative yir4 protein	0.312779819	Y	-	-	-	-	-	
PY03315	putative yir4 protein	0.213637052	Y	-	-	-	-	R	
PY03330	putative yir4 protein	0.129974024	Y	Y	-	-	-	R	
PY04419	putative yir4 protein	0.379996222	Y	-	-	-	-	-	
PY04914	putative yir4 protein	0.448628269	Y	-	-	-	-	-	
PY05507	putative yir4 protein	0.490040231	Y	-	-	-	-	-	
PY06410	putative yir4 protein	0.269392346	-	-	-	-	-	-	
PY06411	putative yir4 protein	0.129678755	Y	-	-	-	-	S	
PY06807	putative yir4 protein	0.056856751	Y	-	-	-	-	R	
PY07281	putative yir4 protein	0.207851647	Y	-	-	-	-	-	
PY07786	pyrroloquinoline quinone biosynthesis protein E-rela	0.452695631	-	-	-	-	-	T	Bacterial
PY07669	Reverse transcriptase, putative	0.031614852	-	-	-	-	-	-	Rodent
PY05142	Ribosomal protein L21e, putative	0.407402992	-	-	-	-	PF14_0240	R	
PY00947	RNA recognition motif, putative	0.498735385	-	-	-	-	PF13_0318	T	
PY07288	RNase H, putative	0.043408497	-	-	-	-	-	-	Rodent
PY07375	similar to L1 retrotransposon, TF subfamily L1MD-TF18	0.026740725	-	-	-	-	-	R	Rodent
PY01520	strong similarity to unknown protein-related	0.294323751	-	-	-	-	PFL0130c	S	
PY07613	titin-L1 fusion protein-related	0.052687048	-	-	-	-	-	-	Rodent
PY07374	Unknown protein	0.062866916	-	-	-	-	-	-	Rodent

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
-------------	---------	------------	----------------------	----------------	-------	-----------------------	-------------	-----------------------	---------

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcript	Comment
PY04930	235 kDa rhostry protein	4.193624993	Y	-	-	nucleotide binding	PFL2520w, PF13_0198, MAL13P1.176	T	
PY01007	93 kDa protein	2.172091589	-	Y	-	-	-	R	
PY00021	Actin	2.57583122	-	-	-	-	PF14_0218	S	
PY01581	apical membrane antigen-1	3.887284227	Y	-	-	-	PF11_0344	T	
PY01120	Arabidopsis thaliana BRAHMA ortholog-related	8.962881638	-	-	-	-	PFB0730w	S	
PY02350	asparagine-rich protein	3.600212528	-	-	-	-	PF13_0283	S	
PY03561	asparagine-rich protein	5.694803055	-	-	-	-	-	R	
PY02085	aspartyl protease-like	3.215999897	Y	Y	-	aspartic-type endopeptidase activity, pepsin A activity	PF13_0133	T	
PY04108	ATP-dependent helicase ddx8	2.171230226	-	-	-	-	MAL13P1.322 , PF10_0294, PF10860c	R	
PY05109	ATPase associated with chromosome architecture/replication	2.858130295	-	-	-	-	PF14_0466	R	
PY00386	CCAAT-box DNA binding protein subunit B	2.15440532	Y	-	-	-	PF14_0154	S	
PY00689	CCAAT-box DNA binding protein subunit B	2.682768694	-	-	-	-	PF14_0271	-	
PY02393	CCAAT-box DNA binding protein subunit B	2.008011048	-	-	-	-	PF11150w	R	
PY05072	CCAAT-box DNA binding protein subunit B	3.827114258	Y	-	-	-	PFF0870w	T	
PY05674	CCAAT-box DNA binding protein subunit B	2.702105824	-	-	-	-	-	T	
PY02289	CCAAT-box DNA binding protein subunit B, putative	4.517824926	-	-	-	-	PFL0530c	R	
PY00653	chromosome segregation protein smc1	8.979870904	-	-	-	-	PF11_0317	T	
PY02565	cir1 protein, putative	4.761021407	Y	-	-	-	-	-	
PY04787	cir1 protein, putative	17.20011404	Y	-	-	-	-	-	
PY06886	cir1 protein, putative	2.730990008	Y	-	-	-	-	-	
PY07826	cir1 protein, putative	3.606584556	Y	-	-	-	-	R	
PY02757	cir2 protein	4.567179214	Y	-	-	-	-	-	
PY03618	coronin binding protein	4.753338104	-	-	-	-	PF14_0673	R	
PY02252	deoxyribose-phosphate aldolase	3.684246313	-	-	-	-	PF10_0210	S	
PY02267	Dictyostellum discoideum HSPC148. 10/100, putative	2.644460929	-	-	-	-	PF07_0091	T	
PY03839	Domain of unknown function, putative	26.62132226	-	-	-	-	PF08_0055	R	
PY04724	Drosophila melanogaster BcDNA.GH02833	3.077979874	-	-	-	-	MAL7P1.113	R	
PY01447	Drosophila melanogaster BcDNA.GH06032	2.510423688	Y	-	-	catalytic activity, ATP binding	PF07_0115	R	
PY07506	Drosophila melanogaster CG12781 gene product	6.234613522	Y	Y	-	-	-	S	
PY04549	Drosophila melanogaster CG15040 gene product	3.409937901	Y	-	-	-	-	T	
PY07267	dynein 14 kDa light chain, flagellar outer arm., putative	2.822387033	-	-	-	-	PF14_0590	S	
PY06583	Endonuclease/Exonuclease/phosphatase family	2.608993917	-	-	-	-	PFA0350w	R	
PY02480	erythrocyte binding protein homolog	2.040963954	-	-	-	-	PF10670w	R	
PY00637	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	2.05059815	-	-	-	-	PF08_0012	T	
PY06125	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	2.596242504	-	-	-	-	PF13_0056	R	
PY07447	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	2.895713687	-	-	-	-	-	T	
PY06959	erythrocyte membrane-associated antigen	2.277952044	Y	Y	-	aspartic-type endopeptidase activity	-	T	
PY00199	Eukaryotic phosphomannomutase	2.624674785	-	-	-	-	PF10_0169	-	
PY03767	Helicase conserved C-terminal domain, putative	2.606330007	-	Y	-	-	PF13_0077	R	
PY05489	histidine kinase DhkG	5.122952308	-	-	-	-	PF13_0237	T	
PY05189	homeobox-containing protein	2.535111587	-	-	-	-	PF10_0045	R	
PY00007	hypothetical protein	2.499447322	-	-	-	-	PF10_0075	R	
PY00143	hypothetical protein	2.814839169	-	Y	-	-	MAL7P1.208	T	
PY00178	hypothetical protein	5.184288373	Y	Y	-	-	PFL0790w	R	
PY00237	hypothetical protein	2.600256289	-	Y	-	-	-	-	
PY00263	hypothetical protein	2.193207468	-	-	-	-	MAL13P1.154	-	
PY00307	hypothetical protein	2.345373541	Y	-	-	-	-	-	
PY00311	hypothetical protein	2.552301779	Y	Y	-	-	-	-	<100nt
PY00408	hypothetical protein	3.365066039	-	-	-	-	-	R	
PY00441	hypothetical protein	3.249257926	-	-	-	-	PFB0475c	T	
PY00476	hypothetical protein	2.276301245	-	-	-	-	PF13_0082	S	
PY00537	hypothetical protein	2.02025075	-	-	-	-	PFL1340c	T	
PY00568	hypothetical protein	2.102231072	-	-	-	-	MAL8P1.80	T	
PY00575	hypothetical protein	3.269496487	Y	Y	-	-	-	-	
PY00576	hypothetical protein	2.337143364	Y	-	-	NRILGDV	-	-	
PY00581	hypothetical protein	2.664054413	Y	-	-	-	PFL1065c	R	
PY00596	hypothetical protein	2.727921314	Y	Y	-	-	PF07_0070	R	
PY00600	hypothetical protein	2.964902779	Y	-	-	-	PF10_0199	R	
PY00628	hypothetical protein	2.818456621	-	-	-	-	PFA0410w	R	
PY00636	hypothetical protein	2.722685407	Y	-	-	-	-	T	
PY00647	hypothetical protein	3.568759582	-	-	-	-	-	R	

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY00677	hypothetical protein	2.108088817	Y	Y	-	-	PFL1415w	T	
PY00682	hypothetical protein	2.137292495	-	-	-	-	-	R	
PY00694	hypothetical protein	13.71774886	-	-	-	-	PFE1435c	R	
PY00723	hypothetical protein	4.864169644	-	-	-	-	PF11_0155	R	
PY00871	hypothetical protein	2.899812819	-	-	-	-	-	-	
PY00890	hypothetical protein	3.158857486	Y	-	-	metal ion transmembrane transporter activity	PF11_0210	-	
PY00936	hypothetical protein	4.02801875	-	-	-	-	-	T	
PY00953	hypothetical protein	5.26309153	-	-	-	-	PF10780w	R	
PY00991	hypothetical protein	2.390071652	-	-	-	-	-	S	<100nt
PY01003	hypothetical protein	2.946783719	-	-	-	-	-	T	
PY01006	hypothetical protein	2.846066473	Y	-	-	-	-	T	
PY01012	hypothetical protein	2.411420629	-	-	-	-	PF10_0358	R	
PY01019	hypothetical protein	4.423034993	-	-	-	-	-	-	
PY01057	hypothetical protein	3.867866709	-	-	-	-	-	-	
PY01073	hypothetical protein	2.656131632	Y	Y	-	-	MAL8P1.130	T	
PY01083	hypothetical protein	2.562563701	Y	-	-	-	-	-	
PY01141	hypothetical protein	2.136184479	Y	Y	-	-	-	-	
PY01160	hypothetical protein	2.836285268	Y	-	-	-	-	R	
PY01176	hypothetical protein	8.389844574	Y	Y	-	-	-	-	
PY01207	hypothetical protein	2.274640304	-	Y	-	-	-	T	
PY01219	hypothetical protein	3.050105289	-	-	-	-	-	T	
PY01234	hypothetical protein	2.280017691	Y	-	-	-	PF13_0235	R	
PY01236	hypothetical protein	4.447739661	Y	Y	-	-	-	-	
PY01262	hypothetical protein	4.843244527	-	-	-	-	-	T	<100nt
PY01280	hypothetical protein	3.779748359	-	-	-	-	PFE1410c	T	
PY01297	hypothetical protein	2.768925479	-	-	-	-	-	S	
PY01308	hypothetical protein	2.933931383	-	-	-	-	PF07_0119	R	
PY01326	hypothetical protein	7.777013521	-	-	-	-	-	-	
PY01387	hypothetical protein	2.301945082	-	-	-	-	MAL8P1.123	R	
PY01401	hypothetical protein	3.81082485	-	-	-	-	-	S	
PY01405	hypothetical protein	3.120953129	-	-	-	-	-	-	
PY01414	hypothetical protein	3.071843032	Y	-	-	-	-	R	
PY01415	hypothetical protein	2.111358798	-	-	-	-	-	R	
PY01443	hypothetical protein	2.853072176	-	-	-	-	-	R	
PY01507	hypothetical protein	2.848701956	Y	-	-	-	PF13_0202	S	
PY01538	hypothetical protein	2.299381226	-	-	-	-	-	-	
PY01588	hypothetical protein	4.641180318	-	-	-	-	-	R	
PY01590	hypothetical protein	2.114583965	-	-	-	-	-	R	
PY01602	hypothetical protein	14.98052386	Y	Y	NRILADA	-	-	T	
PY01650	hypothetical protein	5.638366671	Y	-	-	-	-	-	
PY01673	hypothetical protein	3.581427206	-	-	-	-	-	-	
PY01686	hypothetical protein	5.645024957	Y	-	-	-	-	-	
PY01691	hypothetical protein	2.17573736	-	-	-	-	MAL13P1.157	-	
PY01715	hypothetical protein	2.196339507	-	-	-	-	-	S	
PY01729	hypothetical protein	2.485597093	Y	Y	-	-	-	R	
PY01747	hypothetical protein	4.939852688	-	-	-	-	-	-	
PY01772	hypothetical protein	2.33471654	-	-	-	-	-	T	
PY01789	hypothetical protein	2.064912656	Y	-	-	-	-	-	<100nt
PY01809	hypothetical protein	2.693745206	Y	-	-	-	PFL2390c	R	
PY01838	hypothetical protein	4.232815819	-	-	-	-	-	-	
PY01843	hypothetical protein	3.351985325	-	Y	-	-	-	-	
PY01875	hypothetical protein	3.110693572	-	-	-	-	-	T	
PY01975	hypothetical protein	3.736155772	Y	-	-	-	-	-	
PY01978	hypothetical protein	9.731616745	Y	-	-	-	-	T	
PY01980	hypothetical protein	2.294297014	-	-	-	-	-	R	
PY01985	hypothetical protein	4.569147646	Y	-	-	-	-	-	
PY02031	hypothetical protein	2.146839954	-	-	-	-	PFF1160w	R	
PY02043	hypothetical protein	2.063701085	Y	Y	-	-	-	-	<100nt
PY02053	hypothetical protein	3.874300166	-	-	-	-	MAL7P1.123	S	
PY02078	hypothetical protein	4.614082662	Y	-	-	-	PF10_0290	T	
PY02103	hypothetical protein	3.800758007	-	-	-	-	PF13_0183	-	
PY02133	hypothetical protein	2.518094804	-	-	-	-	-	T	
PY02152	hypothetical protein	3.284771443	Y	-	-	-	MAL8P1.13, PF11_0172	-	
PY02154	hypothetical protein	2.736050773	-	-	-	-	PF11_0170	R	

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY02184	hypothetical protein	31.44809341	Y	-	-	-	-	-	<100nt
PY02199	hypothetical protein	2.324730584	Y	-	-	binding	PFD0720w	S	
PY02211	hypothetical protein	3.05954398	Y	-	-	-	PFF1075w	R	
PY02301	hypothetical protein	2.465201146	-	Y	-	-	PF14_0344	-	
PY02313	hypothetical protein	3.72389538	-	-	-	-	-	-	
PY02356	hypothetical protein	12.88096066	-	Y	-	-	-	-	
PY02366	hypothetical protein	2.112136506	-	-	-	-	-	T	
PY02382	hypothetical protein	2.376131252	-	Y	-	-	-	T	
PY02392	hypothetical protein	2.197184933	Y	Y	-	-	-	-	
PY02399	hypothetical protein	2.528146419	Y	Y	-	-	-	-	
PY02434	hypothetical protein	2.986925681	-	-	-	-	PF11_0388	S	
PY02465	hypothetical protein	2.57782214	-	-	-	-	PFD1030c	T	
PY02467	hypothetical protein	5.194837568	-	-	-	-	-	T	
PY02470	hypothetical protein	2.379940709	-	-	-	-	PFL1190c	-	
PY02489	hypothetical protein	3.860978174	-	-	-	-	PF11050c	T	
PY02546	hypothetical protein	3.141987987	-	-	-	-	-	T	
PY02586	hypothetical protein	2.694830496	-	-	-	-	PF14_0389	R	
PY02593	hypothetical protein	2.034480763	-	-	-	-	MAL13P1.33	S	
PY02628	hypothetical protein	5.061426733	-	-	-	-	-	-	<100nt
PY02630	hypothetical protein	2.643801024	-	Y	-	-	-	-	<100nt
PY02684	hypothetical protein	2.399286154	-	-	-	-	-	T	
PY02714	hypothetical protein	2.202839865	-	-	-	-	PF11_0398	R	
PY02731	hypothetical protein	2.006309963	Y	Y	-	-	-	-	<100nt
PY02789	hypothetical protein	2.436876991	-	-	-	-	PF11_0078	-	
PY02792	hypothetical protein	3.525838633	-	-	-	-	PFE0240c	T	
PY02793	hypothetical protein	2.899796412	-	-	-	-	-	T	
PY02895	hypothetical protein	12.36593308	-	-	-	-	PFE1060c	R	
PY02927	hypothetical protein	2.453192854	-	-	-	-	-	T	
PY02934	hypothetical protein	3.218045712	Y	-	-	-	PFE0260w	T	
PY02952	hypothetical protein	3.164422897	-	-	-	-	-	R	
PY02954	hypothetical protein	2.416460596	-	-	-	-	-	-	
PY02989	hypothetical protein	2.453868773	Y	-	-	-	PFB0580w	R	
PY03032	hypothetical protein	2.572592383	-	-	-	-	-	T	<100nt
PY03061	hypothetical protein	2.02159752	Y	-	-	-	-	-	
PY03084	hypothetical protein	4.952028864	-	-	-	-	PF10_0146	-	
PY03085	hypothetical protein	3.356459116	-	Y	-	-	-	-	<100nt
PY03129	hypothetical protein	2.160029526	Y	-	-	-	PF10465c	T	
PY03243	hypothetical protein	2.525596283	-	-	-	-	PFF0985c	T	
PY03257	hypothetical protein	3.921760753	-	-	-	-	PFE0455w	T	
PY03260	hypothetical protein	10.38729491	-	-	-	-	-	S	
PY03322	hypothetical protein	5.100787834	Y	-	-	-	-	S	
PY03340	hypothetical protein	16.88239963	Y	-	-	-	-	-	
PY03345	hypothetical protein	2.091315435	-	-	-	-	-	-	<100nt
PY03352	hypothetical protein	3.449717705	-	-	-	-	PF14_0089	S	
PY03366	hypothetical protein	2.130632917	-	-	-	-	-	-	
PY03369	hypothetical protein	2.630189269	Y	Y	-	-	PF11_0073	R	
PY03404	hypothetical protein	2.433842597	-	-	-	-	-	-	
PY03405	hypothetical protein	4.399655869	Y	Y	NRILGDV	-	-	S	
PY03407	hypothetical protein	2.443249842	Y	-	-	-	MAL13P1.87	S	
PY03430	hypothetical protein	10.76698408	Y	-	-	-	-	T	
PY03435	hypothetical protein	2.373056676	-	-	-	-	-	-	<100nt
PY03458	hypothetical protein	2.936945243	-	-	-	-	PFC0701w	R	
PY03474	hypothetical protein	6.241573753	-	-	-	-	-	S	<100nt
PY03484	hypothetical protein	2.18023513	Y	-	-	protein binding	PFL0505c	S	
PY03495	hypothetical protein	3.405220667	-	-	-	-	-	T	
PY03568	hypothetical protein	15.98069309	-	-	-	-	-	-	
PY03586	hypothetical protein	2.300067809	-	-	-	-	MAL8P1.20	T	
PY03590	hypothetical protein	2.011107828	Y	-	-	-	PFL1820w	R	
PY03608	hypothetical protein	3.038259141	-	-	-	-	-	-	
PY03621	hypothetical protein	3.720929829	-	-	-	-	-	-	<100nt
PY03624	hypothetical protein	19.4793995	-	-	-	-	-	-	<100nt
PY03657	hypothetical protein	12.6272077	Y	Y	-	-	-	R	
PY03708	hypothetical protein	4.121133664	Y	-	-	-	-	R	
PY03759	hypothetical protein	3.593152174	-	-	-	-	-	R	
PY03762	hypothetical protein	3.178675147	Y	-	-	-	-	-	

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY03836	hypothetical protein	5.59855155	-	-	-	-	-	R	
PY03848	hypothetical protein	18.20364092	Y	-	-	-	-	-	
PY03950	hypothetical protein	5.816604282	-	-	-	-	-	R	
PY03951	hypothetical protein	19.17413793	-	-	-	-	-	T	
PY03952	hypothetical protein	9.449036031	Y	-	-	-	-	-	
PY03959	hypothetical protein	2.052166214	Y	Y	-	-	PF14_0094	T	
PY03968	hypothetical protein	4.077111647	Y	-	-	-	-	R	
PY03975	hypothetical protein	2.201121563	-	-	-	-	-	-	<100nt
PY03993	hypothetical protein	3.136701602	Y	-	-	-	PFE0580w	T	
PY03997	hypothetical protein	14.37574583	-	-	-	-	-	R	
PY04004	hypothetical protein	2.034098601	-	-	-	-	PFL1365w	R	
PY04022	hypothetical protein	2.713740119	Y	Y	-	-	-	R	
PY04043	hypothetical protein	2.528543938	-	-	-	-	PF11_0140	S	
PY04078	hypothetical protein	2.696950383	-	-	-	-	PF11_0439	R	
PY04109	hypothetical protein	3.450176436	Y	-	-	-	-	-	
PY04123	hypothetical protein	3.060525376	Y	-	-	-	-	-	
PY04137	hypothetical protein	3.63377705	-	Y	-	-	PF14_0792	S	
PY04152	hypothetical protein	6.226516342	-	Y	-	-	-	-	
PY04210	hypothetical protein	3.161799461	Y	-	-	-	-	R	
PY04227	hypothetical protein	6.059211971	-	-	-	-	PFF0990c	T	
PY04279	hypothetical protein	4.108602279	-	-	-	-	-	S	
PY04333	hypothetical protein	2.54866414	-	Y	-	-	-	R	
PY04348	hypothetical protein	2.024665921	Y	-	-	-	-	S	
PY04355	hypothetical protein	2.352247529	Y	-	-	-	PF11295c, PFB0465c	S	
PY04397	hypothetical protein	6.339690469	-	-	-	-	-	R	
PY04398	hypothetical protein	3.01204546	Y	Y	-	-	-	-	
PY04401	hypothetical protein	2.857620214	Y	-	-	-	PF11_0176	S	
PY04418	hypothetical protein	2.17591325	-	-	-	-	-	R	
PY04420	hypothetical protein	7.441219759	Y	Y	-	-	-	-	
PY04550	hypothetical protein	7.024641952	-	-	-	-	PFA0325w	T	
PY04553	hypothetical protein	2.78147182	-	-	-	-	-	-	
PY04558	hypothetical protein	2.584068716	-	-	-	-	PF13_0221	S	
PY04679	hypothetical protein	2.512294155	-	-	-	-	-	R	
PY04683	hypothetical protein	3.870597673	Y	-	-	-	PF08_0090	-	
PY04709	hypothetical protein	2.701127133	-	-	-	-	-	R	
PY04734	hypothetical protein	2.909396866	Y	-	NRILGDV	-	-	-	
PY04827	hypothetical protein	2.300540698	-	-	-	-	-	-	
PY04839	hypothetical protein	2.154075045	-	-	-	-	-	R	
PY04842	hypothetical protein	2.820645577	-	-	-	-	PF10650c	R	
PY04847	hypothetical protein	2.180913793	-	-	-	-	-	R	<100nt
PY04866	hypothetical protein	2.751102071	Y	Y	-	-	PFD0555c	R	
PY04876	hypothetical protein	2.396597919	-	-	-	-	-	R	
PY04932	hypothetical protein	2.951712996	-	-	-	-	-	R	
PY05012	hypothetical protein	2.463471397	-	-	-	-	-	-	
PY05013	hypothetical protein	2.296052272	-	-	-	-	-	-	
PY05041	hypothetical protein	5.27298613	-	-	-	-	PFB0280w	R	
PY05042	hypothetical protein	4.606499127	-	-	-	-	-	S	<100nt
PY05093	hypothetical protein	2.647902068	-	-	-	-	-	-	
PY05097	hypothetical protein	2.861786396	-	-	-	-	-	-	<100nt
PY05134	hypothetical protein	3.661884405	-	-	-	-	-	-	
PY05147	hypothetical protein	3.079769955	-	-	-	-	-	R	<100nt
PY05167	hypothetical protein	5.271726621	-	-	-	-	PF08_0058	S	
PY05195	hypothetical protein	8.590889315	-	-	-	-	PF11190w	R	
PY05202	hypothetical protein	2.873327579	Y	-	-	-	PF10_0188	R	
PY05219	hypothetical protein	3.255363794	-	-	-	-	MAL13P1.333	R	
PY05240	hypothetical protein	2.929572773	Y	-	-	-	-	-	
PY05291	hypothetical protein	6.380475499	-	-	-	-	-	-	<100nt
PY05323	hypothetical protein	2.19387098	Y	-	-	-	PFD0595w	T	
PY05333	hypothetical protein	2.419889503	-	-	-	-	-	-	<100nt
PY05347	hypothetical protein	3.331712148	Y	Y	-	-	MAL13P1.49	T	
PY05351	hypothetical protein	4.142016087	-	-	-	-	-	-	<100nt
PY05360	hypothetical protein	2.17791204	-	-	-	-	PF14_0689	T	
PY05456	hypothetical protein	10.11632134	-	-	-	-	-	R	<100nt
PY05466	hypothetical protein	2.001986004	-	Y	-	-	-	R	

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY05479	hypothetical protein	2.023621582	-	-	-	-	PFD0925w	R	
PY05541	hypothetical protein	4.066958193	Y	Y	-	protein-S-isoprenylcysteine O-methyltransferase activity	PFL1780w	R	
PY05589	hypothetical protein	2.125936587	-	-	-	-	PF11365w	T	
PY05594	hypothetical protein	4.468095337	-	-	-	-	PF14_0047	-	
PY05635	hypothetical protein	5.141648402	-	-	-	-	-	R	
PY05663	hypothetical protein	4.691738852	Y	-	-	-	MAL8P1.28	S	
PY05704	hypothetical protein	2.037269195	Y	-	-	-	-	-	
PY05723	hypothetical protein	2.739903367	Y	-	-	-	-	R	
PY05759	hypothetical protein	2.013491411	-	-	-	-	-	-	<100nt
PY05775	hypothetical protein	6.101192015	-	-	-	-	PFA0305c	T	
PY05783	hypothetical protein	6.695320482	-	-	-	-	PF10_0251	T	
PY05795	hypothetical protein	3.30666677	-	-	-	-	PF14_0252	T	
PY05813	hypothetical protein	2.298123142	-	-	-	-	PF10_0243	S	
PY05917	hypothetical protein	2.358607646	-	-	-	-	-	S	
PY05998	hypothetical protein	2.658756108	-	-	-	-	-	T	
PY06021	hypothetical protein	10.99074468	Y	Y	-	-	-	-	
PY06025	hypothetical protein	2.245541539	-	-	-	-	-	T	
PY06092	hypothetical protein	2.845757722	-	-	-	-	-	R	
PY06093	hypothetical protein	2.028908894	Y	Y	-	-	-	R	
PY06103	hypothetical protein	2.784370599	Y	-	-	-	-	-	
PY06128	hypothetical protein	2.148484692	-	-	-	-	PFC0425w	R	
PY06159	hypothetical protein	5.152751551	Y	-	-	-	-	R	
PY06174	hypothetical protein	2.520815789	-	-	-	-	-	-	<100nt
PY06193	hypothetical protein	8.448567233	-	-	-	-	PF08_0016	R	
PY06202	hypothetical protein	2.625525918	-	-	-	-	PFL1665c	R	
PY06229	hypothetical protein	6.125565338	-	-	-	-	-	R	
PY06317	hypothetical protein	5.835851861	Y	-	-	-	-	T	
PY06353	hypothetical protein	8.285341658	-	-	-	-	-	S	
PY06384	hypothetical protein	7.207123642	-	Y	-	-	-	-	<100nt
PY06494	hypothetical protein	2.236093117	-	Y	-	-	-	R	
PY06510	hypothetical protein	2.419431365	-	-	-	-	-	-	
PY06570	hypothetical protein	3.497891786	-	-	-	-	-	-	<100nt
PY06579	hypothetical protein	5.204588139	-	-	-	-	-	S	
PY06581	hypothetical protein	10.99321748	Y	Y	NRILGDV	-	-	R	
PY06582	hypothetical protein	2.062427948	-	-	-	-	-	-	
PY06594	hypothetical protein	3.790178927	-	-	-	-	-	R	<100nt
PY06616	hypothetical protein	2.809227812	-	-	-	-	-	-	<100nt
PY06617	hypothetical protein	2.484498673	Y	-	-	-	-	-	
PY06664	hypothetical protein	2.890082381	Y	-	-	-	PF10_0205	T	
PY06690	hypothetical protein	2.53594744	-	-	-	-	-	T	
PY06731	hypothetical protein	3.802156603	Y	-	-	-	-	-	
PY06781	hypothetical protein	3.001758454	-	-	-	-	PFL0445w	S	
PY06793	hypothetical protein	4.841578656	-	-	-	-	PF14_0200	S	
PY06806	hypothetical protein	3.34173163	-	-	-	-	-	R	<100nt
PY06815	hypothetical protein	5.002835983	-	-	-	-	-	R	
PY06850	hypothetical protein	12.62334394	-	-	-	-	-	T	
PY06879	hypothetical protein	3.454360573	-	-	-	-	-	-	
PY06881	hypothetical protein	2.98329324	Y	-	-	-	-	R	
PY06892	hypothetical protein	2.327697627	-	-	-	-	-	-	
PY06940	hypothetical protein	2.67215176	-	-	-	-	-	-	<100nt
PY06958	hypothetical protein	3.169236599	-	-	-	-	PF11_0269	-	
PY06989	hypothetical protein	2.460744621	-	-	-	-	-	R	<100nt
PY07034	hypothetical protein	2.604149852	Y	-	-	-	-	T	
PY07057	hypothetical protein	2.058102642	-	Y	NRLLFET	-	-	R	
PY07087	hypothetical protein	3.499038381	Y	Y	-	-	-	-	
PY07109	hypothetical protein	8.287273927	-	-	-	-	-	T	
PY07126	hypothetical protein	2.140700876	-	-	-	-	-	-	<100nt
PY07133	hypothetical protein	6.593726915	-	-	-	-	-	-	<100nt
PY07167	hypothetical protein	4.610953291	Y	Y	-	-	-	-	<100nt
PY07188	hypothetical protein	3.216013896	-	-	-	-	-	-	
PY07213	hypothetical protein	2.484285429	-	-	-	-	-	-	<100nt
PY07225	hypothetical protein	3.613774449	-	-	-	-	PF14_0680	R	
PY07231	hypothetical protein	10.13570747	Y	-	-	-	-	-	
PY07234	hypothetical protein	2.970801148	-	-	-	-	-	-	
PY07235	hypothetical protein	2.156657652	Y	-	-	-	-	R	

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY07245	hypothetical protein	2.406810824	-	-	-	-	-	S	
PY07362	hypothetical protein	3.432712473	-	-	-	-	MAL13P1.327	R	
PY07378	hypothetical protein	18.21603255	-	-	-	-	-	T	
PY07393	hypothetical protein	5.349912795	-	-	-	-	-	R	<100nt
PY07451	hypothetical protein	12.17975826	-	-	-	-	-	R	<100nt
PY07458	hypothetical protein	3.782823138	-	Y	-	-	-	-	
PY07478	hypothetical protein	4.861621066	-	-	-	-	-	-	
PY07534	hypothetical protein	6.506183344	-	-	-	-	-	-	<100nt
PY07548	hypothetical protein	8.173122023	-	-	-	-	-	-	<100nt
PY07559	hypothetical protein	9.857725551	Y	Y	-	-	-	-	
PY07562	hypothetical protein	2.830295832	Y	-	-	-	-	T	
PY07608	hypothetical protein	4.27514434	Y	Y	-	-	-	-	
PY07633	hypothetical protein	3.036467271	Y	-	-	-	-	-	
PY07651	hypothetical protein	3.153863047	-	-	-	-	PF11_0476	S	
PY05438	Leishmania major ppg3	15.24052859	Y	-	-	-	-	R	
PY01144	Leucine Rich Repeat, putative	2.697772596	-	-	-	-	-	R	
PY06687	Leucine Rich Repeat, putative	3.383432038	-	-	-	-	-	T	
PY03256	lipase precursor gehm. 6/101	2.46816383	-	-	-	-	PFE0460c	T	
PY03020	maebi	2.206528202	-	-	-	-	PF11_0207	R	
PY07153	Maf-like protein, putative	2.058176784	-	-	-	-	PF10310w	T	
PY02239	mature-parasite-infected erythrocyte surface antigen	5.159775445	-	-	-	-	-	T	
PY06415	merozoite surface protein precursor	5.603459479	Y	-	-	-	PFE0120c	R	
PY02559	methionine aminopeptidase, type II, putative	4.985639224	-	-	-	-	PF14_0327	R	
PY05981	MORN repeat, putative	2.948395449	-	-	-	-	PF10_0306	S	
PY04548	para-aminobenzoic acid synthetase	2.99998873	-	-	-	-	PF11100w	R	
							PFF0340c, PFC0205c, PF07_0036	S	
PY07597	Plasmodium falciparum CG6	2.917408866	-	-	-	-	-	R	
PY02647	Plasmodium vivax PV1H14195_P-related	6.243270295	-	-	-	-	-	-	
PY05888	Plasmodium vivax PV1H14220_P	3.088310922	Y	-	-	ubiquinol-cytochrome-c reductase activity	PFC0262c	T	
PY03610	Predicted membrane protein	3.073077341	Y	Y	-	-	-	T	
PY04363	probable tRNA modification GTPase trme	3.535425193	Y	Y	-	GTP binding	-	-	
PY00927	prolyl-tRNA synthetase-related	2.606208206	-	Y	-	-	PF11240c	R	
PY02094	proteasome subunit alpha type 5	2.586534314	-	-	-	-	PF07_0112	R	
PY02685	proteasome subunit beta type 1-related	2.793057115	-	-	-	-	PFE0915c	S	
PY00120	putative 26S proteasome subunits	4.531902649	-	-	-	-	PF14_0025	S	
PY00090	putative bir1 protein	3.446903401	Y	-	-	-	-	-	
PY00175	putative bir1 protein	14.61581989	Y	-	-	-	-	-	
PY00400	putative bir1 protein	16.02652966	Y	-	-	-	-	-	
PY01110	putative bir1 protein	20.38014389	Y	-	-	-	-	R	
PY01481	putative bir1 protein	7.826927479	Y	-	-	-	-	-	
PY02131	putative bir1 protein	3.686132391	Y	-	-	-	-	-	
PY02354	putative bir1 protein	14.26359726	Y	-	-	-	-	R	
PY02357	putative bir1 protein	10.23339054	Y	-	-	-	-	-	
PY02368	putative bir1 protein	19.26593045	Y	-	-	-	-	R	
PY02449	putative bir1 protein	11.44305276	Y	-	-	-	-	-	
PY02544	putative bir1 protein	6.849741662	Y	-	-	-	-	-	
PY02760	putative bir1 protein	45.95122737	Y	-	-	-	-	S	
PY02928	putative bir1 protein	5.055959367	Y	-	-	-	-	T	
PY03018	putative bir1 protein	5.906504699	Y	-	-	-	-	R	
PY03043	putative bir1 protein	8.09733022	Y	-	-	-	-	R	
PY03044	putative bir1 protein	3.260097464	Y	Y	-	-	-	R	
PY03359	putative bir1 protein	15.22235835	Y	-	-	-	-	-	
PY03399	putative bir1 protein	2.233083224	Y	-	-	-	-	-	
PY03400	putative bir1 protein	8.109820075	Y	-	-	-	-	-	
PY03436	putative bir1 protein	14.87627836	Y	-	-	-	-	-	
PY03695	putative bir1 protein	10.37559101	Y	-	-	-	-	-	
PY03809	putative bir1 protein	4.244710359	Y	-	-	-	-	R	
PY04321	putative bir1 protein	7.466936913	Y	-	-	-	-	-	
PY04379	putative bir1 protein	2.470698191	Y	-	-	-	-	-	
PY04758	putative bir1 protein	5.23152069	-	-	-	-	-	-	
PY04770	putative bir1 protein	4.677272172	Y	-	-	-	-	-	
PY05362	putative bir1 protein	33.94713365	Y	-	-	-	-	-	
PY05820	putative bir1 protein	4.942256478	Y	-	-	-	-	-	
PY06141	putative bir1 protein	4.649048402	Y	-	-	-	-	-	

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY06220	putative bir1 protein	19.62779602	Y	-	-	-	-	R	
PY06485	putative bir1 protein	2.877914298	Y	-	-	-	-	-	
PY06760	putative bir1 protein	3.162522904	Y	-	-	-	-	-	
PY06830	putative bir1 protein	24.56762301	Y	Y	-	-	-	-	
PY06988	putative bir1 protein	3.986977217	Y	-	-	-	-	-	
PY07012	putative bir1 protein	10.49019642	Y	-	-	-	-	-	
PY07187	putative bir1 protein	14.69563181	Y	-	-	-	-	-	
PY07226	putative bir1 protein	3.451862962	Y	-	-	-	-	-	
PY07246	putative bir1 protein	6.03629262	Y	-	-	-	-	T	
PY07252	putative bir1 protein	7.198908335	Y	-	-	-	-	-	
PY07356	putative bir1 protein	3.736883574	Y	-	-	-	-	-	
PY07516	putative bir1 protein	14.78247097	Y	-	-	-	-	-	
PY06451	putative clathrin assembly protein	2.212302676	-	-	-	-	PFD1090c	R	
PY03455	putative dual-specificity protein phosphatase	3.832313426	-	-	-	-	PFC0380w	S	
PY05402	putative HSP protein	2.39353708	-	Y	-	-	MAL13P1.540	R	
PY04859	putative initiation factor E4	2.851436043	Y	-	-	RNA binding, translation initiation factor activity	PFC0635c	R	
PY02994	putative methionyl-tRNA synthetase	2.725710236	-	-	-	-	PF14_0401	-	
PY01852	putative phosphatidylethanolamine-binding protein	2.286738775	-	-	-	-	PFL0955c	S	
PY03456	putative protein kinase	2.098762272	-	-	-	-	PFC0385c	T	
PY07229	putative small nuclear ribonucleoprotein polypeptide G	3.469197702	-	-	-	-	MAL8P1.48	R	
PY05764	putative ubiquinone biosynthesis protein	5.920207152	-	-	-	-	PF11_0128	R	
PY00590	putative ubiquitin-conjugating enzyme 16	2.022781414	-	Y	-	-	MAL13P1.227	R	
PY00268	putative yir1 protein	6.06285384	Y	-	-	-	-	-	
PY00279	putative yir1 protein	8.104563785	Y	-	-	-	-	R	
PY00301	putative yir1 protein	4.126710248	Y	Y	-	-	-	-	
PY00635	putative yir1 protein	5.216631491	Y	-	-	-	-	-	
PY00989	putative yir1 protein	15.03431409	Y	-	-	-	-	-	
PY01174	putative yir1 protein	8.336754489	Y	-	-	-	-	R	
PY01406	putative yir1 protein	2.366891156	Y	Y	-	-	-	-	
PY01417	putative yir1 protein	3.183654794	Y	-	-	-	-	T	
PY01466	putative yir1 protein	4.155704549	Y	-	-	-	-	-	
PY01595	putative yir1 protein	5.363529956	Y	-	-	-	-	R	
PY01769	putative yir1 protein	15.56193076	Y	-	-	-	-	-	
PY02038	putative yir1 protein	4.927618669	Y	-	-	-	-	-	
PY02042	putative yir1 protein	2.392718773	Y	-	-	-	-	-	
PY02137	putative yir1 protein	2.701302857	Y	-	-	-	-	-	
PY02163	putative yir1 protein	5.83644147	Y	-	-	-	-	-	
PY02629	putative yir1 protein	2.232161177	Y	Y	-	-	-	-	
PY02827	putative yir1 protein	9.777876193	Y	-	-	-	-	-	
PY03341	putative yir1 protein	10.0197046	Y	-	-	-	-	R	
PY03355	putative yir1 protein	8.232102027	Y	-	-	-	-	-	
PY03428	putative yir1 protein	7.046533157	Y	-	-	-	-	-	
PY03579	putative yir1 protein	5.20707516	Y	-	-	-	-	-	
PY03976	putative yir1 protein	8.225840714	Y	-	-	-	-	-	
PY03983	putative yir1 protein	8.443013333	Y	-	-	-	-	-	
PY03985	putative yir1 protein	7.12653963	Y	-	-	-	-	R	
PY04151	putative yir1 protein	4.790117875	Y	-	-	-	-	-	
PY04334	putative yir1 protein	7.860433533	Y	-	-	-	-	-	
PY04396	putative yir1 protein	3.432815779	Y	-	-	-	-	-	
PY04414	putative yir1 protein	7.293935863	Y	-	-	-	-	-	
PY04601	putative yir1 protein	16.92677944	Y	-	-	-	-	-	
PY04735	putative yir1 protein	24.5422012	Y	-	-	-	-	-	
PY04805	putative yir1 protein	17.01764544	Y	-	-	-	-	R	
PY05010	putative yir1 protein	3.24121267	Y	-	-	-	-	-	
PY05243	putative yir1 protein	11.10264417	Y	-	-	-	-	-	
PY05244	putative yir1 protein	13.73994221	Y	-	-	-	-	R	
PY05301	putative yir1 protein	7.978309363	Y	-	-	-	-	-	
PY05307	putative yir1 protein	2.543813577	Y	-	-	-	-	-	
PY05363	putative yir1 protein	7.05759755	Y	-	-	-	-	T	
PY05758	putative yir1 protein	6.451475497	Y	Y	-	-	-	-	
PY06113	putative yir1 protein	4.481702444	Y	-	-	-	-	-	
PY06153	putative yir1 protein	15.21289291	Y	-	-	-	-	-	
PY06155	putative yir1 protein	7.076431827	Y	-	-	-	-	-	
PY06188	putative yir1 protein	25.65282517	Y	-	-	-	-	R	
PY06231	putative yir1 protein	3.538035751	Y	-	-	-	-	-	

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY06508	putative yir1 protein	4.517651299	-	-	-	-	-	R	
PY06547	putative yir1 protein	18.64248735	Y	-	-	-	-	-	
PY06567	putative yir1 protein	9.323827457	Y	-	-	-	-	-	
PY06585	putative yir1 protein	7.26830196	Y	-	-	-	-	-	
PY06678	putative yir1 protein	3.563454414	Y	-	-	-	-	R	
PY06754	putative yir1 protein	2.442071197	Y	-	-	-	-	R	
PY06803	putative yir1 protein	11.52655533	Y	-	-	-	-	-	
PY06805	putative yir1 protein	7.172584861	Y	-	-	-	-	-	
PY06818	putative yir1 protein	4.184316322	Y	-	-	-	-	-	
PY06870	putative yir1 protein	5.203660168	Y	-	-	-	-	-	
PY06891	putative yir1 protein	5.711679002	Y	-	-	-	-	R	
PY06897	putative yir1 protein	2.53511747	Y	-	-	-	-	-	
PY06995	putative yir1 protein	17.55303711	Y	-	-	-	-	-	
PY07125	putative yir1 protein	14.72275458	Y	-	-	-	-	R	
PY07240	putative yir1 protein	2.163661283	Y	-	-	-	-	-	
PY07269	putative yir1 protein	3.709538718	Y	-	-	-	-	-	
PY07354	putative yir1 protein	8.822187728	Y	-	-	-	-	R	
PY07386	putative yir1 protein	6.981310542	Y	-	-	-	-	-	
PY07524	putative yir1 protein	24.56945075	Y	-	-	-	-	R	
PY07628	putative yir1 protein	7.571303124	Y	-	-	-	-	R	
PY00236	putative yir2 protein	3.949992001	Y	-	-	-	-	-	
PY00350	putative yir2 protein	10.60845096	Y	-	-	-	-	R	
PY00578	putative yir2 protein	3.863439043	Y	-	-	-	-	-	
PY00612	putative yir2 protein	3.770116384	Y	-	-	-	-	R	
PY00642	putative yir2 protein	4.122383734	Y	-	-	-	-	-	
PY01161	putative yir2 protein	3.454976784	Y	-	-	-	-	-	
PY02107	putative yir2 protein	4.964894131	Y	-	-	-	-	R	
PY02924	putative yir2 protein	7.839205668	Y	-	-	-	-	-	
PY03041	putative yir2 protein	4.77378729	Y	-	-	-	-	R	
PY03542	putative yir2 protein	3.55703544	Y	Y	-	-	-	-	
PY03569	putative yir2 protein	7.193910195	Y	-	-	-	-	-	
PY03580	putative yir2 protein	42.57598657	Y	-	-	-	-	R	
PY03679	putative yir2 protein	2.687792469	Y	-	-	-	-	R	
PY03758	putative yir2 protein	3.437959778	Y	-	-	-	-	R	
PY03939	putative yir2 protein	2.234734513	Y	-	-	-	-	R	
PY03972	putative yir2 protein	8.488541452	Y	-	-	-	-	-	
PY04052	putative yir2 protein	7.584659998	Y	-	-	-	-	R	
PY04461	putative yir2 protein	7.484274659	Y	-	-	-	-	R	
PY04667	putative yir2 protein	13.97070921	Y	-	-	-	-	T	
PY05021	putative yir2 protein	3.973653083	Y	-	-	-	-	R	
PY05410	putative yir2 protein	36.27270964	Y	-	-	-	-	R	
PY05442	putative yir2 protein	27.64851878	Y	-	-	-	-	R	
PY05455	putative yir2 protein	14.27112846	Y	-	-	-	-	R	
PY05656	putative yir2 protein	10.32786748	Y	-	-	-	-	-	
PY05792	putative yir2 protein	6.61525673	Y	-	-	-	-	-	
PY05942	putative yir2 protein	9.715144037	Y	-	-	-	-	R	
PY06279	putative yir2 protein	6.285580932	Y	-	-	-	-	R	
PY06340	putative yir2 protein	10.40993297	Y	-	-	-	-	R	
PY06621	putative yir2 protein	2.29944236	Y	-	-	-	-	-	
PY06821	putative yir2 protein	5.832241128	Y	-	-	-	-	-	
PY06880	putative yir2 protein	2.933444819	Y	-	-	-	-	-	
PY06934	putative yir2 protein	4.260007558	Y	-	-	-	-	-	
PY07036	putative yir2 protein	17.52099153	Y	-	-	-	-	R	
PY07052	putative yir2 protein	2.013566511	Y	-	-	-	-	R	
PY07072	putative yir2 protein	4.596268594	-	-	-	-	-	-	
PY07127	putative yir2 protein	10.117710509	Y	-	-	-	-	R	
PY07466	putative yir2 protein	14.07318839	Y	Y	-	-	-	T	
PY07499	putative yir2 protein	8.961788913	-	-	-	-	-	-	
PY00001	putative yir3 protein	2.789133741	Y	-	-	-	-	-	
PY00089	putative yir3 protein	12.82762855	Y	-	-	-	-	R	
PY00171	putative yir3 protein	11.33325838	Y	Y	-	-	-	-	
PY00213	putative yir3 protein	6.108729502	Y	-	-	-	-	-	
PY00305	putative yir3 protein	2.926770559	Y	-	-	-	-	-	
PY00392	putative yir3 protein	13.92184338	Y	-	-	-	-	-	
PY00428	putative yir3 protein	3.167658308	Y	Y	-	-	-	-	

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY00485	putative yir3 protein	5.473270788	Y	-	-	-	-	-	
PY00486	putative yir3 protein	18.50803161	Y	Y	-	-	-	-	
PY00579	putative yir3 protein	6.930598041	-	-	-	-	-	-	
PY00580	putative yir3 protein	20.37930097	Y	-	-	-	-	R	
PY01079	putative yir3 protein	14.4132165	Y	Y	-	-	-	-	
PY01278	putative yir3 protein	2.510786916	Y	Y	-	-	-	-	
PY01318	putative yir3 protein	11.94558766	Y	-	-	-	-	R	
PY01331	putative yir3 protein	13.17921893	Y	-	-	-	-	-	
PY01404	putative yir3 protein	2.757413741	Y	-	-	-	-	R	
PY01720	putative yir3 protein	8.251308048	Y	-	-	-	-	-	
PY01810	putative yir3 protein	6.143591014	Y	-	-	-	-	R	
PY01842	putative yir3 protein	6.908464201	Y	-	-	-	-	R	
PY01872	putative yir3 protein	23.67843318	Y	-	-	-	-	R	
PY02011	putative yir3 protein	15.38472125	Y	-	-	-	-	R	
PY02364	putative yir3 protein	3.676213751	Y	-	-	-	-	-	
PY02831	putative yir3 protein	4.685287049	Y	-	-	-	-	-	
PY02972	putative yir3 protein	4.733140822	Y	-	-	-	-	-	
PY02997	putative yir3 protein	2.669455171	Y	Y	-	-	-	R	
PY03042	putative yir3 protein	23.39147235	Y	-	-	-	-	R	
PY03632	putative yir3 protein	3.783532587	Y	Y	-	-	-	-	
PY03636	putative yir3 protein	6.275798633	Y	-	-	-	-	R	
PY03760	putative yir3 protein	12.71207169	Y	-	-	-	-	-	
PY03835	putative yir3 protein	2.683138949	Y	-	-	-	-	-	
PY03849	putative yir3 protein	3.03332196	Y	-	-	-	-	-	
PY03973	putative yir3 protein	2.226700333	Y	-	-	-	-	-	
PY03986	putative yir3 protein	13.03726032	Y	-	-	-	-	-	
PY04006	putative yir3 protein	6.330147681	Y	Y	-	-	-	-	
PY04084	putative yir3 protein	8.424864413	Y	-	-	-	-	-	
PY04144	putative yir3 protein	8.104300065	Y	-	-	-	-	R	
PY04233	putative yir3 protein	3.769389526	Y	-	-	-	-	-	
PY04276	putative yir3 protein	10.59027125	Y	-	-	-	-	-	
PY04430	putative yir3 protein	6.005976262	Y	-	-	-	-	-	
PY04552	putative yir3 protein	2.287778458	Y	Y	-	-	-	-	
PY04696	putative yir3 protein	4.501805725	Y	-	-	-	-	-	
PY04795	putative yir3 protein	4.928278621	Y	-	-	-	-	-	
PY04845	putative yir3 protein	8.079567715	Y	-	-	-	-	-	
PY04909	putative yir3 protein	2.869218477	Y	-	-	-	-	-	
PY04939	putative yir3 protein	3.855622558	Y	-	-	-	-	T	
PY05121	putative yir3 protein	2.986795322	Y	-	-	-	-	R	
PY05146	putative yir3 protein	3.506248091	Y	-	-	-	-	R	
PY05150	putative yir3 protein	7.577010529	Y	-	-	-	-	-	
PY05343	putative yir3 protein	4.828390563	Y	-	-	-	-	-	
PY05349	putative yir3 protein	17.22585051	Y	-	-	-	-	-	
PY05350	putative yir3 protein	3.33293153	Y	-	-	-	-	-	
PY05403	putative yir3 protein	7.608511785	-	-	-	-	-	-	
PY05428	putative yir3 protein	3.541788592	Y	Y	-	-	-	R	
PY05437	putative yir3 protein	4.209006118	Y	-	-	-	-	R	
PY05636	putative yir3 protein	9.701103278	Y	-	-	-	-	-	
PY05637	putative yir3 protein	11.29756279	Y	-	-	-	-	-	
PY05638	putative yir3 protein	5.896302131	Y	Y	-	-	-	-	
PY05673	putative yir3 protein	10.0745849	Y	-	-	-	-	S	
PY05708	putative yir3 protein	3.366129687	Y	-	-	-	-	R	
PY05709	putative yir3 protein	7.311585259	Y	-	-	-	-	-	
PY05794	putative yir3 protein	8.054504412	Y	-	-	-	-	R	
PY05828	putative yir3 protein	4.248164047	Y	-	-	-	-	-	
PY05830	putative yir3 protein	2.177398626	Y	-	-	-	-	-	
PY05943	putative yir3 protein	6.29421276	Y	Y	-	-	-	-	
PY06126	putative yir3 protein	5.177085496	Y	-	-	-	-	-	
PY06131	putative yir3 protein	4.478329068	Y	Y	-	-	-	R	
PY06151	putative yir3 protein	5.301400554	Y	-	-	-	-	-	
PY06171	putative yir3 protein	2.315306077	Y	Y	-	-	-	-	
PY06172	putative yir3 protein	5.055747002	Y	-	-	-	-	-	
PY06185	putative yir3 protein	5.894483059	Y	-	-	-	-	-	
PY06244	putative yir3 protein	3.259533034	Y	-	-	-	-	-	
PY06383	putative yir3 protein	6.24848421	Y	-	-	-	-	R	

Appendix 1: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY06428	putative yir3 protein	9.759341994	Y	Y	-	-	-	-	
PY06569	putative yir3 protein	6.157720018	Y	-	-	-	-	-	
PY06619	putative yir3 protein	3.289876922	Y	-	-	-	-	-	
PY06666	putative yir3 protein	7.270952713	Y	-	-	-	-	R	
PY06667	putative yir3 protein	4.623167399	Y	-	-	-	-	-	
PY06674	putative yir3 protein	5.509997933	Y	-	-	-	-	-	
PY06679	putative yir3 protein	4.871700137	Y	-	-	-	-	-	
PY06710	putative yir3 protein	2.198591793	Y	-	-	-	-	-	
PY06743	putative yir3 protein	11.25921413	Y	-	-	-	-	R	
PY06794	putative yir3 protein	13.55195479	Y	-	-	-	-	T	
PY06795	putative yir3 protein	2.102179819	Y	-	-	-	-	R	
PY06861	putative yir3 protein	7.162005385	Y	Y	-	-	-	R	
PY06982	putative yir3 protein	5.184590498	Y	-	-	-	-	-	
PY06997	putative yir3 protein	4.105542663	Y	Y	-	-	-	-	
PY07028	putative yir3 protein	52.38870693	Y	-	-	-	-	-	
PY07042	putative yir3 protein	7.500334545	Y	-	-	-	-	-	
PY07044	putative yir3 protein	8.680315324	Y	-	-	-	-	-	
PY07099	putative yir3 protein	2.816327268	Y	-	-	-	-	-	
PY07163	putative yir3 protein	11.23742075	Y	-	-	-	-	-	
PY07171	putative yir3 protein	7.831960876	Y	-	-	-	-	-	
PY07198	putative yir3 protein	4.978900889	Y	-	-	-	-	-	
PY07233	putative yir3 protein	11.04294168	Y	-	-	-	-	-	
PY07261	putative yir3 protein	12.16817598	Y	Y	-	-	-	-	
PY07284	putative yir3 protein	2.530314025	Y	Y	-	-	-	-	
PY07315	putative yir3 protein	5.834397415	Y	Y	-	-	-	-	
PY07413	putative yir3 protein	3.160635843	Y	Y	-	-	-	-	
PY07501	putative yir3 protein	7.291683557	Y	-	-	-	-	R	
PY07510	putative yir3 protein	6.935371689	Y	-	-	-	-	T	
PY07518	putative yir3 protein	10.85791017	Y	-	-	-	-	R	
PY07533	putative yir3 protein	16.72775522	Y	-	-	-	-	-	
PY07546	putative yir3 protein	4.094477002	Y	-	-	-	-	-	
PY07635	putative yir3 protein	17.37157608	Y	-	-	-	-	-	
PY07666	putative yir3 protein	4.75491124	Y	-	-	-	-	-	
PY07843	putative yir3 protein	10.15970916	Y	-	-	-	-	-	
PY07849	putative yir3 protein	5.565905315	Y	-	-	-	-	-	
PY00303	putative yir4 protein	8.593128578	Y	-	-	-	-	-	
PY00494	putative yir4 protein	7.11468088	Y	-	-	-	-	R	
PY00574	putative yir4 protein	10.8252291	Y	-	-	-	-	-	
PY00577	putative yir4 protein	3.534027648	Y	-	-	-	-	-	
PY00610	putative yir4 protein	17.23266057	Y	-	-	-	-	R	
PY00661	putative yir4 protein	7.591147418	Y	Y	-	-	-	S	
PY00843	putative yir4 protein	4.004187334	Y	-	-	-	-	R	
PY01441	putative yir4 protein	13.53847575	Y	-	-	-	-	R	
PY01774	putative yir4 protein	3.958160097	Y	-	-	-	-	R	
PY01871	putative yir4 protein	5.647993643	Y	-	-	-	-	-	
PY01873	putative yir4 protein	5.023186792	Y	-	-	-	-	-	
PY02183	putative yir4 protein	3.781977608	Y	-	-	-	-	R	
PY02572	putative yir4 protein	6.879301539	Y	-	-	-	-	-	
PY02687	putative yir4 protein	11.56549285	Y	-	-	-	-	-	
PY02688	putative yir4 protein	4.692763571	Y	-	-	-	-	-	
PY02702	putative yir4 protein	8.658441173	Y	-	-	-	-	-	
PY02970	putative yir4 protein	8.117849177	Y	-	-	-	-	R	
PY02971	putative yir4 protein	14.86500926	Y	-	-	-	-	R	
PY02973	putative yir4 protein	14.92112141	Y	-	-	-	-	R	
PY02996	putative yir4 protein	6.323406776	Y	-	-	-	-	S	
PY03010	putative yir4 protein	12.57108308	Y	-	-	-	-	T	
PY03360	putative yir4 protein	3.004510458	Y	Y	-	-	-	-	
PY03402	putative yir4 protein	4.905851912	Y	-	-	-	-	R	
PY03403	putative yir4 protein	8.011570033	Y	-	-	-	-	-	
PY03540	putative yir4 protein	6.27188887	Y	-	-	-	-	-	
PY03583	putative yir4 protein	4.921173862	Y	-	-	-	-	-	
PY03747	putative yir4 protein	3.314894923	Y	-	-	-	-	R	
PY03837	putative yir4 protein	4.231345465	Y	-	-	-	-	R	
PY04083	putative yir4 protein	4.819355344	Y	-	-	-	-	-	
PY04267	putative yir4 protein	14.83557627	Y	-	-	-	-	-	

Appendix I: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY04319	putative yir4 protein	10.69582494	Y	-	-	-	-	R	
PY04412	putative yir4 protein	3.208154883	Y	-	-	-	-	R	
PY04423	putative yir4 protein	10.24848591	Y	-	-	-	-	R	
PY04466	putative yir4 protein	3.389266854	Y	-	-	-	-	R	
PY04537	putative yir4 protein	5.27065964	Y	-	-	-	-	-	
PY04541	putative yir4 protein	3.276457733	Y	-	-	-	-	-	
PY04554	putative yir4 protein	4.17565763	Y	-	-	-	-	-	
PY04806	putative yir4 protein	5.574320488	Y	-	-	-	-	R	
PY04855	putative yir4 protein	4.008653377	Y	-	-	-	-	R	
PY04931	putative yir4 protein	15.95364503	Y	-	-	-	-	-	
PY04979	putative yir4 protein	10.77816031	Y	-	-	-	-	-	
PY04988	putative yir4 protein	10.05534654	Y	-	-	-	-	-	
PY04998	putative yir4 protein	4.808837772	Y	-	-	-	-	-	
PY05044	putative yir4 protein	7.771527772	Y	-	-	-	-	-	
PY05276	putative yir4 protein	2.06834114	Y	-	-	-	-	-	
PY05289	putative yir4 protein	4.188743295	Y	-	-	-	-	-	
PY05624	putative yir4 protein	8.819404683	Y	-	-	-	-	R	
PY05675	putative yir4 protein	7.682711329	Y	-	-	-	-	-	
PY05796	putative yir4 protein	2.930626705	Y	Y	-	-	-	-	
PY05822	putative yir4 protein	10.29111486	Y	-	-	-	-	R	
PY05873	putative yir4 protein	15.00277554	Y	-	-	-	-	R	
PY05895	putative yir4 protein	3.099994666	Y	Y	-	-	-	-	
PY05988	putative yir4 protein	2.054901587	Y	-	-	-	-	R	
PY06499	putative yir4 protein	4.2601342	Y	Y	-	-	-	R	
PY06755	putative yir4 protein	3.215662227	Y	-	-	-	-	-	
PY06763	putative yir4 protein	7.392703537	Y	-	-	-	-	R	
PY06796	putative yir4 protein	29.55485212	Y	-	-	-	-	-	
PY06829	putative yir4 protein	2.115207205	Y	-	-	-	-	R	
PY06990	putative yir4 protein	3.753555254	Y	-	-	-	-	-	
PY07027	putative yir4 protein	11.14732809	Y	-	-	-	-	R	
PY07035	putative yir4 protein	4.862413333	Y	-	-	-	-	-	
PY07053	putative yir4 protein	21.66264155	Y	-	-	-	-	R	
PY07149	putative yir4 protein	10.27222898	Y	-	-	-	-	R	
PY07190	putative yir4 protein	27.88010067	Y	-	-	-	-	-	
PY07305	putative yir4 protein	2.102916513	Y	-	-	-	-	R	
PY07325	putative yir4 protein	14.50923297	Y	-	-	-	-	R	
PY07342	putative yir4 protein	3.794140702	Y	-	-	-	-	R	
PY07406	putative yir4 protein	6.66415346	Y	-	-	-	-	-	
PY07450	putative yir4 protein	8.527587426	Y	-	-	-	-	-	
PY07522	putative yir4 protein	10.70380049	Y	-	-	-	-	-	
PY07528	putative yir4 protein	4.986337934	Y	-	-	-	-	-	
PY07656	putative yir4 protein	2.034340112	Y	-	-	metalloendopeptidase activity, metal ion binding	-	R	
PY04496	replication factor C, 140 kDa subunit	3.09388425	-	-	-	-	PFB0895c	T	
PY06945	rhoptry protein	2.955228887	-	-	-	-	PFE1045c	R	
PY01016	RING-finger protein	3.04582708	-	Y	-	-	-	S	
PY02527	ring-infested erythrocyte surface antigen, putative	2.172014559	-	-	-	-	PF07_0108	-	
PY00973	RNA recognition motif, putative	3.62090316	-	-	-	-	PFC0865w	S	
PY05280	RNA recognition motif, putative	2.284872917	-	-	-	-	PF14_0056	R	
PY04762	rpl2	2.02919107	-	-	-	-	-	S	
PY06442	Saccharomyces cerevisiae ORF 2310-related	3.465037997	-	-	-	-	-	S	
PY00094	Sec7 domain, putative	2.09159847	Y	-	-	binding, ARF guanyl-nucleotide exchange factor ac	PF14_0407	R	
PY06928	serine/threonine-protein kinase, putative	2.213623243	-	-	-	-	MAL13P1.196	R	
PY00908	similar to S. cerevisiae SIK1	2.091188084	-	-	-	-	PF11_0191	R	
PY03663	small GTP-binding protein domain, putative	2.098753632	-	-	-	-	PFC0565w	R	
PY05606	Smt3 activating enzyme 1	10.30447058	-	-	-	-	PF13_0264	R	
PY04835	Streptococcus pyogenes AMV156, putative	3.017136949	Y	-	-	-	PFA0175w	R	
PY05218	synthetic antigen of P.falciparum, putative	3.42949271	-	-	-	-	MAL13P1.334	-	
PY01151	TBC domain, putative	2.544271158	-	-	-	-	PF10195c	R	
PY03247	TERT gene	2.475735638	Y	-	-	-	PFF0965c	T	
PY00825	tRNA-guanine transglycosylase-related	2.317748263	-	-	-	-	PFL2030w, PF07_0071	R	
PY03314	tRNA-pseudouridine synthase I	2.259276768	Y	-	-	tRNA binding, tRNA-pseudouridine synthase activity	PFE0815w	R	
PY02024	type II proteins geranylgeranyltransferase beta subunit	4.171739711	Y	-	-	catalytic activity	PFL0695c	R	
PY04254	tyrosyl-tRNA synthetase 1-related	3.779354403	-	Y	-	-	PF11_0181	R	
PY00468	Ubiquitin-conjugating enzyme, putative	2.68048357	-	-	-	-	PF10_0330	R	

Appendix 1: Polymorphic genes in YA with respect to YM

Appendices

Py Acc. No.	Py Name	Mean YM:YA	transmembrane domain	Signal Peptide	PEXEL	Predicted GO function	Pf ortholog	Peak Py transcription	Comment
PY02784	uncharacterized domain 2, putative	2.168756315	-	-	-	-	PFE1470w	R	
PY02799	unnamed protein product	2.545199234	-	-	-	-	MAL13P1.52	R	
PY06180	unnamed protein product	2.001377706	Y	-	-	GTP binding	PF14_0292	R	
PY04969	unnamed protein product-related	2.654492577	-	-	-	-	PF13_0208	S	
PY02528	uroporphyrinogen decarboxylase, putative	3.925004788	-	-	-	-	PFF0360w	T	
PY05083	Zinc finger, C2H2 type, putative	2.860668088	-	-	-	-	PF14_0707	S	

Stage-specific expression denoted by R: rings; T: trophozoites; S: schizonts.

Py Gene	Name	Rings	Trophozoites	Schizonts
PY02002	1 beta dynein heavy chain		-2.218	-2.389
PY04739	1 beta dynein heavy chain		-1.927	-2.023
PY04285	1-cys peroxidoxin	-2.087	-3.215	
PY02086	15 kDa protein-related	-5.250	-2.182	-1.841
PY04647	235 kDa rhoptry protein	1.562	3.113	3.981
PY04930	235 kDa rhoptry protein	8.225	25.454	18.989
PY00522	28 kDa ookinete surface protein	-4.003	-1.507	-2.470
PY05918	63231-5922		-2.913	-1.297
PY04829	69948-6867, putative	1.864	2.512	3.654
PY02722	6S ribosomal protein L6, putative	3.357	2.799	2.788
PY04072	86 domains.-related		-1.592	-2.461
PY01007	93 kDa protein	-6.810	-6.770	-9.770
PY05666	A/G-specific adenine glycosylase, putative	1.333	2.207	2.682
PY00207	ABC transporter-like protein	1.772	2.096	2.722
PY00319	aconitate hydratase 1	-2.217	-1.780	-1.844
PY00021	Actin	2.037	2.843	2.848
PY01545	actin ii			-6.847
PY02606	amine oxidase, flavin-containing, putative	-1.052	-1.258	-2.068
PY02060	Arabidopsis thaliana At1g233/T6A9_12-related	1.638	2.619	2.264
PY06640	Arabidopsis thaliana At1g7127/F3I17_8	3.241	5.775	5.315
PY03143	Arabidopsis thaliana At3g451/T14D3_4-related	-2.122	-1.589	-1.392
PY04349	Arabidopsis thaliana At5g2481/F6A4_2	-2.036	-1.249	-2.131
PY04243	Arabidopsis thaliana At5g355/C32EPL23M	-3.252	-2.104	
PY00676	Arabidopsis thaliana BRAHMA ortholog	-2.144	-1.346	
PY01120	Arabidopsis thaliana BRAHMA ortholog-related	6.353	11.808	200.444
PY06992	Arabidopsis thaliana MKA23.5	-1.644	-2.390	-2.781
PY06105	Arabidopsis thaliana T27G7.17, putative	1.335	1.779	2.276
PY01549	arabinogalactan protein	-1.760	-1.682	-2.006
PY03744	arabinogalactan protein	-1.075	-1.061	-2.301
PY04251	arabinogalactan protein		-7.027	-10.314
PY02198	arsenite transport subunit A	3.239	5.683	6.727
PY02350	asparagine-rich protein	1.802	2.439	2.200
PY03561	asparagine-rich protein	1.940	2.594	2.271
PY05558	asparagine-rich protein	-1.195	-1.115	-2.341
PY02085	aspartyl protease-like	1.404	2.117	1.770
PY01271	ATP-dependent DNA helicase recQ-related	-1.583	-1.322	-2.071
PY05109	ATPase associated with chromosome architecture/replication	2.023	2.501	2.981
PY02067	ATPase, AAA family, putative		-2.218	-1.403
PY04544	axonemal dynein heavy chain 8-related	-1.324	-1.193	-2.502
PY03705	C2 domain, putative	-1.010	-1.122	-2.162
PY06394	calcium-dependent protein kinase-3	-1.598	2.122	1.764
PY07430	calcium/calmodulin-stimulated cyclic nucleotide phosphodiesterase, putative	-5.002	-1.153	-1.675
PY01677	calmodulin-like protein		-2.443	-4.087
PY06257	carbamoyl phosphate synthetase II-related	-5.287		
PY00253	CCAAT-box DNA binding protein subunit B			-4.962
PY01974	CCAAT-box DNA binding protein subunit B		-1.131	-3.514
PY02177	CCAAT-box DNA binding protein subunit B	2.730	1.873	1.526
PY02242	CCAAT-box DNA binding protein subunit B	-43.197		
PY02393	CCAAT-box DNA binding protein subunit B	2.300	4.033	3.593
PY02413	CCAAT-box DNA binding protein subunit B	-1.196	-0.009	-2.275
PY05072	CCAAT-box DNA binding protein subunit B	3.279	5.867	5.953
PY05674	CCAAT-box DNA binding protein subunit B	1.627	2.081	1.207
PY06762	CCAAT-box DNA binding protein subunit B	1.871	3.250	1.975
PY00834	CCAAT-box DNA binding protein subunit B, putative	-3.774		
PY00879	CCAAT-box DNA binding protein subunit B, putative	-1.490	-1.124	-2.077
PY02289	CCAAT-box DNA binding protein subunit B, putative	60.309	76.884	62.885
PY03317	cdc-2 related kinase 1	2.076	1.953	-1.239
PY03919	cell division cycle protein 27 homolog	-1.273	-1.035	-2.146
PY00809	cell division protein	1.878	3.065	3.025
PY02937	chaperonin containing TCP-1 delta subunit	2.793	4.236	5.200
PY00008	chitinase	-1.275	-1.431	-2.551
PY07200	chitinase-related	1.068	-1.060	-2.130
PY00653	chromosome segregation protein smc1	37.671	132.017	226.132
PY06797	cir1 protein-related	7.964		
PY02415	conjugation stage-specific protein, putative	1.978	2.243	2.372
PY03618	coronin binding protein	10.392	10.321	5.761
PY02686	Cyclin, putative	-1.209	-1.340	-2.362
PY05616	Cyclin, putative	-2.287	-1.382	-1.036
PY01468	cytoplasmic dynein 1-related		-2.475	-1.702
PY05459	delta-aminolevulinic acid synthetase	-2.865	1.032	2.327
PY02252	deoxyribose-phosphate aldolase	2.793	3.443	5.357

PY07180	Dictyostelium discoideum SgcA	5.842	6.739	7.022
PY06281	dimethyladenosine transferase, putative	1.597	1.609	2.716
PY06905	DNA repair protein rad1, putative	-1.100	1.513	2.089
PY05156	DNA-directed RNA polymerase beta chain	1.628	3.231	3.554
PY01286	DnaJ domain, putative	-1.431	-1.732	-2.053
PY04382	DnaJ domain, putative	-2.257	-1.614	-1.666
PY03839	Domain of unknown function, putative	14.619	16.651	15.709
PY01447	Drosophila melanogaster BcDNA.GH632		7.974	6.120
PY06558	Drosophila melanogaster CG11779 gene product, putative	2.088	1.776	1.443
PY00497	Drosophila melanogaster CG1493 gene product	2.196	2.075	2.559
PY00878	Drosophila melanogaster CG154 gene product	-1.189	-2.429	-3.415
PY02028	Drosophila melanogaster CG154 gene product		-1.468	-2.131
PY03071	Drosophila melanogaster CG154 gene product	1.770	2.642	1.383
PY03689	Drosophila melanogaster CG154 gene product	-1.346	-1.352	-3.354
PY04549	Drosophila melanogaster CG154 gene product	3.859	9.398	7.945
PY06441	Drosophila melanogaster CG154 gene product	-2.066	-1.516	-2.402
PY07318	Drosophila melanogaster CG154 gene product	-2.210	-1.402	-2.056
PY04714	Drosophila melanogaster CG4392 gene product	2.837	5.665	6.936
PY02386	Drosophila melanogaster GH13165p		-2.988	-2.280
PY03619	Drosophila melanogaster GH21295p	-3.047	-1.311	-2.233
PY02486	Drosophila melanogaster GM165p	2.477	2.878	2.582
PY01301	Drosophila melanogaster LD1529p	5.707	7.955	7.291
PY01757	Drosophila melanogaster LP1394p			-3.181
PY00209	dynein axonemal intermediate chain		-2.322	-2.090
PY00078	dynein beta chain, ciliary			-2.257
PY01469	dynein light chain-related	-1.469	-1.730	-2.027
PY02433	EF hand, putative		-2.269	-3.398
PY07687	elongation factor-1 alpha		-3.273	
PY06583	Endonuclease/Exonuclease/phosphatase family	2.202	1.913	2.215
PY07668	Endonuclease/Exonuclease/phosphatase family		-3.468	
PY06826	Enoyl-CoA hydratase, putative	-2.648	-1.487	-1.102
PY00837	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	2.195	2.434	2.696
PY01183	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	-2.103	-1.313	-2.980
PY01667	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	-2.063	-2.511	-2.215
PY03698	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	-3.205	-1.288	-2.058
PY04922	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	-9.899	-5.340	-11.632
PY06052	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	-1.828	1.038	-2.313
PY00469	Eukaryotic aspartyl protease	-3.056		-2.317
PY04269	eukaryotic translation initiation factor 3 subunit 7	-6.509	-9.675	-7.409
PY07488	expressed protein, putative	1.110	-1.185	-2.397
PY01855	Formin Homology 2 Domain, putative	-1.125	-1.097	-2.059
PY06292	fulmal1	1.562	2.672	2.281
PY06284	gamma-gliadin	1.205	3.031	2.498
PY04261	glutamate dehydrogenase	-2.746	-3.433	-5.329
PY06496	glutamine-asparagine rich protein	1.778	2.229	
PY04253	GTP-binding nuclear protein ran/tc4		-1.488	-4.381
PY00951	guanylyl cyclase enzyme-related			-2.212
PY02999	guanylyl cyclase enzyme-related			-2.130
PY05489	histidine kinase DhkG	9.414	28.580	19.475
PY02915	homeobox-containing protein	-1.077	-1.028	-2.260
PY03288	homeobox-containing protein	2.192	1.553	1.408
PY04274	homeobox-containing protein			-3.056
PY05178	homeobox-containing protein	3.648	4.115	4.031
PY05189	homeobox-containing protein	1.442	2.243	2.188
PY05961	homeobox-containing protein	2.000	3.324	4.051
PY07066	homeobox-containing protein	-2.339	1.471	1.985
PY07227	homeobox-containing protein	-6.166	-1.321	-2.803
PY07542	Homo sapiens A_IG2N1.1 gene product, putative	2.505	3.263	3.596
PY05294	Homo sapiens dJ19E24.4-related		-2.020	-1.366
PY02273	Homo sapiens dJ261G23.2.2	-6.216	-4.017	-2.632
PY04140	Homo sapiens GAJ, putative	-33.941	-6.832	-6.217
PY03811	Homo sapiens KIAA135 protein-related			-2.925
PY00048	hypothetical protein	1.461	-1.025	2.341
PY00075	hypothetical protein	-2.958		-4.787
PY00126	hypothetical protein			-4.152
PY00127	hypothetical protein			-2.959
PY00146	hypothetical protein	2.969		
PY00151	hypothetical protein	2.257	-1.564	-1.976
PY00174	hypothetical protein	3.715	2.699	3.355
PY00178	hypothetical protein	2.554	6.759	7.312
PY00179	hypothetical protein		-2.796	-1.421
PY00181	hypothetical protein		-4.477	-1.697

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY00193	hypothetical protein	5.144	6.671	6.176
PY00198	hypothetical protein	4.351	4.359	3.483
PY00218	hypothetical protein		-8.023	
PY00219	hypothetical protein	-1.209		-2.070
PY00228	hypothetical protein	-5.082	-1.688	-1.592
PY00233	hypothetical protein	-1.670	1.029	-2.224
PY00237	hypothetical protein	5.490		
PY00251	hypothetical protein	-2.636	-1.191	-2.173
PY00258	hypothetical protein	2.383	3.298	4.561
PY00259	hypothetical protein	2.175	2.777	2.048
PY00281	hypothetical protein	-2.644	-3.067	-3.602
PY00285	hypothetical protein	-10.489	-1.025	-1.283
PY00307	hypothetical protein	2.106	2.046	2.866
PY00324	hypothetical protein	-1.185	-1.145	-20.727
PY00329	hypothetical protein	-1.366	-1.633	-2.181
PY00347	hypothetical protein	2.739	5.571	6.332
PY00356	hypothetical protein	-1.630	-2.407	-1.380
PY00358	hypothetical protein	1.291	1.948	2.372
PY00380	hypothetical protein	-2.100	-1.383	-2.340
PY00383	hypothetical protein	-1.409	-1.250	-2.082
PY00393	hypothetical protein	-3.087	-3.487	-3.362
PY00394	hypothetical protein	-2.769	-3.103	
PY00399	hypothetical protein	2.858	5.738	4.259
PY00401	hypothetical protein	3.036		
PY00402	hypothetical protein	1.024	1.016	2.248
PY00408	hypothetical protein	1.143	1.203	2.227
PY00426	hypothetical protein	1.568	2.173	2.663
PY00439	hypothetical protein	-208.133	-191.203	-233.242
PY00453	hypothetical protein			2.624
PY00455	hypothetical protein		1.682	2.301
PY00462	hypothetical protein	1.675	1.590	2.395
PY00464	hypothetical protein		-1.847	-2.188
PY00466	hypothetical protein	-30.769		
PY00470	hypothetical protein	-1.460	-1.195	-2.142
PY00472	hypothetical protein			-4.007
PY00484	hypothetical protein	-2.666		
PY00492	hypothetical protein	-3.252	-5.407	-3.929
PY00521	hypothetical protein		2.922	1.220
PY00528	hypothetical protein	3.556	3.269	3.482
PY00532	hypothetical protein	2.952	5.017	3.101
PY00551	hypothetical protein	7.269	4.339	7.344
PY00555	hypothetical protein	3.231	2.206	
PY00568	hypothetical protein	1.091	2.262	1.709
PY00576	hypothetical protein	2.136	1.889	2.265
PY00581	hypothetical protein	1.308	2.323	2.381
PY00585	hypothetical protein		-2.201	-1.241
PY00591	hypothetical protein	-5.444		1.170
PY00600	hypothetical protein	7.719	15.472	9.527
PY00608	hypothetical protein	-3.129		-2.155
PY00640	hypothetical protein			-2.005
PY00641	hypothetical protein	1.059	-1.254	-2.068
PY00647	hypothetical protein	3.448	5.792	6.626
PY00677	hypothetical protein	1.868	3.611	2.798
PY00680	hypothetical protein			-4.577
PY00694	hypothetical protein	48.715	4.738	6.127
PY00697	hypothetical protein	2.174	3.065	3.156
PY00698	hypothetical protein	-5.165	-2.563	-7.460
PY00715	hypothetical protein	1.855	2.920	2.843
PY00719	hypothetical protein	1.719	2.507	4.000
PY00730	hypothetical protein		-4.569	-1.923
PY00734	hypothetical protein	-1.352	-2.594	-2.357
PY00740	hypothetical protein			-2.094
PY00752	hypothetical protein		2.961	1.607
PY00760	hypothetical protein			-6.487
PY00811	hypothetical protein	-1.269	-1.308	-2.283
PY00813	hypothetical protein	2.437	4.346	4.535
PY00816	hypothetical protein	2.062	2.475	2.204
PY00820	hypothetical protein			-2.126
PY00822	hypothetical protein	2.104		
PY00831	hypothetical protein	3.991	1.859	1.406
PY00833	hypothetical protein	3.150	5.636	4.157
PY00845	hypothetical protein	-17.640	-28.287	-43.271

PY00859	hypothetical protein	-1.328	-1.321	-2.849
PY00860	hypothetical protein	-4.366	-1.403	-2.254
PY00867	hypothetical protein	2.370	2.262	2.146
PY00888	hypothetical protein	1.308	1.902	2.723
PY00889	hypothetical protein	1.653	2.910	3.344
PY00892	hypothetical protein	1.713	-1.079	-3.161
PY00905	hypothetical protein	-1.741	-1.271	-2.200
PY00906	hypothetical protein	-1.880	-1.250	-2.044
PY00910	hypothetical protein	1.022	-1.020	-2.145
PY00921	hypothetical protein	5.649	6.997	6.574
PY00930	hypothetical protein	-1.499	1.004	-2.062
PY00931	hypothetical protein	-2.431	-1.425	-2.654
PY00932	hypothetical protein	2.616	4.534	3.614
PY00936	hypothetical protein	1.389	2.050	1.540
PY00942	hypothetical protein	-2.751	-1.118	-1.575
PY00955	hypothetical protein	3.042	2.487	2.759
PY00958	hypothetical protein	2.853	1.463	
PY00964	hypothetical protein	-2.105	-3.003	-2.149
PY00965	hypothetical protein	-1.177	-1.916	-2.491
PY00978	hypothetical protein	-1.960	-1.292	-2.428
PY00991	hypothetical protein		3.470	1.639
PY01003	hypothetical protein		136.744	12.871
PY01006	hypothetical protein	1.794	3.056	3.452
PY01012	hypothetical protein	2.034	2.773	2.634
PY01021	hypothetical protein	4.620	2.087	
PY01033	hypothetical protein	-1.082	-1.061	-2.270
PY01042	hypothetical protein	-1.838	-1.130	-2.457
PY01049	hypothetical protein	-1.689	1.064	-2.603
PY01057	hypothetical protein	2.697	5.787	5.689
PY01082	hypothetical protein	-3.241	-3.056	-2.883
PY01113	hypothetical protein	5.844	6.359	9.367
PY01121	hypothetical protein	-1.486	-1.589	-4.015
PY01137	hypothetical protein	-4.924	-1.131	-1.308
PY01149	hypothetical protein	-2.858	-1.482	-2.102
PY01153	hypothetical protein		2.214	1.259
PY01154	hypothetical protein	-1.065	1.116	-2.399
PY01159	hypothetical protein	2.124	1.701	1.388
PY01170	hypothetical protein	1.973	3.173	2.894
PY01175	hypothetical protein	2.053	1.182	1.526
PY01188	hypothetical protein	-1.517	-2.034	-2.206
PY01197	hypothetical protein		-1.645	-2.237
PY01226	hypothetical protein		3.821	3.549
PY01235	hypothetical protein	-1.048	1.417	2.663
PY01250	hypothetical protein	-2.096	-1.554	-1.414
PY01254	hypothetical protein		-1.446	-2.230
PY01259	hypothetical protein	1.849	2.336	1.252
PY01262	hypothetical protein	1.627	2.451	1.518
PY01263	hypothetical protein	-1.939	1.214	-3.694
PY01265	hypothetical protein	-1.277	-0.066	-3.131
PY01277	hypothetical protein		-1.440	-2.055
PY01296	hypothetical protein	-5.431	-1.467	-2.178
PY01297	hypothetical protein		-1.337	-2.208
PY01298	hypothetical protein	-2.140	-1.617	-2.146
PY01299	hypothetical protein	-3.346	-1.876	-2.586
PY01308	hypothetical protein	3.781	7.251	8.323
PY01310	hypothetical protein	-2.017	-2.040	-1.745
PY01313	hypothetical protein	2.471	1.766	1.669
PY01326	hypothetical protein	1.335	2.396	1.709
PY01360	hypothetical protein	-1.939	-2.485	-2.392
PY01380	hypothetical protein	-2.920	-2.208	-2.986
PY01381	hypothetical protein	2.326	1.561	1.512
PY01395	hypothetical protein	1.793	1.992	2.574
PY01397	hypothetical protein	7.352	5.367	11.078
PY01401	hypothetical protein	2.229	4.626	3.740
PY01414	hypothetical protein	1.874	5.174	6.351
PY01415	hypothetical protein	12.357		4.405
PY01416	hypothetical protein	6.225	5.996	11.212
PY01433	hypothetical protein	-1.780	-2.422	-2.640
PY01437	hypothetical protein	-1.296	1.408	2.404
PY01443	hypothetical protein	1.340	2.099	2.298
PY01444	hypothetical protein	-2.406		
PY01458	hypothetical protein	1.628	1.679	2.118

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY01470	hypothetical protein		-2.269	
PY01476	hypothetical protein		-2.772	-6.439
PY01478	hypothetical protein	1.586	2.077	2.265
PY01501	hypothetical protein	3.611	5.232	4.917
PY01507	hypothetical protein	1.848	2.015	2.417
PY01512	hypothetical protein	1.186	-1.285	-2.370
PY01518	hypothetical protein			-2.879
PY01522	hypothetical protein	3.632	7.891	7.649
PY01535	hypothetical protein			-2.996
PY01538	hypothetical protein	4.422	7.712	6.002
PY01568	hypothetical protein	-3.147	-1.696	-2.202
PY01582	hypothetical protein	-1.421	-1.255	-2.804
PY01585	hypothetical protein	-2.139	-1.274	-2.016
PY01588	hypothetical protein	4.147	4.709	4.052
PY01590	hypothetical protein	5.275	3.726	7.067
PY01602	hypothetical protein	12.817	42.364	52.889
PY01613	hypothetical protein	3.767	1.793	-2.157
PY01614	hypothetical protein	2.669	1.399	-2.128
PY01624	hypothetical protein	-3.565	-1.690	-2.198
PY01625	hypothetical protein	2.619	2.968	3.335
PY01640	hypothetical protein	2.122	3.148	3.013
PY01673	hypothetical protein	1.433	1.945	2.010
PY01675	hypothetical protein			-3.236
PY01676	hypothetical protein			-3.829
PY01684	hypothetical protein	-2.838	-3.010	-1.624
PY01687	hypothetical protein	3.301	4.818	3.900
PY01694	hypothetical protein	-1.614	-2.076	
PY01702	hypothetical protein	2.008	1.894	1.781
PY01729	hypothetical protein			-2.737
PY01749	hypothetical protein	1.851	2.710	3.143
PY01751	hypothetical protein	8.562	8.985	6.240
PY01760	hypothetical protein	1.026	1.154	2.071
PY01766	hypothetical protein	-3.200	-2.025	
PY01767	hypothetical protein	-1.521	-2.200	-2.832
PY01768	hypothetical protein	-1.407	-2.188	-2.883
PY01771	hypothetical protein	6.519	9.389	14.526
PY01772	hypothetical protein	1.202	4.120	6.144
PY01799	hypothetical protein	-1.260	1.025	-2.924
PY01806	hypothetical protein	-1.506	-2.048	-1.183
PY01814	hypothetical protein	17.290	29.221	34.908
PY01825	hypothetical protein		2.855	1.280
PY01840	hypothetical protein	-3.898		
PY01859	hypothetical protein	-1.142	-1.181	-2.104
PY01877	hypothetical protein	1.207	1.192	-2.592
PY01880	hypothetical protein	-1.938	-2.048	-1.686
PY01887	hypothetical protein	-1.967	-1.524	-2.348
PY01888	hypothetical protein	-5.114	-1.584	-1.627
PY01899	hypothetical protein	1.517		-5.979
PY01900	hypothetical protein			-3.571
PY01905	hypothetical protein			-2.833
PY01908	hypothetical protein	1.708	2.591	2.986
PY01915	hypothetical protein	-2.853	-1.440	-2.683
PY01937	hypothetical protein	-1.087	-1.237	-2.167
PY01956	hypothetical protein	-2.078	-1.718	-1.687
PY01968	hypothetical protein	-28.358	-5.565	-8.391
PY01972	hypothetical protein	3.437	3.234	3.321
PY01979	hypothetical protein			-2.141
PY01980	hypothetical protein	-3.284		
PY02001	hypothetical protein			-2.211
PY02007	hypothetical protein	-1.349	-1.167	-2.582
PY02013	hypothetical protein		-2.361	-1.701
PY02014	hypothetical protein	-1.694	-2.011	-1.707
PY02020	hypothetical protein	1.295	2.142	3.718
PY02032	hypothetical protein	-2.303	-1.518	-1.167
PY02037	hypothetical protein	-5.899	-2.994	-3.273
PY02039	hypothetical protein	-2.452	-2.363	-2.750
PY02040	hypothetical protein	-2.053	-2.080	
PY02054	hypothetical protein	-2.575	-1.253	-2.234
PY02055	hypothetical protein	-3.126	1.124	1.024
PY02061	hypothetical protein	-2.028	-1.304	-1.715
PY02075	hypothetical protein	-2.333	-2.041	-1.856
PY02078	hypothetical protein		12.875	7.995

PY02087	hypothetical protein			-2.153
PY02100	hypothetical protein	57.170	27.270	21.283
PY02112	hypothetical protein	-1.049	1.184	2.450
PY02133	hypothetical protein	1.322	2.213	2.539
PY02135	hypothetical protein	-10.454		
PY02138	hypothetical protein	2.038	2.224	3.046
PY02141	hypothetical protein	1.894	2.285	2.443
PY02144	hypothetical protein	-3.782	-3.827	-3.768
PY02145	hypothetical protein	-1.892	-2.468	-1.617
PY02152	hypothetical protein	5.362	6.390	5.240
PY02154	hypothetical protein	1.321	1.985	2.143
PY02166	hypothetical protein	2.008	1.240	1.067
PY02167	hypothetical protein	-3.254	-1.481	-2.095
PY02168	hypothetical protein		-2.861	-2.148
PY02184	hypothetical protein	4.397		
PY02199	hypothetical protein	1.758	2.397	2.672
PY02201	hypothetical protein		2.119	2.560
PY02207	hypothetical protein	-1.055	1.653	2.210
PY02235	hypothetical protein	2.130	1.785	-1.979
PY02249	hypothetical protein	2.822	2.289	1.864
PY02251	hypothetical protein	2.109	1.780	2.411
PY02261	hypothetical protein	5.653	16.788	6.717
PY02271	hypothetical protein	-2.460	-3.445	
PY02274	hypothetical protein	-1.101	-1.111	-2.217
PY02299	hypothetical protein	-3.335	-5.356	-5.717
PY02301	hypothetical protein	16.322	23.974	29.131
PY02308	hypothetical protein	-2.585	-1.334	-2.132
PY02330	hypothetical protein		-1.519	-2.254
PY02346	hypothetical protein	1.082		-2.530
PY02356	hypothetical protein	10.772	20.797	60.608
PY02359	hypothetical protein	-1.495	-1.798	-2.921
PY02371	hypothetical protein	1.051	1.176	-2.154
PY02382	hypothetical protein	1.729	-1.065	2.328
PY02391	hypothetical protein	-1.496	-1.043	-2.005
PY02392	hypothetical protein	10.763	24.188	20.865
PY02394	hypothetical protein		-2.543	
PY02396	hypothetical protein	1.793	1.608	2.590
PY02429	hypothetical protein	-1.192	-2.011	-1.824
PY02434	hypothetical protein	-1.181	-1.257	-2.088
PY02451	hypothetical protein	-2.066	-2.275	-2.075
PY02465	hypothetical protein		2.046	-1.302
PY02470	hypothetical protein	-1.013	2.074	1.594
PY02489	hypothetical protein	4.205	8.535	9.303
PY02491	hypothetical protein	-11.084	-9.675	
PY02515	hypothetical protein	3.265	4.187	2.580
PY02518	hypothetical protein	2.058	1.990	2.484
PY02529	hypothetical protein	-1.110	-1.311	-2.094
PY02546	hypothetical protein	1.708	3.553	2.459
PY02556	hypothetical protein	1.140	-1.291	-2.090
PY02594	hypothetical protein	2.276	3.045	2.843
PY02596	hypothetical protein	-3.261	-2.263	-4.889
PY02597	hypothetical protein		-2.523	-4.518
PY02604	hypothetical protein	3.876	3.915	5.285
PY02611	hypothetical protein			-3.122
PY02617	hypothetical protein	-1.373	-1.342	-2.205
PY02633	hypothetical protein		-1.651	-2.602
PY02634	hypothetical protein		-1.773	-2.626
PY02636	hypothetical protein			-2.038
PY02651	hypothetical protein	-2.157	-1.562	-2.206
PY02653	hypothetical protein	-1.020	-1.337	-2.071
PY02658	hypothetical protein	3.573	10.473	7.878
PY02662	hypothetical protein	-3.274	-2.667	-2.121
PY02678	hypothetical protein	-3.129	-3.441	-1.392
PY02684	hypothetical protein	1.488	2.229	2.141
PY02691	hypothetical protein	1.897	1.890	2.581
PY02692	hypothetical protein	1.452	-5.535	
PY02717	hypothetical protein	3.959	3.268	4.795
PY02743	hypothetical protein	5.202	23.287	6.555
PY02754	hypothetical protein	-2.988	-1.145	-2.318
PY02755	hypothetical protein	-2.103	-1.061	-2.095
PY02758	hypothetical protein	-3.213	-2.958	-3.351
PY02765	hypothetical protein	-2.326	-1.601	-2.047

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY02769	hypothetical protein	-3.972	-1.874	-2.632
PY02770	hypothetical protein	-3.311	-1.368	-2.889
PY02862	hypothetical protein	1.295	2.140	2.711
PY02863	hypothetical protein	2.127	2.163	2.720
PY02865	hypothetical protein	1.738	3.801	4.243
PY02871	hypothetical protein	-5.525	-1.723	-2.483
PY02878	hypothetical protein	-4.795	-1.331	-2.306
PY02879	hypothetical protein	-8.518	-4.569	-5.306
PY02888	hypothetical protein		-11.265	-4.792
PY02895	hypothetical protein	14.502	40.773	33.541
PY02934	hypothetical protein	1.136	2.727	2.961
PY02939	hypothetical protein		3.040	6.657
PY02944	hypothetical protein	-2.006	-2.363	
PY02952	hypothetical protein	1.943	2.829	2.829
PY02968	hypothetical protein	2.760	3.106	3.552
PY02981	hypothetical protein	1.486	2.228	-1.565
PY03004	hypothetical protein	1.108	-1.381	-2.060
PY03005	hypothetical protein	-1.295	-1.601	-2.062
PY03006	hypothetical protein	-2.002	-1.443	-1.858
PY03024	hypothetical protein	-32.794	-52.928	-49.132
PY03047	hypothetical protein			-3.008
PY03054	hypothetical protein			-3.454
PY03056	hypothetical protein	1.656	1.952	2.328
PY03074	hypothetical protein	1.022	-1.082	-2.105
PY03079	hypothetical protein	-4.035	-1.181	-1.736
PY03084	hypothetical protein	2.470	6.698	3.683
PY03085	hypothetical protein	6.233	8.450	5.397
PY03102	hypothetical protein	-1.494	-1.101	-2.388
PY03105	hypothetical protein	-12.796	-3.727	
PY03123	hypothetical protein	1.734	2.040	2.180
PY03129	hypothetical protein	12.041	9.006	4.639
PY03134	hypothetical protein	1.591	1.998	2.859
PY03145	hypothetical protein		-1.275	-2.157
PY03162	hypothetical protein	-1.100	-2.203	-1.390
PY03167	hypothetical protein	-5.937		-9.015
PY03185	hypothetical protein	1.208	-1.014	-3.126
PY03202	hypothetical protein	-3.459	-4.817	-3.863
PY03213	hypothetical protein	-2.488	-2.124	
PY03243	hypothetical protein	-1.099	2.720	3.340
PY03248	hypothetical protein	2.770	2.832	2.202
PY03260	hypothetical protein	6.112	6.364	21.960
PY03273	hypothetical protein	-1.293	-1.062	-2.371
PY03276	hypothetical protein	5.309		21.389
PY03291	hypothetical protein			-3.035
PY03294	hypothetical protein		-1.324	-2.181
PY03297	hypothetical protein	2.879	4.469	3.459
PY03303	hypothetical protein		2.497	2.033
PY03340	hypothetical protein	45.625	50672.286	6.933
PY03346	hypothetical protein	-4.558	-2.580	-1.884
PY03347	hypothetical protein	-6.774		
PY03348	hypothetical protein	-9.464	-3.640	-3.258
PY03356	hypothetical protein	-3.252	-3.266	-2.858
PY03366	hypothetical protein	57.874	12.751	
PY03369	hypothetical protein	1.992	2.069	2.311
PY03380	hypothetical protein	-29.626		
PY03388	hypothetical protein	-4.490	-2.437	-3.238
PY03389	hypothetical protein	-3.895	-2.332	-3.091
PY03392	hypothetical protein			-2.176
PY03405	hypothetical protein	8.833	15.685	28.119
PY03416	hypothetical protein	-3.404	-3.506	
PY03422	hypothetical protein	-2.250	-1.254	-2.166
PY03430	hypothetical protein	4.329	7.500	7.072
PY03431	hypothetical protein	-6.509		
PY03434	hypothetical protein	2.503	3.163	2.816
PY03435	hypothetical protein	4.913		
PY03437	hypothetical protein	6.994		
PY03440	hypothetical protein			4.460
PY03459	hypothetical protein	-1.288	-1.297	-2.718
PY03460	hypothetical protein	2.518	1.594	1.821
PY03467	hypothetical protein	1.705	2.737	2.460
PY03484	hypothetical protein	2.128	3.440	2.609
PY03487	hypothetical protein	5.002	4.982	9.081

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY03490	hypothetical protein	-2.013	-1.209	-1.012
PY03495	hypothetical protein	1.635	2.245	1.886
PY03507	hypothetical protein	-2.344	1.080	
PY03516	hypothetical protein	-2.143	-2.225	-3.383
PY03527	hypothetical protein	-2.970	-1.008	-2.524
PY03530	hypothetical protein	-1.716	1.138	-2.172
PY03551	hypothetical protein	1.013	-1.167	-2.749
PY03568	hypothetical protein	36.829		
PY03624	hypothetical protein	45.086	23.693	40.242
PY03643	hypothetical protein	-11.463		
PY03647	hypothetical protein	1.040	1.291	2.276
PY03648	hypothetical protein	1.075	2.520	4.651
PY03657	hypothetical protein	13.333	27.566	78.449
PY03662	hypothetical protein	-2.165	-2.854	-2.254
PY03670	hypothetical protein	-1.616	-2.317	-2.099
PY03675	hypothetical protein	-2.571	-2.630	-2.980
PY03690	hypothetical protein	-2.182	-1.361	-3.125
PY03697	hypothetical protein	-3.169	-1.564	-2.177
PY03714	hypothetical protein			-2.736
PY03718	hypothetical protein	-3.394	-1.004	-2.200
PY03728	hypothetical protein	-11.219	-7.420	
PY03757	hypothetical protein	4.826	2.816	5.684
PY03759	hypothetical protein	28.524	27.632	39.077
PY03771	hypothetical protein	-2.568	-1.150	-2.080
PY03788	hypothetical protein	1.088	1.139	-2.872
PY03795	hypothetical protein	1.036	-1.168	-2.215
PY03800	hypothetical protein	2.567	5.367	6.055
PY03804	hypothetical protein	-25.000		
PY03805	hypothetical protein	-4.793	-6.153	-8.055
PY03836	hypothetical protein	4.342	4.227	2.822
PY03860	hypothetical protein	-1.328	-2.898	-5.111
PY03866	hypothetical protein	1.980	2.371	2.604
PY03891	hypothetical protein	-7.145		
PY03907	hypothetical protein	-2.707	-2.569	-1.885
PY03908	hypothetical protein	1.997	1.702	2.108
PY03915	hypothetical protein	-2.268	-1.321	1.262
PY03934	hypothetical protein		-2.253	-3.076
PY03936	hypothetical protein	1.696	2.512	-1.223
PY03937	hypothetical protein	1.458	2.101	-1.414
PY03938	hypothetical protein	1.368	2.149	-1.545
PY03943	hypothetical protein	-1.145	-1.193	2.167
PY03950	hypothetical protein	62.698	69.019	51.225
PY03951	hypothetical protein	25.600	144.536	217.071
PY03955	hypothetical protein	1.767	3.776	8.546
PY03968	hypothetical protein	1.619	1.522	3.785
PY03997	hypothetical protein	5.569	10.266	13.656
PY04004	hypothetical protein	4.550	5.068	3.837
PY04007	hypothetical protein	3.951	3.500	3.386
PY04020	hypothetical protein	-8.134	-7.650	-7.736
PY04022	hypothetical protein	-2.916	-2.877	-1.902
PY04032	hypothetical protein		-2.091	-1.575
PY04033	hypothetical protein		-2.012	-1.463
PY04036	hypothetical protein	1.593	-1.094	-2.018
PY04037	hypothetical protein	1.163	-1.413	-3.048
PY04041	hypothetical protein	-1.345	-1.113	-2.486
PY04043	hypothetical protein	2.455	3.373	3.656
PY04109	hypothetical protein	2.334	1.656	1.308
PY04126	hypothetical protein	-2.009		
PY04129	hypothetical protein	-2.714	-3.475	-3.003
PY04147	hypothetical protein	1.785	2.158	2.838
PY04149	hypothetical protein	1.777	3.489	3.249
PY04152	hypothetical protein	1.539	2.607	2.959
PY04154	hypothetical protein	-1.709	-2.144	
PY04159	hypothetical protein	2.367	2.193	
PY04163	hypothetical protein			-9.973
PY04165	hypothetical protein	-2.341	-1.297	-2.613
PY04166	hypothetical protein	-4.454	-1.205	-1.753
PY04175	hypothetical protein	-1.483	-2.190	-3.819
PY04176	hypothetical protein	-2.193	-2.025	-4.310
PY04183	hypothetical protein	-1.851	-1.358	-2.222
PY04184	hypothetical protein	-2.513	-1.465	-2.507
PY04187	hypothetical protein		-3.009	

PY04202	hypothetical protein	-2.961	-1.236	-2.010
PY04203	hypothetical protein	-1.690	-1.503	-2.378
PY04213	hypothetical protein	-2.415	-2.618	-2.908
PY04220	hypothetical protein	-36.552		
PY04222	hypothetical protein	1.628	1.895	2.094
PY04227	hypothetical protein	30.685	10.704	14.132
PY04228	hypothetical protein	1.434	2.523	2.859
PY04229	hypothetical protein	-1.734	-3.149	-2.551
PY04234	hypothetical protein	2.148	3.113	3.112
PY04237	hypothetical protein	3.389	6.080	
PY04240	hypothetical protein		-1.404	-2.071
PY04244	hypothetical protein	-2.840	-4.395	-6.700
PY04252	hypothetical protein	1.249	1.120	2.027
PY04256	hypothetical protein			-34.895
PY04258	hypothetical protein	-2.355	-1.281	
PY04264	hypothetical protein	-2.260		
PY04279	hypothetical protein	2.297	5.581	3.268
PY04280	hypothetical protein	-2.449	-5.568	-2.843
PY04290	hypothetical protein	-22.086	-11.474	-26.630
PY04292	hypothetical protein	-2.400	-1.048	-1.838
PY04293	hypothetical protein	-7.244	-4.267	-6.070
PY04303	hypothetical protein	-5.309	-3.117	
PY04305	hypothetical protein	-17.172	-12.684	-13.315
PY04313	hypothetical protein	-6.754	-23.479	-29.609
PY04315	hypothetical protein	-64.513	-55.547	-92.132
PY04317	hypothetical protein	-6.000	-7.010	-14.295
PY04320	hypothetical protein	2.803	5.366	7.644
PY04331	hypothetical protein	-2.466	-1.156	-1.861
PY04333	hypothetical protein	-1.421	-2.435	
PY04338	hypothetical protein	-2.231		
PY04350	hypothetical protein	-1.415	-1.052	-2.342
PY04356	hypothetical protein	-1.565		-3.535
PY04377	hypothetical protein	3.870	6.479	10.013
PY04397	hypothetical protein	2.721	2.269	5.513
PY04399	hypothetical protein	-4.138	-4.806	-4.956
PY04407	hypothetical protein	-7.931		
PY04413	hypothetical protein	16.414	25.938	32.121
PY04418	hypothetical protein	-5.470	-5.566	-7.036
PY04474	hypothetical protein	2.128	2.728	2.802
PY04486	hypothetical protein			-2.859
PY04490	hypothetical protein	1.604	2.161	2.286
PY04497	hypothetical protein		-3.368	-1.339
PY04502	hypothetical protein	-4.277	-1.334	-2.003
PY04520	hypothetical protein	1.634	3.865	5.154
PY04521	hypothetical protein	1.543	3.409	5.073
PY04550	hypothetical protein	13.479	55.658	39.650
PY04557	hypothetical protein	-4.653	-1.266	-2.281
PY04579	hypothetical protein	-3.385	-2.139	-1.597
PY04628	hypothetical protein	-1.496	-1.926	-2.274
PY04634	hypothetical protein	1.019	-1.132	-2.508
PY04655	hypothetical protein	-4.758		-1.651
PY04663	hypothetical protein	21.424	105.958	47.543
PY04672	hypothetical protein	2.760	1.331	1.398
PY04693	hypothetical protein	-1.470	-1.644	-2.295
PY04703	hypothetical protein	2.256	2.068	1.929
PY04709	hypothetical protein	2.289	1.313	-1.058
PY04734	hypothetical protein	2.776	2.423	3.516
PY04738	hypothetical protein	-2.056	-1.698	-1.626
PY04743	hypothetical protein	1.945	3.159	3.376
PY04749	hypothetical protein			-4.339
PY04775	hypothetical protein	-2.527	-2.638	-1.410
PY04803	hypothetical protein	-2.512	-2.057	-2.336
PY04807	hypothetical protein	5.099	3.118	
PY04809	hypothetical protein	1.908	2.139	2.173
PY04811	hypothetical protein	1.743	1.406	2.124
PY04827	hypothetical protein	2.849	3.914	9.879
PY04839	hypothetical protein	1.712	2.094	2.287
PY04843	hypothetical protein	46.204	48.043	96.280
PY04864	hypothetical protein	-4.787	-1.566	-2.078
PY04866	hypothetical protein	2.368	3.562	3.990
PY04876	hypothetical protein	2.450	2.976	2.920
PY04890	hypothetical protein	-1.709	-1.564	-2.035

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY04893	hypothetical protein	1.170	1.449	2.009
PY04897	hypothetical protein			-2.062
PY04899	hypothetical protein	1.063	-1.106	-2.354
PY04915	hypothetical protein	2.864	5.559	3.497
PY04924	hypothetical protein	-2.638	-3.349	-3.162
PY04932	hypothetical protein		6.422	4.951
PY04940	hypothetical protein	1.587	1.566	2.170
PY04950	hypothetical protein	2.076	1.892	1.781
PY04958	hypothetical protein	-2.210	-2.654	-1.230
PY04963	hypothetical protein	-22.146	-16.178	
PY04989	hypothetical protein	38.781	15.964	153.453
PY04991	hypothetical protein	-1.462	-1.552	-3.114
PY04997	hypothetical protein	2.879	2.966	3.158
PY05009	hypothetical protein	2.396	2.700	2.758
PY05012	hypothetical protein	3.445	3.571	3.500
PY05013	hypothetical protein	-2.097	-2.361	-1.839
PY05022	hypothetical protein			-2.960
PY05024	hypothetical protein	1.794	2.042	2.203
PY05031	hypothetical protein	-2.044		-1.763
PY05036	hypothetical protein	-3.036		-2.397
PY05040	hypothetical protein	2.149	2.099	2.165
PY05041	hypothetical protein	8.317	15.599	11.855
PY05042	hypothetical protein	4.483	5.962	5.383
PY05074	hypothetical protein	1.811	2.870	2.226
PY05080	hypothetical protein	-2.096		-2.533
PY05081	hypothetical protein	-3.629	-2.721	-2.671
PY05092	hypothetical protein	-3.971	-3.369	-2.765
PY05093	hypothetical protein	-2.736	-1.745	-1.461
PY05094	hypothetical protein	-6.407	-4.906	-3.948
PY05113	hypothetical protein			-2.324
PY05114	hypothetical protein		-3.376	-2.388
PY05118	hypothetical protein	-4.238	-5.583	
PY05125	hypothetical protein	2.246		
PY05167	hypothetical protein	1.746	5.685	5.042
PY05172	hypothetical protein	-6.053	-5.543	-5.724
PY05173	hypothetical protein	-3.802	-3.250	-3.501
PY05190	hypothetical protein	3.050	5.544	4.354
PY05194	hypothetical protein	1.867	2.339	1.917
PY05195	hypothetical protein	3.439	7.320	7.198
PY05219	hypothetical protein	5.625	11.304	7.162
PY05228	hypothetical protein	-2.249	-1.330	-1.961
PY05232	hypothetical protein	19.253	24.179	20.013
PY05234	hypothetical protein			-4.397
PY05240	hypothetical protein	5.408	4.708	
PY05256	hypothetical protein			5.829
PY05260	hypothetical protein	-1.980	-2.332	-2.926
PY05261	hypothetical protein	1.015	1.007	-2.030
PY05273	hypothetical protein			-2.828
PY05274	hypothetical protein		-3.883	-2.771
PY05275	hypothetical protein			-2.698
PY05279	hypothetical protein	1.810	4.503	2.786
PY05293	hypothetical protein	-1.555	-1.342	-2.264
PY05299	hypothetical protein	17.893	40.571	46265.393
PY05300	hypothetical protein	-1.943	-2.483	-1.238
PY05310	hypothetical protein	2.428	1.143	1.119
PY05312	hypothetical protein	3.411	1.443	2.113
PY05340	hypothetical protein	-2.171	-1.510	-2.844
PY05352	hypothetical protein	2.131	4.177	3.576
PY05360	hypothetical protein	3.580	5.568	6.787
PY05391	hypothetical protein	-1.012		-4.126
PY05404	hypothetical protein	3.343		
PY05406	hypothetical protein	-13.773	-6.523	
PY05408	hypothetical protein	-5.650	-7.553	-14.618
PY05409	hypothetical protein	-9.687	-10.453	-20.893
PY05412	hypothetical protein	5.366	7.218	6.594
PY05413	hypothetical protein	-1.024	-1.044	-2.397
PY05424	hypothetical protein	1.652	2.035	1.171
PY05427	hypothetical protein	-1.922	-3.261	-3.205
PY05433	hypothetical protein	21.073	17.859	31.338
PY05454	hypothetical protein	-47.325		
PY05456	hypothetical protein	62.840	55.646	60.423
PY05464	hypothetical protein	-1.071	-1.238	-2.491

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY05465	hypothetical protein	2.058	1.405	1.980
PY05476	hypothetical protein	-2.363	-1.605	-2.612
PY05479	hypothetical protein	3.974	6.025	4.540
PY05499	hypothetical protein	-4.633	-2.847	-2.289
PY05528	hypothetical protein	-1.849	-1.307	-3.007
PY05533	hypothetical protein	-1.985		-2.522
PY05534	hypothetical protein	2.044	2.592	2.637
PY05541	hypothetical protein	4.236	3.442	4.428
PY05547	hypothetical protein	-3.731	-2.975	-2.529
PY05568	hypothetical protein	-2.237	1.004	-1.523
PY05569	hypothetical protein	-2.168	1.032	-1.487
PY05608	hypothetical protein	2.137	3.706	4.329
PY05615	hypothetical protein	-2.312	-1.672	-1.315
PY05635	hypothetical protein	33.172	33.903	32.118
PY05643	hypothetical protein			-4.235
PY05685	hypothetical protein	-2.123	-1.916	-1.891
PY05686	hypothetical protein	-2.111	-1.971	-1.522
PY05706	hypothetical protein	1.370	1.647	2.053
PY05738	hypothetical protein	1.541	2.236	1.899
PY05775	hypothetical protein	1.646	2.816	2.209
PY05783	hypothetical protein	10.570	9.295	6.407
PY05784	hypothetical protein	-1.157	-1.630	-3.149
PY05785	hypothetical protein	2.202	3.232	2.831
PY05795	hypothetical protein	1.783	3.315	2.408
PY05806	hypothetical protein	1.082	-1.871	-4.017
PY05807	hypothetical protein			-4.864
PY05809	hypothetical protein	3.428	3.418	3.681
PY05819	hypothetical protein	1.239	6.819	10.887
PY05835	hypothetical protein	-1.262	-1.309	-2.884
PY05848	hypothetical protein	1.836	2.377	3.601
PY05852	hypothetical protein	4.178	6.599	7.031
PY05867	hypothetical protein	-55.902		
PY05874	hypothetical protein	1.591	3.366	3.293
PY05882	hypothetical protein		-2.890	
PY05917	hypothetical protein	-2.058	-2.378	-1.508
PY05926	hypothetical protein	-1.169	-1.687	-2.404
PY05929	hypothetical protein	-2.768	-4.077	
PY05956	hypothetical protein		-43.290	
PY05957	hypothetical protein	-2.040	-2.246	-1.794
PY05964	hypothetical protein	-1.357	-1.369	-2.152
PY05966	hypothetical protein	-2.864	-2.239	-1.146
PY05980	hypothetical protein	-1.840	-3.583	
PY05998	hypothetical protein	19.355	2.186	1.520
PY06028	hypothetical protein	-4.810		
PY06032	hypothetical protein	-2.222	-1.947	-1.268
PY06034	hypothetical protein	5.689	4.206	12.857
PY06038	hypothetical protein	-1.116	-1.123	-2.304
PY06039	hypothetical protein	1.840	2.386	2.674
PY06045	hypothetical protein	-1.402	-1.535	-2.433
PY06047	hypothetical protein			-29.763
PY06048	hypothetical protein	1.567	2.577	-2.236
PY06053	hypothetical protein	2.122	2.684	2.927
PY06069	hypothetical protein	-2.625		-4.912
PY06070	hypothetical protein			-2.439
PY06071	hypothetical protein			-2.140
PY06084	hypothetical protein		2.837	
PY06085	hypothetical protein		-2.827	-2.739
PY06092	hypothetical protein	2.590	1.687	2.422
PY06093	hypothetical protein	-2.426	-3.049	-2.617
PY06103	hypothetical protein		4.010	
PY06122	hypothetical protein	-1.167	1.880	3.326
PY06127	hypothetical protein	-2.005	-1.907	
PY06154	hypothetical protein	-3.309	-3.140	-2.368
PY06159	hypothetical protein	13.734	13.267	16.113
PY06190	hypothetical protein	-2.692		
PY06193	hypothetical protein	18.604	27.226	12.900
PY06197	hypothetical protein	-8.123		
PY06202	hypothetical protein	3.708	8.527	10.022
PY06216	hypothetical protein	2.079	2.058	2.204
PY06226	hypothetical protein			-12.293
PY06229	hypothetical protein	33.261	83.134	24.624
PY06266	hypothetical protein	-1.854	-2.035	

PY06283	hypothetical protein	-2.660	-3.076	-1.161
PY06288	hypothetical protein	-2.740	-3.335	-3.341
PY06293	hypothetical protein	1.813	1.908	2.075
PY06303	hypothetical protein	-2.421		
PY06317	hypothetical protein	4.846	13.546	6.892
PY06332	hypothetical protein	-1.984	-1.252	-2.725
PY06353	hypothetical protein	2.086	1.458	1.087
PY06354	hypothetical protein	1.706	1.233	2.122
PY06355	hypothetical protein	-2.951	-2.855	-2.578
PY06356	hypothetical protein	-2.396	-1.760	-1.550
PY06358	hypothetical protein	-5.977	-1.558	-2.407
PY06364	hypothetical protein		-3.046	
PY06365	hypothetical protein		-2.845	
PY06366	hypothetical protein	-1.147	-2.034	-2.253
PY06372	hypothetical protein		-3.061	-2.641
PY06376	hypothetical protein	-1.783	-2.109	-2.526
PY06392	hypothetical protein	2.565	5.036	3.714
PY06408	hypothetical protein	2.579	3.682	2.810
PY06419	hypothetical protein	-1.904	-3.151	
PY06426	hypothetical protein	2.461	2.500	3.078
PY06469	hypothetical protein			-2.614
PY06470	hypothetical protein	1.427	-1.318	-2.684
PY06471	hypothetical protein			-2.633
PY06476	hypothetical protein			-2.201
PY06494	hypothetical protein	1.471	1.621	2.349
PY06498	hypothetical protein	1.364	2.821	3.180
PY06500	hypothetical protein	2.116	3.088	3.049
PY06510	hypothetical protein	-2.324	-2.798	-2.314
PY06518	hypothetical protein			-5.605
PY06553	hypothetical protein	-2.183	-1.092	-1.993
PY06555	hypothetical protein	1.483	1.917	2.028
PY06579	hypothetical protein	11.394	40.606	30.061
PY06581	hypothetical protein	13.310	14.707	12.067
PY06594	hypothetical protein	1.435	2.478	2.373
PY06596	hypothetical protein	-2.335	-4.772	-4.344
PY06598	hypothetical protein	4.840	4.099	3.295
PY06603	hypothetical protein	-1.682	-1.606	-2.721
PY06609	hypothetical protein		-2.601	-1.533
PY06612	hypothetical protein	-2.081	-2.231	-1.941
PY06623	hypothetical protein	-1.178	1.595	2.430
PY06628	hypothetical protein	-1.287	-1.238	-2.196
PY06633	hypothetical protein	-1.840	-2.010	-1.484
PY06664	hypothetical protein	1.925	3.176	4.778
PY06671	hypothetical protein	-2.408	-1.640	-1.432
PY06676	hypothetical protein	-2.517	-4.102	-5.783
PY06690	hypothetical protein	-1.661	-1.176	-2.055
PY06703	hypothetical protein	2.235	1.495	1.496
PY06716	hypothetical protein			-2.375
PY06738	hypothetical protein	2.977	5.896	6.937
PY06751	hypothetical protein	1.687	3.786	2.561
PY06752	hypothetical protein	1.059	2.013	-1.744
PY06756	hypothetical protein		-111.111	-19.231
PY06765	hypothetical protein	-2.723		
PY06773	hypothetical protein	1.115	1.845	2.008
PY06779	hypothetical protein	-2.377	-1.743	-2.496
PY06781	hypothetical protein	4.892	9.759	7.778
PY06784	hypothetical protein	150.516	178.493	179.048
PY06793	hypothetical protein	1.699	3.617	4.187
PY06798	hypothetical protein	12.206		
PY06815	hypothetical protein	6.932	8.931	6.140
PY06816	hypothetical protein	2.532	3.041	2.846
PY06828	hypothetical protein	1.309	2.019	-1.845
PY06839	hypothetical protein	1.556	1.255	2.039
PY06847	hypothetical protein	-5.525	-14.512	
PY06850	hypothetical protein	125.972	122.205	118.780
PY06856	hypothetical protein	3.588	5.223	5.290
PY06879	hypothetical protein	5.111	11.513	46223.877
PY06892	hypothetical protein	6.374	4.902	3.266
PY06894	hypothetical protein	-2.505	-1.080	-1.879
PY06912	hypothetical protein	1.070		-2.130
PY06914	hypothetical protein	2.594	6.023	6.286
PY06917	hypothetical protein			-2.658

PY06926	hypothetical protein		-2.317	
PY06929	hypothetical protein	-3.350	1.065	-1.399
PY06931	hypothetical protein	1.745	2.236	2.408
PY06942	hypothetical protein	-2.208	-1.005	-1.229
PY06978	hypothetical protein	-2.478	-1.127	-2.161
PY06989	hypothetical protein	3.202		
PY07009	hypothetical protein	3.618	4.226	4.056
PY07015	hypothetical protein		-2.272	-8.800
PY07034	hypothetical protein	2.226	4.391	4.358
PY07050	hypothetical protein	8.113	22.293	20.057
PY07065	hypothetical protein	-1.431	-2.027	-1.594
PY07070	hypothetical protein	11.340	5.555	5.683
PY07076	hypothetical protein	-1.504	-3.075	-2.068
PY07109	hypothetical protein		8.853	13.670
PY07119	hypothetical protein	-1.345	-1.466	-2.529
PY07120	hypothetical protein	-1.214	-1.394	-2.617
PY07122	hypothetical protein	-3.352		
PY07144	hypothetical protein			-2.737
PY07155	hypothetical protein	-1.520	-1.303	-2.139
PY07167	hypothetical protein	15.472	40.606	44.609
PY07170	hypothetical protein	-7.579	-4.755	-23.394
PY07186	hypothetical protein	-1.576	-1.141	-2.171
PY07209	hypothetical protein	-3.167	-2.473	-2.846
PY07217	hypothetical protein		27.589	10.066
PY07230	hypothetical protein	-2.035	-2.368	-2.409
PY07231	hypothetical protein	1.505	2.237	1.552
PY07232	hypothetical protein	-1.951	-1.617	-2.378
PY07238	hypothetical protein	-2.187	-1.874	1.013
PY07241	hypothetical protein	3.451		
PY07242	hypothetical protein	2.081	2.152	2.949
PY07245	hypothetical protein	21.749	32.360	29.772
PY07257	hypothetical protein	2.745	2.505	2.039
PY07270	hypothetical protein	-3.617	-3.975	-2.973
PY07290	hypothetical protein	1.345	2.818	2.509
PY07300	hypothetical protein	-1.269	-1.283	-2.014
PY07301	hypothetical protein	-3.025	-2.407	-2.871
PY07308	hypothetical protein	2.577	4.738	3.223
PY07310	hypothetical protein	1.902	2.758	2.309
PY07319	hypothetical protein	-3.145	-1.393	-2.411
PY07324	hypothetical protein	1.588	2.012	1.912
PY07328	hypothetical protein	2.366	4.051	4.020
PY07335	hypothetical protein	2.955	4.206	6.635
PY07346	hypothetical protein	-55.902		
PY07347	hypothetical protein	1.638	3.101	5.790
PY07366	hypothetical protein		-3.449	
PY07367	hypothetical protein		-2.365	-3.021
PY07376	hypothetical protein		-2.604	-8.380
PY07378	hypothetical protein	33.179	842.996	169.013
PY07379	hypothetical protein	1.654	2.154	2.205
PY07393	hypothetical protein	22.046	25.974	63.501
PY07396	hypothetical protein	2.070	2.445	3.731
PY07398	hypothetical protein			-168.660
PY07407	hypothetical protein	-2.005		
PY07446	hypothetical protein	-2.090	-1.227	-1.837
PY07451	hypothetical protein	46.703	24.476	23.327
PY07452	hypothetical protein	1.648	2.225	1.827
PY07458	hypothetical protein	-3.876	-5.109	-2.635
PY07464	hypothetical protein	2.755	2.182	2.886
PY07478	hypothetical protein	24.693		
PY07479	hypothetical protein	4.179	3.940	2.027
PY07490	hypothetical protein	2.130	1.575	1.063
PY07503	hypothetical protein	-3.511	-1.821	
PY07504	hypothetical protein	-7.092		
PY07519	hypothetical protein	178.346	116.831	87.661
PY07530	hypothetical protein	-1.063	1.001	-2.788
PY07547	hypothetical protein	27.670	2.429	1.372
PY07552	hypothetical protein	-5.590	-2.726	-21.999
PY07556	hypothetical protein	-2.424	1.256	1.074
PY07559	hypothetical protein	4.371	27.268	198.332
PY07562	hypothetical protein	1.772	3.413	3.625
PY07564	hypothetical protein	-4.728		
PY07565	hypothetical protein	-24.096		

PY07567	hypothetical protein	-5.579		
PY07568	hypothetical protein	-5.160		
PY07572	hypothetical protein	-5.836		
PY07573	hypothetical protein	-2.437		
PY07585	hypothetical protein	-2.486		
PY07589	hypothetical protein	-2.830		
PY07591	hypothetical protein	-7.664		
PY07592	hypothetical protein	-5.588		
PY07602	hypothetical protein		-2.706	
PY07603	hypothetical protein	1.001	-1.960	-2.422
PY07604	hypothetical protein		-2.598	-8.234
PY07605	hypothetical protein	-1.576	-2.474	-3.732
PY07608	hypothetical protein	2.933	5.085	6.682
PY07612	hypothetical protein		-3.162	
PY07617	hypothetical protein		-2.086	-1.924
PY07621	hypothetical protein	-4.899		
PY07649	hypothetical protein	-6.524		
PY07650	hypothetical protein	-1.377	-1.518	-2.739
PY07651	hypothetical protein			9.458
PY07678	hypothetical protein	-2.105	1.190	-1.074
PY07682	hypothetical protein	1.111	-1.018	-2.022
PY07710	hypothetical protein			-8.258
PY07714	hypothetical protein	-25.685		
PY07821	hypothetical protein	2.513	2.006	4.343
PY07836	hypothetical protein	1.318	2.437	2.765
PY07842	hypothetical protein		-2.057	
PY07847	hypothetical protein	-2.632		
PY07855	hypothetical protein	-8.254		
PY06556	inositol hexakisphosphate kinase		-1.994	-2.026
PY04232	Insulinase, putative	-1.639	-3.086	-3.575
PY07014	integrase-related	-1.771	-2.036	-2.414
PY07371	isoleucyl-tRNA synthetase-related	-1.185	1.816	3.884
PY05788	isoleucyl-tRNA synthetase	-1.291	1.124	2.226
PY06600	KED	-1.254	-1.095	-2.307
PY03681	KED, putative	-1.459	-1.737	-2.092
PY00403	kinase Akt/PKB-related	3.702	3.637	2.108
PY06543	Kinesin motor domain, putative	-1.782	-1.332	-2.398
PY02372	kinesin-like protein-related	1.251	1.475	-2.637
PY00972	kinesin-related protein	1.878	2.948	2.017
PY06984	L1 transposable element		-3.328	-12.341
PY06196	latent nuclear antigen-related			-2.280
PY05554	LCCL domain-containing protein CCP2		-3.439	-1.429
PY05090	LCCL domain-containing protein CCP2-related	-4.326	-1.557	-1.819
PY04170	Lecithin:cholesterol acyltransferase, putative	-8.518	-7.845	
PY05438	Leishmania major ppg3			3.891
PY06687	Leucine Rich Repeat, putative	-1.001	2.932	1.491
PY03256	lipase precursor gehm. 6/11	2.299	3.108	3.005
PY01663	M83163 circumsporozoite protein-related	-2.298	-2.068	-1.575
PY03397	malate dehydrogenase-related		-2.143	-1.006
PY03922	malate dehydrogenase-related		-2.784	-1.357
PY02427	mannosyltransferase-related	-1.756	-1.091	-2.138
PY06554	MAP kinase-related	-2.806	-1.212	-1.921
PY00474	mature parasite-infected erythrocyte surface antigen	-2.805	-1.150	-1.961
PY00902	mature parasite-infected erythrocyte surface antigen		-2.391	-2.093
PY01460	mature parasite-infected erythrocyte surface antigen	-2.268	-1.428	-3.359
PY02239	mature-parasite-infected erythrocyte surface antigen	15.421	93.048	56.752
PY05855	Mechanosensitive ion channel, putative	-1.896	-1.495	-2.574
PY02147	Meloidogyne incognita COL-1-related	90.264	199.700	212.607
PY03535	membrane skeletal protein IMC1	-2.768	-1.526	-2.125
PY06415	merozoite surface protein precursor	50.052	46.325	36.879
PY02559	methionine aminopeptidase, type II, putative	6.592	9.010	7.661
PY07681	micronemal protein WARP	-1.982	-1.271	-2.587
PY00731	mitogen-activated protein kinase 2			-2.314
PY00208	MORN repeat, putative	1.281	-1.055	-2.812
PY00425	MORN repeat, putative	1.197		-3.310
PY02591	MORN repeat, putative		-1.415	-2.641
PY04141	MORN repeat, putative		-2.480	-1.480
PY05981	MORN repeat, putative	1.446	2.547	2.477
PY05972	Mus musculus RIKEN cDNA 53346L14 gene-related	-2.182	-1.773	-2.130
PY01096	MutS homolog 2-related	-1.703	-2.749	-3.087
PY06951	Myb1 protein-related	-1.611	-2.064	-1.301
PY00409	myosin A tail domain interacting protein MTIP	-2.335	1.207	-1.763

PY05907	NAD(P) transhydrogenase beta subunit, putative	-1.551	-1.453	-2.066
PY01048	neural polypyrimidine tract binding protein	-3.207	-3.825	
PY05730	neurofilament-H	-2.009	-1.105	-1.150
PY03136	nhp2-like protein 1-related	1.765	1.243	2.075
PY01658	Nuclear matrix protein NMP238		-2.162	1.125
PY00424	nuclear transport factor 2			-10.222
PY04487	NUDIX domain	-1.978	-2.205	-2.064
PY05065	O2	-1.488	-1.394	-2.552
PY07176	OGG1 protein type 2e-related	-2.082	-1.137	1.025
PY00523	ookinete surface antigen-like protein Pfs28	-3.924	-1.355	-1.209
PY04188	oxysterol-binding protein-related protein 2	-2.112	-1.823	-2.297
PY06381	p235 rhoptry protein E5	-17.667	-474.303	-1045.699
PY00292	Papain family cysteine protease, putative	2.604	11.028	9.755
PY04205	Pb47	-44.382	-16.178	-8.535
PY04395	Pb47-related		-3.370	-1.288
PY02721	PCI domain, putative	2.290	2.927	3.026
PY05336	PDI-like protein-related		1.655	2.114
PY04145	Peptidyl-tRNA hydrolase domain, putative		-3.696	
PY04898	pf2 homolog	-1.241	-1.432	-3.177
PY03857	PfS23 paralog-related	-1.442	-1.797	-2.165
PY04774	PfS23 paralog-related	-1.874	-1.991	-2.484
PY03088	Pfs77 protein-related	-2.017	-1.379	-2.709
PY04329	PfSUB-1	-8.799	-4.532	
PY00609	PH domain, putative	-1.480	-1.107	-2.065
PY00216	phosphatase	1.122	-1.005	-2.208
PY00102	phosphatidylinositol transfer protein 2	2.349	2.985	3.111
PY07597	Plasmodium falciparum CG6	11.020	10.018	8.881
PY00599	Plasmodium falciparum CPW-WPC domain, putative	-3.457	-1.405	-2.187
PY03515	Plasmodium falciparum CPW-WPC domain, putative	-2.165	-1.926	-2.880
PY04297	Plasmodium falciparum CPW-WPC domain, putative	-7.702	-15.664	-2.584
PY02647	Plasmodium vivax PV1H14195_P-related	6836.819	13061.140	648.593
PY05888	Plasmodium vivax PV1H1422_P	2.443	5.505	5.264
PY00501	Plasmodium vivax PV1H1495_P-related	1.178	1.945	2.221
PY04248	pre-mRNA cleavage complex ii protein clp1-related	-8.984		
PY00830	probable ATP-dependent RNA helicase has1	-1.871	-2.463	-2.057
PY01660	probable citrate synthase, mitochondrial precursor	-2.059	-1.627	-1.454
PY02817	probable DNA helicase ii homolog		-2.537	-1.670
PY07258	probable DNA-directed RNA polymerase iii 34 kDa polypeptide	-2.674	-2.781	-6.253
PY00385	probable glycine C-acetyltransferase, putative		-3.817	-3.091
PY00475	probable lipote-protein ligase a-related	1.498	2.145	2.290
PY04363	probable tRNA modification GTPase trme	2.161	2.460	1.447
PY00927	prolyl-tRNA synthetase-related	3.150	6.823	6.269
PY04190	proteasome subunit alpha Type 6-B	-3.644		
PY02685	proteasome subunit beta type 1-related	3.406	4.820	4.813
PY04296	protein disulfide isomerase precursor	-3.066	-2.295	-2.362
PY04198	Protein kinase domain, putative	-2.528	-7.310	-6.146
PY04195	protein Saccharomyces cerevisiae YLL18c-a-related	-1.641	-1.943	-2.485
PY04112	protein with 5'-3' exonuclease domain			-4.564
PY00120	putative 26S proteasome subunits	4.541	9.711	12.810
PY05756	putative 6S Ribosomal protein L7	2.153	2.105	2.053
PY02368	putative bir1 protein	17.137		
PY04804	putative bir1 protein	-37.037		
PY06804	putative bir1 protein	5.081		
PY07516	putative bir1 protein	104.124	6.714	25.199
PY03455	putative dual-specificity protein phosphatase	2.579	2.250	-1.245
PY02994	putative methionyl-tRNA synthetase	4.091	3.688	3.787
PY05818	Putative Methyltransferase-related	1.121	3.265	5.208
PY01852	putative phosphatidylethanolamine-binding protein		1.484	5.272
PY02573	putative phospho-ser/thr phosphatase	-1.398	-1.067	-2.185
PY01745	putative ribosomal protein L11	6.060	8.948	11.002
PY07229	putative small nuclear ribonucleoprotein polypeptide G	15.941	18.017	4.239
PY05710	Putative snoRNA binding domain, putative	2.776	3.241	3.773
PY05764	putative ubiquinone biosynthesis protein	1.748	3.223	3.747
PY00590	putative ubiquitin-conjugating enzyme 16	3.383	4.074	4.896
PY00495	putative yir1 protein	2.688	2.129	2.662
PY00552	putative yir1 protein	-2.731	-5.018	
PY01174	putative yir1 protein	86.069	50.202	202.095
PY01595	putative yir1 protein	17.319		
PY02824	putative yir1 protein	12.452	10.314	6.372
PY03486	putative yir1 protein	6.142	6.419	3.535
PY04785	putative yir1 protein	-2.404	-2.519	
PY04923	putative yir1 protein	-3.456		

PY05363	putative yir1 protein	5.627	2.679	1.075
PY05758	putative yir1 protein	7.059	3.040	
PY06547	putative yir1 protein	5.708		
PY06805	putative yir1 protein	9.294	32.872	10.068
PY03041	putative yir2 protein	11.498	18.348	28.719
PY03569	putative yir2 protein	11.746		
PY03758	putative yir2 protein	4.273	3.507	3.879
PY04782	putative yir2 protein	2.281	2.242	2.891
PY05942	putative yir2 protein	41.173	31.355	35.591
PY00498	putative yir3 protein	5.307	4.477	6.189
PY00579	putative yir3 protein	12.672		
PY05349	putative yir3 protein	2.254	1.413	1.044
PY05403	putative yir3 protein	2.316		
PY05636	putative yir3 protein	3.511	4.081	4.005
PY06131	putative yir3 protein	46.097	7.521	10.499
PY06795	putative yir3 protein	100.154	63.317	105.659
PY07473	putative yir3 protein	2.671		
PY07574	putative yir3 protein	-1.739	-2.214	-1.468
PY07831	putative yir3 protein	2.762		
PY00846	putative yir4 protein	-32.365	-47.801	-25.316
PY00924	putative yir4 protein	-125.606	-107.720	-59.982
PY01101	putative yir4 protein	-32.258		
PY01692	putative yir4 protein	-1.901	-2.150	-1.736
PY01981	putative yir4 protein	-14.153	-12.366	-12.155
PY03539	putative yir4 protein	14.020		
PY05148	putative yir4 protein	20.805	1.646	-1.003
PY05875	putative yir4 protein			-17.857
PY06807	putative yir4 protein	-66.799	-40.000	-40.000
PY00819	pyruvate dehydrogenase E1 alpha subunit	2.354	4.253	4.951
PY04691	R27-2 protein	-1.737	-1.988	-2.390
PY06249	related to CG16984 PROTEIN	1.961	1.549	-3.541
PY04214	repeat organellar protein-related	2.146	4.602	3.671
PY03130	repetitive proline rich protein, putative	-1.690	-1.539	-2.303
PY04496	replication factor C, 14 kDa subunit	2.581	5.272	5.651
PY01014	retinitis pigmentosa GTPase regulator-like protein	1.517	2.369	1.746
PY07669	Reverse transcriptase, putative		-3.160	
PY01365	rhophry protein	8.297	10.774	5.022
PY01526	rhophry protein	-4.413	-3.571	-3.048
PY01674	rhophry protein			-3.213
PY02635	rhophry protein		-2.407	-2.269
PY06945	rhophry protein	3.225	6.683	5.427
PY04318	rhophry protein, putative	-13.333		-8.322
PY03473	ribonucleoside-diphosphate reductase large chain	-1.291	-1.309	-2.098
PY03671	Ribonucleotide reductase, small chain	-2.013	-2.440	-1.536
PY06787	Ribosomal protein L36e	23.099	15.368	16.731
PY05324	ribosomal protein S12, putative	1.276	1.307	2.302
PY06059	ribosomal protein S2, putative		-3.279	-1.547
PY04916	ribosomal protein var1, putative	26.693	65.694	70.149
PY01016	RING-finger protein	4.421	15.772	20.472
PY02260	ring-infested erythrocyte surface antigen	1.318	-1.141	-2.403
PY02385	RNA polymerase Rpb7, N-terminal domain, putative	-2.088	-2.292	-1.521
PY04257	RNA pseudouridylate synthase, putative	-1.939	-2.038	-2.096
PY00973	RNA recognition motif, putative	7.083	8.418	6.608
PY00604	roadblock	-2.758	-2.234	-2.917
PY06442	Saccharomyces cerevisiae ORF 231-related	1.904	2.162	4.939
PY00291	SERA-3	3.441	9.975	11.087
PY00480	signal peptidase 18 subunit-related	1.889	3.105	4.725
PY03587	similar ATP-dependent RNA Helicase	2.052	1.775	1.310
PY07375	similar to L1 retrotransposon, TF subfamily L1MD-TF18	-1.028	-2.202	-2.886
PY06726	similar to S. cerevisiae RRP45, putative	1.612	2.207	2.518
PY00908	similar to S. cerevisiae SIK1	8.580	11.466	7.505
PY03646	similar to unknown protein	-2.105	-2.243	-1.466
PY00566	small heat shock protein	-1.814	-1.474	-2.233
PY05606	Smt3 activating enzyme 1	5.285	5.015	5.525
PY07291	splicing factor 3a subunit 3	3.685	3.593	3.183
PY04835	Streptococcus pyogenes AMV156, putative	1.589	2.225	2.828
PY07007	syntaxin 5	2.868	1.407	1.862
PY01032	synthetic antigen of P.falciparum, putative		-3.197	-1.156
PY02773	synthetic antigen of P.falciparum, putative		-2.129	-1.915
PY07275	t-complex protein 1, alpha subunit	1.624	1.972	2.842
PY06698	TERT gene	-1.091	1.081	-2.342
PY00479	TERT gene, putative	-1.173	1.724	2.251

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY05543	Tetrahymena thermophila Igr3p	-1.245	-1.042	-2.207
PY04185	thioredoxin	-2.494	-3.084	-1.944
PY06980	Thioredoxin, putative	-2.108	-1.700	-1.158
PY02498	thrombospondin-related protein-1		-26.092	
PY00129	TPR Domain, putative			-4.374
PY02959	transmission-blocking target antigen s23 precursor	1.008	-1.939	-2.122
PY03856	transmission-blocking target antigen s23 precursor		-2.150	-2.323
PY02417	transport protein particle component Bet3p-like protein	2.013	1.647	2.056
PY00825	tRNA-guanine transglycosylase-related	2.610	3.046	2.614
PY02818	Tubulin-tyrosine ligase family, putative		-1.447	-2.145
PY04254	tyrosyl-tRNA synthetase 1-related	5.387	11.365	7.767
PY06467	ubiquitin activating enzyme E1-like protein-related	-4.577	-3.055	-2.718
PY00183	Ubiquitin family, putative	6.240	6.845	5.791
PY02784	uncharacterized domain 2, putative	2.291	1.770	1.966
PY05575	undecaprenyl pyrophosphate synthetase-like-related	1.905	4.871	3.631
PY03841	Unknown protein	1.676	2.742	3.268
PY00828	unknown protein-related	1.373	3.525	3.492
PY05379	unknown protein-related	-4.566		-2.626
PY04139	Unknown-related	-2.494		-3.351
PY05160	Unknown, putative		3.636	
PY02328	unnamed protein product	1.181	1.901	2.028
PY02799	unnamed protein product	5.364	4.849	5.900
PY03577	unnamed protein product	-1.977	-1.325	-2.145
PY06180	unnamed protein product	4.667	3.948	3.963
PY01131	unnamed protein product-related		1.643	2.128
PY04969	unnamed protein product-related	3.178	5.311	6.430
PY02801	unnamed protein product, putative			-2.656
PY02946	unnamed protein product, putative	1.357	2.484	1.143
PY02528	uroporphyrinogen decarboxylase, putative	1.623	1.991	2.420
PY00513	vacuolar-type H ⁺ pumping pyrophosphatase-related	-10.439	-10.422	-11.690
PY05367	vps26 protein homolog	-2.238	-1.410	-1.327
PY01251	VPS45-like protein	3.239	4.629	5.662
PY01095	ymf77	1.653	2.593	1.845
PY05083	Zinc finger, C2H2 type, putative	3.067	2.322	1.987
PY00050	YIR		-2.295	
PY00085	YIR	-1.400	-2.000	
PY00089	YIR	22.799	30.172	36.482
PY00090	YIR	147.476		
PY00171	YIR	15.159	3.452	
PY00173	YIR	4.007	5.056	5.821
PY00175	YIR	18.737	6.206	19.304
PY00177	YIR	-39.087	-57.289	-77.555
PY00268	YIR	51.470		
PY00279	YIR	21.667	22.210	19.904
PY00300	YIR	4.988	4.499	3.641
PY00301	YIR	2.738	2.864	4.206
PY00303	YIR	34.383	55.769	26.405
PY00350	YIR	29.964	26.295	44.282
PY00390	YIR	-5.273	-4.999	-4.047
PY00392	YIR	8.535		
PY00400	YIR	19.752	7.074	15.010
PY00429	YIR		-75.911	
PY00488	YIR	2.663	2.790	3.530
PY00494	YIR	7.252	9.739	10.297
PY00500	YIR	4.618	4.281	5.267
PY00574	YIR		-2.329	-8.216
PY00578	YIR	7.405	11.954	14.709
PY00580	YIR	16.890	15.573	6.402
PY00610	YIR	61.586	45.576	61.038
PY00612	YIR	1.921	2.708	3.029
PY00635	YIR	7.383		
PY00642	YIR	17.209		
PY00814	YIR	-2.044	-2.422	-2.353
PY00842	YIR	3.146	4.974	4.640
PY00843	YIR	13.286	18.428	21.066
PY00849	YIR	2.923		
PY00989	YIR	21.221		
PY01079	YIR	7.461	25.195	16.608
PY01080	YIR	2.870	5.316	7.701
PY01099	YIR	-1.841	-2.126	-1.788
PY01110	YIR	27.905	6.529	121.547
PY01162	YIR	-2.185	-2.309	-1.979

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY01303	YIR	4.809	5.340	6.034
PY01318	YIR	5.749	2.695	1.672
PY01394	YIR	2.524	4.031	4.989
PY01404	YIR	5.186	3.965	4.139
PY01406	YIR	2.684	2.776	2.713
PY01417	YIR	3.254	4.988	4.224
PY01441	YIR	38.863	22.469	12.339
PY01442	YIR	2.313	1.806	1.964
PY01466	YIR	17.091	17.770	8.984
PY01541	YIR	-10.000	-9.042	
PY01598	YIR	3.177	2.909	3.351
PY01693	YIR	-2.283	-2.819	
PY01720	YIR	15.692	5.899	24.837
PY01774	YIR	24.355	45.559	34.102
PY01787	YIR	2.899	2.960	4.588
PY01788	YIR	3.823		
PY01810	YIR	12.915	18.853	15.372
PY01837	YIR	-1.995	-2.211	
PY01842	YIR	44.179	32.057	34.911
PY01865	YIR	-5.069		
PY01871	YIR	6.866		
PY01872	YIR	97.271	139.117	689.328
PY01873	YIR	26.814	94.790	8.386
PY01961	YIR	-16.077		
PY02011	YIR	38.026		
PY02038	YIR	16.687		
PY02107	YIR	3.312	1.780	1.720
PY02113	YIR	-1.562		-4.199
PY02131	YIR	98.173	22.604	19.764
PY02183	YIR	14.331	9.034	
PY02278	YIR	-7.411	-16.015	-9.290
PY02298	YIR	-72.918	-36.364	
PY02354	YIR		4.267	8.762
PY02542	YIR	29.634		
PY02544	YIR	9.588	3.258	
PY02560	YIR	9.141	12.380	12.790
PY02565	YIR	2.631		
PY02612	YIR	51.056	43.818	14.447
PY02687	YIR	5.800		
PY02688	YIR	10.971	17.270	21.003
PY02689	YIR	2.030	2.430	3.472
PY02727	YIR	-1.816	-2.048	-1.575
PY02730	YIR	2.701	2.381	2.411
PY02734	YIR	111.007	144.906	123.523
PY02757	YIR	7.551		
PY02762	YIR	-1.131	1.913	3.706
PY02828	YIR	-2.386	-4.884	
PY02831	YIR	46.037	89.421	13.672
PY02924	YIR	18.458	12.484	10.884
PY02940	YIR	-2.705	-2.085	-2.165
PY02970	YIR	2.196	1.734	2.154
PY02971	YIR	11.029	9.022	5.501
PY02972	YIR	25.087	16.322	5.839
PY02973	YIR	64.208	77.909	58.215
PY02976	YIR	-2.294	-2.587	-1.572
PY02996	YIR	17.892	8.602	5.407
PY03018	YIR	2.401	2.598	2.755
PY03023	YIR	-2.210	-2.396	-1.859
PY03042	YIR	44.894	109.089	57.705
PY03043	YIR	11.885	10.160	9.282
PY03094	YIR	-26.076		
PY03166	YIR	11.312		
PY03195	YIR	-37.616	-52.587	-43.523
PY03315	YIR	-4.622	-3.106	-3.124
PY03330	YIR	-19.356	-17.622	
PY03341	YIR	45.348	38.618	14.818
PY03358	YIR	19.689	6.659	5.806
PY03359	YIR	4.983	2.705	2.400
PY03360	YIR	4.492	3.603	5.715
PY03362	YIR	2.845	2.306	
PY03402	YIR	2.507	4.088	4.684
PY03403	YIR	3.951		

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY03428	YIR	33.283		
PY03436	YIR	6.871		
PY03438	YIR	1.935	2.042	1.692
PY03540	YIR	2.559	3.659	2.493
PY03542	YIR	3.913		
PY03564	YIR	3.090		
PY03580	YIR	108.851	127.423	55.057
PY03632	YIR	9.510	4.053	
PY03636	YIR	32.962	43.861	29.014
PY03679	YIR	-2.196		
PY03729	YIR	-22.183	-41.667	
PY03747	YIR	1.526	2.145	2.637
PY03784	YIR	2.213		
PY03826	YIR	1.690	2.692	3.013
PY03835	YIR	2.152		
PY03837	YIR	86.575	79.202	59.237
PY03849	YIR	1.341	2.133	
PY03858	YIR	41.178	17.644	10.337
PY03906	YIR	1.533		3.116
PY03912	YIR	-3.911		-1.152
PY03939	YIR	3.313		
PY03946	YIR	-15.913	-11.415	
PY03966	YIR	-31.180	-30.600	-23.272
PY03967	YIR	22.206	16.741	4.897
PY03974	YIR	3.048	2.251	2.446
PY03976	YIR	19.212		
PY03979	YIR	1.661	1.657	2.393
PY03983	YIR	25.231	5.517	
PY03985	YIR	16.101	10.967	12.417
PY03991	YIR	-2.817	-2.205	-1.907
PY03992	YIR	-2.523	-2.142	-1.727
PY04021	YIR	-9.998	-9.856	-9.211
PY04052	YIR	38.606	26.794	43.362
PY04053	YIR	-2.197	-2.494	-1.840
PY04082	YIR	1.825	2.082	2.372
PY04083	YIR	4.645	4.189	5.073
PY04084	YIR	2.779		
PY04144	YIR	-2.073	-1.205	1.485
PY04157	YIR	4.365	3.449	3.459
PY04208	YIR	12.668	11.652	46225.714
PY04209	YIR	2.425	2.151	1.663
PY04215	YIR	-9.599	-15.917	-16.063
PY04267	YIR	18.013	12.788	8.670
PY04276	YIR	31.348		
PY04319	YIR	27.543	20.952	29.361
PY04334	YIR	89.355		
PY04379	YIR	3.301	1.710	1.254
PY04396	YIR	3.996		
PY04412	YIR	51.136	36.013	27.260
PY04414	YIR	49.982		
PY04423	YIR	43.523	32.657	26.542
PY04466	YIR	5.072	4.829	4.932
PY04511	YIR	2.083	1.620	1.697
PY04541	YIR	-10.642	-5.762	
PY04552	YIR	19.227	8.975	
PY04667	YIR	2.969	1.621	1.108
PY04696	YIR	6.344		
PY04735	YIR	2.187		
PY04737	YIR	-35.936	-35.593	-20.928
PY04783	YIR	18.806	19.822	17.183
PY04786	YIR	-2.708		
PY04806	YIR	5.007	3.699	2.342
PY04808	YIR	8.087	10.203	14.927
PY04845	YIR	18.765		
PY04855	YIR	2.831	4.042	2.660
PY04918	YIR	-2.994		
PY04931	YIR	161.553	8.333	7.770
PY04939	YIR	1.945	2.369	2.064
PY04964	YIR	-35.421	-66.578	-144.565
PY04979	YIR	81.763	47.911	49539.492
PY04980	YIR	5.717	8.805	13.170
PY04988	YIR	74.799	128.864	25.959

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY04996	YIR	6.155	4.795	2.816
PY04998	YIR	15.091	5.917	6.949
PY05010	YIR	1.105	2.913	4.145
PY05021	YIR	2.889	5.798	6.776
PY05146	YIR	3.021	4.185	4.178
PY05150	YIR	21.596	9.331	6.737
PY05243	YIR	24.596	12.827	116.396
PY05244	YIR	11.090	6.884	3.479
PY05301	YIR	89.949	8.975	
PY05307	YIR	3.743		
PY05317	YIR	-15.346		
PY05410	YIR	59.402	30.069	18.191
PY05428	YIR	8.948	12.601	10.804
PY05437	YIR	31.984	57.626	102.185
PY05442	YIR	39.893	32.670	43.461
PY05455	YIR	81.847	75.373	105.423
PY05475	YIR	2.010	1.956	
PY05586	YIR	7.444		
PY05624	YIR	4.582	10.877	7.907
PY05637	YIR	79.600		
PY05638	YIR	4.735	6.753	6.109
PY05648	YIR	-2.085		
PY05656	YIR	6.286	9.091	11.146
PY05673	YIR	7.665	2.206	1.683
PY05675	YIR	19.188	25.378	12.470
PY05698	YIR	-3.737	-3.061	
PY05708	YIR	64.416	287.707	69.277
PY05709	YIR	54393.474	19.551	14.823
PY05731	YIR	-28.571		
PY05740	YIR	4.041	2.578	4.175
PY05770	YIR	-2.103		
PY05781	YIR	-4.422	-4.296	-3.877
PY05792	YIR	11.899		
PY05794	YIR	73.305	68.098	49499.974
PY05822	YIR	268.114	55109.005	28.402
PY05861	YIR	-14.153	-46.993	
PY05871	YIR	-1.558	-2.449	
PY05873	YIR	6.685	7.038	3.393
PY05889	YIR	2.652		
PY05894	YIR	6.252		
PY05943	YIR	21.105	9.182	8.593
PY06056	YIR	1.971	1.591	2.328
PY06113	YIR	2.436		
PY06126	YIR	5.908	3.108	
PY06129	YIR	12.073	10.258	12.145
PY06130	YIR	-78.978		
PY06141	YIR	18.163	10.704	7.507
PY06149	YIR	-5.168	-3.853	-4.912
PY06166	YIR	3.027	4.015	
PY06171	YIR	14.276	15.716	11.246
PY06188	YIR	54.372	27.210	86.170
PY06215	YIR	27.483		
PY06220	YIR	54.383	142.559	32.625
PY06244	YIR	12.448		
PY06271	YIR	-2.460	-2.298	
PY06279	YIR	2.796	5.959	6.795
PY06324	YIR	96.751	54.785	18.704
PY06335	YIR	3.455	1.772	1.702
PY06339	YIR	25.832	69.588	
PY06340	YIR	26.550	87.389	24.840
PY06352	YIR	57.735	418.389	35.705
PY06383	YIR	2.784	4.633	
PY06411	YIR	-2.661		
PY06428	YIR	5.296		
PY06482	YIR	4.547		
PY06497	YIR	54416.781	11.474	12.467
PY06499	YIR	11.204	10.783	4.976
PY06508	YIR	7.941	3.874	3.941
PY06536	YIR	-2.853	-2.695	-2.625
PY06549	YIR	-35.552	-48.929	-28.493
PY06569	YIR	105225.500		
PY06574	YIR	-11.314		

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY06611	YIR	-2.418	-2.473	-1.646
PY06619	YIR	10.390		
PY06621	YIR	103.471	9.048	
PY06649	YIR	-4.844	-5.828	-4.088
PY06666	YIR	12.117	54758.607	33.392
PY06667	YIR	13.121	7.629	
PY06674	YIR	12.115		
PY06678	YIR	76.634	54772.791	20.003
PY06700	YIR	5.781	3.840	3.309
PY06721	YIR	61.823	31.933	22.270
PY06732	YIR	-2.529	-3.758	
PY06743	YIR	58.414	48.943	18.115
PY06754	YIR	3.595	1.850	1.407
PY06755	YIR	29.579	13.173	
PY06763	YIR	7.828	9.649	22.174
PY06782	YIR	-8.485	-152.203	-11.120
PY06796	YIR	167.949	6.121	
PY06822	YIR	-7.986	-1.525	
PY06829	YIR	6.870	3.760	2.620
PY06830	YIR	27.035		
PY06831	YIR	18.859	10.678	
PY06834	YIR	3.341	2.846	4.551
PY06851	YIR	5.183		
PY06861	YIR	4.540	4.078	2.446
PY06891	YIR	63.816	649.290	52.743
PY06897	YIR	3.934	2.612	
PY06920	YIR	-2.255		
PY06934	YIR	6.735		
PY06982	YIR	30.166	26.298	22.606
PY06988	YIR	-2.733	-4.566	
PY06993	YIR	56.924	132.061	78.190
PY07027	YIR	46.087	39.635	39.464
PY07035	YIR	11.932		
PY07036	YIR	39.731	89.446	53.662
PY07052	YIR	3.411	3.737	4.418
PY07053	YIR	7.495	2.043	1.531
PY07090	YIR	-61.052		
PY07099	YIR	3.874	2.818	2.272
PY07125	YIR	3.068	4.405	4.134
PY07127	YIR	3.816	5.474	5.752
PY07149	YIR	76.378	10.471	4.426
PY07171	YIR	10.000		
PY07198	YIR	23.834	11.514	
PY07226	YIR	20.804		18.610
PY07246	YIR	5.696		
PY07261	YIR	2.624		
PY07281	YIR	-39.597		
PY07293	YIR	2.326	2.060	2.834
PY07305	YIR	3.963	4.901	6.074
PY07309	YIR		-18.978	-33.805
PY07315	YIR	13.048		
PY07325	YIR	134.180	75.037	102.747
PY07327	YIR	2.837	1.924	1.887
PY07342	YIR	2.542	2.381	2.107
PY07391	YIR	7.433		
PY07406	YIR	3.574		
PY07421	YIR	-2.279	-2.467	
PY07422	YIR	1.502	1.445	2.677
PY07450	YIR		19.846	154.528
PY07477	YIR	-2.492	-2.828	-2.367
PY07500	YIR	-14.649	-5.919	
PY07501	YIR	6.753	9.203	14.945
PY07512	YIR	-2.180		
PY07518	YIR	41.901	43.104	33.991
PY07524	YIR	73.825	27.273	42.048
PY07528	YIR	104.762	16.810	
PY07546	YIR	46.771	35.786	48.069
PY07577	YIR	-2.099	-2.386	-2.017
PY07594	YIR	-25.316		
PY07609	YIR	2.236		
PY07618	YIR	-6.377	-15.478	
PY07623	YIR	-2.124		

Appendix J: Differentially transcribed genes in YM with respects to YA

Appendices

PY07628	YIR	9.988	7.361	4.161
PY07635	YIR	26.964	37.199	103.440
PY07637	YIR	1.978	1.461	2.585
PY07639	YIR	2.415		
PY07645	YIR	2.432	3.284	2.145
PY07656	YIR	10.905		
PY07666	YIR	7.818		

Values are expressed as the relative fold expression of each gene where a positive value indicates a higher expression in YM while a negative value indicates a higher expression in YA.

Py Gene	Name	Fold difference TM
PY02416	3-oxoacyl-acyl-carrier protein reductase precurs	-1.295 -
PY02917	Arabidopsis thaliana T13D8.31	-1.681 -
PY06257	carbamoyl phosphate synthetase II-related	-5.287 -
PY00739	CCAAT-box DNA binding protein subunit B	1.210 -
PY02242	CCAAT-box DNA binding protein subunit B	-43.197 -
PY06770	CCAAT-box DNA binding protein subunit B	1.164 -
PY00834	CCAAT-box DNA binding protein subunit B, puta	-3.774 -
PY06797	cir1 protein-related	7.964 Y
PY01027	Drosophila melanogaster LD3351p	1.110 -
PY04120	ERYTHROCYTE MEMBRANE PROTEIN PFEMP3	-1.375 Y
PY02777	Ferrochelatase	1.312 -
PY00145	hypothetical protein	1.203 -
PY00146	hypothetical protein	2.969 Y
PY00214	hypothetical protein	-1.995 -
PY00235	hypothetical protein	-1.453 Y
PY00237	hypothetical protein	5.490 -
PY00338	hypothetical protein	-1.697 Y
PY00401	hypothetical protein	3.036 -
PY00466	hypothetical protein	-30.769 -
PY00484	hypothetical protein	-2.666 -
PY00627	hypothetical protein	-1.744 -
PY00822	hypothetical protein	2.104 -
PY00885	hypothetical protein	-1.679 -
PY01147	hypothetical protein	-1.291 -
PY01189	hypothetical protein	-1.294 -
PY01288	hypothetical protein	-1.192 -
PY01384	hypothetical protein	-1.172 -
PY01422	hypothetical protein	1.655 -
PY01444	hypothetical protein	-2.406 -
PY01796	hypothetical protein	1.756 Y
PY01840	hypothetical protein	-3.898 Y
PY01980	hypothetical protein	-3.284 -
PY02043	hypothetical protein	1.537 Y
PY02135	hypothetical protein	-10.454 -
PY02184	hypothetical protein	4.397 Y
PY02361	hypothetical protein	1.534 -
PY02424	hypothetical protein	-1.534 -
PY02468	hypothetical protein	1.130 -
PY02520	hypothetical protein	-1.732 -
PY02615	hypothetical protein	1.068 -
PY02630	hypothetical protein	1.382 -
PY02703	hypothetical protein	-1.175 Y
PY02759	hypothetical protein	-1.030 -
PY02998	hypothetical protein	1.807 Y
PY03221	hypothetical protein	-1.497 Y
PY03347	hypothetical protein	-6.774 -
PY03380	hypothetical protein	-29.626 -
PY03431	hypothetical protein	-6.509 -
PY03435	hypothetical protein	4.913 -
PY03437	hypothetical protein	6.994 Y
PY03567	hypothetical protein	1.364 -
PY03568	hypothetical protein	36.829 -

PY03584	hypothetical protein	-1.732 -
PY03643	hypothetical protein	-11.463 Y
PY03651	hypothetical protein	1.440 Y
PY03655	hypothetical protein	1.040 Y
PY03660	hypothetical protein	1.351 -
PY03785	hypothetical protein	-1.123 Y
PY03796	hypothetical protein	1.122 -
PY03804	hypothetical protein	-25.000 -
PY03891	hypothetical protein	-7.145 -
PY03977	hypothetical protein	1.075 -
PY03978	hypothetical protein	1.633 -
PY04126	hypothetical protein	-2.009 -
PY04138	hypothetical protein	-1.712 -
PY04216	hypothetical protein	-1.625 -
PY04220	hypothetical protein	-36.552 Y
PY04264	hypothetical protein	-2.260 -
PY04316	hypothetical protein	1.232 Y
PY04338	hypothetical protein	-2.231 -
PY04407	hypothetical protein	-7.931 -
PY04429	hypothetical protein	-1.518 Y
PY04553	hypothetical protein	1.341 -
PY04669	hypothetical protein	-1.776 Y
PY04788	hypothetical protein	-1.061 Y
PY04810	hypothetical protein	-1.225 Y
PY04836	hypothetical protein	1.076 Y
PY04846	hypothetical protein	-1.605 -
PY04847	hypothetical protein	1.280 -
PY04999	hypothetical protein	-1.068 -
PY05011	hypothetical protein	1.704 Y
PY05122	hypothetical protein	1.787 Y
PY05125	hypothetical protein	2.246 -
PY05159	hypothetical protein	-1.147 -
PY05334	hypothetical protein	1.024 -
PY05376	hypothetical protein	1.200 -
PY05404	hypothetical protein	3.343 -
PY05454	hypothetical protein	-47.325 Y
PY05655	hypothetical protein	1.388 Y
PY05862	hypothetical protein	-1.131 -
PY05867	hypothetical protein	-55.902 Y
PY05872	hypothetical protein	-1.522 -
PY05924	hypothetical protein	1.643 -
PY05944	hypothetical protein	1.824 Y
PY05989	hypothetical protein	-1.301 Y
PY06028	hypothetical protein	-4.810 -
PY06173	hypothetical protein	-1.724 -
PY06190	hypothetical protein	-2.692 Y
PY06197	hypothetical protein	-8.123 -
PY06303	hypothetical protein	-2.421 -
PY06487	hypothetical protein	-1.358 -
PY06744	hypothetical protein	1.408 Y
PY06745	hypothetical protein	-1.496 -
PY06765	hypothetical protein	-2.723 -
PY06798	hypothetical protein	12.206 -

PY06881	hypothetical protein	1.498 Y
PY06919	hypothetical protein	1.251 Y
PY06974	hypothetical protein	-1.144 Y
PY06989	hypothetical protein	3.202 -
PY06991	hypothetical protein	-1.333 -
PY07069	hypothetical protein	1.725 -
PY07122	hypothetical protein	-3.352 -
PY07241	hypothetical protein	3.451 -
PY07346	hypothetical protein	-55.902 -
PY07407	hypothetical protein	-2.005 Y
PY07478	hypothetical protein	24.693 -
PY07504	hypothetical protein	-7.092 Y
PY07564	hypothetical protein	-4.728 -
PY07565	hypothetical protein	-24.096 -
PY07567	hypothetical protein	-5.579 -
PY07568	hypothetical protein	-5.160 -
PY07572	hypothetical protein	-5.836 -
PY07573	hypothetical protein	-2.437 -
PY07585	hypothetical protein	-2.486 -
PY07589	hypothetical protein	-2.830 -
PY07591	hypothetical protein	-7.664 -
PY07592	hypothetical protein	-5.588 Y
PY07619	hypothetical protein	-1.467 -
PY07621	hypothetical protein	-4.899 Y
PY07649	hypothetical protein	-6.524 Y
PY07652	hypothetical protein	1.282 -
PY07839	hypothetical protein	-1.851 -
PY07844	hypothetical protein	1.097 Y
PY07846	hypothetical protein	-1.969 -
PY07847	hypothetical protein	-2.632 -
PY07855	hypothetical protein	-8.254 -
PY04248	pre-mRNA cleavage complex ii protein clp1-rela	-8.984 -
PY02547	Predicted membrane protein	-1.110 Y
PY03610	Predicted membrane protein	1.059 Y
PY04190	proteasome subunit alpha Type 6-B	-3.644 -
PY01773	putative bir1 protein	1.403 Y
PY02368	putative bir1 protein	17.137 Y
PY04804	putative bir1 protein	-37.037 Y
PY05703	putative bir1 protein	-1.110 -
PY06804	putative bir1 protein	5.081 Y
PY01595	putative yir1 protein	17.319 Y
PY03924	putative yir1 protein	-1.373 Y
PY04923	putative yir1 protein	-3.456 Y
PY06547	putative yir1 protein	5.708 Y
PY06740	putative yir1 protein	1.107 -
PY07093	putative yir1 protein	1.114 -
PY03569	putative yir2 protein	11.746 Y
PY07415	putative yir2 protein	-1.051 Y
PY00579	putative yir3 protein	12.672 -
PY05403	putative yir3 protein	2.316 -
PY07473	putative yir3 protein	2.671 Y
PY07831	putative yir3 protein	2.762 Y
PY00754	putative yir4 protein	1.237 Y

PY01101	putative yir4 protein	-32.258 -
PY03488	putative yir4 protein	-1.021 -
PY03539	putative yir4 protein	14.020 Y
PY05482	similar to RIKEN cDNA 4931428F4 gene	-1.920 -
PY04306	triosephosphate isomerase	-1.766 -
PY00321	YgbB family, putative	-1.741 Y
PY00090	YIR	147.476 Y
PY00268	YIR	51.470 Y
PY00392	YIR	8.535 Y
PY00490	YIR	1.521 Y
PY00559	YIR	-1.077 Y
PY00635	YIR	7.383 Y
PY00642	YIR	17.209 Y
PY00849	YIR	2.923 Y
PY00989	YIR	21.221 Y
PY01190	YIR	-1.177 Y
PY01649	YIR	-1.878 Y
PY01706	YIR	1.819 Y
PY01788	YIR	3.823 Y
PY01865	YIR	-5.069 Y
PY01871	YIR	6.866 Y
PY01961	YIR	-16.077 Y
PY02011	YIR	38.026 Y
PY02038	YIR	16.687 Y
PY02541	YIR	-1.255 Y
PY02542	YIR	29.634 Y
PY02565	YIR	2.631 Y
PY02687	YIR	5.800 Y
PY02757	YIR	7.551 Y
PY02782	YIR	1.850 Y
PY02943	YIR	-1.433 Y
PY02974	YIR	-1.008 Y
PY03094	YIR	-26.076 Y
PY03166	YIR	11.312 Y
PY03194	YIR	1.103 Y
PY03263	YIR	1.119 Y
PY03399	YIR	1.132 Y
PY03400	YIR	-1.155 Y
PY03403	YIR	3.951 Y
PY03415	YIR	1.451 Y
PY03428	YIR	33.283 Y
PY03436	YIR	6.871 Y
PY03542	YIR	3.913 Y
PY03564	YIR	3.090 Y
PY03679	YIR	-2.196 Y
PY03699	YIR	1.197 Y
PY03784	YIR	2.213 Y
PY03835	YIR	2.152 Y
PY03939	YIR	3.313 Y
PY03976	YIR	19.212 Y
PY04084	YIR	2.779 Y
PY04221	YIR	-1.059 Y
PY04233	YIR	1.857 Y

PY04276	YIR	31.348 Y
PY04310	YIR	1.242 Y
PY04321	YIR	1.991 Y
PY04334	YIR	89.355 Y
PY04396	YIR	3.996 Y
PY04414	YIR	49.982 Y
PY04425	YIR	-1.773 Y
PY04505	YIR	1.382 Y
PY04696	YIR	6.344 Y
PY04721	YIR	1.172 Y
PY04733	YIR	1.068 Y
PY04735	YIR	2.187 Y
PY04786	YIR	-2.708 Y
PY04845	YIR	18.765 Y
PY04918	YIR	-2.994 Y
PY05166	YIR	-1.837 Y
PY05307	YIR	3.743 Y
PY05317	YIR	-15.346 Y
PY05318	YIR	1.229 Y
PY05415	YIR	-1.770 Y
PY05586	YIR	7.444 Y
PY05637	YIR	79.600 Y
PY05648	YIR	-2.085 Y
PY05731	YIR	-28.571 Y
PY05770	YIR	-2.103 Y
PY05792	YIR	11.899 Y
PY05796	YIR	1.559 Y
PY05816	YIR	1.914 Y
PY05889	YIR	2.652 Y
PY05894	YIR	6.252 Y
PY06031	YIR	1.615 Y
PY06113	YIR	2.436 Y
PY06130	YIR	-78.978 Y
PY06153	YIR	-1.051 Y
PY06215	YIR	27.483 Y
PY06227	YIR	-1.611 Y
PY06244	YIR	12.448 Y
PY06302	YIR	-1.100 Y
PY06326	YIR	-1.048 Y
PY06411	YIR	-2.661 Y
PY06428	YIR	5.296 Y
PY06482	YIR	4.547 Y
PY06495	YIR	-1.467 Y
PY06569	YIR	105225.500 Y
PY06574	YIR	-11.314 Y
PY06619	YIR	10.390 Y
PY06674	YIR	12.115 Y
PY06710	YIR	-1.991 Y
PY06760	YIR	-1.885 Y
PY06830	YIR	27.035 Y
PY06851	YIR	5.183 Y
PY06853	YIR	-1.072 Y
PY06862	YIR	1.126 Y

PY06863	YIR	-1.223 Y
PY06880	YIR	1.288 Y
PY06918	YIR	-1.304 Y
PY06920	YIR	-2.255 Y
PY06934	YIR	6.735 Y
PY06946	YIR	-1.192 Y
PY07035	YIR	11.932 Y
PY07090	YIR	-61.052 Y
PY07111	YIR	1.081 Y
PY07171	YIR	10.000 Y
PY07246	YIR	5.696 Y
PY07261	YIR	2.624 Y
PY07269	YIR	1.193 Y
PY07281	YIR	-39.597 Y
PY07314	YIR	1.141 Y
PY07315	YIR	13.048 Y
PY07356	YIR	1.209 Y
PY07391	YIR	7.433 Y
PY07397	YIR	-1.033 Y
PY07406	YIR	3.574 Y
PY07474	YIR	-1.828 Y
PY07512	YIR	-2.180 Y
PY07517	YIR	1.527 Y
PY07594	YIR	-25.316 Y
PY07609	YIR	2.236 Y
PY07623	YIR	-2.124 Y
PY07624	YIR	-1.051 Y
PY07639	YIR	2.415 Y
PY07656	YIR	10.905 Y
PY07666	YIR	7.818 Y

Values are expressed as the relative fold expression of each gene where a positive value indicates a lower expression in YA while a negative value indicates a higher expression in YA.

Legend:

TM: transmembrane domain

SP: signal peptide

SP	Comment
Y	-
-	-
-	-
Y	tRNA synthetase
-	no domain
-	no domain
-	no domain
-	-
-	-
-	no domain
-	-
-	-
-	-
-	<100nt
Y	-
Y	-
-	-
-	-
Y	-
-	-
-	-
-	<100nt
-	-
-	-
-	-
-	<100nt
-	-
-	-
Y	-
Y	<100nt
-	-
Y	<100nt
-	-
-	<100nt
-	<100nt
-	-
-	-
-	-
-	<100nt
Y	<100nt
Y	-
-	-
Y	-
-	-
-	-
-	-
-	-
-	<100nt
Y	<100nt
-	-
-	-

-	<100nt
-	-
Y	-
-	-
-	-
-	-
-	-
-	<100nt
-	-
-	<100nt
-	-
-	-
-	<100nt
-	-
-	-
-	-
-	-
Y	-
-	-
Y	-
-	-
-	-
Y	-
-	<100nt
-	<100nt
-	-
Y	-
Y	-
-	-
-	-
-	<100nt
-	-
-	-
Y	-
Y	-
-	<100nt
-	<100nt
-	<100nt
-	-
Y	<100nt
-	-
-	-
-	-
-	-
-	-
-	<100nt
-	-
-	-
-	<100nt

[illegible]

Y	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
Y	-
-	-
-	-
-	-
Y	-
-	-
-	-
Y	-
-	-
-	-
-	-
-	-
-	-
-	-
-	-
Y	-
-	-
-	-

e indicates a higher expression in YM