Chapter 9

Standardization in Generating and Reporting Genetically Modified Rodent Malaria Parasites: The RMgmDB Database

Shahid M. Khan, Hans Kroeze, Blandine Franke-Fayard, and Chris J. Janse

Abstract

Genetically modified *Plasmodium* parasites are central gene function reagents in malaria research. The Rodent Malaria genetically modified DataBase (RMgmDB) (www.pberghei.eu) is a manually curated Web - based repository that contains information on genetically modified rodent malaria parasites. It provides easy and rapid access to information on the genotype and phenotype of genetically modified mutant and reporter parasites. Here, we provide guidelines for generating and describing rodent malaria parasite mutants. Standardization in describing mutant genotypes and phenotypes is important not only to enhance publication quality but also to facilitate cross-linking and mining data from multiple sources, and should permit information derived from mutant parasites to be used in integrative system biology approaches. We also provide guidelines on how to submit information to RMgmDB on non-published mutants, mutants that do not exhibit a clear phenotype, as well as negative attempts to disrupt/mutate genes. Such information helps to prevent unnecessary duplication of experiments in different laboratories, and can provide indirect evidence that these genes are essential for blood-stage development.

Key words: *Plasmodium*, *Plasmodium berghei*, Malaria, Rodent, Genetically modified mutants, Database, RMgmDB, Standardization, Genotype, Phenotype

1. Introduction

Reverse genetic technologies are the frontline methodologies used to gain an understanding into the function of malaria parasite genes (1–3). The availability of efficient reverse genetic technologies for the rodent malaria parasites, *Plasmodium berghei* and *Plasmodium yoelii*, and the fact that these can be combined with analyses of rodent parasites throughout their complete life cycle, both in vitro and in vivo, have made these parasites the most frequently used

Robert Ménard (ed.), *Malaria: Methods and Protocols*, Methods in Molecular Biology, vol. 923,
DOI 10.1007/978-1-62703-026-7_9,

models for gene function analysis (1). Targeted disruption (knockout) or mutation of genes coupled with protein tagging has been extensively used to investigate gene function in rodent parasite species. Reverse genetics has been largely applied to loss-of-function studies and is now increasingly being used for the generation of the so-called reporter parasites. These reporter parasites have heterologous genes introduced into their genome, frequently genes encoding fluorescent or luminescent proteins that are expressed under the control of a variety of parasite stage-specific or constitutive promoters.

The number of different rodent parasite mutants is increasing rapidly, year on year. In order to bring together essential information on all these rodent parasite lines, a Web-based repository has been developed that contains genotypic and phenotypic data from mutant and reporter parasites (4). Most information in this database (the Rodent Malaria genetically modified DataBase; RMgmDB; www.pberghei.eu) has been manually curated from scientific literature retrieved from Medline searches. The database is constantly updated and collates all the latest data from Medline searches. RMgmDB also contains information on unpublished mutants without a clear phenotype and negative trials to disrupt or mutate genes. Information in the database can be searched using predefined key features, such as phenotype, life cycle stage, gene model, gene tags, and mutations (Fig. 1). The database can also be accessed via corresponding gene pages in the *Plasmodium* genome resources PlasmoDB (www.plasmodb.org) and GeneDB (www.genedb.org).

The full utility of RMgmDB for the research community is dependent on the quality of the information available in the database and how this information is presented and how it can be searched. In turn the accuracy of the genotype and phenotype information is principally dependent on the quality of the description of the mutants in the corresponding scientific report. Below we show a set of considerations, distilled from a number of publications as a "standardized template" in the generation, analysis, and description of mutant rodent malaria parasites. These criteria, which have been described by Janse et al. (4), are based on generally implemented and now increasingly widely accepted procedures used in a number of different laboratories. In addition, we provide guidelines and templates for submitting information to RMgmDB on non published mutants, mutants without a clear phenotype, and negative attempts to disrupt/mutate genes. It is believed that the use of these guidelines might not only improve the quality of the generation and description of mutants but also can facilitate the comparison of data from different studies. In addition, it will enhance searching the database for relevant information and integrating RMgmDB data with data from other resources.

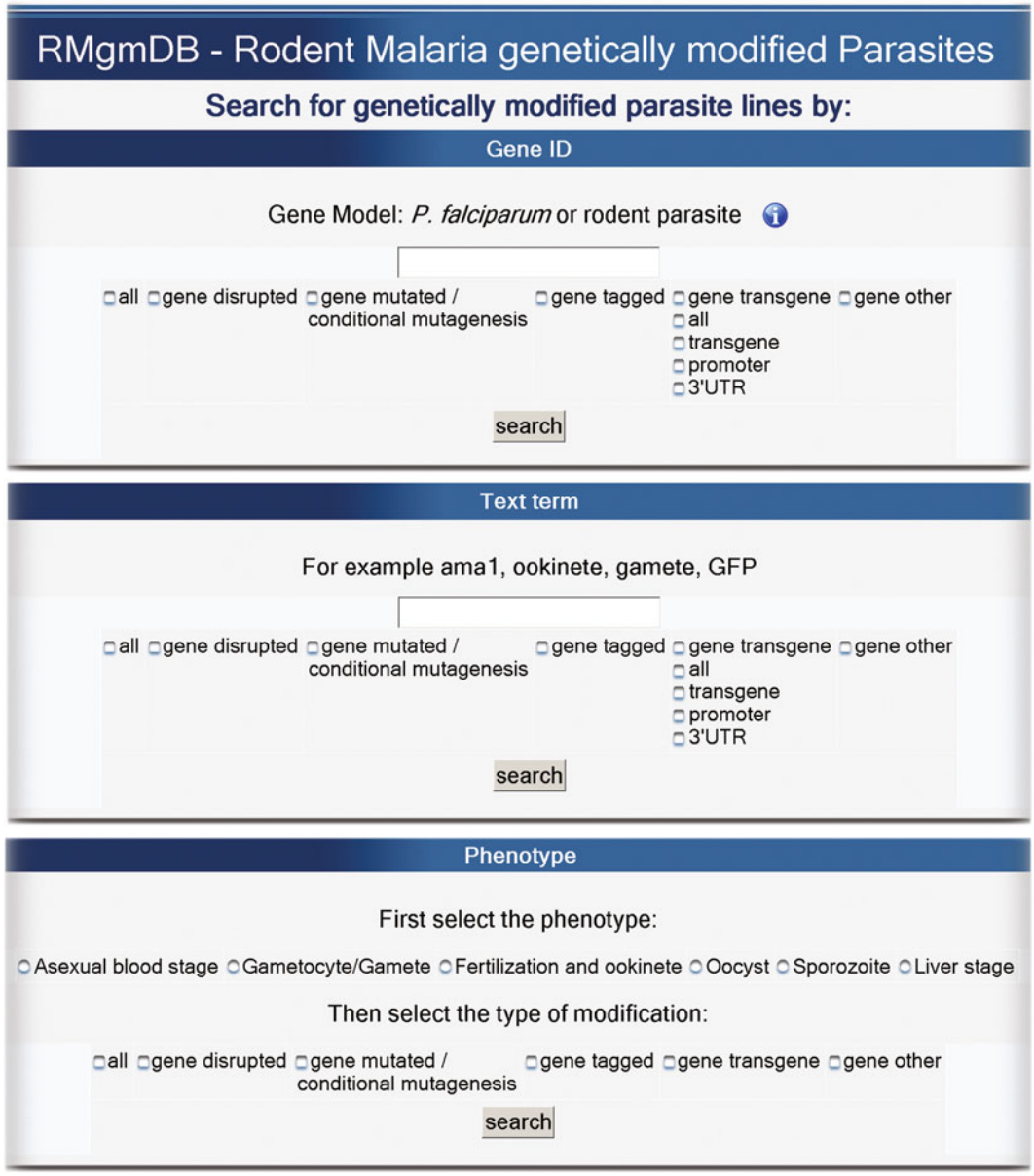

Fig. 1. Search fields for querying RMgmDB. Searches on mutant information can be performed based on the *Plasmodium* Gene ID, a text term, or a life cycle phenotype, and these searches can be restricted to the type of mutation (gene disruption, mutation, tagging, etc.).

2. Materials and Methods

We provide guidelines for generating rodent malaria parasites mutants as well as on how to report the genotype of these mutants.

2.1. Standardization of Mutant Generation and Mutant Genotype Description

2.1.1. Parasite

Required information:

– Rodent malaria parasite species (*P. berghei*, *P. yoelii*, *Plasmodium chabaudi*) and strain/isolate used to generate the mutant.

Additional information:

– Further information if the genetic modification is made in a specific clone or a mutant parasite line (e.g., in a parasite line that expresses reporter proteins such as GFP or luciferase).

2.1.2. Target Gene for Disruption, Mutation, or Tagging (See Note 1)

Required information:

– GeneDB/PlasmoDB gene model.

– Schematic representation of the gene and information necessary to understand the mutant genotype (e.g., restriction sites, location of probes and primers) (see Fig. 2).

Additional information:

– In the case where multiple mutants all have a similar mutation/modification (e.g., gene disruption) and have been analyzed using an identical approach, it might be sufficient to provide the gene models and the sequence of the appropriate primers (e.g., used to amplify the target regions for homologous recombination) accompanied with a general schematics.

2.1.3. Introduction of a Transgene (See Note 1)

Required information:

– GeneDB/PlasmoDB gene model of a *Plasmodium* transgene, e.g., where a *Plasmodium falciparum* gene was integrated in a rodent parasite genome.

– Standard name of a "non-*Plasmodium*" transgene, including other commonly used synonyms, and other details (e.g., origin).

– Method of introduction of the transgene into the parasite: episomal or, if integrated into the genome, by double crossover (DXO) or single crossover (SXO) homologous recombination.

– GeneDB/PlasmoDB gene model of the target locus where the transgene is integrated.

– GeneDB/PlasmoDB models of genes providing the promoter (5′UTR) and terminator (3′UTR) sequences that drive transgene expression.

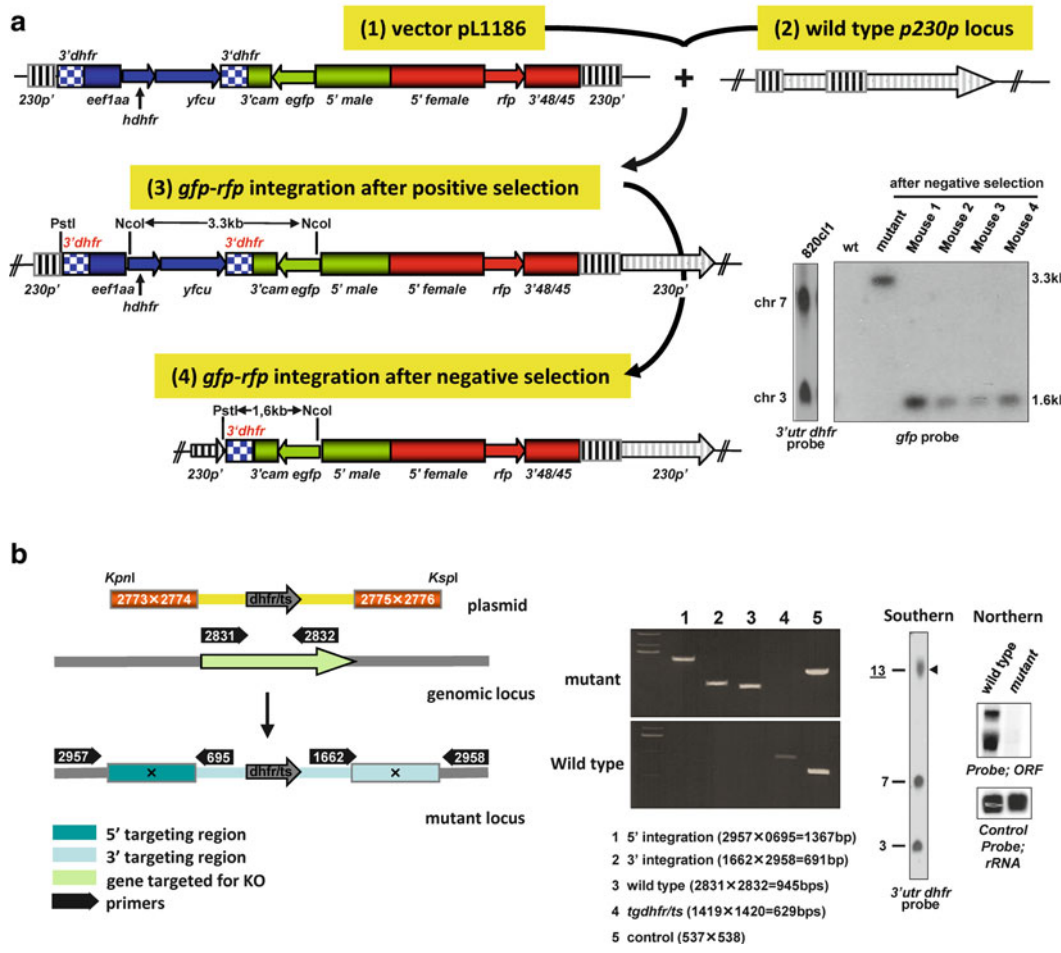

Fig. 2. Graphical representations showing the generation of parasite mutants and analysis of their genotype (adapted from (14)). (**a**) Generation and analysis of a *P. berghei* reference reporter line that stably expresses GFP in male gametocytes and RFP in female gametocytes. Schematic representations of (1) the vector used to introduce the *gfp/rfp* male/female expression cassette into the *p230p* locus, (2) the *p230p* genomic locus, (3) the resulting genomic locus after integration of the construct and after positive selection with pyrimethamine, and (4) the genomic locus after negative selection with 5–fluorocytosine (5FC). Integration of pL1186 into the genome occurs by double crossover homologous recombination resulting in a 1 kb deletion of the nonessential *p230p* gene in parasites selected with pyrimethamine. After negative selection with 5FC, parasites are selected in which the positive/negative selectable marker cassette (*hdfr–yfcu*) has been excised from the integrated construct by a recombination event between the two *3′dhfr* sequences (*chequered boxes*). *Arrows* indicate the position and size of expected restriction site fragments in Southern analysis. Southern analysis of separated chromosomes and restricted DNA shows the correct integration of pL1186 in the *p230p* locus on chromosome 3. In addition Southern analysis of genomic DNA of parasites is shown after positive selection and after negative selection in four mice (m1–m4). Southern analysis of digested DNA shows the presence of the GFP–positive DNA fragment with a reduced size (1.6 kb) in parasites after 5FC treatment (resulting from excision of the selectable marker). (**b**) Schematic representation of the generation and characterization of a *P. berghei* gene disruption mutant. The targeting construct (plasmid) containing the selectable marker gene, *tgdhfr/ts*, is used to disrupt the genomic locus/gene (arrow on genomic locus). Primers used for generating the targeting regions (at the end of the plasmid) and for diagnostic PCR are shown (not to scale). In this example we have chosen a gene deletion mutant generated in a *P. berghei* reporter (*gfp*) background line. PCR analysis shows correct integration of the plasmid into the genomic locus; additional PCR include amplification of the *tgdhfr/ts* gene, the wild-type gene, and a control reaction. Southern analysis of separated chromosomes of the mutant parasite shows correct integration of the targeting plasmid into chromosome 13. Hybridization with a 3′UTR *dhfr/ts* probe results in additional signals in chromosome 7 (endogenous *dhfr/ts* gene) and 3 (the reporter *gfp* gene containing a *dhfr/ts* 3′UTR). Northern analysis of wild-type and mutant parasites shows the absence of transcripts in the mutant line when RNA was hybridized to a probe recognizing the open reading frame (ORF) of the target gene and as a positive control the same RNA was hybridized to total ribosomal RNA (rRNA).

2.1.4. DNA Construct

Required information:

- Schematic representation of the targeting construct and information necessary to understand the mutant genotype (e.g., restriction sites, location of probes and primers) (see Fig. 2).
- Type of construct: episomal or linear for SXO or DXO integration.
- Sequence information for all sequences/primers used to disrupt or mutate the target gene. Primer information must include the amplified region and the type of genetic event intended in the parasite genome, e.g., disruption or mutation of a target gene or introduction of a transgene (see Note 2).
- Selectable marker cassette: drug selectable marker, e.g., the pyrimethamine-resistant *tgdhfr/ts* gene, and promoter and terminator sequences of the marker.

2.1.5. Mutated Gene

Required information:

- Schematic representation of the gene and information to understand the mutant genotype (e.g., restriction enzyme sites, location of probes, sizes of restriction fragments in Southern analysis, PCR primers to confirm the correct integration of the DNA construct).

2.1.6. Mutation: Procedure for Gene Disruption

Required information:

- Type of disruption: complete or partial gene deletion, i.e., removal of all or some of the open reading frame (ORF), using a construct that targeted the gene by either SXO or DXO integration (see Note 3).

Additional information:

- In case of partial ORF disruption: Northern and/or Western analysis should prove the absence of, or reduced, gene expression.
- In case of disruption by SXO: genotype and phenotype analyses should show an absence of wild-type parasites.

2.1.7. Mutation: Procedure for Gene Mutation or Gene Tagging

Required information:

- Type of mutation or tagging: integration of a construct through either DXO or SXO homologous recombination.

2.1.8. Mutation: Procedure for Introducing a Transgene

Required information:

- Location of the transgene, i.e., maintained as an episome or integrated into the genome by DXO or SXO homologous recombination.

Additional information:

- If the transgene is introduced using a *Plasmodium* artificial chromosome (PAC), which can be retained stably in the parasite without drug selection (5): Details of the PAC construct and initial selection should be described (see Note 4).

2.1.9. Selection of Mutants

Required information:

– Procedure to select the mutants from the transfected, parent population: Drug selection or flow-sorting (FACS).

2.1.10. Cloning of Mutants

Required information:

– Procedure to clone the mutant: limiting dilution of blood stages or otherwise (e.g., via merosomes).

Additional information:

– If mutants are not cloned, genotype and phenotype analyses should show the absence of wild-type parasites (see Note 5).

2.1.11. Genotype Analysis

Required information:

– PCR and Southern data (on digested genomic DNA or separated chromosomes) showing the correct disruption, mutation, or tagging event (see Fig. 2).
– In case of gene disruption: transcriptional and/or protein expression evidence of gene knockout.
– In case of gene tagging: evidence of tagged-protein expression.
– In case of transgene introduction: evidence of transgene expression and accordance with promoter activity.

Additional information:

– If the procedure for introducing the mutation was complex involving multiple steps (e.g., using negative selection to remove drug selectable markers), a Southern analysis of the targeted locus is necessary (see Fig. 2a).

2.1.12. Phenotype Analysis (See Note 6)

As of yet no guidelines have been agreed upon or exist on standardized phenotype analysis in *Plasmodium* research that might facilitate a uniform reporting of mutant phenotypes (see Note 7).

2.2. Submitting (Additional) Information on Published Mutants, Unpublished Mutants Without a Clear Phenotype, or Negative Attempts to Disrupt/Mutate Genes to RMgmDB

The RMgmDB is intended to be a dynamic and responsive resource, and researchers are encouraged to update the information on phenotypes of the mutants they have generated and to cross-reference their mutants to other relevant mutants that are available in the database. Importantly, much unpublished data exist on *Plasmodium* mutants that were generated but did not show a clear phenotype or where experimenters attempted on multiple occasions to disrupt a parasite gene but were unsuccessful. This information, while difficult to publish, is of significant value and can also prevent an unnecessary duplication of effort. Moreover, the existence of null-mutants without a distinct phenotype might provide information about the functional redundancy of the target gene (see Note 8). Similarly, the lack of an observable phenotype might also be the result of assays that are currently inadequate or, as yet, too insensitive to reveal a phenotypic effect of the genetic modification. Further

analysis of such mutants in improved phenotype assays might reveal novel aspects of gene function. Below, a template is provided for submitting information to RMgmDB on unpublished mutants and negative trials to disrupt or mutate genes.

2.2.1. Submission of (Additional) Information on Mutants Already Available in RMgmDB

– For all mutants a "comment box" is available where additional information can be submitted: corrections, comments, or suggestions for improving the description of the mutants or associations to other, relevant mutants in the database.

2.2.2. Submitting Information on Unpublished Mutants Without a Clear Phenotype

Information on new mutants without a clear phenotype can be submitted in spreadsheet (Excel) format. This must include a minimal set of information on the constructs used and the mutant as shown in Table 1 (left-hand column).

Table 1
Information required for publication in RMgmDB of either unpublished mutants without a clear phenotype (left-hand column) or unsuccessful attempts to disrupt/mutate genes (right-hand column)

Mutants without a clear phenotype	Unsuccessful attempts to disrupt/mutate genes
– Rodent parasite species	– Rodent parasite species
– Parent strain/isolate/line	– Parent strain/isolate/line
– Name PI/Researcher	– Name PI/Researcher
– Name Group/Department	– Name Group/Department
– Name Institute	– Name Institute
– Gene Model of Rodent Parasite (as cited in GeneDB)	– Gene Model of Rodent Parasite (as cited in GeneDB)
– Plasmid/construct—double or single crossover	– Plasmid/construct—double or single crossover
– Partial or complete disruption ORF	– Partial or complete disruption ORF
– Other details of mutation, if applicable	– Other details of mutation, if applicable
– Selection method (drug treatment, FACS, etc.)	– Selection method (drug treatment, FACS, etc.)
– Selectable marker	– Selectable marker
– Promoter selectable marker	– Promoter selectable marker
– Drug used to select	– Drug used to select
– Primer information (= sequence and name) Primer sequence target region 1a Primer sequence target region 1b Primer sequence target region 2a Primer sequence target region 2b	– Primer information (= sequence and name) Primer sequence target region 1a Primer sequence target region 1b Primer sequence target region 2a Primer sequence target region 2b
– Mutant cloned? (yes or no)	– Not applicable
– Not applicable	– Number of transfection attempts

2.2.3. Submitting Information to RMgmDB on Negative Attempts to Disrupt or Mutate Plasmodium Genes

Information on negative attempts to disrupt or mutate genes can be submitted in spreadsheet (Excel) format. This must include a minimal set of information on the constructs generated and number of attempts to disrupt the gene of interest as shown in Table 1 (right-hand column).

3. Notes

1. Genes heterologous to the modified genome, either from other organisms (e.g., GFP, luciferase) or from other *Plasmodium* species, can be episomally expressed or stably integrated into the genome at a "silent locus" (a locus that when disrupted does not alter the wild-type phenotype of the parasite). PLEASE NOTE: If genes from other organisms or *Plasmodium* species replace their homolog/ortholog in the rodent parasite genome, they are classified as "mutated genes."

   Some *Plasmodium* genes have been targeted by conditional mutagenesis (e.g., at distinct parasite life cycle stages) to remove parasite genes using the yeast Flp/FRT recombination system (6). These genes are classified in RMgmDB as "Mutated genes" and not as "Disrupted genes." The Flp/FRT site-specific recombination system works by a recombinase (Flipase, Flp)-mediated recognition and excision between two identically orientated 34-nucleotide *FRT* ("Flirted") sites, these sites being previously introduced into a contiguous region of the parasite genome (referred to as the FRTed sequence). In *Plasmodium* these genetic modifications have been achieved by introducing two flanking *FRT* ("FliRTed") around the gene of interest and gene removal then occurs only after expression of the yeast Flp recombinase, which is regulated from a stage-specific promoter, that then excises parasite DNA sequences contained between the *FRT* sites (6, 7).

   RMgmDB contains entries of "mutants" created by random insertional mutagenesis (i.e., *piggyBac* transposon) into the *P. berghei* genome. It has been shown that the transposase *piggyBac* integrates randomly into the *Plasmodium* genome (8–10). Random *piggyBac* - mediated insertion of plasmid DNA into *P. berghei* genome has been achieved in two ways, described in (8). Analysis of insertions of these DNA elements revealed a large number of genes that had been disrupted, suggesting that these genes were amenable to modification and, possibly, silencing. The only *piggyBac* insertions that have been included in RMgmDB were those that were found in the ORF/CDS, 5′UTR (0–500 bp from the start codon), or within introns of genes, where the insertion was expected to disrupt gene expression. The presence of an insert may there-

fore provide indirect evidence that the gene is not essential for blood-stage development. We have thus included this information in RMgmDB. The *piggyBac* inserts were identified using a TAIL-PCR approach from mixed populations of transfected parasites and the individual mutants have not been selected by cloning. Therefore, neither the expression nor the phenotype of the mutant containing the insert has been analyzed.

2. If multiple, near-identical constructs have been generated that differ only in the target gene or have different transgenes, describe one construct as above and proceed with the other constructs by describing only those regions (i.e., primers, gene models, etc.) that differ between them; a general schematics is often useful in these cases.
3. In general, gene deletion/replacement by DXO integration is preferred over gene disruption by SXO integration. With SXO integration of the construct, the possibility exists of reversion to a wild-type genotype through a recombination event that removes the integrated construct and restores the gene locus to its original configuration (1).
4. When introducing transgenes on episomes, loss of episomes occurs during propagation of the blood stages if continuous drug selection is not maintained (5, 11, 12). Drug selection/pressure is less easy to apply during mosquito/liver propagation of the transgenic parasites, resulting in the loss of the transgene-containing episomes (5).
5. Where episomal transfection has been performed, e.g., during promoter expression studies or protein tagging, cloning is usually not performed. However, drug selection needs to be maintained in order to retain plasmids in the parasites.
6. Confirmation of the mutant phenotype: In the case of gene disruption or mutation, confirmation of the mutant phenotype should include either (1) confirmation through the analysis of a second, independently derived mutant or (2) genetic complementation of the mutant such that the wild-type copy of the gene (and wild-type phenotype) is restored. Only (2) demonstrates that the phenotype is the result of the intended mutation and is not due to unrelated mutations or alterations in the parasite genome (13). While complementation is desirable, such experiments in rodent malaria parasites are often difficult to perform because of the limited set of selectable markers available. This is especially apparent when a reported phenotype is based on a set of different mutants or is the consequence of a set of successive transfections performed in a single parasite (e.g., deletions of more than one gene). In this situation, an alternative requirement is that a mutant phenotype is confirmed through the analysis of at least a second, independently

derived, and genetically identical mutant. Up to now, most studies have used this strategy and analyzed phenotypes of two independent mutants, and to our knowledge it has not been proven that an incorrect phenotype has been reported when this analytical strategy was applied. However, a few published studies have reported phenotypes based on studies of only a single mutant and with no complementation studies. When dealing with such mutants, or mutants where there might be insufficient reported information on the genotype or phenotype in RMgmDB, it is indicated that the analysis of these RMgm entries is incomplete.

7. Standardization of analysis and description of phenotypes: No guidelines exist in *Plasmodium* research for the use of standardized assays to analyze mutant phenotypes or standardized vocabularies for describing mutant phenotypes. The lack of standardized vocabularies for phenotypes and gene functions limits searches of the database, for example searching for mutants with a comparable phenotype or for genes with a similar cellular location or function. Moreover, it reduces possibilities of integrating information on gene function and protein location from RMgmDB with gene data from other resources and therefore limits the use of this information with larger scale analyses of gene function. Only through the use of defined vocabularies to describe gene function and protein location will databases, such as RMgmDB, be able to easily facilitate both the cross-linking and mining of these different datasets (expression, metabolic, experimental, etc.), which then would allow for a fully integrative "systems biology" program in *Plasmodium*. Such approaches will help to better understand not only gene function and gene regulatory networks in *Plasmodium* but also their involvement in infection and disease.
8. Although the failure to disrupt or mutate a gene does not by itself prove that the target gene is not amenable to genetic modification, it does provide indirect evidence that the gene might be essential for blood-stage development. *Plasmodium* has a haploid genome (during blood-stage development); therefore disruption of an essential gene results in parasites that are incapable of completing development in erythrocytes. However, other reasons for failing to disrupt a gene may include the following: (1) the genomic locus is not "accessible" for the genetic modification or (2) the genetic modification results in (unintentional) disruption or down-regulation of an additional gene which for example can occur if the gene of interest is very close to a neighboring gene or when gene annotation is incorrect and results in the targeting of multiple genes. To address accessibility of the target gene, a modification of

the target gene using a "tagging-construct" can be used to determine if the locus is amenable to disruption. Successful tagging of a gene in combination with the failure to disrupt the gene in two or three independent transfection experiments provides a strong indication that the gene of interest is essential during asexual blood-stage development. Since unsuccessful attempts to disrupt a gene may be due to incorrect gene annotation, it is essential that for all unsuccessful attempts the sequence of primers (used to amplify the target regions of the gene) is provided.

Acknowledgments

We would like to thank Gunnar Mair for the use of the schematics shown in Fig. 2, and Senad Mededovic and Auke van Wigcheren for building and maintaining RMgmDB. The work was supported by European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement N° 242095. We would like to thank members of PlasmoDB and GeneDB (Welcome Trust Sanger Institute) for their input and support of RMgmDB.

References

1. Carvalho TG, Menard R (2005) Manipulating the *Plasmodium* genome. Curr Issues Mol Biol 7:39–55
2. Gardiner DL et al (2003) Malaria transfection and transfection vectors. Trends Parasitol 19:381–383
3. Balu B, Adams JH (2007) Advancements in transfection technologies for *Plasmodium*. Int J Parasitol 37:1–10
4. Janse CJ et al (2011) A genotype and phenotype database of genetically modified malaria parasites. Trends Parasitol 27:31–39
5. Iwanaga S et al (2010) Functional identification of the *Plasmodium* centromere and generation of a *Plasmodium* artificial chromosome. Cell Host Microbe 7:245–255
6. Combe A et al (2009) Clonal conditional mutagenesis in malaria parasites. Cell Host Microbe 5:386–396
7. Falae A et al (2010) Role of *Plasmodium berghei* cGMP-dependent protein kinase in late liver stage development. J Biol Chem 285: 3282–3288
8. Fonager J et al (2011) Development of the *piggyBac* transposable system for *Plasmodium berghei* and its application for random mutagenesis in malaria parasites. BMC Genomics 12:155
9. Balu B et al (2010) A genetic screen for attenuated growth identifies genes crucial for intraerythrocytic development of *Plasmodium falciparum*. PLoS One 5:e13282
10. Balu B et al (2005) High-efficiency transformation of *Plasmodium falciparum* by the lepidopteran transposable element piggyBac. Proc Natl Acad Sci USA 8:16391–16396
11. O'Donnell RA et al (2001) An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected *Plasmodium falciparum* parasites. Nucleic Acids Res 29:716–724
12. van Dijk MR et al (1997) Replication, expression and segregation of plasmid-borne DNA in genetically transformed malaria parasites. Mol Biochem Parasitol 86:155–162
13. Goldberg DE et al (2011) Has the time come for us to complement our malaria parasites? Trends Parasitol 27:1–2
14. Mair GR et al (2010) Universal features of post-transcriptional gene regulation are critical for *Plasmodium* zygote development. PLoS Pathog 6:e1000767