



# Evolutionary genetics of malaria

Kristan Alexander Schneider<sup>1\*</sup> and Carola Janette Salas<sup>2</sup>

<sup>1</sup>Department of Applied Computer- and Biosciences, University of Applied Sciences Mittweida, Mittweida, Germany, <sup>2</sup>Department of Parasitology, U.S. Naval Medical Research Unit No 6 (NAMRU-6), Lima, Peru

## OPEN ACCESS

### EDITED BY

Rongling Wu,  
The Pennsylvania State University (PSU),  
United States

### REVIEWED BY

Chenqi Wang,  
University of South Florida,  
United States  
Edith Christiane Bougouma,  
Groupe de Recherche Action en Santé  
(GRAS), Burkina Faso

### \*CORRESPONDENCE

Kristan Alexander Schneider,  
kristan.schneider@hs-mittweida.de

### SPECIALTY SECTION

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 29 August 2022

ACCEPTED 26 September 2022

PUBLISHED 03 November 2022

### CITATION

Schneider KA and Salas CJ (2022),  
Evolutionary genetics of malaria.  
*Front. Genet.* 13:1030463.  
doi: 10.3389/fgene.2022.1030463

### COPYRIGHT

© 2022 Schneider and Salas. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Many standard-textbook population-genetic results apply to a wide range of species. Sometimes, however, population-genetic models and principles need to be tailored to a particular species. This is particularly true for malaria, which next to tuberculosis and HIV/AIDS ranks among the economically most relevant infectious diseases. Importantly, malaria is not one disease—five human-pathogenic species of *Plasmodium* exist. *P. falciparum* is not only the most severe form of human malaria, but it also causes the majority of infections. The second most relevant species, *P. vivax*, is already considered a neglected disease in several endemic areas. All human-pathogenic species have distinct characteristics that are not only crucial for control and eradication efforts, but also for the population-genetics of the disease. This is particularly true in the context of selection. Namely, fitness is determined by so-called fitness components, which are determined by the parasites live-history, which differs between malaria species. The presence of hypnozoites, i.e., dormant liver-stage parasites, which can cause disease relapses, is a distinct feature of *P. vivax* and *P. ovale* sp. In *P. malariae* inactivated blood-stage parasites can cause a recrudescence years after the infection was clinically cured. To properly describe population-genetic processes, such as the spread of anti-malarial drug resistance, these features must be accounted for appropriately. Here, we introduce and extend a population-genetic framework for the evolutionary dynamics of malaria, which applies to all human-pathogenic malaria species. The model focuses on, but is not limited to, the spread of drug resistance. The framework elucidates how the presence of dormant liver stage or inactivated blood stage parasites that act like seed banks delay evolutionary processes. It is shown that, contrary to standard population-genetic theory, the process of selection and recombination cannot be decoupled in malaria. Furthermore, we discuss the connection between haplotype frequencies, haplotype prevalence, transmission dynamics, and relapses or recrudescence in malaria.

### KEYWORDS

complexity of infection (COI), co-infection, mixed-species infection, recrudescence, relapse, seed bank, hypnozoites, multiplicity of infection (MOI)

## 1 Introduction

After a decade of declining incidence the number of annual malaria infections rises since 2018, challenging the WHO goal to reduce malaria incidence by at least 90% by 2030 (WHO, 2021a). This is partly attributed to the rapid emergence and spread of anti-malarial drug resistance, an evolutionary-genetic process whose understanding is a global health priority (WHO, 2021b).

Malaria is caused in humans and animals by *Plasmodium* parasites. These unicellular, haploid eukaryotes are transmitted by numerous species of female *Anopheles* mosquitoes. Both the parasite and vector species are adapted to specific human or animal hosts. Five species of *Plasmodium* are pathogenic to humans, which can be transmitted by over 100 *Anopheles* species (Nicoletti, 2020). Over 95% of the 240 million annual infections and 620,000 deaths worldwide are attributed to *P. falciparum*. Although, the WHO recommended the use of RTS,S, the first approved malaria vaccine, in children to prevent *P. falciparum* infections in areas of moderate to high transmission, the vaccine's efficacy is low and malaria control depends strongly on reliable diagnostics and drug treatments to cure acute infections (Greenwood et al., 2021). While the second most relevant species, *Plasmodium vivax*, receives considerable attention, the other species *P. ovale* sp., *P. malariae*, and *P. knowlesi* are somewhat neglected, due to an outdated distinction between harmful and harmless malaria species (Lover et al., 2018).

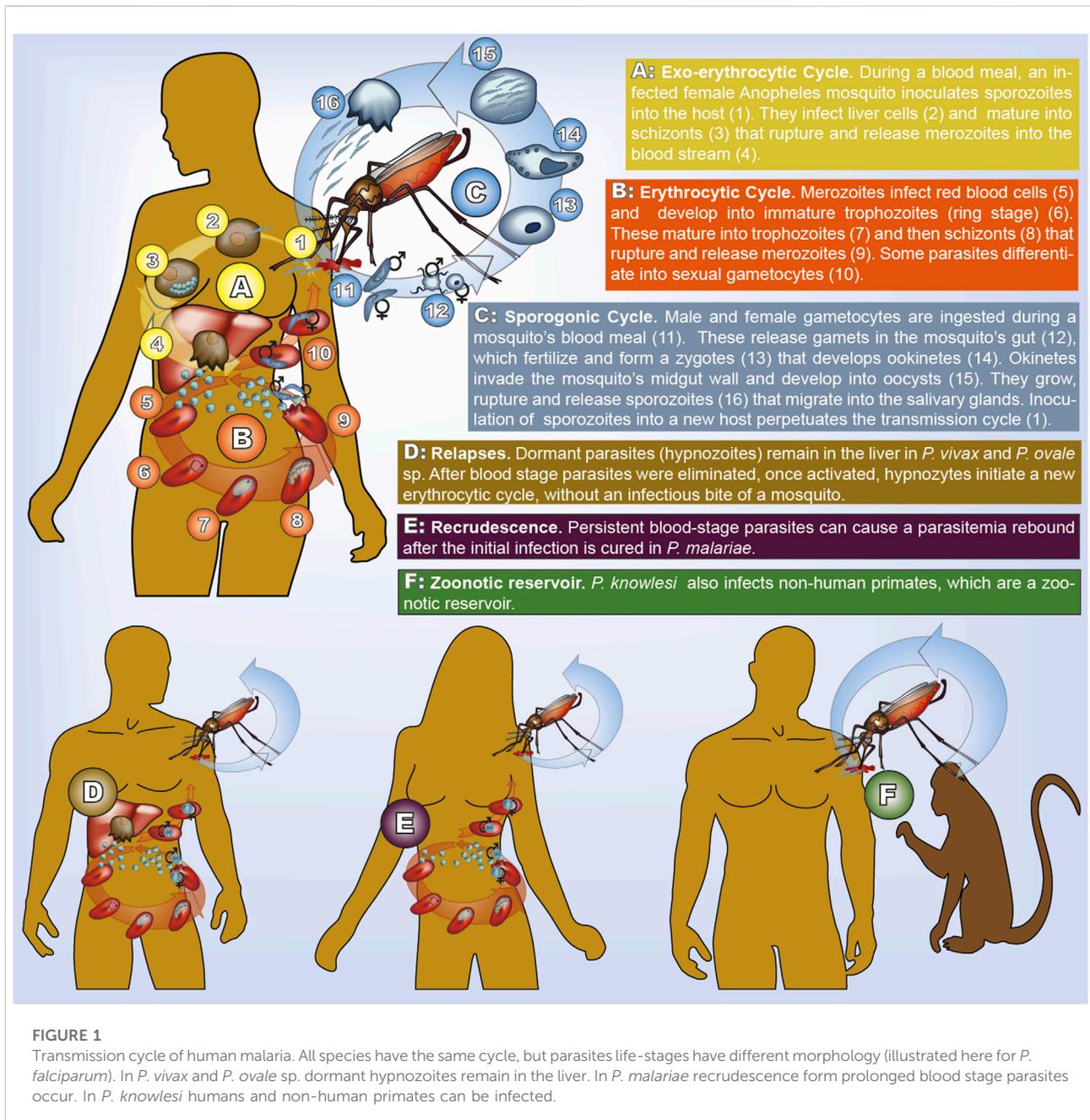
The spread of deletions in the histidine-rich protein 2 and 3 (HRP2/3) genes of *P. falciparum*, which encode for the antigens targeted by rapid diagnostic tests (RDTs) as well as drug-resistant *P. falciparum* and *P. vivax* haplotypes substantially challenge successful malaria control. These evolutionary genetic processes are tightly linked to the pathogen's complex transmission cycle, which besides some species-specific differences, is commonly shared among all *Plasmodia* (Su et al., 2019; Beshir et al., 2022).

The transmission cycle starts with an infected mosquito taking her blood meal. She inoculates parasites in the form of sporozoites from her salivary glands into the human body. This is followed by the exo-erythrocytic cycle, during which sporozoites reach the liver to infect hepatocytes. In the infected liver cells parasites mature into schizonts. The erythrocytic cycle is initiated when the schizonts rupture and merozoites are released into the bloodstream. Erythrocytes are invaded by merozoites, which form ring stage trophozoites and then mature into schizonts. Once they rupture, new merozoites are released into the bloodstream. During this step of asexual reproduction, some parasites differentiate into male or female gametocytes, which do not reproduce in the human host. Once a mosquito ingests male and female gametocytes, the sporogonic cycle is initiated. Gametes released by male and female gametocytes fertilize and form zygotes. Following a step of meiosis, and hence recombination, the zygote becomes

tetraploid and develops into ookinetes, which migrate through the midgut wall and transform into oocysts. In the oocyst sporozoite budding occurs in the haploid state. Division of each oocyst produces thousands of sporozoites that move into the mosquito salivary glands, completing the transmission cycle. Because gametocytes immediately release gametes, only parasites exiting the same host recombine, potentially leading to a high degree of inbreeding during the sexual reproduction of the parasite (Ngwa et al., 2016).

Species-specific differences occur in the number of parasites within an infection (parasitemia and gametocytemia counts), and the duration of the various phases in the transmission cycle. The replication of merozoites in 72-hour- rather than 48-hour-cycles distinguishes *P. ovale* sp. from other species. The onset of gametocytogenesis and the longevity of gametocytes were argued to accelerate drug-resistance evolution in *P. falciparum* compared to *P. vivax* (Schneider and Escalante, 2013). Dormant liver-stage parasites (hypnozoites), can result in disease relapses weeks, months, or even years after the clearance of blood stage parasites and occur only in *P. vivax* and *P. ovale* sp. Currently primaquine (PQ) and tafenoquine (TQ) are the only approved drugs to clear hypnozoites (Watson et al., 2021). Unfortunately, patients with (glucose-6 phosphate dehydrogenase) G6PD deficiency, which is widespread in many malaria-endemic areas, cannot be treated with these drugs (Baird et al., 2018; Dean et al., 2020). Extremely prolonged carriage of blood-stage parasites causing recrudescences occur in *P. malariae* (Collins and Jeffery, 2007). It is commonly accepted, although not completely ruled out, that the rebound of parasitaemia in *P. malariae* is not caused by quiescent pre-erythrocytic stages such as hypnozoites. Because of relapses occurring in *P. vivax*, *P. ovale* sp., and prolonged blood stage parasite carriage in *P. malariae*, these species are resilient in areas in which *P. falciparum* transmission cannot be sustained. While all other human malaria species can—at least in theory—be eradicated by concentrating on the human host, this is not possible for *P. knowlesi*, which is characterized by zoonotic transmission. It became the predominant species in several endemic countries in Southeast Asia, which shifted from malaria control toward elimination (Sutherland, 2016).

The characteristics of the transmission cycle render the application of standard textbook population-genetic results incorrect. Particularly it was shown that the process of selection acting on parasites in the human hosts (including selection for drug resistance) and recombination cannot be separated (Schneider and Kim, 2010). Hence, population-genetic theory and models have to be tailored to the malaria transmission cycle. This has been done mainly for *P. falciparum*. Because a clear path to eradication has been chartered only for *P. falciparum*, the other malaria species gain more importance due to their resilient nature (Lover et al., 2018). This requires to further adapt population-genetic theory to the characteristics of other human-pathogenic malaria species.



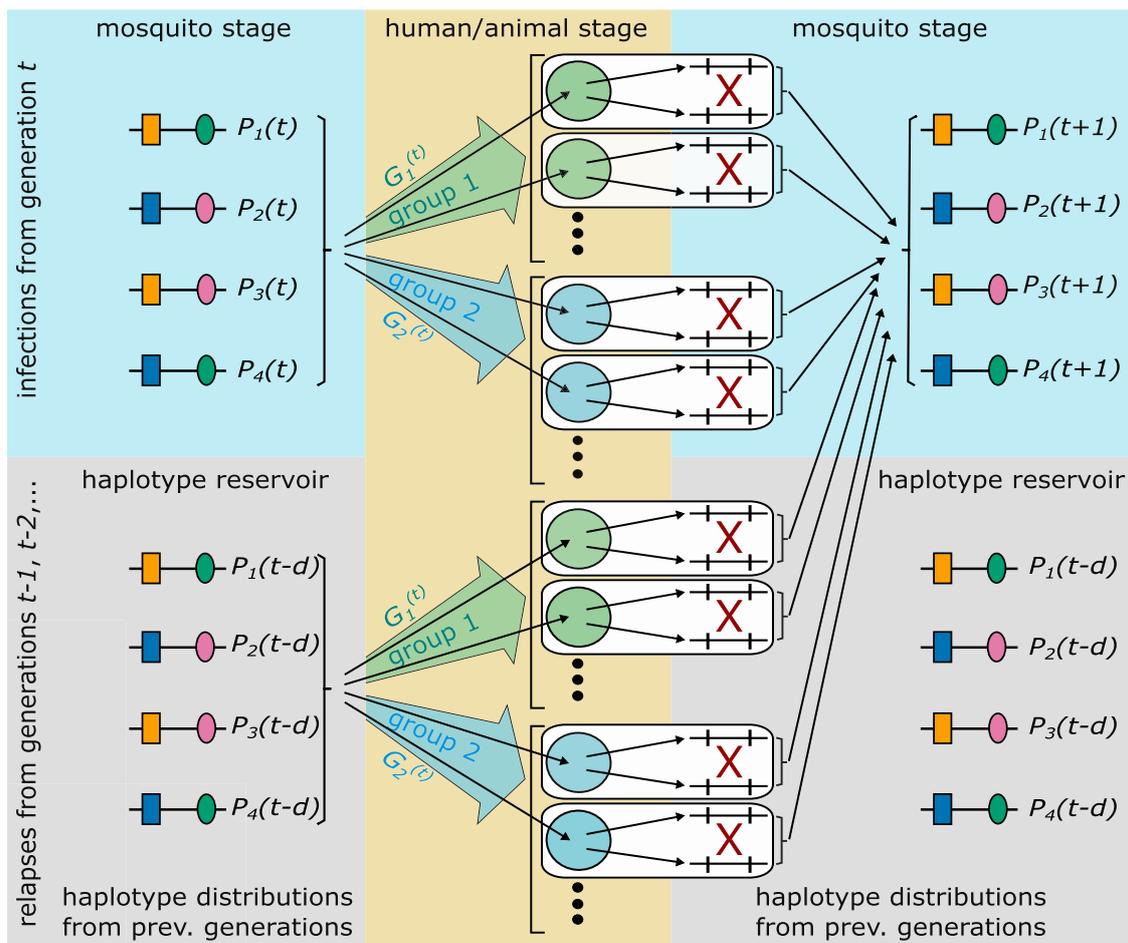
Here, we extend a population-genetic framework, originally developed for *P. falciparum*, to be applicable to all other malaria species.

We exemplify the importance of species-specific differences by clarifying the role of hypnozoites in the evolution of drug resistance in *P. vivax* vs. *P. falciparum*. We also clarify, how haplotype frequencies (i.e., their relative abundance in the parasite population) and prevalence (i.e., the likelihood that a given haplotype occurs in an infection) are affected by relapses/recrudescence in other malaria species. Based on this framework,

we discuss past and current developments with relevance for the evolutionary genetics of malaria.

## 2 Methods

We extend the population-genetic framework of (Schneider and Kim, 2010; Schneider and Kim, 2011; Schneider, 2021) that describes the temporal change in the distribution of parasite haplotypes due to recombination and selection in generations of



**FIGURE 2**  
 Illustrated is the idealization of the malaria transmission cycle underlying the population-genetic framework. The illustrated genetic architecture of malaria haplotypes assumes two biallelic loci, leading to four possible haplotypes. Furthermore, two groups of hosts are illustrated. Each host is infected by randomly drawing haplotypes from generation  $t$ , or a relapse/recrudescence from a previous generation occurs, which corresponds to randomly draw parasites from a previous generation (haplotype reservoir). With probability  $G_g^{(t)}$  a host belongs to group  $g$  in generation  $t$ . The selective environment is different in the two groups. Recombination occurs exclusively between haplotypes exiting the same host. After recombination, haplotypes in the mosquitoes are pooled together to derive their distribution in generation  $t + 1$ .

transmission cycles. While the original framework was tailored to *P. falciparum*, the extension captures the characteristics of all human-pathogenic malaria species.

The model is based on an idealization of the complex malaria transmission cycle (cf. Figure 1), which is illustrated in Figure 2. Although, pathogen, mosquito vector, human hosts (and, in the case of *P. knowlesi* the animal host) are involved in transmission, the framework does not require to model transmission dynamics (i.e., the interaction of mosquito vectors and human or animal hosts) explicitly. This conceptual advantages arise, because haplotype frequencies are considered at the end of the sporogonic cycle (cf. Figure 2). Thus, the frequency distribution of parasite haplotypes in the mosquitoes' salivary glands, which are ready for vector-host transmission, is followed.

Host and vector populations are assumed to be sufficiently large and malaria infections sufficiently frequent to justify a deterministic description of the evolutionary dynamics. Steps of full transmission cycles correspond to steps of sexual reproduction, because only one step of sexual reproduction occurs during one full transmission cycle, namely inside the mosquito vector. Many steps of asexual reproduction occur inside the vectors and hosts.

### 2.1 Genetic architecture of haplotypes

The genetic architecture of haplotypes is determined by their allelic configuration at one or several loci. We denote the number

of all possible haplotypes by  $H$ . E.g.,  $L$  biallelic loci lead to  $H = 2^L$  haplotypes. In general, if haplotypes are determined by  $L$  loci, and  $n_l$  alleles are segregating at locus  $l$ ,  $H = n_1 \cdot n_2 \cdot \dots \cdot n_L$ . The frequency of haplotype  $h$  in generation  $t$  is denoted by  $P_h^{(t)}$ . Collectively, the vector of haplotype frequencies is  $\mathbf{P}_t = (P_1^{(t)}, \dots, P_H^{(t)})$ .

## 2.2 Idealizing the transmission cycle

The idealized transmission cycles allows to describe the evolutionary genetics of malaria in generations of full transmission cycles (Figure 2). In generation  $t$ , it is assumed that all hosts are infected (or have a relapse or recrudescence) at the same time. Moreover, host-vector transmission is also synchronized. Inside the mosquito, parasites, which were ingested by the mosquitoes, can recombine during one step of sexual reproduction. This determines the distribution of haplotypes in the mosquitoes' salivary glands of the parasite (sporozoite) population in generation  $t + 1$ .

### 2.2.1 Heterogeneity

Disease exposure and transmission intensities are heterogeneous in endemic areas and change over time (e.g. in the context of seasonal transmission) (Bousema et al., 2011; Selvaraj et al., 2018). Moreover, hosts are heterogeneous regarding their level of genetic and naturally acquired immunity, number of co-morbidities, or the drug treatment they receive to cure the infection (in case they receive any), etc. (Hedrick, 2011; Gonzales et al., 2020). All of these factors can be addressed by modeling hosts in different groups (strata). Let  $G_g^{(t)}$  be the probability that a host, in which an infection occurs in generation  $t$ , belongs to group  $g$ . Hence,  $G_1^{(t)} + \dots + G_S^{(t)} = 1$  for every generation  $t$ .

The number of groups,  $S$ , has to be chosen to capture the features important to the specific application of the framework. For instance, when considering drug resistance evolution, a simple distinction would be between treated and untreated infections, i.e.,  $S = 2$ . In the case of *P. knowlesi* different groups can model human and animal hosts. In the simplest case one would have just two groups ( $S = 2$ ), namely humans and animals.

### 2.2.2 Relapses and recrudescence

Hosts are not modelled explicitly. This becomes relevant when considering relapses (in *P. vivax* and *P. ovale* sp.) and recrudescence in *P. malariae*. In the following we use relapse and recrudescence synonymously, unless a distinction is necessary.

In the idealized transmission cycle, a relapse in generation  $t$ , which occurs after a delay of  $d$  generations, is equivalent to a new infection from the sporozoite population from  $d$  generations in the past, i.e., from generation  $t - d$ . Let  $R_d^{(t)}$  be the probability that an infection in generation  $t$  is a relapse, with a delay of  $d$

generations, where  $R_0^{(t)}$  is the probability of a new infection at time  $t$ . Assuming the maximum possible delay is  $D$ , the relation  $\sum_{d=0}^D R_d^{(t)} = 1$  for all  $t$ , and  $1 - R_0^{(t)}$  is the probability that a relapse occurs at time  $t$ .

The framework models the haplotype distribution in generations of transmission cycles not in real-time. The higher the transmission intensities, the more transmission cycles occur per year. The choice of the distribution of relapses has to take this into account (see Results section The effect of recrudescences and relapses). Moreover, the timing of relapses depends on the *Plasmodium* species (White, 2011).

Importantly, a host might have been exposed differently to the disease in the past, i.e., the host might belong to different groups in generations  $t - d$  and  $t$ . Let  $G_{g',g}^{(t-d,t)}$  be the probability that a host, who belonged to group  $g'$  in generation  $t - d$ , belongs to group  $g$  in generation  $t$  ( $d \geq 0$ ). Marginalisation yields

$$G_g^{(t)} = \sum_{g'=1}^S G_{g',g}^{(t-d,t)} \quad (1)$$

for all  $t, d, g$ . Hence, the probability that a relapse occurs in generation  $t$  in a host in group  $g$  after a delay of  $d$  generations, when he belonged to group  $g'$ , is given by

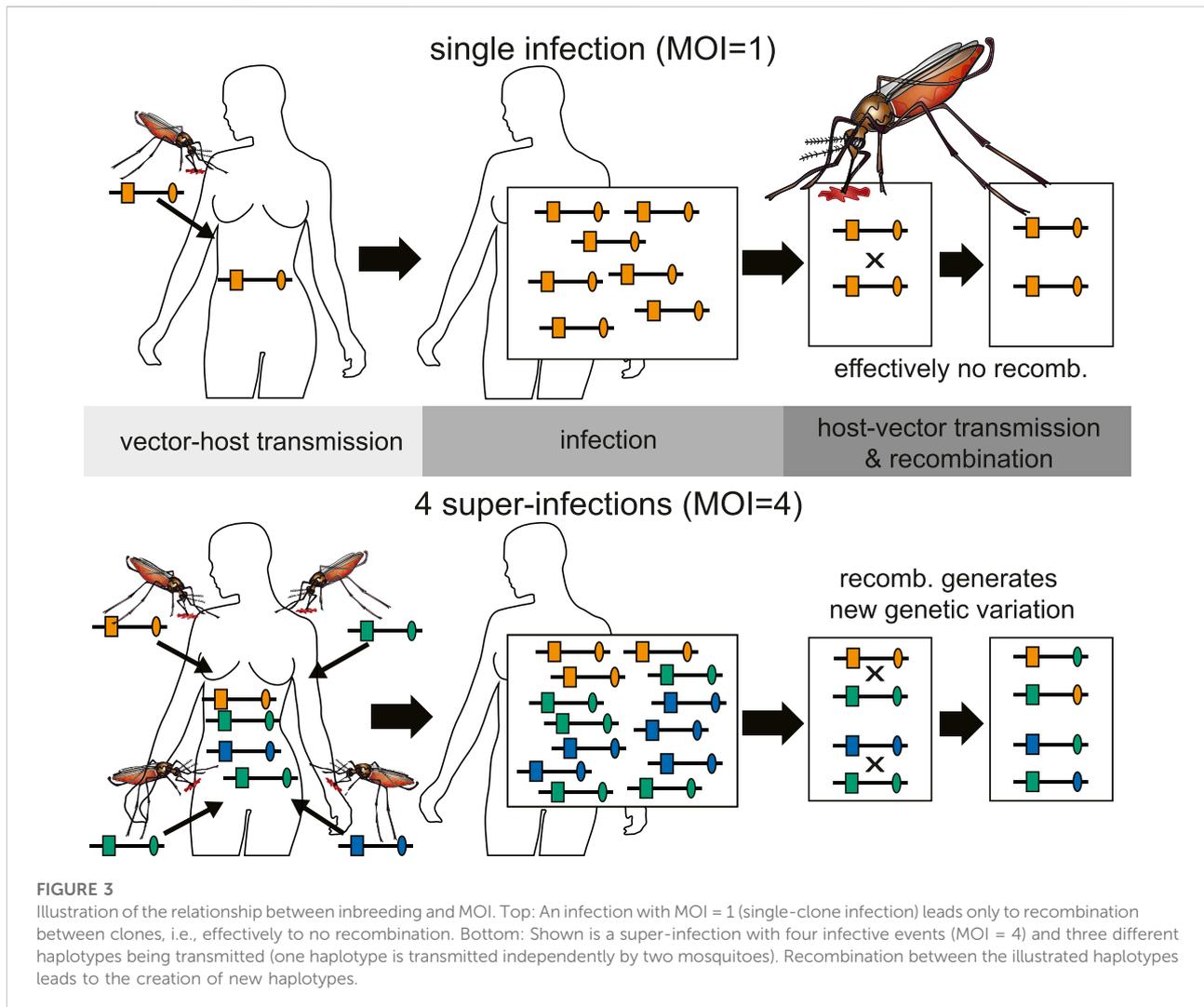
$$R_d^{(t)} G_{g',g}^{(t-d,t)}.$$

## 2.3 Vector-host transmission and multiplicity of infection

The presence of multiple genetically distinct parasite haplotypes within an infection is frequently referred to as multiplicity of infection (MOI) or complexity of infections (COI) and considered important in malaria. The terms MOI and COI are ambiguously defined in the literature (see (Schneider et al., 2022) for a comprehensive review). Although, it is unclear whether MOI is affecting the clinical pathogenesis of malaria, or whether different parasite haplotypes are competing within infections (intra-host competition), MOI mediates the amount of meiotic recombination and scales with transmission intensities (Pacheco et al., 2020) (see Figure 3).

Different parasite haplotypes can occur within an infection, because they are 1) sequentially transmitted (during the course of one disease episode) by different mosquitoes (super-infection); 2) co-transmitted by one mosquito (co-infection); 3) mixed up with parasites from previous infections by relapses or recrudescence.

Concerning models of MOI, the focus was mainly on super-infections. More recently, the importance of co-infections is being emphasized. Namely, more parasite genomics data is being generated, which has enough resolution to study genetic relatedness of parasites. Such data is appropriate for molecular surveillance of transmission routes (Ndiaye et al., 2021). Formal population-genetic frameworks to describe the evolutionary



genetics of malaria that consider relapses do not exist. Mathematical models describing relapses in *P. vivax* and *P. ovale* sp. are limited to epidemiological models, e.g., the compartmental model of (Chamchod and Beier, 2013), which neglects parasite genetics. A population-genetic framework applicable to all human-pathogenic malaria species has to be flexible enough to accommodate super-infections, co-infections, relapses, and recrudescence.

To set up the framework an infection is identified by a vector  $\mathbf{m} = (m_1, \dots, m_H)$ , where  $m_h$  is the number of times haplotype  $h$  is infecting. Hence,  $m_h = 0$  or  $m_h > 0$  if haplotype  $h$  is absent or present in the infection, respectively. The number  $m_h$  accounts for super-infections with the same haplotype. Moreover, it can be interpreted as the “concentration” of haplotype  $h$  if several haplotypes are co-infecting, etc.

Let  $\Pr[\mathbf{m}|t]$  be the probability of an infection with configuration  $\mathbf{m}$  given generation  $t$ . The infection might be a new infection or a relapse. The probability of infection  $\mathbf{m}$ , given it

occurs in generation  $t$ , when the host belongs to group  $g$ , and given it is a relapse with a delay of  $d$  generations, when the host belonged to group  $g'$ , is denoted by  $\Pr[\mathbf{m}|t-d, g'; t, g]$ . Hence, the probability of infection  $\mathbf{m}$  occurring in a host in group  $g$  in generation  $t$ , which is a relapse from generation  $t-d$ , when the host belonged to group  $g'$ , is

$$\Pr[\mathbf{m}; t-d, g'; t, g] = \Pr[\mathbf{m}|t-d, g'; t, g] R_d^{(t)} G_{g',g}^{(t-d,t)}. \quad (2)$$

The conditional probability  $\Pr[\mathbf{m}|t-d, g'; t, g]$  reflects the model of super- and co-infections. There are many possible models. Super- and co-infections are both notoriously difficult to address. Namely, knowledge about the vector dynamics and the distribution of haplotype combinations in the mosquito population must be known. This is a difficult task and research on the topic is currently expanding, (cf. Nkhoma et al., 2012; Wong et al., 2018; Zhu et al., 2019; Nkhoma et al., 2020; Dia and Cheeseman, 2021; Neafsey et al., 2021).

### 2.3.1 A model for super-infections

Many approaches to estimate MOI or COI by Bayesian or maximum-likelihood methods (e.g. (Hill and Babiker, 1995; Stephens et al., 2001; Rastas et al., 2005; Li et al., 2007; Hastings and Smith, 2008; Druet and Georges, 2010; Ross et al., 2012; Wigger et al., 2013; Taylor et al., 2014; Galinsky et al., 2015; Ken-Dror and Hastings, 2016; Schneider, 2018; Hashemi and Schneider, 2021)) are based on a model, which assumes only super-infections, but no co-infections. The number of super-infections  $m$  is referred to as multiplicity of infection (MOI; see Figure 3).

Let  $M_m^{(t,g)}$  be the probability that a host belonging to group  $g$  is super-infected exactly  $m$  times in generation  $t$ . This is a probability distribution, hence

$$\sum_{m=1}^{\infty} M_m^{(t,g)} = 1 \tag{3}$$

for all  $t$  and  $g$ . At each infectious event, exactly one haplotype is randomly drawn from the mosquito population, i.e., the haplotype distribution  $\mathbf{P}_t$ . Hence, given MOI  $m$  in generation  $t$ , the infection  $\mathbf{m} = (m_1, \dots, m_H)$ , which indicates how many times haplotype  $h$  was transmitted, follows a multinomial distribution with parameters  $m$  and  $\mathbf{P}_t$ , i.e.,

$$\Pr[\mathbf{m}|m; t] = \binom{m}{\mathbf{m}} \mathbf{P}_t^m, \tag{4}$$

where  $\binom{m}{\mathbf{m}} := \frac{m!}{\prod_{h=1}^H m_h!}$  is a multinomial coefficient, and  $\mathbf{P}_t^m := \prod_{h=1}^H P_h^{(t) m_h}$ . Clearly, the constraint  $|\mathbf{m}| := \sum_{h=1}^H m_h = m$  must hold. If an infection is a relapse with a delay of  $d$  generations, the haplotypes have to be drawn according to the distribution  $\mathbf{P}_{t-d}$ .

Therefore, the probability of infection  $\mathbf{m}$  given it has MOI  $m = |\mathbf{m}|$  and occurs in generation  $t$ , when the host belongs to group  $g$ , from a relapse with a delay of  $d$  generations, when the host belonged to group  $g'$ , is given by

$$\Pr[\mathbf{m}, m|t - d, g'; t, g] = M_m^{(t-d,g')} \binom{m}{\mathbf{m}} \mathbf{P}_{t-d}^m, \tag{5}$$

where  $M_m^{(t-d,g')}$  is the probability of MOI  $m$  in generation  $t - d$  of a host in group  $g'$ . This model makes the expression (WHO, 2021b) much more explicit.

### 2.3.2 Choices for the distribution of super-infections

The model (WHO, 2021b) becomes even more explicit for specific choices of the distribution of MOI. A popular choice emerges from the assumption of rare and independent infections, namely that MOI is conditionally Poisson distributed (cf. Schneider, 2021), i.e.,

$$M_m^{(t,g)} = \frac{1}{\exp(\lambda_{t,g}) - 1} \frac{\lambda_{t,g}^m}{m!}, \tag{6}$$

where  $\lambda_{t,g} > 0$  is the Poisson parameter of group  $g$  in generation  $t$  and  $m = 1, 2, \dots$

Another popular choice is the conditional negative-binomial distribution. It is similar to the Poisson distribution but over-dispersed (cf. 17).

## 2.4 The exo-erythrocytic and erythrocytic cycles

Assume an infection subsumed by the vector  $\mathbf{m}$  having MOI  $m = |\mathbf{m}|$ . Since all steps of reproduction are clonal inside the host, it is not necessary to model the different parasite stages explicitly. Rather, it suffices to model the change in haplotype frequencies inside the host as a single step.

If the host belongs to group  $g$ , the ‘absolute’ frequency of haplotype  $h$  is  $\frac{m_h}{m} W_{\mathbf{m},h}^{(t,g)}$ . Here,  $W_{\mathbf{m},h}^{(t,g)}$  is the fitness in generation  $t$  of haplotype  $h$  in infection  $\mathbf{m}$  of a host belonging to group  $g$ . It is interpreted as the expected number of gametocyte descendants of a single copy of haplotype  $h$  infecting the host at the time a mosquito takes her blood meal.

### 2.4.1 Host-vector transmission

Concerning host-vector transmission, a mosquito ingests a fraction  $f$  of male and female gametocytes at her blood meal. The gametocyte haplotypes ingested are assumed to be proportional to the haplotype frequencies within the host. More precisely,  $f \frac{m_h}{m} W_{\mathbf{m},h}^{(t,g)}$  male and female haplotype  $h$  are ingested from infection  $\mathbf{m}$  in group  $g$ . (Note different fractions  $f$  can also be assumed for male and female gametocytes, reflecting an unequal sex ratio.)

### 2.4.2 Sporogonic cycle

Recombination occurs immediately after the blood meal (see Figure 1), and only parasites descending from the same host can recombine (see Figure 3). Assuming the mosquito bite a host from group  $g$  with infection  $\mathbf{m}$ , the probability that a male gamete of haplotype  $h$  fertilizes a female  $i$ -gamete is the product of their relative frequencies in the mosquito’s gut, i.e.,

$$\frac{f \frac{m_h}{m} W_{\mathbf{m},h}^{(t,g)}}{f W_{\mathbf{m}}^{(t,g)}} \cdot \frac{f \frac{m_i}{m} W_{\mathbf{m},i}^{(t,g)}}{f W_{\mathbf{m}}^{(t,g)}} = \frac{m_h W_{\mathbf{m},h}^{(t,g)} m_i W_{\mathbf{m},i}^{(t,g)}}{m^2 W_{\mathbf{m}}^{(t,g)} 2}, \tag{7}$$

where

$$f W_{\mathbf{m}}^{(t,g)} := f \sum_{j=1}^H \frac{m_j}{m} W_{\mathbf{m},j}^{(t,g)} \tag{8}$$

is the total amount of parasites in the mosquito’s gut. Therefore, the absolute number of such matings is obtained by multiplying the probability of the mating by the total amount of parasites, i.e.,

$$f A_{h,i}^{(t,g)} \tag{9}$$

where

$$A_{m,h,i}^{(t,g)} := \frac{m_i W_{m,h}^{(t,g)} m_i W_{m,i}^{(t,g)}}{m^2 W_m^{(t,g)}}. \tag{10}$$

The absolute frequency of haplotype  $h$  in the population of mosquitoes, which descends from infections with configuration  $\mathbf{m}$ , given 1) MOI  $m = |\mathbf{m}|$ , 2) the infections occur in generation  $t$ , 3) in hosts in group  $g$ , which 4) are either novel infections (delay  $d = 0$ ) or relapses with a delay of  $d$  generations, is

$$\Pr[\mathbf{m}|\mathbf{m}; t - d; t, g] \sum_{j,l=1}^H f A_{m,j,l}^{(t,g)} r(jl \rightarrow h), \tag{11}$$

where  $r(jl \rightarrow h)$  is the probability that a mating between gametes with haplotypes  $j$  and  $l$  lead to offspring of haplotype  $h$ .

The absolute number of haplotype  $h$  in the mosquito population, which descend from hosts in group  $g$  with MOI  $m$ , is calculated from the theorem of total probability, i.e., by ‘averaging’ over all possible infections  $\mathbf{m}$  with MOI  $m$ . Incorporating all relapses it is given by

$$P_h^{*(g,m)}(t) = \sum_{d=0}^D R_d^{(t)} \sum_{m: |\mathbf{m}|=m} \Pr[\mathbf{m}|\mathbf{m}; t - d; t, g] \times \sum_{j,l=1}^H f A_{m,j,l}^{(t,g)} r(jl \rightarrow h). \tag{12}$$

If an infection in generation  $t$  is a relapse from generation  $t - d$  the host might have belonged to a different group  $g'$  then. Noting, that

$$\Pr[\mathbf{m}|t - d; t, g] = \sum_{g'=1}^S G_{g',g}^{(t-d,t)} \Pr[\mathbf{m}|\mathbf{m}; t - d, g'; t, g] \tag{13}$$

equation (Su et al., 2019) can be rewritten as

$$P_h^{*(g,m)}(t) = \sum_{d=0}^D R_d^{(t)} \sum_{g'=1}^S G_{g',g}^{(t-d,t)} \sum_{m: |\mathbf{m}|=m} \Pr[\mathbf{m}|\mathbf{m}; t - d, g'; t, g] \times \sum_{j,l=1}^H f A_{m,j,l}^{(t,g)} r(jl \rightarrow h). \tag{14}$$

### 2.5 Evolutionary dynamics

To determine the number of haplotypes  $h$  in generation  $t + 1$ , equation (Ngwa et al., 2016) has to be averaged over all possible groups and values of MOI. Hence, the absolute frequency of haplotype  $h$  in the next generation’s sporozoite population is

$$P_h^*(t + 1) = f \sum_{d=0}^D R_d^{(t)} \sum_{g,g'=1}^S G_{g',g}^{(t-d,t)} \sum_{m=1}^{\infty} \times \sum_{m: |\mathbf{m}|=m} \Pr[\mathbf{m}, \mathbf{m}|t - d, g'; t, g] \times \sum_{j,l=1}^H A_{m,j,l}^{(t,g)} r(jl \rightarrow h). \tag{15}$$

The relative frequency of haplotype  $h$  in the sporozoite population in generation  $t + 1$  is hence

$$P_h(t + 1) = \frac{P_h^*(t)}{\sum_{i=1}^H P_i^*(t)}. \tag{16}$$

The dynamics (Watson et al., 2021) are extremely flexible. They allow to model, e.g., temporal changes in selection pressures (for instance changing treatment policies in the context of drug-resistance evolution, temporally varying transmission intensities, intra-host competition of parasites, super- and co-infections, relapses, recrudescences etc.). This however requires to specify the model more explicitly.

Next, we show how this is done if only super-infections but no co-infections are considered.

### 2.6 Evolutionary dynamics with super-infections

We introduce a couple of simplifying assumptions, which make the model more explicit. First, only super- but no co-infections are assumed. I.e., the super-infection model (Lover et al., 2018) applies and is substituted into (Schneider and Escalante, 2013). Thus, (Schneider and Escalante, 2013), becomes

$$P_h^*(t + 1) = f \sum_{d=0}^D R_d^{(t)} \sum_{g,g'=1}^S G_{g',g}^{(t-d,t)} \sum_{m=1}^{\infty} M_m^{(t-d,g')} \sum_{m: |\mathbf{m}|=m} \binom{m}{\mathbf{m}} P_{t-d}^m \times \sum_{j,l=1}^H A_{m,j,l}^{(t,g)} r(jl \rightarrow h). \tag{17}$$

### 3 Results

The framework is appropriate to investigate numerous evolutionary-genetics aspects in malaria. It would be far too comprehensive to exemplify the full flexibility. Hence, only special cases are illustrated here. We assume that only super-infections but no co-infections occur, i.e., the dynamics (Baird et al., 2018) are assumed. First, we clarify the difference between haplotype frequency and prevalence. Then we focus on a simple model of drug resistance. Although it is applicable to all malaria species, primarily it shall illustrate the differences between *P. falciparum* and *P. vivax*, because there

were no reports on drug resistance in any of the other species (Tseha and Tyagi, 2021).

### 3.1 Frequency and prevalence

The evolutionary genetics of malaria are described as the time-change in the frequency distribution of parasite haplotypes. For instance, monitoring the frequencies of haplotypes, which confer drug resistance is essential. However, concerning the clinical pathogenesis, the occurrence of resistance-conferring haplotypes in infections is more relevant. Due to super- and co-infections the frequency of a haplotype  $h$ , i.e., its relative abundance among sporozoites in the mosquito population does not coincide with the probability that haplotype  $h$  occurs in an infection. The latter is referred to as the haplotype's prevalence.

If only super-infections are considered, the prevalence of haplotype  $h$  in generation  $t$ , denoted by  $q_h^{(t)}$  is derived in section Prevalence in the Mathematical Appendix. It is given by

$$q_h^{(t)} = 1 - \sum_{d=0}^D R_d^{(t)} \sum_{g'=1}^S G_{g'}^{(t-d)} U_{g'}^{(t-d)} (1 - P_h^{(t-d)}), \quad (18)$$

where  $U_{g'}^{(t-d)}(x)$  is the probability generating function of the MOI distribution in group  $g'$  in generation  $t - d$ . This function characterizes transmission in group  $g'$  in generation  $t - d$ . From the above expression it is clear that prevalence depends on (i) the frequency of haplotype  $h$ , (ii) the distributions of MOI in the various groups, and (iii) the distribution of relapses/recrudescence. If no relapses or recrudescences occur, as it is the case for *P. falciparum* and *P. knowlesi*, the prevalence simplifies to

$$q_h^{(t)} = 1 - \sum_{g=1}^S G_g^{(t)} U_g^{(t)} (1 - P_h^{(t-d)}). \quad (19)$$

Hence, for *P. falciparum* and *P. knowlesi* prevalence is characterized by the haplotype frequency distribution in  $t$ , the distribution of groups, and the MOI distributions in the groups. We illustrate the effect of relapses on prevalence in a simple example below.

### 3.2 Selection at a single locus without intra-host competition

Assume drug resistance is determined by a single locus. This is a reasonable assumption since often drug resistance is determined mainly by mutations at one locus. For instance, in *P. falciparum* resistance to chloroquine is determined by mutations at the *Pfcr* locus, while resistance artemisinin is determined by mutations in the Kelch-13 propeller region (Cui et al., 2015). The assumption is even justified in sulfadoxine-pyrimethamine resistance, determined by the

*Pfdhfr* and *Pfdhps* loci, because mutations at the *Pfdhfr* locus seem to have a much stronger effect (McCollum et al., 2012).

Assume  $n$  alleles  $A_1, \dots, A_n$  are segregating at the selected locus. The  $n$  different alleles confer different levels of drug resistance. All other alleles are assumed to be neutral. Thus, the number of possible haplotypes,  $H$ , is a multiple of  $n$ , i.e.,  $H = nN$ . Hence,  $N$  is the number of all possible haplotypes when the resistance-conferring locus is disregarded. Let us assume that the haplotypes are ordered such that haplotypes  $h = (a - 1)N + 1, \dots, aN$  carry allele  $A_a$  at the resistance-conferring locus. Therefore, the frequency of allele  $A_a$  at time  $t + 1$ , denoted by  $p_a^{(t+1)}$  is given by

$$p_a^{(t)} = \sum_{h=(a-1)N+1}^{aN} P_h^{(t)}. \quad (20)$$

Cumulatively, we denote the vector of allele frequencies in generation  $t$  by  $\mathbf{p}_t$ .

Under the assumption of no intra-host competition of parasites these dynamics can be made more explicit. In an infection characterized by  $\mathbf{m}$  of a host in group  $g$ , no intra-host competition means that the fitness of an infecting haplotype  $h$  is independent of what other haplotypes are present in the infection, i.e., it is independent of  $\mathbf{m}$ , or formally

$$W_{\mathbf{m},h}^{(t,g)} = W_h^{(t,g)}. \quad (21)$$

Furthermore, because fitness is only determined by the resistance-conferring locus, the fitness of haplotype  $h$  depends only on its allele at this locus. Let the fitness of haplotypes carrying allele  $A_a$  at the resistance-conferring locus be denoted by  $w_a^{(t,g)}$ , i. e.,

$$w_a^{(t,g)} = W_h^{(t,g)} = W_{\mathbf{m},h}^{(t,g)} \quad \text{for } h = (a - 1)N + 1, \dots, aN \quad \text{and for all } \mathbf{m}. \quad (22)$$

Moreover, let the average fitness of allele  $A_a$  in generation  $t$  be

$$w_a^{(t)} = \sum_{g=1}^S w_a^{(t,g)} G_g^{(t)}. \quad (23)$$

As shown in the Mathematical Appendix the dynamics of the allele frequencies are given by

$$p_a^{(t+1)} = \frac{w_a^{(t)} \sum_{d=0}^D R_d^{(t)} p_a^{(t-d)}}{\sum_{b=1}^n w_b^{(t)} \sum_{d=0}^D R_d^{(t)} p_b^{(t-d)}}. \quad (24)$$

As in the case without relapses/recrudescence (cf. 17), these dynamics are independent of the distribution of MOI. This holds because no intra-host competition occurs and because only super-infections are considered. Even without intra-host competition the dynamics of the allele frequencies at the selected locus might depend on MOI, depending on the assumed model for co-infections; a general statement cannot be made.

Further, the dynamics (Bousema et al., 2011) depend only on the average fitnesses of the alleles  $w_a^{(t)}$ . This implies that the stratification of the host population into different groups does not need to be modelled explicitly, when considering selection at a single locus.

Note that the average fitnesses can be scaled by any constant without affecting the dynamics (Bousema et al., 2011). Hence, it suffices to consider relative fitnesses, and fitness can be normalized such that  $w_1^{(t)} = 1$  in every generation.

### 3.2.1 The effect of recrudescences and relapses

In the dynamics of the allele frequencies (Bousema et al., 2011) the effect of relapses or recrudescence is clearly visible. In the case of no relapses or recrudescence, i.e.,  $R_d^{(t)} = 0$  for  $d \geq 0$  the dynamics simplify to

$$p_a^{(t+1)} = \frac{w_a^{(t)} p_a^{(t)}}{\sum_{b=1}^n w_b^{(t)} p_b^{(t)}} \tag{25}$$

In this situation, the allele frequencies in generation  $t + 1$  are solely determined by the fitnesses and the allele frequencies in generation  $t$ . Once relapses or recrudescences are considered, the allele frequencies in generation  $t + 1$ , depend also on the allele frequencies in previous generations. This is intuitively clear, because relapses/recrudescence are equivalent to infections from the sporozoite population from previous generations (see Figure 2). Hence, relapses/recrudescence act as “seed banks”. Intuitively, this will delay the evolutionary dynamics, because the allele frequencies are averaged over several previous generations.

To further discuss the effect of relapses/recrudescence we impose some additional assumptions. First, we assume that the selective environment does not change over time, i.e.,  $w_a^{(t)} = w_a$  for all  $t$ . This is a reasonable assumption when considering drug resistance evolution over a time period in which treatment policies do not change. In this case, the change in allele frequencies can be solved explicitly only in the absence of relapses/recrudescence. Namely, the dynamics become

$$p_a^{(t+1)} = \frac{w_a^{t+1} p_a^{(0)}}{\sum_{b=1}^n w_b^{t+1} p_b^{(0)}} \tag{26}$$

where  $p_a^{(0)}$  are the initial allele frequencies in generation  $t = 0$ . From these dynamics it follows that the average fitnesses  $w_a$  can be estimated from longitudinal data of allele frequencies by fitting a straight-line regression (see 48, 17).

Once relapses/recrudescence are considered, the dynamics can no longer be solved explicitly, but need to be calculated recursively from the frequencies of the last  $D + 1$  generations, i.e., they become

$$p_a^{(t+1)} = \frac{w_a \sum_{d=0}^D R_d^{(t)} p_a^{(t-d)}}{\sum_{b=1}^n w_b \sum_{d=0}^D R_d^{(t)} p_b^{(t-d)}} \tag{27}$$

Importantly, to be able to iterate these dynamics, initial frequencies need to be known from  $D$  generations in the past. Hence, to calculate the frequencies in generation  $t = 1$ , initial frequencies  $p_a^{(0)}, p_a^{(-1)}, \dots, p_a^{(-D)}$  need to be specified. Moreover, the distribution  $R_d^{(t)}$  needs to be known. In practice, the distribution of relapses might change over time. For instance, changes in control policies impact malaria transmission and hence the proportion of new infection in comparison to relapses. If transmission intensities decrease, relapses amount for a larger fraction of infections. Also the number of transmission cycles during 1 year decrease. Because the distribution of the time to relapse measured in years will not change, the time distribution measured in units of transmission cycles will change. In the simplest case the distribution of relapses remains constant over time, i.e.,  $R_d^{(t)} = R_d$ , the change of allele frequencies is given by

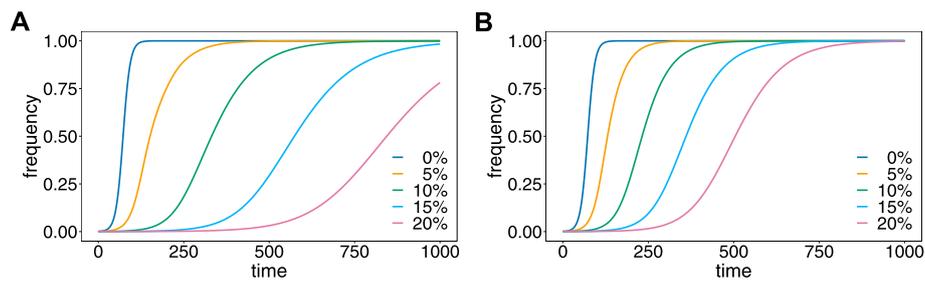
$$p_a^{(t+1)} = \frac{w_a \sum_{d=0}^D R_d p_a^{(t-d)}}{\sum_{b=1}^n w_b \sum_{d=0}^D R_d p_b^{(t-d)}} \tag{28}$$

Unfortunately, even if the distribution of relapses is constant, the average fitnesses can no longer be estimated by a linear regression.

The distribution of relapses depends crucially on the specific parasite strain (White, 2011). Consider the following example of drug-resistance evolution, with just two alleles: allele  $A_1$  being the drug sensitive wildtype and  $A_2$  the mutant allele conferring drug resistance. The mutant allele first occurs in generation  $t = 0$  at frequency  $p_2^{(0)} = 0.001$ . Let  $w_2 = 1 + s$ , where  $s$  is the selective advantage of the drug resistant allele  $A_2$ . We assume  $s = 0.1$ , i.e., the fitness is increased by 10%, which is strong selection for population-genetic processes, but reasonable for selection for drug-resistance.

Regarding the distribution of relapses, we assume a situation in which 1 year corresponds to 10 transmission cycles. Relapses often occur in periodic patterns (White, 2011). We first assume a pattern which resembles the relapse pattern described by (Hankey et al., 1953) in temperate zones of Korea. Namely, let  $v$  be the probability that a malaria episode relapses, i.e.,  $R_0 = 1 - v$ . We assume the first relapse can occur after 10 transmission cycles, and all further relapses after 4 further transmission cycles for a maximum delay of  $D = 90$ . More precisely,  $R_d = \frac{v}{21}$  for  $d = 10, 14, 18, 22, \dots, 90$  and  $R_d = 0$  else. As a comparison we assume a simple second pattern of relapses, in which relapses occur 4–50 generations after the initial infection with equal probability, i.e.,  $R_d = \frac{v}{43}$  for  $d = 4, \dots, 50$ . Compared with the first pattern, relapses occur more frequently and earlier.

The evolutionary dynamics are illustrated in Figure 4. Without relapses  $v = 0$ , the resistance-conferring allele spreads in approximately 110 generations, which corresponds to 11 years, under the assumed number of 10 transmission cycles per year. Relapses substantially slow down the spread of resistance. The



**FIGURE 4**

Effect of relapses on the evolutionary dynamics. Shown is the frequency of the resistance-conferring allele as a function of time assuming different proportions,  $\nu$  of relapses (colors) for the first (A) and second (B) patterns of relapses.

reason is that relapses act like ‘seed banks’ which retain the frequency distribution of previous generations. For the first pattern (Figure 4A), 5% relapses already substantially delay the spread of resistance to about 400 generations or 40 years. With 20% relapses, the frequency of the mutant allele is just 75% after 1,000 generations corresponding to 100 years. For the second pattern (Figure 4B), the results are qualitatively similar, but relapses have a less profound effect, because they occur with shorter delay after the original infection.

These results provide formal evidence that drug resistance spreads faster in *P. falciparum*, where no relapses occur, than in *P. vivax*, where relapses are common. In fact, while drug resistance is a major concern in *P. falciparum*, it is less common in *P. vivax* (Schneider and Escalante, 2013).

The pattern of relapses depends on 1) genetic factors mediating the frequency of their occurrence; 2) transmission intensities determining the number of malaria generations (transmission cycles per year); 3) the fractions of new infections and relapses; and 4) treatment policies. Particularly, if a drug is partnered with primaquine (PQ) or tafenoquine (TQ) for radical cure, the fraction of relapses reduces, accelerating the spread of resistance to the primary treatment. However, since PQ or TQ also act on gametocytes, they prevent transmission and reduce the selective advantage of drug resistance (cf. 23).

### 3.2.2 Prevalence

Next consider the prevalences corresponding to the evolutionary dynamics illustrated in Figure 4. The evolutionary dynamics are determined by the average fitnesses across the groups of hosts and the distribution of relapses. Consequently, it was not necessary to specify the groups explicitly. However, prevalence given by (Collins and Jeffery, 2007) depends on the generating functions of MOI in the different groups. In the simplest case, which we consider here, the whole population consists of only one group ( $S = 1$ ). Furthermore, we assume that the MOI distribution does not change over time, and follows a conditional Poisson distribution

(cf. Eq. (6)) with parameter  $\lambda$ . The generating function of this distribution is given by

$$U(x) = \frac{\exp(\lambda x) - 1}{\exp(\lambda) - 1} \quad (29)$$

(cf. 17).

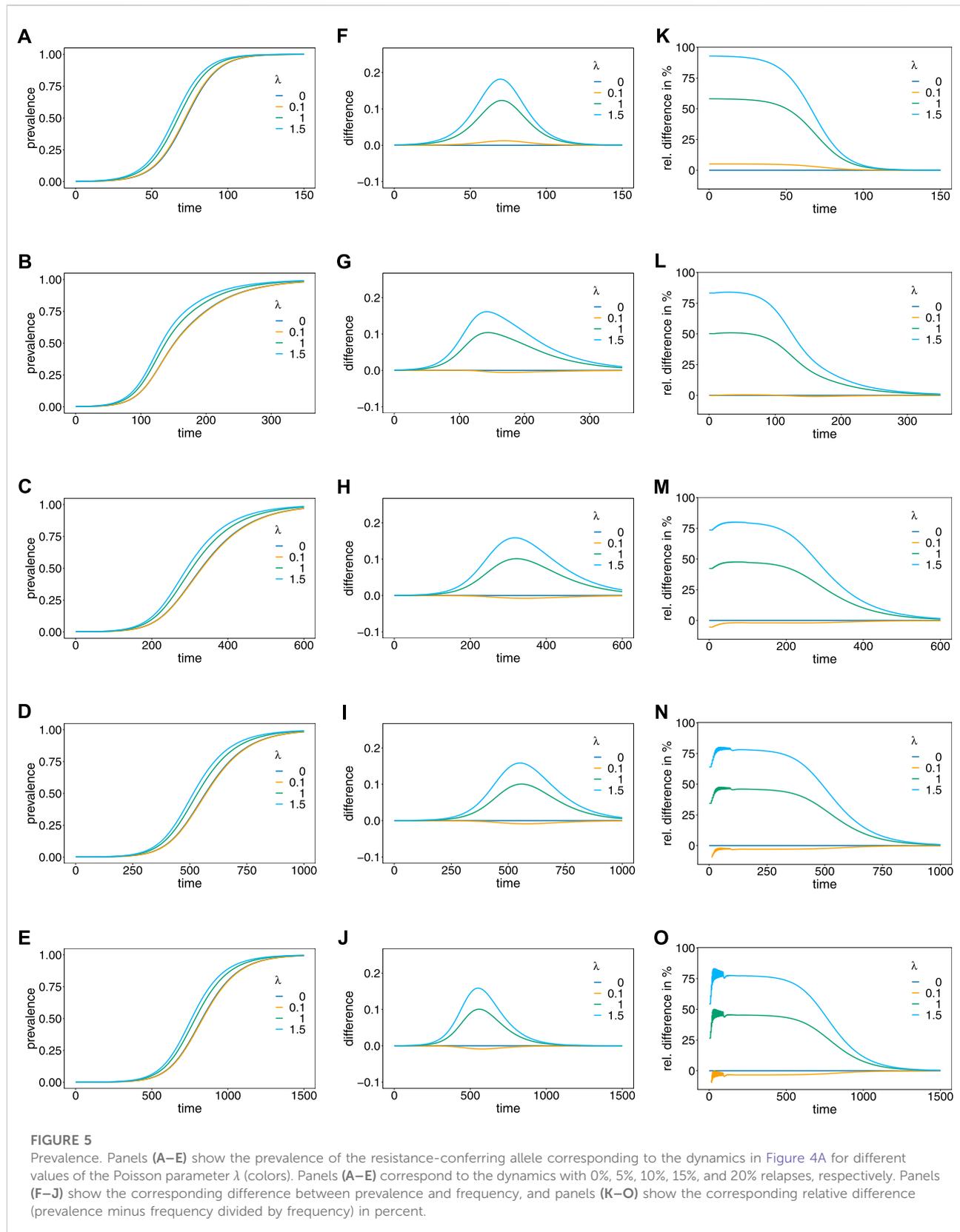
The prevalence of the resistance-conferring allele is obtained from (Collins and Jeffery, 2007) by assuming that haplotypes are characterized by a single locus. Hence,

$$q_2^{(t)} = 1 - \sum_{d=0}^D R_d U(1 - p_2^{(t-d)}) = \sum_{d=0}^D R_d \frac{1 - \exp(-\lambda p_2^{(t-d)})}{1 - \exp(-\lambda)}. \quad (30)$$

The prevalences corresponding to the dynamics illustrated in Figure 4A, are depicted in Figure 5, assuming different values of the Poisson parameter  $\lambda$ , corresponding to different transmission intensities.

The case  $\lambda = 0$ , implies that only ‘single-infection’ (one infective event) occurs, in which case prevalence and frequency coincide. As shown in (Schneider, 2021) prevalence always exceeds frequency in the case in which no relapses occur (Figures 5A,F,K). This is intuitive, because the likelihood to observe a parasite variant in an infection increases as the average number of super-infections increase. If transmission intensities are intermediate to high ( $\lambda \geq 1$ ), prevalence is considerably higher than frequency (Figure 5F). If the frequency of the resistance-conferring allele is small, the difference between frequency and prevalence is small in absolute terms, but high in relative terms (compare Figure 5F with Figure 5K).

If relapses occur, the pattern is similar, however, prevalence can be lower than frequency (see Figures 5F–J). The reason is that prevalence is also determined by the frequency distribution of past generations. This occurs only if the average number of super-infections is small ( $\lambda$  slightly larger than 0) and is increasingly pronounced if relapses are more frequent. In general, the difference between frequency and prevalence becomes smaller in absolute and relative terms as the fraction



of relapses increase. If this fraction is high ( $v = 0.15$  or  $v = 0.2$ ) the particular pattern of relapses leads to oscillations in the relative difference between prevalence and frequency, if the frequency of the resistance-conferring allele is low (see Figures 5N,O).

## 4 Discussion

We introduced a general framework to model evolutionary-genetic processes in malaria, which is flexible enough to capture the characteristics of all human-pathogenic *Plasmodium* species. Such a framework is justified since standard population-genetic theory can only be approximately applied to malaria. The reason is rooted in the malaria transmission cycle, which involves one step of sexual reproduction in the mosquito vectors. A high degree of selfing occurs during this step, because only parasites which descend from the same human host can recombine (cf. Figure 3). The framework extends the one introduced in (Schneider and Kim, 2010; Schneider and Kim, 2011; Schneider, 2021), which is only applicable to *P. falciparum*, because it ignores relapses from dormant liver stages as they occur in *P. vivax* and *P. ovale* sp., and recrudescence from prolonged blood stage parasites as they occur in *P. malariae*. These previously widely neglected species are resilient because of relapses and recrudescence, and hence are gaining more importance in the context of malaria eradication. We demonstrated the importance of relapses/recrudescence by contrasting drug resistance-evolution in *P. vivax* and *P. falciparum*.

The necessity to extend the population-genetic framework toward other malaria species is clearly justified by the results presented here. Even in the simplest case of resistance being determined by a single locus, relapses have a profound effect on the evolutionary dynamics, when assuming the same hypothetical drug pressure in both species. Namely, relapses substantially delay the spread of resistance, because they are equivalent—at least in the idealization of the model—to infections with regard to past parasite frequency distributions. In other words, relapses act as seed banks. Dormancy by seed banks is known in evolutionary biology as a bet-hedging strategy that allows organisms to survive through sub-optimal conditions (Shoemaker and Lennon, 2018)—in the case of malaria the absence of the vector. Seed banks are also known to slow down evolutionary processes and influence recombination (Živković and Tellier, 2012; Koopmann et al., 2017; Tellier, 2019). This is no exception in malaria. Although exploring the effect of relapses/recrudescence on recombination was beyond the scope of this work, the effect is rather obvious. Because relapses/recrudescence slow down the evolutionary dynamics, more genetic variation is maintained, leading to a higher level of recombination. In fact, in *P. vivax* higher levels of genetic variations than in *P. falciparum* are a common empirical observation (e.g. (Pacheco et al., 2020)).

Our results have to be understood in a qualitative rather than a quantitative context. Namely, the pattern of relapses have a substantial influence on the evolutionary dynamics. Hence, for adequately predict the spread of resistance, good empirical estimates on the pattern of

relapses are necessary. However, empirically distinguishing reinfections (consecutive independent infectious), recrudescence (a rebound of parasitaemia due to incomplete clearance of merozoites), and relapses are notoriously difficult. With more advanced molecular methods becoming available to produce deep-sequencing data (e.g. (Zhong et al., 2018; Gruenberg et al., 2019)), heuristic methods to distinguish recrudescence from reinfections have been proposed (Lin et al., 2015). Also haplotype-based statistical models have been proposed (e.g. (Plucinski et al., 2015)). In principle the framework here can be used to further develop statistical methods to distinguish reinfections from relapses.

To obtain quantitative predictions it is also important to estimate other model parameters. In the context of drug resistance, this includes fitness parameters, metabolic costs for resistance, and the proportion of asymptomatic or untreated infections. The latter can be achieved by routine diagnostics using reliable methods such as ultra-sensitive PCR (e.g. (Gruenberg et al., 2020)). However, also the transmission potential, determined by the abundance of gametocytes has to be determined (cf. 9). Selection parameters of drug-resistant haplotypes can be determined from longitudinal molecular data by a linear regressions in *P. falciparum* (McCullum et al., 2012; Schneider, 2021). Disentangling the fitness parameters into metabolic costs and selective advantages of resistance is more difficult. Namely, costs and selective advantages as found in *in vitro* studies (cf. Cortese and Plowe, 1998) do not linearly scale with *in vivo* observations. In principle, costs can be achieved by contrasting different populations with different drug usage. Comparing such results with *in vitro* studies helps to identify the functional relationship between *in vitro* measurements and *in vivo* observations. Notably, fitness estimates from a linear regression apply mainly to *P. falciparum*. For other malaria species the estimates have to be adapted to the evolutionary dynamics which account for relapses/recrudescence.

Note that the application to modelling drug resistance here had only the purpose of contrasting the absence and presence of relapses. Therefore, only a simplistic model was assumed for drug resistance, i.e., resistance was assumed to be determined by a single biallelic locus. The examples here did not exhibit the full flexibility of the model. If drug resistance occurs in a stepwise fashion as it is found in sulfadoxine-pyrimethamine resistant *P. falciparum* haplotypes (Cortese and Plowe, 1998), where resistance is caused by mutations at several codons in the *Pfdhfr* and *Pfdhps* loci. To capture this situations, resistance-conferring haplotypes have to be modelled by two multiallelic loci, where each two-locus haplotype is associated with its own metabolic costs and fitness advantage. Moreover, the mutation haplotypes have to be introduced into the model at different time points. A simple example can be found in (Schneider, 2021).

Relapses are irrelevant in *P. falciparum*, and recrudescences can be neglected, because they occur shortly after the initial infection and do not need to be modeled explicitly. Nevertheless, if transmission intensities are high, which is mainly relevant for *P. falciparum*, the assumption of non-overlapping generations (transmission cycles) are questionable. In the extended

framework, relapses can be reinterpreted to mimic overlapping generations. This explains, at least partially, why drug resistance in *P. falciparum* does not necessarily spread first in areas of high transmission (as they occur in Africa) with many more transmission cycles per year.

Reinterpreting relapses in the framework is also important when applied to *P. knowlesi*, which is primarily pathogenic to non-human primates, but became the dominant human-pathogenic malaria species in some endemic areas (Sutherland, 2016). The zoonotic animal-host reservoir renders *P. knowlesi* resilient. Different transmission dynamics between humans and animal hosts can mediate the duration of a transmission cycle. If the number of transmission cycles per year differs among human and non-human primate hosts, this discrepancy can be compensated by modeling overlapping generations by relapses.

We also discussed the differences of frequency and prevalence of parasite haplotypes. The former is the relative abundance of a haplotype in the parasite population, the latter the likelihood that the haplotype occurs in an infection. Studying the haplotype frequency distribution over time is the aim of evolutionary genetics. From a clinical or epidemiological point of view, prevalence is more relevant. The latter is determined by the haplotype frequency distribution and the distribution of super- or co-infections. This was already emphasized in the context of seasonal malaria transmission in (Schneider, 2021) for *P. falciparum*. It was shown that the prevalence of a haplotype always exceeds its frequency. This changes if relapses/recrudescence occur and was exemplified here by the hypothetical dynamics of drug-resistance evolution.

The applications of the framework introduced here are manifold. For instance, in the context of drug resistance, the framework allows to investigate the evolution of multi-drug resistance determined by several loci and changing drug-treatment policies. Also patterns of selection, e.g., genetic hitchhiking, can be studied using this framework. The illustrated applications were only under the simplest assumptions, e.g., of no intra-host competition and super- but no co-infections.

Intra-host competition plays an important role in the spread of HRP2/3 gene deletions associated with false-negative malaria rapid diagnostic tests (RDTs) (Gamboa et al., 2010). Namely, if the treatment guidelines require to verify suspected infections by RDTs before treatment with artemisinin combination therapies (ACTs), as recommended by the WHO (World Health Organization, 2017), false-negative results can lead to delayed treatment. Similarly intra-host competition seems relevant when considering selection on merozoite surface proteins (Goh et al., 2021).

Intra-host dynamics enter the model *via* the definition of fitness. It is not necessary to define an evolutionary-genetic model which captures two timescales, the evolutionary dynamics in terms of generations of transmission cycles, and the timescale of an infectious episode in the same model, as it was done, e.g., in (Kim et al., 2014). Rather, the framework can be used in a multi-scale model, which takes input from a separate intra-host model.

Similarly, the framework does not require to model the mosquito dynamics explicitly. They rather enter *via* the distribution of super- and co-infections. Considering only super-infections has the conceptual advantage, that it is a well-defined model. It is frequently used in statistical approaches to estimate haplotype frequency distributions and MOI (*cf.* e.g. Hill and Babiker, 1995; Stephens et al., 2001; Li et al., 2007; Hastings and Smith, 2008; Wigger et al., 2013; Schneider, 2018; Hashemi and Schneider, 2021). Ignoring co-infections is justified if the distribution of haplotypes in the mosquitoes is uncorrelated or when considering only few loci. However, if one aims to include genetic relatedness, it is important to specify a model for co-infections. This becomes increasingly popular as more high-quality genomic data is becoming available in malaria, which has enough resolution to study genetic relatedness (*cf.* Nkhoma et al., 2012; Wong et al., 2018; Zhu et al., 2019; Nkhoma et al., 2020; Dia and Cheeseman, 2021; Neafsey et al., 2021).

Although the framework is very general, it also has several limitations. First, it ignores mutations. This is not a strong restriction, because in many applications one is interested in *de novo* mutations which occur at discrete time points. This is captured by the model, by introducing new haplotypes (i.e., extending the model) at certain times. However, constant mutation rates, e.g., to study mutation-selection balance, can be easily introduced. Another limitation is the deterministic nature of the framework. When aiming to study stochastic effects such as genetic drift, it is rather straightforward to develop a stochastic version of the framework. Third, the model ignores mitotic recombination during merozoite production inside the host. This plays an important role in some applications, particularly in the structural rearrangement of Var genes (Claessens et al., 2014). These hypervariable genes are responsible to generate important antigen profiles for parasite-host interactions (Warimwe et al., 2009). In any case the framework introduced here allows studying manifold evolutionary-genetic aspects of malaria. Importantly, it allows us to specify benchmark scenarios. More empirical evidence is required to refine relevant parametrizations of the framework.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Author contributions

KS conceptualized the work, developed the mathematical model, performed the mathematical analysis, produced all figures, and wrote the manuscript. CS assisted to conceptualize the work and wrote the manuscript.

## Funding

This study was funded by the Armed Forces Health Surveillance Division (AFHSD), Global Emerging Infections Surveillance (GEIS) Branch, ProMIS ID P0082\_22\_N6. This work was also supported by grants of the German Academic Exchange (DAAD; <https://www.daad.de/de/>; Project-ID 57417782, Project-ID: 57599539), the Sächsisches Staatsministerium für Wissenschaft, Kultur und Tourismus and Sächsische Aufbaubank—Förderbank (SMWK-SAB; <https://www.smwksachsen.de/>; <https://www.sab.sachsen.de/>; project “Innovationsvorhaben zur Profilschärfung an Hochschulen für angewandte Wissenschaften”, Project-ID 100257255; project “Innovationsvorhaben zur Profilschärfung 2022”, Project-ID: 100613388), the Federal Ministry of Education and Research (BMBF) and the DLR (Project-ID 01DQ20002; <https://www.bmbf.de/>; <https://www.dlr.de/>).

## Acknowledgments

The authors gratefully acknowledge the constructive comments of the two reviewers and the editor.

## References

- Baird, J. K., Battle, K. E., and Howes, R. E. (2018). Primaquine ineligibility in anti-relapse therapy of plasmodium vivax malaria: The problem of g6pd deficiency and cytochrome p-450 2d6 polymorphisms. *Malar. J.* 17, 42–46. doi:10.1186/s12936-018-2190-z
- Beshir, K. B., Parr, J. B., Cunningham, J., Cheng, Q., and Rogier, E. (2022). Screening strategies and laboratory assays to support plasmodium falciparum histidine-rich protein deletion surveillance: Where we are and what is needed. *Malar. J.* 21, 201–212. doi:10.1186/s12936-022-04226-2
- Bousema, T., Kreuels, B., and Gosling, R. (2011). Adjusting for heterogeneity of malaria transmission in longitudinal studies. *J. Infect. Dis.* 204, 1–3. doi:10.1093/infdis/jir225
- Chamchod, F., and Beier, J. C. (2013). Modeling plasmodium vivax: Relapses, treatment, seasonality, and g6pd deficiency. *J. Theor. Biol.* 316, 25–34. doi:10.1016/j.jtbi.2012.08.024
- Claessens, A., Hamilton, W. L., Kekre, M., Otto, T. D., Faizullahb, A., Rayner, J. C., et al. (2014). Generation of antigenic diversity in plasmodium falciparum by structured rearrangement of var genes during mitosis. *PLoS Genet.* 10, e1004812. doi:10.1371/journal.pgen.1004812
- Collins, W. E., and Jeffery, G. M. (2007). Plasmodium malariae: Parasite and disease. *Clin. Microbiol. Rev.* 20, 579–592. doi:10.1128/CMR.00027-07
- Cortese, J. F., and Plowe, C. V. (1998). Antifolate resistance due to new and known plasmodium falciparum dihydrofolate reductase mutations expressed in yeast. *Mol. Biochem. Parasitol.* 94, 205–214. doi:10.1016/s0166-6851(98)00075-9
- Cui, L., Mharakurwa, S., Ndiaye, D., Rathod, P. K., and Rosenthal, P. J. (2015). Antimalarial drug resistance: Literature review and activities and findings of the icemr network. *Am. J. Trop. Med. Hyg.* 93, 57–68. doi:10.4269/ajtmh.15-0007
- Dean, L., and Kane, M. (2020). “Tafenoquine therapy and g6pd genotype,” in *Medical genetics summaries*. Editors V. M. Pratt, S. A. Scott, M. Pirmohamed, B. Esquivel, M. S. Kane, and B. L. Kattman (United States: National Center for Biotechnology Information).
- Dia, A., and Cheeseman, I. H. (2021). Single-cell genome sequencing of protozoan parasites. *Trends Parasitol.* 37, 803–814. doi:10.1016/j.pt.2021.05.013
- Druet, T., and Georges, M. (2010). A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* 184, 789–798. doi:10.1534/genetics.109.108431
- Galinsky, K., Valim, C., Salmier, A., de Thoisy, B., Musset, L., Legrand, E., et al. (2015). Coil: A methodology for evaluating malarial complexity of infection using

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1030463/full#supplementary-material>

- likelihood from single nucleotide polymorphism data. *Malar. J.* 14, 4. doi:10.1186/1475-2875-14-4
- Gamboa, D., Ho, M. F., Bendezu, J., Torres, K., Chiodini, P. L., Barnwell, J. W., et al. (2010). A large proportion of p. falciparum isolates in the amazon region of Peru lack pfrp2 and pfrp3: Implications for malaria rapid diagnostic tests. *PLoS one* 5, e8091. doi:10.1371/journal.pone.0008091
- Goh, X. T., Lim, Y. A., Lee, P. C., Nissapatorn, V., and Chua, K. H. (2021). Diversity and natural selection of merozoite surface protein-1 in three species of human malaria parasites: Contribution from south-east Asian isolates. *Mol. Biochem. Parasitol.* 244, 111390. doi:10.1016/j.molbiopara.2021.111390
- Gonzales, S. J., Reyes, R. A., Braddom, A. E., Batugedara, G., Bol, S., and Bunnik, E. M. (2020). Naturally acquired humoral immunity against plasmodium falciparum malaria. *Front. Immunol.* 11, 594653. doi:10.3389/fimmu.2020.594653
- Greenwood, B., Cairns, M., Chaponda, M., Chico, R. M., Dicko, A., Ouedraogo, J. B., et al. (2021). Combining malaria vaccination with chemoprevention: A promising new approach to malaria control. *Malar. J.* 20, 361–367. doi:10.1186/s12936-021-03888-8
- Gruenberg, M., Lerch, A., Beck, H. P., and Felger, I. (2019). Amplicon deep sequencing improves plasmodium falciparum genotyping in clinical trials of antimalarial drugs. *Sci. Rep.* 9, 17790. doi:10.1038/s41598-019-54203-0
- Gruenberg, M., Moniz, C. A., Hofmann, N. E., Koepfli, C., Robinson, L. J., Nate, E., et al. (2020). Utility of ultra-sensitive qPCR to detect plasmodium falciparum and plasmodium vivax infections under different transmission intensities. *Malar. J.* 19, 319. doi:10.1186/s12936-020-03374-7
- Hankey, D. D., Jones, R., Jr, Coatney, G. R., Alving, A. S., Coker, W. G., Garrison, P. L., et al. (1953). Korean vivax malaria. i. natural history and response to chloroquine. *Am. J. Trop. Med. Hyg.* 2, 958–969. doi:10.4269/ajtmh.1953.2.958
- Hashemi, M., and Schneider, K. A. (2021). Bias-corrected maximum-likelihood estimation of multiplicity of infection and lineage frequencies. *PLoS one* 16, e0261889. doi:10.1371/journal.pone.0261889
- Hastings, I. M., and Smith, T. A. (2008). MalHaploFreq: A computer programme for estimating malaria haplotype frequencies from blood samples. *Malar. J.* 7, 130. doi:10.1186/1475-2875-7-130
- Hedrick, P. W. (2011). Population genetics of malaria resistance in humans. *Heredity* 107, 283–304. doi:10.1038/hdy.2011.16
- Hill, W. G., and Babiker, H. A. (1995). Estimation of numbers of malaria clones in blood samples. *Proc. Biol. Sci.* 262, 249–257. doi:10.1098/rspb.1995.0203

- Ken-Dror, G., and Hastings, I. M. (2016). Markov chain Monte Carlo and expectation maximization approaches for estimation of haplotype frequencies for multiply infected human blood samples. *Malar. J.* 15, 430. doi:10.1186/s12936-016-1473-5
- Kim, Y., Escalante, A. A., and Schneider, K. A. (2014). A population genetic model for the initial spread of partially resistant malaria parasites under anti-malarial combination therapy and weak intrahost competition. *PLOS ONE* 9, e101601–e101615. doi:10.1371/journal.pone.0101601
- Koopmann, B., Müller, J., Tellier, A., and Živković, D. (2017). Fisher–wright model with deterministic seed bank and selection. *Theor. Popul. Biol.* 114, 29–39. doi:10.1016/j.tpb.2016.11.005
- Li, X., Foulkes, A. S., Yucel, R. M., and Rich, S. M. (2007). An expectation maximization approach to estimate malaria haplotype frequencies in multiply infected children. *Stat. Appl. Genet. Mol. Biol.* 6, 33. doi:10.2202/1544-6115.1321
- Lin, J. T., Hathaway, N. J., Saunders, D. L., Lon, C., Balasubramanian, S., Kharabora, O., et al. (2015). Using amplicon deep sequencing to detect genetic signatures of plasmodium vivax relapse. *J. Infect. Dis.* 212, 999–1008. doi:10.1093/infdis/jiv142
- Lover, A. A., Baird, J. K., Gosling, R., and Price, R. N. (2018). Malaria elimination: Time to target all species. *Am. J. Trop. Med. Hyg.* 99, 17–23. doi:10.4269/ajtmh.17-0869
- McCollum, A. M., Schneider, K. A., Griffing, S. M., Zhou, Z., Kariuki, S., Ter-Kuile, F., et al. (2012). Differences in selective pressure on dhps and dhfr drug resistant mutations in Western Kenya. *Malar. J.* 11, 77. doi:10.1186/1475-2875-11-77
- Ndiaye, Y. D., Hartl, D. L., McGregor, D., Badiane, A., Fall, F. B., Daniels, R. F., et al. (2021). Genetic surveillance for monitoring the impact of drug use on plasmodium falciparum populations. *Int. J. Parasitol. Drugs Drug Resist.* 17, 12–22. doi:10.1016/j.ijpddr.2021.07.004
- Neafsey, D. E., Taylor, A. R., and MacInnis, B. L. (2021). Advances and opportunities in malaria population genomics. *Nat. Rev. Genet.* 22, 502–517. doi:10.1038/s41576-021-00349-5
- Ngwa, C. J., Rosa, T., and Pradel, G. (2016). *The biology of malaria gametocytes*. Rijeka, Croatia: IntechOpen.
- Nicoletti, M. (2020). Three scenarios in insect-borne diseases. *Insect-Borne Dis. 21st Century* 2020, 99–251. doi:10.1016/b978-0-12-818706-7.00005-x
- Nkhoma, S. C., Nair, S., Cheeseman, I. H., Rohr-Allegrini, C., Singlam, S., Nosten, F., et al. (2012). Close kinship within multiple-genotype malaria parasite infections. *Proc. Biol. Sci.* 279, 2589–2598. doi:10.1098/rspb.2012.0113
- Nkhoma, S. C., Trevino, S. G., Gorena, K. M., Nair, S., Khoswe, S., Jett, C., et al. (2020). Co-Transmission of related malaria parasite lineages shapes within-host parasite diversity. *Cell Host Microbe* 27, 93–103. e4. doi:10.1016/j.chom.2019.12.001
- Pacheco, M. A., Forero-Peña, D. A., Schneider, K. A., Chavero, M., Gamardo, A., Figuera, L., et al. (2020). Malaria in Venezuela: Changes in the complexity of infection reflects the increment in transmission intensity. *Malar. J.* 19, 176. doi:10.1186/s12936-020-03247-z
- Plucinski, M. M., Morton, L., Bushman, M., Dimbu, P. R., and Udhayakumar, V. (2015). Robust algorithm for systematic classification of malaria late treatment failures as recrudescence or reinfection using microsatellite genotyping. *Antimicrob. Agents Chemother.* 59, 6096–6100. doi:10.1128/AAC.00072-15
- Rastas, P., Koivisto, M., Mannila, H., and Ukkonen, E. (2005). “A hidden markov technique for haplotype reconstruction,” in *Algorithms in bioinformatics*. Editors R. Casadio and G. Myers (Berlin, Heidelberg: Springer), 140–151. Lecture Notes in Computer Science. doi:10.1007/11557067\_12
- Ross, A., Koepfli, C., Li, X., Schoepflin, S., Siba, P., Mueller, I., et al. (2012). Estimating the numbers of malaria infections in blood samples using high-resolution genotyping data. *Plos One* 7, e42496. doi:10.1371/journal.pone.0042496
- Schneider, K. A. (2021). *Charles Darwin meets ronald Ross: A population-genetic framework for the evolutionary dynamics of malaria*. (Cham: Springer International Publishing), 149–191. chap. 6. doi:10.1007/978-3-030-50826-5\_6
- Schneider, K. A., and Escalante, A. A. (2013). Fitness components and natural selection: Why are there different patterns on the emergence of drug resistance in plasmodium falciparum and plasmodium vivax? *Malar. J.* 12, 15–11. doi:10.1186/1475-2875-12-15
- Schneider, K. A., and Kim, Y. (2010). An analytical model for genetic hitchhiking in the evolution of antimalarial drug resistance. *Theor. Popul. Biol.* 78, 93–108. doi:10.1016/j.tpb.2010.06.005
- Schneider, K. A., and Kim, Y. (2011). Approximations for the hitchhiking effect caused by the evolution of antimalarial-drug resistance. *J. Math. Biol.* 62, 789–832. doi:10.1007/s00285-010-0353-9
- Schneider, K. A. (2018). Large and finite sample properties of a maximum-likelihood estimator for multiplicity of infection. *PloS one* 13, e0194148. doi:10.1371/journal.pone.0194148
- Schneider, K. A., Tsoungui Obama, H. C. J., Kamanga, G., Kayanula, L., and Adil Mahmoud Yousif, N. (2022). The many definitions of multiplicity of infection. *Front. Epidemiol.* 2, 961593. doi:10.3389/fepeid.2022.961593
- Selvaraj, P., Wenger, E. A., and Gerardin, J. (2018). Seasonality and heterogeneity of malaria transmission determine success of interventions in high-endemic settings: A modeling study. *BMC Infect. Dis.* 18, 413–414. doi:10.1186/s12879-018-3319-y
- Shoemaker, W. R., and Lennon, J. T. (2018). Evolution with a seed bank: The population genetic consequences of microbial dormancy. *Evol. Appl.* 11, 60–75. doi:10.1111/eva.12557
- Stephens, M., Smith, N. J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989. doi:10.1086/319501
- Su, Xz, Lane, K. D., Xia, L., Sá, J. M., and Welles, T. E. (2019). Plasmodium genomics and genetics: New insights into malaria pathogenesis, drug resistance, epidemiology, and evolution. *Clin. Microbiol. Rev.* 32, e00019. doi:10.1128/CMR.00019-19
- Sutherland, C. J. (2016). Persistent parasitism: The adaptive biology of malariae and ovale malaria. *Trends Parasitol.* 32, 808–819. doi:10.1016/j.pt.2016.07.001
- Taylor, A. R., Flegg, J. A., Nsoyba, S. L., Yeka, A., Kanya, M. R., Rosenthal, P. J., et al. (2014). Estimation of malaria haplotype and genotype frequencies: A statistical approach to overcome the challenge associated with multiclonal infections. *Malar. J.* 13, 102. doi:10.1186/1475-2875-13-102
- Tellier, A. (2019). Persistent seed banking as eco-evolutionary determinant of plant nucleotide diversity: Novel population genetics insights. *New Phytol.* 221, 725–730. doi:10.1111/nph.15424
- Tseha, S. T. (2021). “Plasmodium species and drug resistance,” in *Plasmodium species and drug resistance*. Editor R. K. Tyagi (Rijeka: IntechOpen). chap. 2. doi:10.5772/intechopen.98344
- Warimwe, G. M., Keane, T. M., Fegan, G., Musyoki, J. N., Newton, C. R., Pain, A., et al. (2009). Plasmodium falciparum var gene expression is modified by host immunity. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21801–21806. doi:10.1073/pnas.0907590106
- Watson, J. A., Nekkab, N., and White, M. (2021). Tafenoquine for the prevention of plasmodium vivax malaria relapse. *Lancet. Microbe* 2, e175–e176. doi:10.1016/S2666-5247(21)00062-8
- White, N. J. (2011). Determinants of relapse periodicity in plasmodium vivax malaria. *Malar. J.* 10, 297. doi:10.1186/1475-2875-10-297
- WHO (2021). *Global technical strategy for malaria 2016–2030*. Geneva, Switzerland: World Health Organization.
- WHO (2021). World malaria report 2020: 20 years of global progress and challenges. Available at: [www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2020](http://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2020).
- Wigger, L., Vogt, J. E., and Roth, V. (2013). Malaria haplotype frequency estimation. *Stat. Med.* 32, 3737–3751. doi:10.1002/sim.5792
- Wong, W., Wenger, E. A., Hartl, D. L., and Wirth, D. F. (2018). Modeling the genetic relatedness of Plasmodium falciparum parasites following meiotic recombination and cotransmission. *PLoS Comput. Biol.* 14, e1005923. doi:10.1371/journal.pcbi.1005923
- World Health Organization, (2017) A framework for malaria elimination.
- Zhong, D., Lo, E., Wang, X., Yewhalaw, D., Zhou, G., Atieli, H. E., et al. (2018). Multiplicity and molecular epidemiology of plasmodium vivax and plasmodium falciparum infections in east Africa. *Malar. J.* 17, 185. doi:10.1186/s12936-018-2337-y
- Zhu, S. J., Hendry, J. A., Almagro-Garcia, J., Pearson, R. D., Amato, R., Miles, A., et al. (2019). The origins and relatedness structure of mixed infections vary with local prevalence of P. falciparum malaria. *eLife* 8, e40845. doi:10.7554/eLife.40845
- Živković, D., and Tellier, A. (2012). Germ banks affect the inference of past demographic events. *Mol. Ecol.* 21, 5434–5446. doi:10.1111/mec.12039