Chapter 1

Introduction

1

Malaria

Malaria is a devastating disease that has already been described by Hippocrates in ancient Greece, roughly 2,500 years ago. For long it was thought that the bad air (mal aria) from marshes was causing the disease. In 1880, Alphonse Laveran, a French surgeon working in Algeria discovered the malarial parasite in the blood of a patient suffering from malaria and in 1897, Ronald Ross, an English doctor born and working in British India demonstrated the transmission of avian malaria parasites by feeding female anopheline mosquitoes, which two years later was confirmed for humans by the Italian investigator Giovanni Grassi. There are four species of parasitic protozoa that cause malaria in humans of which *Plasmodium falciparum* is the most devastating and is responsible for the majority of deaths. The second most important malaria parasite for humans is *Plasmodium vivax*, which is found mainly in South-East Asia and South America but which is absent from large parts of Africa. In 1955, the World Health Organization (WHO) initiated an ambitious and intensive eradication programme. With a combination of mosquito control using dichlorodiphenyl-trichloroethane (better known as DDT) to prevent transmission and extensive treatment of malaria cases using anti-malarial drugs such as the highly effective and affordable drug chloroquine, one hoped to be able to deal with the malaria problem once and for all. Despite all the efforts, tropical countries are dealing with a strong resurgence of malaria during the past decades, such that more people are now suffering from malaria than ever before[1]. Different aspects have contributed to this resurgence, including (i) the emergence and rapid spread of drug-resistant malaria parasites[2-4] and insecticide-resistant mosquitoes[5]; (ii) factors that affect the public-health system, such as continuing political instability and war, unrelenting poverty and natural disasters; and (iii) more frequent transmission due to an increased (more than doubled) human population[6]. A recent extensive survey has shown that, in 2002, roughly 2.2 billion people were at risk of contracting *P. falciparum*, while a conservative estimate of 515 million became infected[1]. The majority of these cases (70%) occurred in Africa, while a significant 25% of the cases were reported in the densely populated South-East Asian region. The same authors estimated that, in 2000, 1.1 million Africans, mainly children under five years old, died from malaria[7], a number only challenged by tuberculosis. Apart from the human suffering, malaria is responsible for a significant economic burden and has been estimated to decrease economic growth by 1.3% annually[8].

Symptoms of the disease are a consequence of proliferation of the parasites in the blood where they infect red blood cells, resulting in complications such as anaemia, hypoglycemia, cerebral and placental malaria. The blood-stage infection is only one part of the complex life cycle shared by all parasites of the genus *Plasmodium*. The malaria life cycle is summarized below; a detailed description is provided in Chapter 2.

Parasites in the form of sporozoites are introduced into the blood with the saliva of a feeding mosquito and rapidly invade a liver cell where they develop and replicate to form over 10,000 daughter parasites (merozoites). Upon release merozoites will infect red blood cells where they will develop and divide into 16-32 new merozoites. This stage of the infection is responsible for the pathology typical

of the disease. Small numbers of the merozoites stop replication after erythrocyte invasion and develop into either a male or female sexually committed cell. After ingestion by another mosquito, these sexually committed parasites escape from the red blood cells transforming into gametes, fertilization takes place and the parasite traverses the midgut epithelium. In the midgut lining an oocyst is formed in which over 10,000 sporozoites develop that migrate to the salivary glands of the mosquito, ready to continue the cycle.

The problem of malaria and the aim of this study

Malaria has been under investigation for over a century, but despite the intensive research efforts no effective vaccine is available yet and people are still dependent on the few cheap and effective drugs in use that are losing efficacy rapidly due to drug resistance. It is vital to continue research efforts to generate drugs against previously successful targets and to identify and exploit new targets. The availability of an effective vaccine is generally seen as an essential tool to successfully combat this devastating infection, while alternative strategies may be developed and also employed to prevent transmission of the parasite.

The *P. falciparum* genome sequencing project was initiated with these goals in mind. The real-time release of partial genome sequences during the course of this 6-year project enabled researchers to identify unique *P. falciparum* genes that can serve as novel drug and vaccine targets[9,10]. The completion of the *P. falciparum* genome provided the malaria research community with an unprecedented opportunity to identify more *P. falciparum*-specific genes or genes that differ sufficiently from the host genes such that they may serve as targets for chemotherapeutic interventions with a decreased risk of side effects. In addition, the genome is predicted to encode a large number of proteins that would be dispersed to the surface of the parasite offering a much expanded range of potential vaccine candidates. Proteomic studies experimentally validated 70% of the predicted genes providing an insight in the evolution of the metabolic pathways utilized by the parasite and its unique features as compared with the human host[11,12]. The comparison of the *P. falciparum* and *Plasmodium yoelii* genomes provided a wealth of information on differences in genome organization stressing the importance of the subtelomeric regions in the generation of diversity in genes that allow the parasite to change and thereby evade recognition by the host immune system. The biggest advance has been made in the understanding of the biology of the parasite, its complex life cycle and the strategies employed for its survival in the variable environments. Whether one studies molecular evolution or gene transcription, population genetics or developmental biology, cellular mechanics or signal transduction, whole-genome information is what defines the playing field[13].

The aim of the studies described in this thesis was to compare and exploit the conservation of organization and gene content of the genome of malaria parasites that infect rodents with those of the human parasite *P. falciparum*. These rodent malaria parasites (RMPs) are widely used as research models and one of these, *Plasmodium berghei*, is the model for malaria research used in the Leiden Malaria Research Group (see also our website, http://www.lumc.nl/1040/research/malaria/malaria.html). A high level of conservation of genome organization and gene

content would validate the use of the RMPs for investigations to identify and characterize new vaccine and drug targets.

RMPs: models in malaria research

There are over 200 different *Plasmodium* species described infecting a wide range of hosts, including reptiles, birds, rodents, non-human primates and humans[14]. Only four species infect humans: *P. falciparum*, *P. vivax*, *Plasmodium ovale* and *Plasmodium malariae*, while the first two are a common cause of infection, the latter two are relatively rare.

It is possible to culture *P. falciparum* to study the disease-causing blood stages of the parasite. Technical and ethical considerations render the study of other stages of the infection, such as invasion of liver cells and transmission through mosquito feeding, virtually impossible. The use of model malaria species, for example infecting birds, rodents or non-human primates, can provide access to these less accessible stages of the malaria life cycle. Additionally, these systems allow the study of the infection *in vivo* and thus with all the complications associated with it, such as cerebral malaria (Ref. [15] for review) or responsiveness of the immune system. For example, the differential effects on the biochemistry, bioenergetics and gene expression in mice brains were examined following infection with cerebral and non-cerebral *P. berghei* strains[16]. Analysis of the competitive ability of *Plasmodium chabaudi* strains of variable virulence suggested that within-host competition is a driving force in parasite evolution where transmission efficiency is related directly to blood-stage parasite numbers, which may explain why many parasites harm their hosts[17]. Four *Plasmodium* species have been identified that have African tree rats as their natural host but which can also infect other rodents such as laboratory mice and rats. Three of these (*P. berghei*, *P. yoelii*, *P. chabaudi*) are widely in use as models in malaria research, mainly because of the relatively low costs and acceptible ethical concerns of *in vivo* experimentation (in comparison with model *Plasmodium* species that infect non-human primates). Many aspects of the biology, life cycle, and morphology of RMPs show a high level of similarity with the human parasites, validating their use as models for human infection. There is a high degree of conservation of metabolic pathways, which is reflected in a similar molecular basis of drug-sensitivity and resistance. In addition, many surface antigens of human parasites that are prime-candidate vaccine targets are also present in RMPs, such as TRAP and CS of sporozoites; CTRP, P25 and P28 of ookinetes; AMA1 and MSP1 of merozoites; and P45/48, P47 and P230 of gametes. *In vitro* culture techniques for large-scale production and manipulation of different life cycle stages are available, including the parts of the life cycle less accessible in the human parasites, such as the liver and mosquito stages. RMPs further allow *in vivo* investigations of parasite-host interactions as well as *in vitro* and *in vivo* drug testing, while the possibility to genetically modify parasites and the availability of well-characterized genetic background of mouse and rat and transgenic lines are invaluable for immunological studies. A high level of conservation of genome organization and gene content between *P. falciparum* and the RMPs would further validate the use of the RMPs for investigations to identify and characterize new vaccine and drug targets in these models. It was expected that a considerable proportion of the genome would be

conserved, reflecting the conservation of the complex life cycle of all *Plasmodium* species that infect mammals. Morphologically, little to no differences can be observed between the corresponding life cycle stages of different *Plasmodium* species. Many of the processes are shared; these include but are not restricted to the invasion of liver cells and red blood cells (though from different hosts), the sexual development necessary for transmission, fertilization, penetrating the mosquito midgut epithelium and migration to the salivary glands. Differences in gene content and genome organization will most likely be related to the adaptation of the parasites to their respective hosts. Relatively small differences between human and mouse or rat liver and red blood cell architecture, but more importantly, differences in the immune defence systems, will have forced the parasites to adapt, thus generating differences that we expect to find back in the genomic organization and gene content.

P. falciparum pre-genomics

The first malaria parasite genes were cloned at the beginning of the 1980s. Many of these genes encoded surface proteins that are exposed to the host immune system. In the early days of recombinant DNA technology, hopes were high that cloning important *Plasmodium* antigens would rapidly lead to development and production of an effective vaccine. Cloning of the first *Plasmodium* antigen, encoding a surface protein of the infective sporozoites (circumsporozoite protein) from a *Plasmodium knowlesi* cDNA library was a milestone in malaria research in 1983[18]. One year later, the *P. falciparum* orthologue of this gene was cloned[19], followed by other genes encoding potential vaccine candidates chiefly of blood stages of the parasite life cycle. Thereafter, there was a rapid increase in the amount of *Plasmodium* DNA sequence available in GenBank (http://www.ncbi.nlm.nih.gov/) - in 1990, there were roughly 70 entries and by 1995 that number had grown to more than 1,000, mainly from *P. falciparum* but also from *Plasmodium vivax* and other model *Plasmodium* parasites.

The individual chromosomes of *Plasmodium* could not be visualized by conventional microscopy, but their separation by pulsed-field gel electrophoresis (PFGE) revealed that the genome comprises 14 linear chromosomes in the size range of 0.5-3.5 Mb, resulting in an early estimate of the total genome size of about 25-30 Mb (~2.5x that of the yeast *Saccharomyces cerevisiae*[20,21]). PFGE analysis also revealed large variations in the size of the subtelomeric regions (Refs. [22,23] for reviews), which contain numerous and varied DNA repeats. The subtelomeric regions also contain species-specific gene families encoding proteins that are transported to the surface of infected erythrocytes and are involved in antigenic variation and immune evasion (Ref. [24] for review). By contrast, initial results of comparative mapping of housekeeping genes on the individual chromosomes revealed a high level of organizational conservation (synteny) between the core regions of the chromosomes of different *Plasmodium* species[25,26]. It was also demonstrated that *Plasmodium* possesses two non-nuclear genomes - a compact but anticipated mitochondrial genome of 6 kb and, surprisingly, a plastid-like 35-kb circular genome that was ultimately shown to reside in an organelle now known as the apicoplast[27].

Despite such advances, problems such as non-protective immune responses evoked by the chosen antigens and difficulties with vaccine antigen production all hindered the production of the hypothesized multi-stage cocktail vaccine[28]. In addition, many areas of the biology of the parasite remained insufficiently characterized. It was generally hoped that the sequencing of the genome might sidestep several of these problems. Following a successful multi-centre genome-mapping exercise[20], a genome sequencing consortium was established in 1997[29] working on the principle of real-time data deposition to allow the scientific community to benefit from the work-in-progress.

The genomics era and comparative genomics

The genomics era for eukaryotes started in 1996 with the publication of the complete genome sequence of the yeast *Saccharomyces cerevisiae*[21]. Since then, roughly 20 eukaryotic genomes have been sequenced, including those of mammals (that can be infected by malaria parasites) such as human[30], rat[31] and mouse[32], the tiger pufferfish[33], the sea squirt[34], insects like the fruit fly *Drosophila melanogaster*[35] and the malaria mosquito *Anopheles gambiae*[36], the nematode worms *Caenorhabditis elegans*[37] and *Caenorhabditis briggsae*[38] and plants including two rice species[39,40] and *Arabidopsis thaliana*[41]. Besides genome sequences of numerous prokaryotes, many of which cause disease in humans, complete genome sequences are now also available for a number of eukaryotic parasites. These include five apicomplexan parasites: *P. falciparum*[42], *Cryptosporidium parvum*[43] (infecting both humans and other mammals), *Cryptosporidium hominis*[44] (restricted to humans), *Theileria parva*[45], and *Theileria annulata*[46] (which both infect African cattle); three kinetoplastid parasites: *Trypanosoma brucei*[47] (which causes African sleeping sickness), *Trypanosoma cruzi*[48] (Chagas disease), and *Leishmania major*[49] (Leishmaniasis); and finally *Entamoeba histolytica*[50]. These data together with substantial amounts of partial genome sequence data, including partial genome sequences of three RMPs, *P. yoelii*[51] (Chapter 3), *P. berghei* and *P. chabaudi*[52] (Chapter 4), have been made publicly available through the websites of the sequencing centres and consortiums. It is expected that the volume of released sequence data will increase rapidly if not exponentially over the coming years.

The sequences of all these individual genomes contain a wealth of information that can be used by the scientific communities that study the different organisms. Analysis of single genomes has enabled the generation of hypotheses such as the whole-genome duplication of yeast[53] but comparative genome analysis has proved to be a powerful tool to test these theories. Comparative genomics can further improve our understanding of the genetic principles underlying the differences between both closely and more distantly related species and their evolutionary relationships by shedding light on the mechanisms that helped reshape their respective genomes including but not limited to: (i) micro-rearrangements such as single gene deletions, inversions and duplications; (ii) gross chromosomal rearrangements like translocations and deletions, inversions and duplications of entire segments resulting in loss of synteny; (iii) whole-genome duplications as shown for the yeast; (iv) ectopic exchanges in the (sub)telomeric regions of the chromosomes; and (v) the presence of recombination hotspots. Furthermore,

comparative genome analysis can significantly improve: (vi) the identification of both highly and less conserved orthologues either through homology or the analysis of syntenic segments; (vii) the identification of species-specific gene content which might be related to specific adaptations to environmental conditions; (viii) the identification of conserved non-coding elements, regulatory or structural; (ix) the annotation of genes, especially of small or complex multi-exon genes; and (x) assigning putative functions to hypothetical proteins. Many of these aspects can also aid in *Plasmodium* research as will be demonstrated below using examples from recently published comparative genome studies.

All comparative genome analysis is based upon the assumption that the two studied genomes originate from a common ancestor and that the respective genome sequences are the result of evolution acting on the ancestral genome sequence, *i.e.* a combination of the accumulation of random mutations and subsequent selection for example due to environmental pressures. Therefore, the resolving power of a two-sided whole-genome comparison to a large extent depends upon the proximity of the phylogenetic relationship between the species.

Global alignment comparisons between vertebrates (450 million years [My] divergence)[33] and comparisons between two Diptera, the fruit fly *D. melanogaster* and malaria mosquito *A. gambiae* (250 My divergences)[54] revealed that roughly 75% and 50% of the genes in the respective genomes being compared have orthologues. Gene contents of the fruit fly were also compared with the more distantly related nematode *C. elegans* and the yeast *S. cerevisiae* demonstrating that nearly 20% of the *D. melanogaster* genes have a putative orthologue in both other species reflecting a core eukaryotic gene set, however, all three genomes also appeared to contain approximately one-third of unique, species-specific genes without a homologue in either of the other species or itself[55]. Comparison between species from a single genus revealed that *C. elegans* and *C. briggsae* share 60% orthologues[38], while for four *Saccharomyces* species the amount of orthologues is as high as 95%[56]. Comparison of the *Plasmodium* gene content with that of the human host may provide insight the unique gene content and biological pathways that may be employed for the development new drugs, while knowledge of the common genes shared by parasite and host should help understand and hopefully prevent potential side effects induced by the treatment. The first comparison of the *P. falciparum* gene content with all sequences available in the public databases revealed that in contrast with other eukaryotes, as much as two-thirds of the *Plasmodium* gene content are unique[42]. This may reflect the larger evolutionary distance between *Plasmodium* and other eukaryotes, increased even more by the exceptionally high AT content of the genome. With the availability of genomes of more closely related species such as the Apicomplexa *C. parvum*[43]-*C. hominis*[44] and *T. parva*[45]-*T. annulata*[46] and more apicomplexan genome sequences like that of *Toxoplasma gondii* underway, this number is expected to decrease considerably. The identification of a *Plasmodium* core gene (Chapters 3 and 4) set should help validate the model species used in malaria research and identify common targets for drug interventions aiming to cure all four different malaria species infecting humans. The differences in gene content between different *Plasmodium* species will hopefully shed light on the molecular basis underlying species-specific traits such as host-specificity, differences in virulence (including

pathologically important phenotypes like sequestration, rosetting, or clumping) or transmission efficiency, hypnozoite formation in *P. vivax* or reticulocyte-preference. This is further exemplified by the comparative analysis of *Listeria monocytogenes*, the etiologic agent of listeriosis, a severe food-borne disease, with the non-pathogenic *Listeria innocua*. The presence of 270 *L. monocytogenes*- and 149 *L. innocua*-specific genes (clustered in 100 and 63 indels, respectively) suggests that virulence in *Listeria* results from multiple gene acquisition and deletion events[57]. Such a clear relation between gene content and virulence is not obvious from the comparison of the genome sequences of *Bacillus cereus*, an opportunistic pathogen causing food poisoning, and the animal and human pathogen *Bacillus anthracis*, which indicated the conservation of numerous factors for invasion, establishment and propagation of bacteria within the host expected for *B. anthracis* but not *B. cereus*[58]. Comparative analysis of the genome *Bordetella bronchiseptica*, which causes a chronic infection of the respiratory tract in a variety of animals, with the genomes of two closely-related bacteria causing whooping cough in humans (*Bordetella pertussis* and *Bordetella parapertussis*) demonstrated relations between genome organization and host-specificity[59]. During evolution of the host-restricted species, there has been extensive gene loss and inactivation. The authors also suggest a link between virulence and loss of regulatory functions.

The closest available pairs of eukaryotic genomes that are most fully sequenced to date are, for multicellular organisms, the free living nematodes, *C. elegans* and *C. briggsae* that diverged 80-110 My ago[38] and, for unicellular species, *S. cerevisiae* and three related yeast species, *Saccharomyces paradoxus*, *Saccharomyces mikatae* and *Saccharomyces bayanus*, that are thought to have a 5-20 My evolutionary distance[56]. In both comparisons the chromosomal ends appear to diverge more rapidly than the core regions of the chromosomes and the compared genomes show extensive co-linearity. The comparison of such closely related eukaryotic genomes reveals dynamic processes such as the constitution and perhaps origins of gene families. In *Caenorhabditis*, 96% of functionally organized gene clusters are conserved and the majority of diverged sequence consists of rapidly evolving repetitive DNA elements, which also account for the 4% difference in genome size. Furthermore, a direct comparison at the nucleotide level suggested a predicted increase of 1,300 *C. elegans* genes based purely on the identification of conserved regions between the two genomes[38]. The analyses of the different *Saccharomyces* genomes suggested the elimination of ~500 gene models of *S. cerevisiae* and further enhanced previous annotations by identifying 43 new putative, small genes (encoding 50-99 amino acids) and by redefining intron-exon boundaries, start and stop codons. Screening of the intergenic regions of the four yeast species for sequence motifs further revealed 72 genome-wide elements, including most known regulatory motifs but various new ones were also defined[56]. On a smaller scale comparative analysis has also been shown to help improve *Plasmodium* gene annotation by identifying genes in a complex and gene-dense region that were initially missed by the annotation algorithms[60]. This study also showed that using genome comparison it was possible to improve definitions of the intron-exon boundaries. The availability of multiple *Plasmodium* genomes can help train the annotation algorithms through identification of short but conserved coding regions.

Micro-rearrangements result from insertions, deletions and duplications of genes, repeat elements or other short DNA segments. The rate at which rearrangements occur seems to depend on the genomic location[32,61]. These local changes are most prominently present in (sub)telomeric and (peri)centromeric regions and happen at a much higher rate than gross chromosomal rearrangements. These highly recombinant regions mainly consist of tandem arrays of repetitive elements, including species-specific transposable elements (like in *D. melanogaster* and *A. gambiae*)[35,36] and recent gene duplications amongst primates. In the case of *P. falciparum*, subtelomeric regions are the main location for members of three gene families involved in immune evasion, the *var*, *rif*, and *stevor* families. Indeed, it has been suggested that the subtelomeric location of the *var* genes is essential for the process of antigenic variation in *P. falciparum*[62]. Though the nature of the repeats varies amongst different organisms, a relation with the genomic instability of these regions seems obvious.

A majority of the synteny breakpoints (SBPs) between human chromosome 19 and the mouse genome are located in regions with many repeats elements or clustered gene families[63]. Similar associations were found when different primate genomes were compared and, strikingly, many of the segmental duplications also seem to play a role in chromosomal rearrangements involved in human genetic diseases and polymorphisms[64,65]. *transfer rna* (*trna*) genes flank inversion breakpoints between four yeast genomes[56] and several repetitive elements in *C. elegans* could be associated with translocation and transposition events[66].

Gross chromosomal rearrangements helped reshape the organization of large synteny blocks (SBs). SBs are regions of conserved gene content and organization between different species, with the exception of micro-rearrangements like gene insertions, deletions or inversions. Within these syntenic regions the resolving power of comparisons can be greater facilitating the identification of both novel genes and conserved non-coding elements that control gene expression. While the genomes of four yeast species exhibit a relatively small number of one to five translocations[56], those of the nematodes *C. elegans* and *C. briggsae* are arranged in as many as 4,837 syntenic clusters[38]. Human and mouse have a predicted gene content that is 80% orthologous[32] arranged in 281 SBs larger than 1 Mb[67]. The presence of a large number of short “hidden” SBs, which are defined by closely located SBPs, led to the suggestion that mammalian genomes are mosaics of fragile regions with high propensity for rearrangements and solid regions with low propensity for rearrangements[68]. It has been estimated that at least 245 rearrangements of these SBs have occurred since the divergence of human and mouse[67]. Establishing if similar fragile regions exist in the genomes of malaria parasites could demonstrate additional mechanisms through which genetic diversity can be created as well as confirm the known generation of genetic diversity in the subtelomeric regions. Alternatively, the conservation of large genome segments between different *Plasmodium* species could indicate that there is a selective disadvantage to these gross chromosomal rearrangements perhaps because of some higher order organization of the genome[69].

Eichler and Sankoff[70] wrote a clear review on chromosomal dynamics of eukaryotic chromosome evolution also containing a synteny map of the mouse genome overlaid on top of the human genome while a more detailed description of

yeast evolution and comparative genomics was published recently by Liti and Louis[71]. Using the general principles set out in these two papers, we attempted to put our findings on the evolution *Plasmodium* genome organization and gene content in the perspective of what is known about eukaryotic genome evolution, noting the remarkable similarities as well as differences that exist between *Plasmodium* genomes and both extremes of eukaryote genome landscape.

* * *

Outline of this thesis

Malaria parasites that infect rodents are widely used models in the study of the biology of human malaria parasites and for the identification and characterization of targets for drugs and vaccines. The value of such studies using RMPs is dependent on the level of similarity between RMPs and the malaria parasites that infect man. The aim of the studies described in this thesis was to investigate the genome organization of the RMPs, with specific emphasis on *P. berghei*, in more detail and compare and exploit the organization and gene content of RMP genomes with those of the human parasite *P. falciparum*.

In **Chapter 2**, a review is given describing the current status of genomic and post-genomic research in *Plasmodium* summarizing the different genome sequencing projects and our understanding of the genome organization of different *Plasmodium* species, including the conclusions from the comparative genome analyses between RMPs and *P. falciparum* resulting from the investigations described in this thesis. In addition, this chapter contains a detailed description of the complex life cycle of the malaria parasite and many useful links to websites containing information on both genome and post-genome research in general and about malaria in particular.

Prior to the publication of the *P. berghei* genome sequence, investigations on the genome organization of *P. berghei* started with the characterization of the 14 chromosomes by separation using PFGE[26]. We in particular focussed on unravelling the genome organization of *P. berghei* chromosome 5 (Pbchr5), since several genes expressed in the sexual stages appeared to be clustered on Pbchr5[60,72]. Using a long-range restriction map of Pbchr5 and 15 markers, a physical map was generated. Simultaneously with the publication of the complete *P. falciparum* genome in 2002, partial sequence data for another RMP *P. yoelii* were released enabling the first-ever comparative genome analysis of genome sequences of two species belonging to the same genus. In this study presented in **Chapter 3**, the physical map of Pbchr5 was used to demonstrate the high level of conservation of the core region of this chromosome of the RMPs *P. yoelii* and *P. berghei* with that of large parts of only two chromosomes (Pfchr4 and 10) of the human parasite *P. falciparum*. This showed for the first time in detail the high level of synteny (conservation gene content and organization with the exception of microrearrangements) between rodent and human malaria parasites. This study also provided the first clues that the subtelomeric regions of chromosomes of RMPs are highly divergent from those of *P. falciparum* and that these regions are

separated by distinct boundaries from the core regions that show a high level of synteny between the rodent and human malaria parasites. Interestingly, in these variable regions many species-specific gene families are located.

After publication of the first RMP genome of *P. yoelii*, the partial genome sequences of two additional RMPs, *P. berghei* and *P. chabaudi*, have been published, which is presented in **Chapter 4**. Comparison of these genome sequences with that of the other RMP *P. yoelii* and that of the human parasite *P. falciparum* showed a high level of conservation of gene content. At least 4,500 of the 5,300 genes of *P. falciparum* have an RMP orthologue (the core *Plasmodium* gene set) and are localized in the core regions of the chromosomes (the central, non-subtelomeric regions). A majority of the 736 *P. falciparum* genes without an RMP orthologue belong to one of the *P. falciparum*-specific gene families; 161 are located within the core regions disrupting synteny while 575 are located in the subtelomeric regions of the chromosomes. These subtelomeric genes could be assembled into 12 distinct gene families only five of which are shared with the RMPs.

The availability of the genome sequences of three RMPs and a completely annotated *P. falciparum* genome made it possible to generate a detailed genome-wide synteny map of four *Plasmodium* species. This study is described in **Chapter 5** and shows that the organization of the core regions of the RMP and *P. falciparum* genomes are highly conserved in as little as 36 SBs. Analysis of these SBs showed that the organization of *P. falciparum* genome could be generated from that of the composite RMP (cRMP) genome in a minimum of 15 chromosomal recombination events and *vice versa*. This relatively low number of only 15 rearrangements suggests that gross chromosomal rearrangements resulting in the loss of or change in synteny is infrequent in *Plasmodium*. Moreover, the locations of both centromeres and boundaries between the conserved core regions and variable subtelomeric regions are conserved between the RMP and *P. falciparum*. The 168 non-subtelomeric *P. falciparum*-specific genes (161 genes reported in Chapter 4 plus seven genes of the newly discovered *vicar* family) disrupting synteny were analysed in more detail. Of these genes, 42 are located between the SBs at SBPs while 126 are located in so-called indels disrupting the syntenic regions. Interestingly, 68% of these genes are potentially exported to the surface of the parasite or infected erythrocyte and several belong to gene families, including two newly discovered gene families. These results show that not only subtelomeric regions but also SBPs and indels can be foci for species-specific genes with a role in host-parasite interaction and immune evasion and suggest involvement of gross chromosomal rearrangements in the generation of *P. falciparum*-specific gene families. This is exemplified by the discovery of *P. falciparum*-specific gene family consisting of 21 copies that encode transforming growth factor β (TGF-β) receptor-like serine/threonine protein kinases (PfTSTK) with only a single syntenic orthologue in the RMPs. Combination of the suggested 15 recombination events with phylogenetic analysis of the TSTK protein sequences provided insights in the mechanisms underlying the generation of this gene family.

The studies described in Chapters 2 to 5 have been initiated with the characterization of the genome organization of Pbchr5 since this chromosome contained a number of genes that are exclusively expressed during sexual development[72]. A detailed analysis of a 13.6-kb region, the B9 locus, of *P. berghei* containing six tightly clustered genes, three of which are exclusively expressed during the sexual stages of the parasite, revealed high levels of conservation with its *P. falciparum* counterpart on Pfchr10. The gene number, organization of the intron-exon boundaries of the four multi-exon genes and expression patterns are entirely conserved[60]. We have been trying to investigate these genes in more detail by gene modification technologies. The results of these studies have not been published yet but will be discussed briefly in Chapter 7. Analysis of the gene content of Pbchr5 revealed the gene encoding α-tubulin II, which is also expressed during sexual development. Malaria parasites have two genes that encode α-tubulins, one of which, α-tubulin I, is expressed constitutively and is located on Pbchr4, while the second one, α-tubulin II on Pbchr5, is highly expressed in male gametocytes and there is evidence for a specific function in the formation of the axoneme of the male gamete. We have characterized both *P. berghei* genes and tried to analyse the precise role of α-tubulin II in sexual development of particularly the male gametocytes by gene modification, which is described in **Chapter 6**. Surprisingly and despite its importance for male gamete formation, α-tubulin II is not exclusively expressed during sexual development but is also essential for normal asexual development of the blood stages.

In **Chapter 7**, the results of our studies on the genome organization of *P. berghei*, that were initiated with investigation of the organization of Pbchr5, and the comparative genomics studies are summarized and discussed. In addition, some studies are mentioned that were aimed at characterization of individual sex-specific genes that are located in the gene-dense B9 locus on Pbchr5 that is highly conserved between the RMP and *P. falciparum*.