

RESOURCE ARTICLE

The population genetics of partial diapause, with applications to the aestivating malaria mosquito *Anopheles coluzzii*

Rita Mwima^{1,2} | Tin-Yu J. Hui³  | Jonathan K. Kayondo¹ | Austin Burt³¹Department of Entomology, Uganda Virus Research Institute (UVRI), Entebbe, Uganda²Department of Biotechnical and Diagnostic Sciences, College of Veterinary Medicine, Animal Resources and Biosecurity (COVAB), Makerere University, Kampala, Uganda³Department of Life Sciences, Imperial College London, Ascot, UK**Correspondence**

Tin-Yu J. Hui, Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, UK.

Email: tin-yu.hui11@imperial.ac.uk**Funding information**

Wellcome Trust, Grant/Award Number: 224487; Bill and Melinda Gates Foundation, Grant/Award Number: INV006610

Handling Editor: Kimberly J. Gilbert**Abstract**

Diapause, a form of dormancy to delay or halt the reproductive development during unfavourable seasons, has evolved in many insect species. One example is aestivation, an adult-stage diapause enhancing malaria vectors' survival during the dry season (DS) and their re-establishment in the next rainy season (RS). This work develops a novel genetic approach to estimate the number or proportion of individuals undergoing diapause, as well as the breeding sizes of the two seasons, using signals from temporal allele frequency dynamics. Our modelling shows the magnitude of drift is dampened at early RS when previously aestivating individuals reappear. Aestivation severely biases the temporal effective population size (N_e), leading to overestimation of the DS breeding size by $1/(1-\alpha)^2$ across 1 year, where α is the aestivating proportion. We find sampling breeding individuals in three consecutive seasons starting from an RS is sufficient for parameter estimation, and perform extensive simulations to verify our derivations. This method does not require sampling individuals in the dormant state, the biggest challenge in most studies. We illustrate the method by applying it to a published data set for *Anopheles coluzzii* mosquitoes from Thierola, Mali. Our method and the expected evolutionary implications are applicable to any species in which a fraction of the population diapauses for more than one generation, and are difficult or impossible to sample during that stage.

KEYWORDS

aestivation, diapause, genetic drift, persistence, temporal allele frequency

1 | INTRODUCTION

Many insect species have evolved the mechanism of diapause, a form of dormancy, which halts development and reproduction under unfavourable environmental conditions (Cloutier et al., 2021; Ojima et al., 2018). Some forms of diapause may also include reduced activity such as flying and feeding. Metabolism is suppressed, and most energy resources are diverted towards

survival. Diapause is pre-programmed and genetically coded, then activated by changing of environments (Ojima et al., 2018). Depending on the species, known stimuli include day length, temperature and access to water. Examples of species that undergo diapause include *Drosophila suzukii*, *Phormia regina* (Queen blow flies), *Danaus plexippus* (Monarch butterfly), multiple moth species, tropical walking-stick insects and many more (Herman & Tatar, 2001; Salman et al., 2019; Tatar & Yin, 2001;

Rita Mwima and Tin-Yu J. Hui should be considered joint first authors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

Toxopeus et al., 2016; Yoder & Denlinger, 1992). While the definition of what being “unfavourable” varies across species, as does the life stage (embryonic, larval or adult) in which diapause occurs, the common goal is to increase organismal survival, and ultimately enable them to re-establish when the next favourable season arrives (Denlinger & Armbruster, 2014). If the duration of the unfavourable season lasts for multiple generations, then diapause has a direct impact on the rate of evolution and the genetic constitution of the population.

In many locations, malaria vector populations exhibit strong seasonal fluctuations in abundance, being present in large numbers during the rainy season (RS) and then drop to low levels or even become undetectable when the breeding sites disappear in the dry season (DS) (Charlwood et al., 1995; Hidalgo et al., 2016; Kabbale et al., 2013). Various hypotheses have been suggested to account for their persistence in the DS and the source of their re-establishment soon after the first rains. For some vector species, like *Anopheles coluzzii*, aestivation, a DS adult-state diapause, is thought to be the primary route (Charlwood et al., 2000; Lehmann et al., 2010; Wang et al., 2011). Aestivation is in itself an interesting biological process which predominantly activated by the absence of water (Depinay et al., 2004). However, using conventional mosquito trapping methods, it is almost impossible to detect aestivators; thus, they are thought to occupy distinct hidden habitats during this state (Charlwood et al., 2000; Lehmann et al., 2017). There are other hypotheses, such as long-distance migration, which assumes (near) extinction of the local DS population then recolonisation by migrants from further afield where breeding sites exist year-round (Adamou et al., 2011; Mwima et al., 2023). Another is local refugia, where populations survive locally but hidden from sampling, and maintain themselves by low-level breeding during the DS, but this implies the constant availability of larval sites (Omer & Cloudsley-Thompson, 1970).

Many studies have focused on finding direct evidence of aestivating mosquitoes, notably via the mark–release–recapture (MRR) experiment. For example, Lehmann et al. (2010) recaptured one marked female *An. coluzzii* after it had endured a 7-month-long DS in the Sahelian region. More recently, a modified MRR using hydrogen isotopes as markers estimated that aestivation contributed at least 20% in census size to the persistence of *An. coluzzii* in the Sahelian villages of Mali until the following RS (Faiman et al., 2022). One obvious challenge of the direct methods is the sampling of aestivators, that these laborious MRR provided a disproportionate amount of evidence with low recapture rates (Epopa et al., 2017). There have been fewer indirect (i.e. genetic) studies. Lehmann et al. (2017) explored the population genetics of several Malian *An. coluzzii* populations through successive seasons, and attempted to interpret the results in the context of possible aestivation, noting that drift should be negligible in aestivating populations due to reproductive arrest, but the qualitative nature of the work did not present a formal model or statistical testing. There have been other studies to test whether populations undergo annual bottlenecks (Lehmann et al., 1998; Luikart et al., 1998), but these drift-based analyses are limited to

the DS only. They did not directly infer aestivation nor provide any means for parameter estimation.

In this work, we first study from a theoretical perspective the magnitude of genetic drift under diapause. If we define temporal effective population size (N_e) as that estimated from the change in allele frequencies between two time points under the idealised Wright–Fisher (WF) model (Waples, 1989), we can address several interesting questions: How does diapause affect temporal N_e estimates, if samples are collected across seasons? What are these N_e estimates referring to? What happens if only a fraction of individuals undergo diapause, while the others continue to breed? Can we estimate the number (or proportion) of individuals entering the diapause stage during the unfavourable season from allele frequency dynamics? Below we give answers by mathematical derivations and computer simulations.

We use aestivation to represent diapause (and the DS as the unfavourable season) in the following paragraphs. It is because the worked example of the genotypic data set primarily concerns *An. coluzzii*. Note that the theory is also applicable to other diapausing species (see Discussion, below).

1.1 | Theory

1.1.1 | Standardised variance of allele frequency

Without other forces (migration, selection, mutation), allele frequency fluctuates randomly due to genetic drift. While the mean allele frequency remains unchanged over time, the variance increases. Under the WF model, the standardised variance of allele frequency V_t (Waples, 1989) increases by approximately $1/2N_e$ per generation:

$$V_{t+1} = V_t + \frac{1}{2N_e} \quad (1)$$

(full derivations of this and other equations in this section can be found in the Appendix S1). A relevant statistic that can be derived from V_t is the temporal F , which is the standardised variance of allele frequency change (Waples, 1989). If two samples are taken Δ_t generations apart, then F is approximately:

$$F \approx \frac{\Delta_t}{2N_e} \quad (2)$$

from which N_e can be solved or estimated (Waples, 1989). If N_e is not a constant, then this method estimates the harmonic mean N_e of all the generations between the two samples (Waples, 2005).

To extend the model to incorporate aestivation, we assume two seasons per cycle (or year, for most organisms), the RS and DS, lasting Δ_R and Δ_D generations, and with breeding population sizes N_R and N_D , respectively. While the population still reproduces under the WF manner between most generations, the transitions between seasons are of particular interest. These include the generations of ERS (E for early, the first generation of RS), LRS (L for late, the last generation of RS) and similarly EDS and LDS. Upon entering EDS, N_D offspring

enter the breeding compartment, and N_A to another compartment to aestivate (N_A is the number that survives aestivation to re-emerge the next ERS; individuals that die during aestivation do not affect allele frequencies and are ignored). The breeding compartment continues to reproduce during DS but with a smaller size, while reproduction halts for the aestivators. The two compartments remain isolated throughout the DS. At the next ERS, those N_A aestivators reappear and unite with the descendants of the breeding individuals from the LDS. This completes the population dynamics of one cycle. The model can be visualised in Figure 1, and the notations used in the paper are found in Table 1.

With this model in mind, we then quantify the effect of aestivation on the temporal allele frequency dynamics, such as how V_t changes over time, the expected temporal F and the interpretations of N_e estimates. Given that the model begins with the ERS, V_t follows the WF for all generations up to the LDS, but under two different breeding sizes. For instance, we have

$$V_{EDS} = V_{ERS} + \frac{\Delta_R - 1}{2N_R} + \frac{1}{2N_D} \tag{3}$$

where Δ_R is the duration of the RS (similarly for Δ_D). The next ERS (denoted as ERS') can be viewed as a mixture of the two compartments, one experiencing no drift (other than that associated with the establishment of the aestivating class) and another with Δ_D generations of drift. Then, $V_{ERS'}$ is calculated from their standardised variances and covariance (see SI):

$$V_{ERS'} = V_{LRS} + \left(\frac{N_R - N_A}{N_R}\right)^2 \left(\frac{\Delta_D}{2N_D}\right) + \frac{1}{2N_R} = V_{LRS} + (1 - \alpha)^2 \left(\frac{\Delta_D}{2N_D}\right) + \frac{1}{2N_R} \tag{4}$$

We further introduce $\alpha = \frac{N_A}{N_R}$ as the fraction of aestivators contributing to ERS'. One can quickly examine the limiting cases: For complete reproductive arrest, $\alpha = 1$, then ERS' is in effect the next generation of LRS hence with only one generation of drift with breeding size N_R . If $\alpha = 0$, then it reduces to the pure WF process with Δ_D generations of drift in the DS, and one generation when the population enters the next RS.

Note that the variance at the beginning of an RS ($V_{ERS'}$) can be lower than that from the previous generation (V_{LDS}):

$$V_{ERS'} - V_{LDS} = -\alpha(2 - \alpha) \left(\frac{\Delta_D}{2N_D}\right) + \frac{1}{2N_R} \tag{5}$$

The difference is negative when

$$\alpha > 1 - \sqrt{1 - \frac{N_D}{N_R \Delta_D}}$$

1.1.2 | Temporal F and N_e estimates

With V_t no longer monotonically increasing we need to examine its impact on temporal N_e estimates. Here, our derivations focus on the following three-time points: ERS, EDS and ERS' (Figure 1), from which three temporal N_e estimates can be calculated. Without much ambiguity, the expected temporal F between ERS and EDS is

$$F_{ERS,EDS} = \frac{\Delta_R - 1}{2N_R} + \frac{1}{2N_D} \tag{6}$$

As there are Δ_R generations between them, the corresponding temporal N_e estimate refers to the harmonic mean of $(\Delta_R - 1)$ generations of size N_R and one generation of size N_D . A more complex scenario is when the two samples are from exactly one cycle apart from successive ERS's (Appendix S1):

$$F_{ERS,ERS'} = \frac{\Delta_R}{2N_R} + (1 - \alpha)^2 \left(\frac{\Delta_D}{2N_D}\right) \tag{7}$$

The temporal N_e estimate between two consecutive ERS is equivalent to that of the WF process with Δ_R generations of size N_R , and Δ_D of $\frac{N_D}{(1 - \alpha)^2}$.

Finally, between EDS and ERS' (Appendix S1):

$$F_{EDS,ERS'} = \frac{2\alpha - 1}{2N_D} + (1 - \alpha)^2 \left(\frac{\Delta_D}{2N_D}\right) + \frac{1}{2N_R} \tag{8}$$

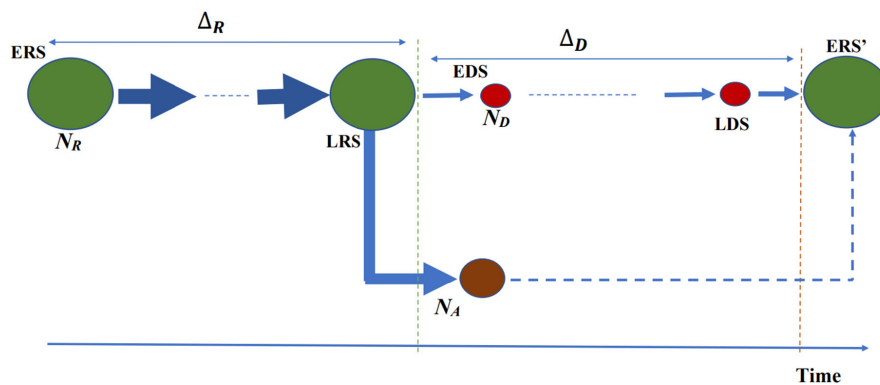


FIGURE 1 Diagram for the aestivation model. The model begins at ERS with breeding size N_R . Within the same season, the population reproduces (horizontal arrows) according to the WF model. Upon entering EDS, the population branches into two compartments: one continues to breed but with a smaller size N_D , another aestivates with N_A individuals. At the next ERS (denoted as ERS'), it is formed by previously aestivating individuals, and the descendants of the breeding compartment from LDS.

TABLE 1 Notation used in this study.

| | |
|------------|---|
| RS | Rainy season |
| DS | Dry season (the unfavourable season for aestivation) |
| ERS (EDS) | The first generation of RS (DS) |
| LRS (LDS) | The final generation of RS (DS) |
| ERS' | ERS of the next cycle |
| N_R | RS breeding size |
| N_D | DS breeding size |
| N_A | Size of the aestivating (diapause) compartment |
| α | N_A / N_R |
| Δ_R | Length of RS in generations |
| Δ_D | Length of DS in generations |
| V_t | Standardised variance of allele frequency, at generation t |
| F | Temporal F , standardised variance of allele frequency change (between two time points) |

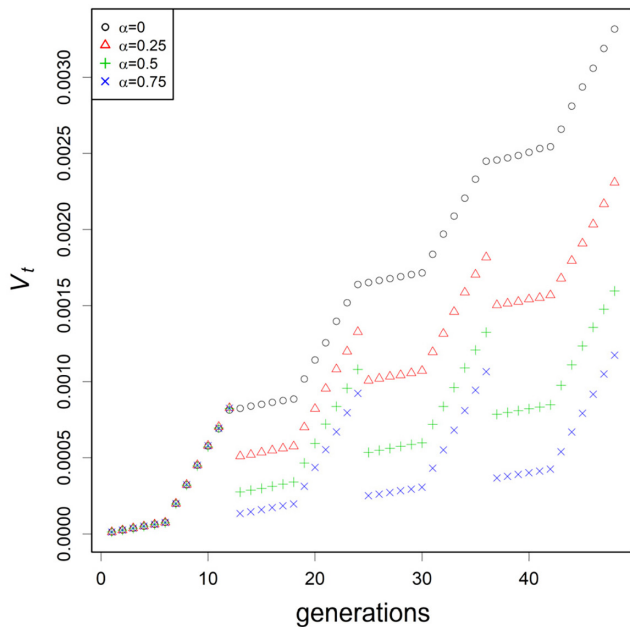


FIGURE 2 Simulated V_t over time under different aestivating proportion α . Four values of α were examined. Note that $\alpha = 0$ is equivalent to WF. Simulation parameters are as follows: $N_R = 10000$, $N_D = 1000$, $N_A = \alpha N_R$, $\Delta_R = \Delta_D = 6$. V_t were calculated from 10,000 independent loci, with an initial frequency of 0.5. Four complete cycles (48 generations) were simulated, starting from an RS.

With Δ_D generations between them, we are unable to express it analytically as any kind of harmonic mean within the sampling horizon. Note that Equations 6–8 are functions of the three parameters, hinting at the potential use of the three temporal F 's for parameter estimation. Another interesting observation is that when there is aestivation ($N_A > 0$), the sum of the two shorter F 's exceeds the overall F , with the difference determined by the three sizes:

$$F_{ERS,EDS} + F_{EDS,ERS'} = F_{ERS,ERS'} + \frac{N_A}{N_R N_D} \quad (9)$$

2 | METHODS

We built a forward WF simulator with breeding and aestivating compartments to verify the above derivations. Biallelic loci were simulated and the allele frequencies of the breeding compartment over time were recorded. The model further assumed non-overlapping, discrete generations, with random mating. The three sizes are non-negative with $N_R \geq N_A + N_D$, and Δ_R and Δ_D are known. The first simulation aimed to visualize the V_t dynamics under different aestivation proportions and compare them to our theoretical expectations. Then, we aimed to validate the temporal F 's we derived, focusing only on ERS, EDS and ERS'.

3 | RESULTS

Simulation results can be found in Figures 2 and 3, and Appendix S1. Figure 2 plots V_t over time for several values of α . Within a cycle starting from the ERS, V_t increases with two different slopes because of the two different breeding sizes. The shape of V_t within one cycle was the same for all α , as the breeding sizes were the same. The only difference was between LDS and the next ERS when aestivating individuals reappear. Larger α gave a bigger drop in V_t between the two generations, concurring with our theory (Equation 4).

In the second simulation, we focused on the three pairs of temporal F 's among ERS, EDS and ERS'. Each box plot and error bar in Figure 3 came from 1000 repeated simulations. Between the first two time points, different values of α produced the same mean F as no aestivation was involved. The widths of the error bars appeared to be similar as well. Both $F_{EDS,ERS'}$ and the overall $F_{ERS,ERS'}$ decreased with α . Because of their inversely proportional relation, larger α gave larger N_e estimates, despite having the same breeding sizes. In short, the results all agreed with our theoretical expectations (Equations 6–8).

4 | PARAMETER ESTIMATION ON A REAL DATA SET

To illustrate how the theory can be used to estimate parameters, we searched for data sets in which the same population had been genotyped at multiple loci in three successive seasons (RS, DS, RS'). The most appropriate we found was the study of Lehmann et al. (2017), though it is not clear it meets all the assumptions of the model (see Discussion, below). We particularly focused on the temporal *An. coluzzii* samples collected between 2008 and 2010 from the Sahelian village of Thierola in Mali. Sample sizes per time point varied from about 30 to 60, and were scored for 738 SNPs.

To estimate parameters, we used Approximate Bayesian Computation (ABC; Beaumont et al., 2002). As the data set spanned 2 years, we analysed the two DS separately, with non-overlapping intervals. For the first DS, the three temporal samples were collected in August 08, May 09 and October 09, and for the second

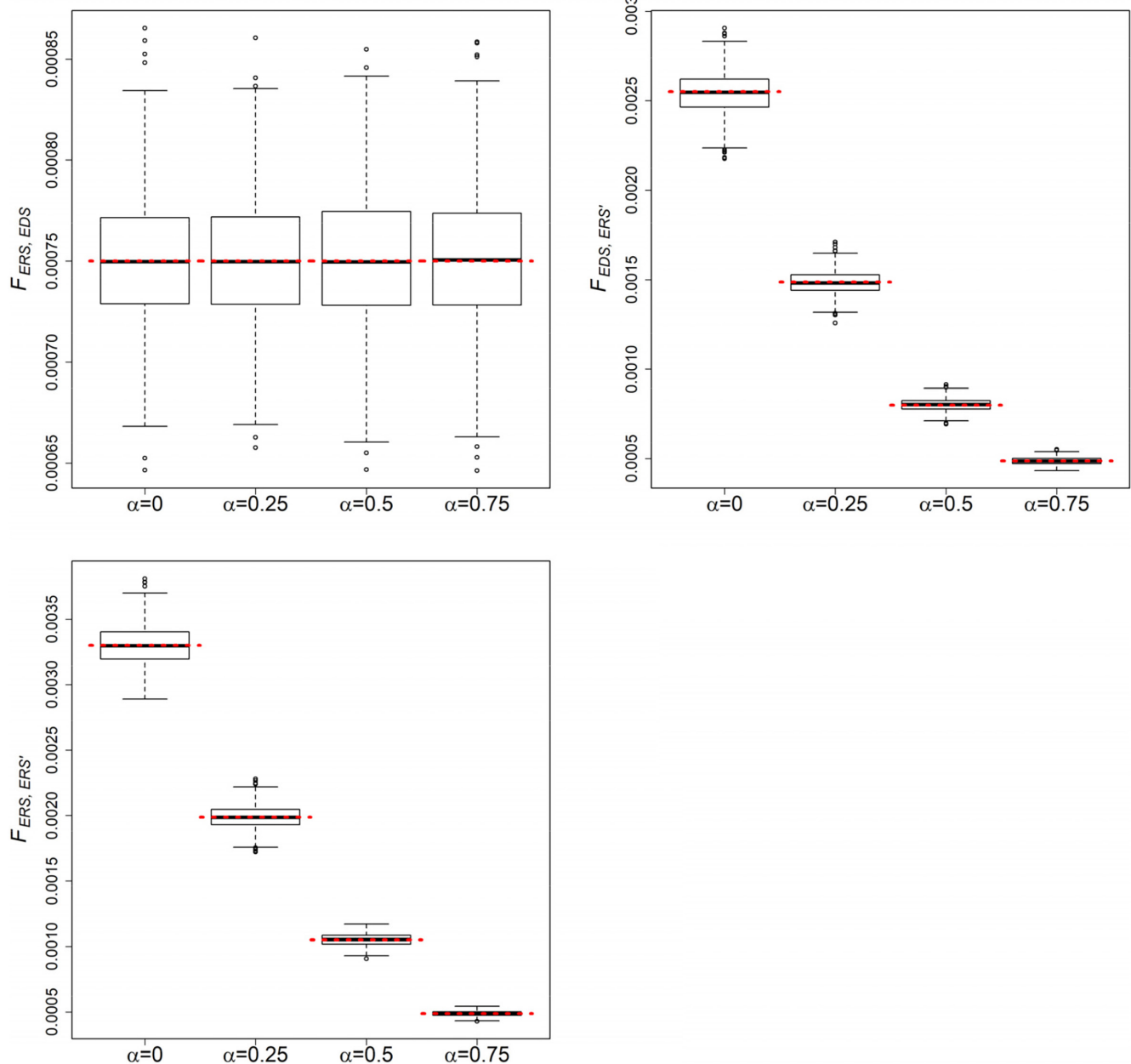


FIGURE 3 Simulated temporal F statistics against α . The parameters were identical to those stated in Figure 1. Each box plot was generated from 1000 repeated simulations. The red dotted lines show the theoretical expectations of F 's (Equations 7–9). Exhaustive sampling was assumed in this simulation.

DS, they were collected in October 09, April 10 and June 10. As the generation time was assumed to be 1 month (Ag1000G, 2017), the sampling horizons for the 2 years were 14 and 8 generations respectively. According to the original authors, the DS runs from December to May (Lehmann et al., 2017). For summary statistics, we used the three temporal F 's, which were computed via F_c to account for finite sampling and missing data (Peel et al., 2013; Pollak, 1983; Waples, 1989, see also Appendix S1).

In each ABC simulation, a burn-in period of 24 generations (2 cycles) was run. The initial allele frequency was sampled from $U(0.2, 0.8)$, where $U(a, b)$ denotes a uniform distribution with bounds a and b . For the priors, N_D was drawn from $U(10, 1000)$ based on preliminary

studies on it being not too large; $N_R | N_D \sim U(N_D + 2, 100,000)$; and $N_A | N_R, N_D \sim U(1, N_R - N_D)$ conditionally. Finally, samples were taken binomially from the population, with sample sizes equal to those of the real data set. The simple Euclidean distance between the observed and simulated summary statistics was calculated. For each year, 2 million ABC simulations were run, and the 1000 (0.05%) with the smallest distances were accepted as realizations of the posteriors. We report the medians from the posterior distributions, and the 90% confidence interval (CI) calculated by excluding the 5% on either extreme.

The inferred posterior distributions from the ABC are shown in Figures 4 and 5. For the first year, the posterior medians for N_R, N_D

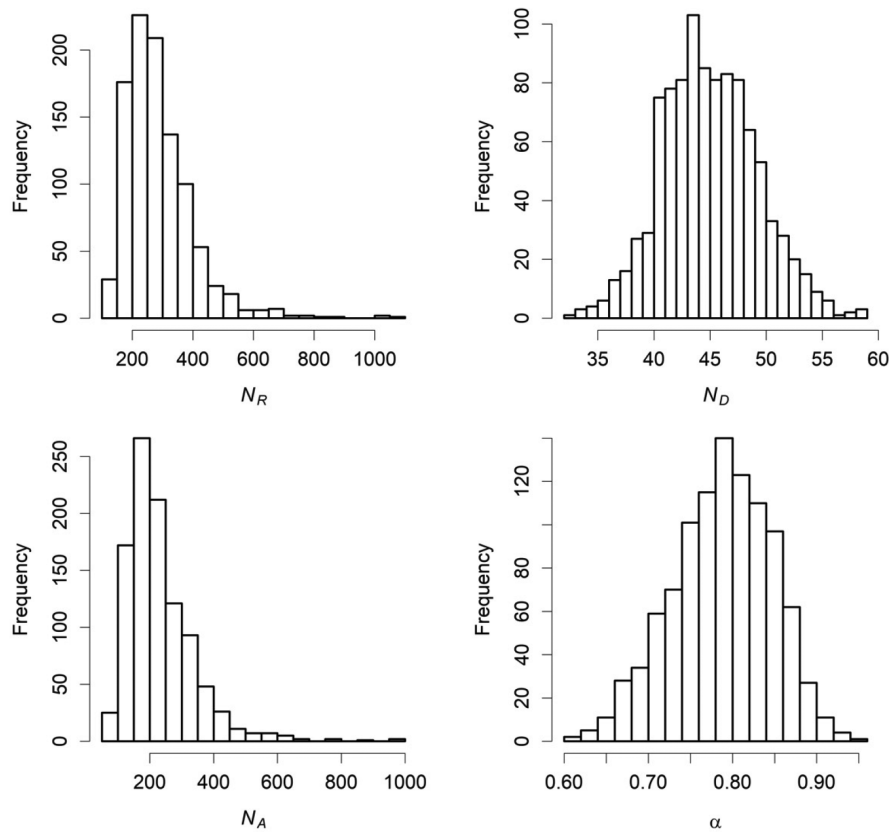


FIGURE 4 Posterior distributions for the first DS (2008–2009). The 90% C.I. were estimated as follows: N_R : [163, 486], N_D : [39, 53], N_A : [113, 420], α : [0.68, 0.88]. The corresponding temporal F estimates were 0.0714, 0.0669 (shorter term) and 0.0290 (overall).

and N_A were 265, 45 and 210, respectively. In addition to the three sizes, the posterior of α was also computed from the pairs of accepted N_A and N_R . The posterior median for α was 0.79. For the second year, the posterior median for N_D and N_A were 578 and 15,972. Note that the posterior N_A was highly skewed. The result for N_R was inconclusive because the posterior was flat (see discussion below). The posterior of α had a sharp peak with a median of 0.39. CIs are reported in the figure legends.

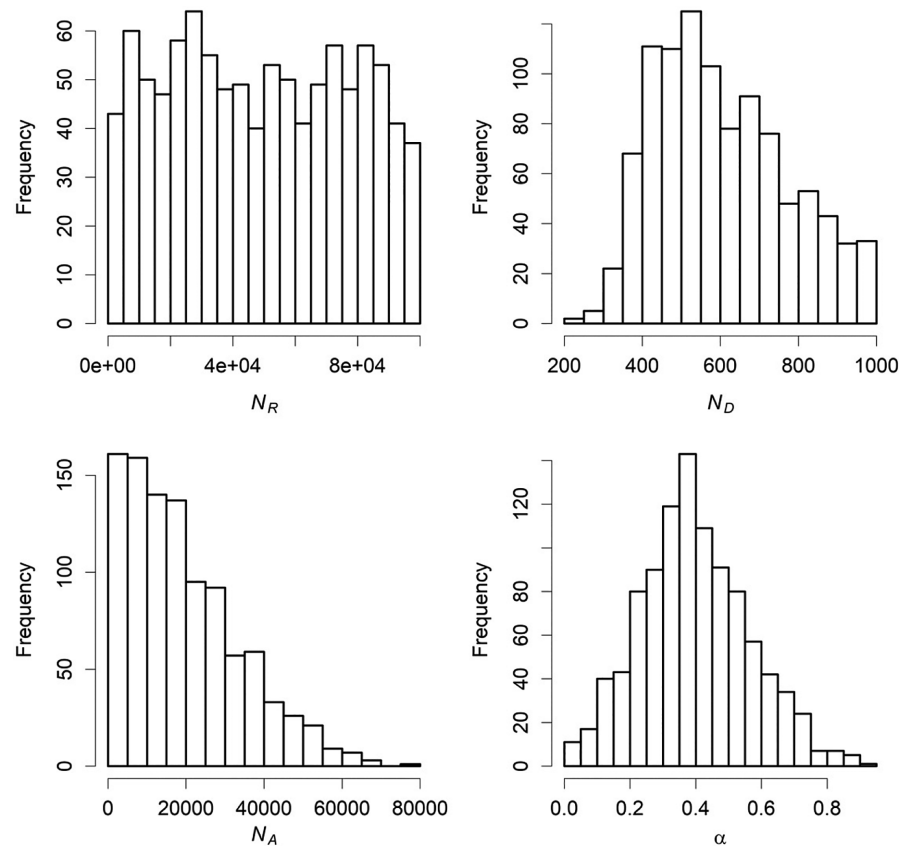
5 | DISCUSSION

Diapause allows many insect species to survive through their unfavourable seasons. Its interesting mechanisms have been studied for model organisms such as *Drosophila melanogaster* (Ojima et al., 2018; Piper & Partridge, 2016). Many genera of mosquito are also known to undergo diapause and at different life stages: egg and larva diapause occurs in *Aedes*, while adult diapause is usually found in *Culex* and *Anopheles* (Denlinger & Armbruster, 2014). The mechanisms by which reproductive arrest is achieved can be different: some take blood meals but do not develop eggs, some do neither (Yaro et al., 2012). Even within the same species, diapause may or may not be activated depending on the climate the populations live in (Lee & Duvall, 2022). For most species, mosquitoes and beyond, the unfavourable season is usually the harsh winter, but for *An. coluzzii*, this becomes the tropical DS in the Sahel region. Aestivation, an adult-stage diapause, provides a route for the populations to persist then rebound when the rain returns. Aestivation, by its very

nature, is difficult to study because the individuals in this state are difficult or impossible to collect (Lehmann et al., 2017). In this work, we have explored how an indirect approach, using allele frequency changes across seasons, can be used to estimate the proportion of aestivation. Our model extends the classical WF model by incorporating two separate compartments in the DS for partial aestivation. Here, the meaning of partial is twofold: First, our model still allows breeding during the DS under a smaller size, as opposed to the more restrictive complete reproductive arrest (Lehmann et al., 2017). Second, $(N_A + N_D)$ can be smaller than N_R such that the total population size can be smaller in the DS. The formulae we derived for V_t and temporal F assumed that only the breeding compartment is observed. One advantage of this method is it does not require sampling of aestivating individuals, which thus far have been difficult or almost impossible to find, while sampling breeding DS individuals is more feasible (Lehmann et al., 2017; Ag1000G, 2020).

Cross-seasonal dynamics are crucial to understanding the population genetic consequences of aestivation. One characteristic is the drop in V_t at every ERS'. Previously aestivating individuals, who have experienced no drift through the DS, reappear and admix with the rest of the population, lowering the combined V_t . If the proportion of the RS population that aestivates (α) is sufficiently large, then V_t is severely dampened or even reset to the level similar to the LRS. Populations with huge seasonal variation in population density, such as *An. coluzzii*, whose N_D/N_R can be as low as 0.01 (Bomblies & Eltahir, 2009; Khatri & Burt, 2019; Mabaso et al., 2007; Minakawa et al., 2002), require only a tiny α to give such an effect (Equation 5).

FIGURE 5 Posterior distributions for the second DS (2009–2010). The 90% C.I. were estimated as follows: N_D : [365, 918], N_A : [1430, 47385], α : [0.13, 0.68]. The estimates for N_R were not reported as its posterior was flat. The corresponding temporal F estimates were 0.0071, 0.0024 (shorter term) and 0.0036 (overall).



Signals from allele frequency over time have been utilised for N_e estimation, often via the well-studied temporal F statistic (Hui et al., 2021; Waples, 1989). Before this work, little was known about the interpretation of these estimates for aestivating populations, especially when the two samples are taken across different seasons or sandwich an ERS. The temporal F across two consecutive ERS overestimates N_D by a factor of $1/(1-\alpha)^2$ (Equation 7). It may therefore give the false impression of having a very large DS breeding size, contradictory to within-season estimates or the ecology of the species of interest. Note that temporal N_e estimate between EDS and ERS' (Equation 8) has more terms than the number of generations between them, presumably because there are two routes linking them, one via the DS breeding compartment, and another via LRS (backward in time) then through the aestivating compartment.

Equations (6–8) show that sampling from three consecutive seasons starting with an RS is sufficient to estimate the three different population sizes (N_R, N_D, N_A), which serves as the basis for our ABC. With aestivation, the sum of the two shorter term F 's is greater than the overall F , and the difference increases with α (Equation 9). As a result, the overall N_e estimate will be larger than either of the two shorter term estimates. While the theory was developed assuming samples were taken at the first generation of each season, it could be modified to consider samples taken in most other generations from their respective seasons, with slightly different expectations. The theory could also be extended to consider populations with overlapping generations (Waples & Yokota, 2007).

The inequalities between the shorter and longer term temporal F 's from the Thierola samples (Equation 9, and the legends of Figures 4 and 5) are indicators of aestivation. The first-year DS effective breeding size was estimated to be about 45, several times smaller than that of the RS. For the second year, N_D was also small compared to its corresponding RS, but was larger than the previous year. This result concurred with the original publication that the first year was genetically less stable (Lehmann et al., 2017). In both years, the C.I. of N_D and α excluded 0, suggesting the coexistence of aestivation and reproduction in the DS (partial aestivation). The N_R estimate for the second year was inconclusive as the posterior was flat and almost identical to the uniform prior. We believe this was caused by two factors: First, the sample size might be inadequate, that sampling noise overwhelmed the relatively weak drift signals (Waples, 1989). This was further supported by the small temporal F estimates (Figure 5). The second possibility was the relatively short sampling interval. It began rather late in RS (October) and ended at the next ERS (June); hence, it may not have covered enough RS generations. The original authors also found very little differentiation among the samples collected from October '09 and beyond, which also explained the challenges we faced when estimating the parameters for the second year. While the N_R estimate was inconclusive, we could still estimate α with precision. This is not uncommon in population genetics, that only composite or ratio parameters can be resolved when information is limited.

The proportion of the population that aestivates is unlikely to be a fixed characteristic of the species, but rather to vary

geographically, depending on the severity of the DS. Indeed, there is likely to be an adaptive cline in aestivation phenotypes, given their heterogeneity between climate regions (Yaro et al., 2012). In areas such as the Equatorial where surface waters are found year-round, α is expected to be 0 (see Appendix S1). In contrast, Sahelian *An. coluzzii* requires specific mechanism like aestivation to persist through the DS (Mwima et al., 2023). We believe partial aestivation occurs along this gradient, but this hypothesis requires further investigation. There are other potential factors that may affect the α estimate, such as the switching of individuals between the two compartments in the DS, or if aestivators begin their dormancy at different DS generations.

As in other temporal N_e studies, perfect knowledge of generation time is usually assumed. Additionally, we require information on seasonality and its transition. The beginning of RS is usually associated with the first rain and thus does not always fall on the exact same date or week every year. Sampling too close to the beginning of each season may risk obtaining individuals that are from the previous one. The proposed sampling regime of RS-DS-RS is only an example, and further investigations are encouraged. As mentioned, our model and derivations are applicable to any diapause scenario that lasts for more than one generation, which is exactly the case for most short-lived insect species.

Key assumptions for our model and method of analysis are (1) that individuals collected during the DS are reproductive, and (2) that the population under study is closed, without significant genetic impacts due to immigration over the course of the study. It is not clear how well the population of *An. coluzzii* from Theriola fits these assumptions. For example, no breeding sites were found during the DS, and it is possible that individuals collected then were aestivators that briefly became active in order to feed before returning to aestivate. It will be worthwhile to expand the model to allow for this possibility, and for possible large-scale immigration at the beginning of the RS, which has also been proposed for these populations (Dao et al., 2014; Huestis et al., 2019). Each of these processes should leave a distinguishable genetic signature which will allow appropriate statistical inference. We hope our work will stimulate and accelerate relevant studies, to shed light on questions such as why aestivating proportions vary between populations and years, and how they interact with vector control programmes. All else being equal, aestivation is expected to slow down the rate of evolution in calendar time, including insecticide resistance. Considering genetic control strategies, a large aestivating population will slow down the spread of a gene drive (North et al., 2020), but can also serve as a reservoir for the gene drive mosquitoes so that they do not locally go extinct after a DS, potentially having an effect similar to repeated releases (Eckhoff et al., 2017). The consequences hence vary from case to case, requiring further analyses. More generally, understanding the persistence mechanisms is essential for all vector control programmes to identify and target the susceptible phases of the vector populations (Lehmann et al., 2017). It is foreseeable that long-distance migration and aestivation, the two main hypotheses of their

persistence and re-establishment in highly seasonal environments, require different vector control strategies. Future genetic investigations will be necessary to help differentiate these processes with quantitative testing, but they will almost certainly require intensive sampling and a combination of genetic parameters, such as linkage disequilibrium, clustering and kinship.

AUTHOR CONTRIBUTIONS

R.M., T-Y.J.H. and A.B. conceived the research. T-Y.J.H. coded the computer models and simulations. R.M. and T-Y.J.H. wrote the manuscript with input from the other authors.

ACKNOWLEDGEMENTS

The authors would like to thank John B. Connolly and Bhavin S. Khatri, for their support and helpful comments towards this work. We also thank Tovi Lehmann and an anonymous reviewer for their feedbacks on the previous draft. All authors receive funding through Target Malaria, which is supported by a grant from the Bill & Melinda Gates Foundation and from the Open Philanthropy Project Fund, an advised fund of the Silicon Valley Community Foundation. This work is supported by Wellcome Trust (224487).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. The funding bodies had no direct role in the design of the study nor in the collection, analysis, interpretation of data or in the writing of the manuscript.

DATA AVAILABILITY STATEMENT

The data set analysed in the paper comes from Lehmann et al. (2017) (URL: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.t5k6f>). Computer scripts are available on Github: <https://github.com/tinyuhui/aestivation>.

ORCID

Tin-Yu J. Hui  <https://orcid.org/0000-0002-1702-803X>

REFERENCES

- Adamou, A., Dao, A., Timbine, S., Kassogu , Y., Yaro, A. S., Diallo, M., Traor , S. F., Huestis, D. L., & Lehmann, T. (2011). The contribution of aestivating mosquitoes to the persistence of *Anopheles gambiae* in the Sahel. *Malaria Journal*, 10(April 2016), 1–9. <https://doi.org/10.1186/1475-2875-10-151>
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4), 2025–2035. <https://doi.org/10.1093/genetics/162.4.2025>
- Bombli s, A., & Eltahir, E. A. B. (2009). Assessment of the impact of climate shifts on malaria transmission in the Sahel. *EcoHealth*, 6(3), 426–437. <https://doi.org/10.1007/s10393-010-0274-5>
- Charlwood, J. D., Kihonda, J., Sama, S., Billingsley, P. F., Hadji, H., Verhave, J. P., Lyimo, E., Luttkhuizen, P. C., & Smith, T. (1995). The rise and fall of *Anopheles arabiensis* (Diptera: Culicidae) in a Tanzanian village. *Bulletin of Entomological Research*, 85(1), 37–44. <https://doi.org/10.1017/S0007485300051993>
- Charlwood, J. D., Vij, R., Billingsley, P. F., Charlwood, D. J., Vij, R., & Billingsley, F. P. (2000). Dry season refugia of malaria-transmitting mosquitoes in a dry savannah zone of east Africa. *American Journal*

- of *Tropical Medicine and Hygiene*, 62(6), 726–732. <https://doi.org/10.4269/ajtmh.2000.62.726>
- Cloutier, C., Guay, J. F., Champagne-Cauchon, W., & Fournier, V. (2021). Overwintering survival of *Drosophila suzukii* (Diptera: Drosophilidae) in temperature regimes emulating partly protected winter conditions in a cold-temperate climate of Québec, Canada. *Canadian Entomologist*, 153(3), 259–278. <https://doi.org/10.4039/tce.2021.6>
- Dao, A., Yaro, A. S., Diallo, M., Timbiné, S., Huestis, D. L., Kassogué, Y., Traoré, A. I., Sanogo, Z. L., Samaké, D., & Lehmann, T. (2014). Signatures of aestivation and migration in Sahelian malaria mosquito populations. *Nature*, 516(7531), 387–390.
- Denlinger, D. L., & Armbruster, P. A. (2014). Mosquito diapause. *Annual Review of Entomology*, 59, 73–93. <https://doi.org/10.1146/annurev-ento-011613-162023>
- Depinay, J. M. O., Mbogo, C. M., Killeen, G., Knols, B., Beier, J., Carlson, J., Dushoff, J., Billingsley, P., Mwambi, H., Githure, J., Toure, A. M., & McKenzie, F. E. (2004). A simulation model of African anopheles ecology and population dynamics for the analysis of malaria transmission. *Malaria Journal*, 3, 1–21. <https://doi.org/10.1186/1475-2875-3-29>
- Eckhoff, P. A., Wenger, E. A., Godfray, H. C. J., & Burt, A. (2017). Impact of mosquito gene drive on malaria elimination in a computational model with explicit spatial and temporal dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 114(2), E255–E264. <https://doi.org/10.1073/pnas.1611064114>
- Epopa, P. S., Millogo, A. A., Collins, C. M., North, A., Tripet, F., Benedict, M. Q., & Diabate, A. (2017). The use of sequential mark-release-recapture experiments to estimate population size, survival and dispersal of male mosquitoes of the *Anopheles gambiae* complex in Bana, a west African humid savannah village. *Parasites and Vectors*, 10(1), 1–15. <https://doi.org/10.1186/s13071-017-2310-6>
- Faiman, R., Yaro, A. S., Dao, A., Sanogo, Z. L., Diallo, M., Samake, D., Yossi, O., Veru, L. M., Graber, L. C., Conte, A. R., Kouam, C., Krajacich, B. J., & Lehmann, T. (2022). Isotopic evidence that aestivation allows malaria mosquitoes to persist through the dry season in the Sahel. *Nature Ecology & Evolution*, 6(11), 1687–1699. <https://doi.org/10.1038/s41559-022-01886-w>
- Herman, W. S., & Tatar, M. (2001). Juvenile hormone regulation of longevity in the migratory monarch butterfly. *Proceedings of the Royal Society B: Biological Sciences*, 268(1485), 2509–2514. <https://doi.org/10.1098/rspb.2001.1765>
- Hidalgo, K., Siauxsat, D., Braman, V., Dabiré, K. R., Simard, F., Mouline, K., & Renault, D. (2016). Comparative physiological plasticity to desiccation in distinct populations of the malarial mosquito *Anopheles coluzzii*. *Parasites and Vectors*, 9(1), 1–13. <https://doi.org/10.1186/s13071-016-1854-1>
- Huestis, D. L., Dao, A., Diallo, M., Sanogo, Z. L., Samake, D., Yaro, A. S., Ousman, Y., Linton, Y.-M., Krishna, A., Veru, L., Krajacich, B. J., Faiman, R., Florio, J., Chapman, J. W., Reynolds, D. R., Weetman, D., Mitchell, R., Donnelly, M. J., Talamas, E., ... Lehmann, T. (2019). Windborne long-distance migration of malaria mosquitoes in the Sahel. *Nature*, 574(1), 404–408. <https://doi.org/10.1038/s41586-019-1622-4>
- Hui, T. Y. J., Brenas, J. H., & Burt, A. (2021). Contemporary N_e estimation using temporally spaced data with linked loci. *Molecular Ecology Resources*, 21(7), 2221–2230. <https://doi.org/10.1111/1755-0998.13412>
- Kabbale, F. G., Akol, A. M., Kaddu, J. B., & Onapa, A. W. (2013). Biting patterns and seasonality of anopheles gambiae sensu lato and anopheles funestus mosquitoes in Kamuli District, Uganda. *Parasites and Vectors*, 6(1), 1–9. <https://doi.org/10.1186/1756-3305-6-340>
- Khatri, S. B., & Burt, A. (2019). Robust estimation of recent effective population size from number of independent origins in soft sweeps. *Molecular Biology and Evolution*, 36(7), 2040–2052.
- Lee, I. H., & Duvall, L. B. (2022). Maternally instigated diapause in aedes albopictus: Coordinating experience and internal state for survival in variable environments. *Frontiers in Behavioral Neuroscience*, 16, 778264.
- Lehmann, T., Dao, A., Yaro, S. A., Adamou, A., Kassogue, Y., Diallo, M., Sekou, T., & Coscaron-Aras, C. (2010). Aestivation of the African malaria mosquito, *Anopheles gambiae* in the Sahel. *The American Journal of Tropical Medicine and Hygiene*, 83(3), 601–606.
- Lehmann, T., Hawley, W. A., Grebert, H., Collins, F. H., Lehmann, T., Hawley, A. W., Grebert, H., & Collins, H. F. (1998). The effective population size of *Anopheles gambiae* in Kenya: Implications for population structure. *Molecular Biology and Evolution*, 15(3), 264–276. <https://doi.org/10.1093/oxfordjournals.molbev.a025923>
- Lehmann, T., Weetman, D., Huestis, D. L., Yaro, A. S., Kassogue, Y., Diallo, M., Donnelly, M. J., & Dao, A. (2017). Tracing the origin of the early wet-season anopheles coluzzi in the Sahel. *Wiley Evolutionary Applications*, 10(1), 704–717.
- Luikart, G., Allendorf, F. W., Cornuet, J. M., & Sherwin, W. B. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, 89(3), 238–247. <https://doi.org/10.1093/jhered/89.3.238>
- Mabaso, M. L. H., Craig, M., Ross, A., & Smith, T. (2007). Environmental predictors of the seasonality of malaria transmission in Africa: The challenge. *American Journal of Tropical Medicine and Hygiene*, 76(1), 33–38. <https://doi.org/10.4269/ajtmh.2007.76.33>
- Minakawa, N., Sonye, G., Mogi, M., Githeko, A., & Yan, G. (2002). The effects of climatic factors on the distribution and abundance of malaria vectors in Kenya. *Population and Community Ecology*, 39(6), 833–841.
- Mwima, R., Hui, T. Y. J., Nanteza, A., Burt, A., & Kayondo, J. K. (2023). Potential persistence mechanisms of the major *Anopheles gambiae* species complex malaria vectors in sub-Saharan Africa: A narrative review. *Malaria Journal*, 22(1), 336.
- North, A. R., Burt, A., & Godfray, H. C. J. (2020). Modelling the suppression of a malaria vector using a CRISPR-Cas9 gene drive to reduce female fertility. *BMC Biology*, 18, 1–14.
- Ojima, N., Hara, Y., Ito, H., & Yamamoto, D. (2018). Genetic dissection of stress-induced reproductive arrest in *Drosophila melanogaster* females. *PLoS Genetics*, 14(6), 1–15. <https://doi.org/10.1371/journal.pgen.1007434>
- Omer, S. M., & Cloudsley-Thompson, J. L. (1970). Survival of female *Anopheles gambiae* Giles through a 9-month dry season in Sudan. *Bulletin of the World Health Organization*, 42(2), 319–330.
- Peel, D., Waples, R. S., Macbeth, G. M., Do, C., & Ovenden, J. R. (2013). Accounting for missing data in the estimation of contemporary genetic effective population size (N_e). *Molecular Ecology Resources*, 13(2), 243–253. <https://doi.org/10.1111/1755-0998.12049>
- Piper, M. D., & Partridge, L. (2016). Protocols to study aging in drosophila. *Methods in Molecular Biology*, 1478, 291–302.
- Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104(3), 531–548. <https://doi.org/10.1093/genetics/104.3.531>
- Salman, M. H., Bonsignore, C. P., El Fels, A. E. A., Giomi, F., Hodar, J. A., Laparie, M., Marini, L., Merel, C., Zalucki, M. P., Zamoum, M., & Battisti, A. (2019). Winter temperature predicts prolonged diapause in pine processionary moth species across their geographic range. *PeerJ*, 7, e6530.
- Tatar, M., & Yin, C. M. (2001). Slow aging during insect reproductive diapause: Why butterflies, grasshoppers and flies are like worms. *Experimental Gerontology*, 36(4–6), 723–738. [https://doi.org/10.1016/S0531-5565\(00\)00238-2](https://doi.org/10.1016/S0531-5565(00)00238-2)
- The *Anopheles gambiae* 1000 Genomes consortium. (2017). Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*, 552, 96–100.
- The *Anopheles gambiae* 1000 Genomes consortium. (2020). Genome variation and population structure among 1142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. *Genome Research*, 30(10), 1533–1546. <https://doi.org/10.1101/GR.262790.120>

- Toxopeus, J., Jakobs, R., Ferguson, L. V., Garipey, T. D., & Sinclair, B. (2016). Reproductive arrest and stress resistance in winter-acclimated *Drosophila suzukii*. *Journal of Insect Physiology*, *89*, 37–51.
- Wang, M.-H. H., Marinotti, O., Vardo-Zalik, A., Boparai, R., & Yan, G. (2011). Genome-wide transcriptional analysis of genes associated with acute desiccation stress in *Anopheles gambiae*. *PLoS One*, *6*(10), 1–8. <https://doi.org/10.1371/journal.pone.0026011>
- Waples, R. S., & Yokota, M. (2007). Temporal estimates of effective population size in species with overlapping generations. *Genetics*, *175*(1), 219–233. <https://doi.org/10.1534/genetics.106.065300>
- Waples, S. R. (1989). A generalised approach for estimating effective population size from temporal changes in allele frequency. *Genetics Society of America*, *121*(1), 379–391.
- Waples, S. R. (2005). Genetic estimates of contemporary effective population size: To what time periods do the estimates apply? *Molecular Ecology*, *14*(1), 3335–3352.
- Yaro, A. S., Traoré, A. I., Huestis, D. L., Adamou, A., Timbiné, S., Kassogué, Y., Diallo, M., Dao, A., Traoré, S. F., & Lehmann, T. (2012). Dry season reproductive depression of *Anopheles gambiae* in the Sahel. *Journal of Insect Physiology*, *58*(8), 1050–1059.
- Yoder, J. A., & Denlinger, D. L. (1992). Water vapour uptake by diapausing eggs of a tropical walking stick. *Physiological Entomology*, *17*(1), 97–103.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mwima, R., Hui, T.-Y., Kayondo, J. K., & Burt, A. (2024). The population genetics of partial diapause, with applications to the aestivating malaria mosquito *Anopheles coluzzii*. *Molecular Ecology Resources*, *24*, e13949. <https://doi.org/10.1111/1755-0998.13949>