bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Close-kin mark-recapture methods to estimate demographic parameters of mosquitoes

Yogita Sharma[1‡], Jared B. Bennett[2‡], Gordana Rašić[3¤], John M. Marshall[1,4*],

1 Divisions of Biostatistics and Epidemiology, School of Public Health, University of California, Berkeley, CA, USA
2 Biophysics Graduate Group, Division of Biological Sciences, College of Letters and Science, University of California, Berkeley, CA, USA
3 Mosquito Genomics, QIMR Berghofer Medical Research Institute, Brisbane, Australia
4 Innovative Genomics Institute, University of California, Berkeley, CA, USA

‡Equal contribution
* Correspondence: john.marshall@berkeley.edu

Abstract

Close-kin mark-recapture (CKMR) methods have recently been used to infer demographic parameters such as census population size and survival for fish of interest to fisheries and conservation. These methods have advantages over traditional mark-recapture methods as the mark is genetic, removing the need for physical marking and recapturing that may interfere with parameter estimation. For mosquitoes, the spatial distribution of close-kin pairs has been used to estimate mean dispersal distance, of relevance to vector-borne disease transmission and novel biocontrol strategies. Here, we extend CKMR methods to the life history of mosquitoes and comparable insects. We derive kinship probabilities for mother-offspring, father-offspring, full-sibling and half-sibling pairs, where an individual in each pair may be a larva, pupa or adult. A pseudo-likelihood approach is used to combine the marginal probabilities of all kinship pairs. To test the effectiveness of this approach at estimating mosquito demographic parameters, we develop an individual-based model of mosquito life history incorporating egg, larva, pupa and adult life stages. The simulation labels each individual with a unique identification number, enabling close-kin relationships to be inferred for sampled individuals. Using the dengue vector *Aedes aegypti* as a case study, we find the CKMR approach provides unbiased estimates of adult census population size, adult and larval mortality rates, and larval life stage duration for logistically feasible sampling schemes. Considering a simulated population of 3,000 adult mosquitoes, estimation of adult parameters is accurate when a total of 1,000 adult females are sampled biweekly-to-fortnightly over a three month period. Estimation of larval parameters is accurate when adult sampling is supplemented with a total of 4,000 larvae sampled biweekly over the same period. As the cost of genome sequencing declines, these methods hold great promise for characterizing the demography of mosquitoes and comparable insects of epidemiological and agricultural significance.

Author summary

Close-kin mark-recapture (CKMR) methods are a genetic analogue of traditional mark-recapture methods in which the frequency of marked individuals in a sample is

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

used to infer demographic parameters such as census population size and mean dispersal distance. In CKMR, the mark is a close-kin relationship between individuals (parents and offspring, siblings, etc.). While CKMR methods have mostly been applied to aquatic species to date, opportunities exist to apply them to insects and other terrestrial species. Here, we explore the application of CKMR to mosquitoes, with *Aedes aegypti*, a primary vector of dengue, chikungunya and yellow fever, as a case study. By analyzing simulated *Ae. aegypti* populations, we find the CKMR approach provides unbiased estimates of adult census population size, adult and larval mortality rates, and larval life stage duration. Optimal sampling schemes are consistent with *Ae. aegypti* ecology and field studies, requiring only minor adjustments to current mosquito surveillance programs. This study represents the first theoretical exploration of the application of CKMR to an insect species, and demonstrates its potential for characterizing the demography of insects of epidemiological and agricultural importance.

1. Introduction

In the last few years, there has been a growth of interest in close-kin mark-recapture (CKMR) methods to characterize the demography of wild populations [1]. These methods are analogous to traditional mark-recapture methods, which estimate census population size and other demographic parameters based on the recapture rates of marked individuals. The advantages of CKMR methods stem from the mark being a genetically-inferred close-kin relationship, removing the need for physical marking and recapturing. Initial applications of these methods have included a wide range of fish species - southern bluefin tuna [2], white sharks [3], brook trout [4] and Atlantic salmon [5]. Fish provide a good case for CKMR because their populations are well-mixing, physical marking and recapturing pose logistical challenges, and there is a willingness to invest in population size estimates given their importance to fisheries and conservation [1]. CKMR studies on fish have also estimated annual juvenile and adult survival probabilities and rates of population growth [2, 3].

As high-throughput genomic sequencing, which enables accurate kinship estimation, becomes cheaper, it is expected that CKMR methods will be applied to an increasing number of species. For insects, two recent studies used the spatial distribution of close-kin pairs to characterize dispersal patterns of *Aedes aegypti* [6, 7], the mosquito vector of dengue, Zika, chikungunya and yellow fever. Both studies were set in urban landscapes - in Malaysia [6] and Singapore [7] - where mosquitoes inhabit high-rise apartment buildings. These locations were chosen to support releases of *Wolbachia*-infected mosquitoes intended for population replacement [8] and suppression [9]. Characterizing mosquito movement is important to understanding the spatial transmission of vector-borne diseases [10], and to designing optimal biocontrol strategies, such as those involving *Wolbachia*, for vector-borne disease control. By analyzing close-kin pairs, these two studies estimated mean dispersal distances in agreement with previous mark-recapture studies [7, 11], and isolated a radius of dispersal specific to *Ae. aegypti* oviposition behavior [6].

In this paper, we extend the CKMR formalism described by Bravington *et al.* [1] to mosquitoes, using *Ae. aegypti* as a case study, in order to derive demographic parameters from close-kin pairs. These methods involve deriving "kinship probabilities" describing the chance that a given individual is related to another in the population. These are calculated as the reproductive output consistent with a given kinship relationship divided by the total reproductive output of all adult females in the population, and depend upon a parameterized model of life history and mating behavior, including egg production and mortality rates. Because the age of adult *Ae. aegypti* mosquitoes is difficult to estimate in the field, age must be accommodated as a

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

latent variable, with marginal kinship probabilities being calculated by considering all consistent event histories. For fish species to which CKMR methods have been applied thus far, full-siblings are rare as adults tend to be polygamous [3]. In contrast, for mosquitoes, full-siblings are common as adult females tend to mate only once, soon after emergence, and lay eggs from this mating event over an extended period. Mosquito half-siblings are also common, and tend to be paternal (i.e. have the same father and different mothers). Taking these considerations into account, we derive kinship probabilities for mother-offspring, father-offspring, full-sibling and half-sibling pairs where either individual in each pair may be a larva, pupa or adult. A pseudo-likelihood approach is used to combine the marginal probabilities of all kinship pairs [1].

To test the effectiveness of this approach at estimating mosquito demographic parameters, we develop an individual-based model of mosquito life history, incorporating egg, larva, pupa and adult life stages. By labeling each individual with a unique identification number (IN) and tracking parental INs, this enables close-kin relationships to be inferred for sampled individuals. As studies of aquatic species have shown, a parsimonious individual-based simulation of life history allows a variety of CKMR sampling schemes to be explored, and for effectiveness at parameter estimation to be assessed [12, 13]. The short generation time of mosquitoes - less than a month for *Ae. aegypti* [14] - means that sampling may take place over a few months, as opposed to several years for long-lived fish species [2]. Open questions regarding sampling schemes for mosquitoes relate to the required sample size, optimal frequency (e.g., daily, biweekly or weekly), duration (i.e., number of months), and distribution of collections across larval, pupal and adult life stages in order to estimate population size, mortality rates, and durations of juvenile life stages. Here, we use our simulation model and CKMR framework to address these questions, and in doing so, provide a case study for CKMR applications to comparable insects of epidemiological and agricultural significance.

2. Materials and methods

2.1. Mosquito population dynamics

We use a discrete-time version of the lumped age-class model [15, 16], applied to mosquitoes [17], as the basis for our population simulation and CKMR analysis (Figure 1). This model considers discrete life history stages - egg (E), larva (L), pupa (P) and adult (A) - with sub-adult stages having defined durations - T_E, T_L and T_P for eggs, larvae and pupae, respectively. We use a daily time-step, since mosquito samples tend to be recorded by day, and this is adequate to model the organism's population dynamics [18]. Daily mortality rates vary according to life stage - μ_E, μ_L, μ_P and μ_A for eggs, larvae, pupae and adults, respectively - and density-dependent mortality occurs at the larval stage. Sex is modeled at the adult stage - half of pupae emerge as females (F), and the other half as males (M). Females mate once upon emergence, and retain the genetic material from that mating event for the remainder of their lives. Males mate at a rate equal to the female emergence rate which, for a population at equilibrium, is equal to the female mortality rate, μ_A. Females lay eggs at a rate, β, which is assumed to be independent of age.

Default life history and demographic parameters for *Ae. aegypti* are listed in Table 1. Given the difficulty of measuring juvenile stage mortality rates in the wild, these are chosen for consistency with observed population growth rates in the absence of density-dependence (see S1 Text §1.1 for formulae and derivations). Larval mortality increases with larval density and, according to the lumped age-class model, reaches a set value when the population is at equilibrium. Although mosquito populations vary seasonally, we assume a constant adult population size, N_A, for the CKMR analysis,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

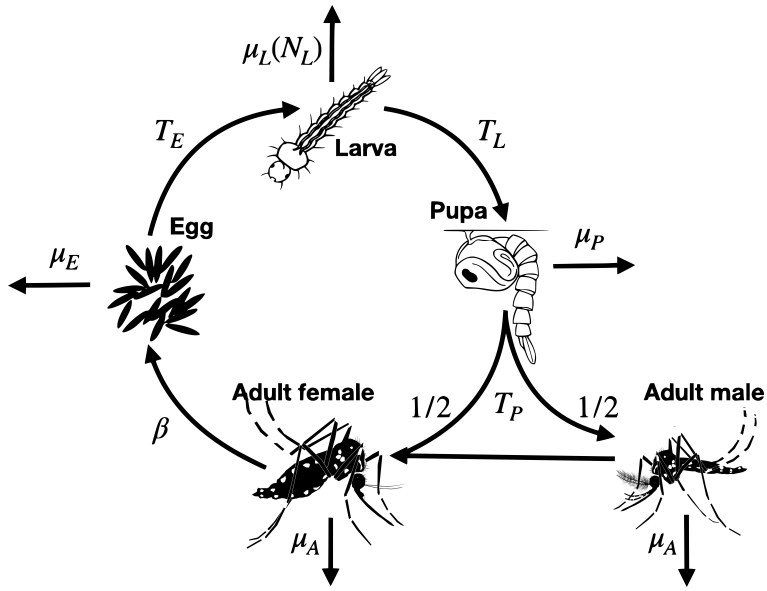

Fig 1. The lumped age-class model of mosquito life history. Mosquitoes are divided into four life stages: egg, larva, pupa and adult. The durations of the sub-adult stages are T_E, T_L and T_P for eggs, larvae and pupae, respectively. Sex is modeled at the adult stage, with half of pupae developing into females and half developing into males. Daily mortality rates vary by life stage - μ_E, μ_L, μ_P and μ_A for eggs, larvae, pupae and adults, respectively. Density-dependent mortality occurs at the larval stage and is a function of the total number of larvae, N_L. Females mate once upon emergence, and retain the genetic material from that mating event for the remainder of their lives. Males mate at a rate equal to the female emergence rate. Females lay eggs at a rate, β.

and restrict sampling to a maximum period of four months, corresponding to a season. Minor population size fluctuations occur in the simulation model due to sampling and stochasticity.

Table 1. Demographic and life history parameters for *Aedes aegypti* mosquitoes.

Parameter:	Definition:	Value:	References:
N_A	Adult population size	3000	[19–21]
μ_A	Adult mortality rate	0.09 / day	[22]
β	Female fecundity	20 / day	[23]
T_E	Duration of egg stage	2 days	[14]
T_L	Duration of larval stage	5 days	[14]
T_P	Duration of pupal stage	1 day	[14]
μ_E	Egg mortality rate	0.175 / day	S1 Text §1, [24]
μ_L	Larval mortality rate	0.554 / day	S1 Text §1
μ_P	Pupal mortality rate	0.175 / day	S1 Text §1, [24]

2.2. Kinship probabilities

Following the methodology of Bravington *et al.* [1], we now derive kinship probabilities for mother-offspring, father-offspring, full-sibling and half-sibling pairs based on the lumped age-class mosquito life history model. Each kinship probability is calculated as

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the reproductive output consistent with that relationship divided by the total reproductive output of all adult females in the population. In each case, we consider two individuals (adult, larva or pupa) sampled at known times, t_1 and t_2, with probability symbols and references to equations listed in Table 2. Important details to note for this analysis are that: i) mosquito sampling is lethal, ii) although age is a latent variable, temporal information is captured in the life stages of sampled individuals, and iii) mosquito mating behaviour, in which females mate once upon emergence and males mate throughout their adult lifespan, is reflected in the calculations.

Table 2. Kinship categories, sampled life stages, sampling times, and probability symbols used in close-kin mark-recapture analysis.

Kinship category:	Sampled life stages:	Probability symbol:	Equations:
Mother-offspring	Adult female (t_1), larva (t_2)	$P_{MOL}(t_1, t_2)$	§2.2.1
	Adult female (t_1), adult (t_2)	$P_{MOA}(t_1, t_2)$	§2.2.1
	Adult female (t_1), pupa (t_2)	$P_{MOP}(t_1, t_2)$	S1 Text §2.1
Father-offspring	Adult male (t_1), larva (t_2)	$P_{FOL}(t_1, t_2)$	S1 Text §2.2
	Adult male (t_1), adult (t_2)	$P_{FOA}(t_1, t_2)$	§2.2.2
	Adult male (t_1), pupa (t_2)	$P_{FOP}(t_1, t_2)$	S1 Text §2.2
Full-siblings	Larva (t_1), larva (t_2)	$P_{FSLL}(t_1, t_2)$	§2.2.3
	Adult (t_1), adult (t_2)	$P_{FSAA}(t_1, t_2)$	§2.2.3
	Larva (t_1), adult (t_2)	$P_{FSLA}(t_1, t_2)$	S1 Text §2.3
	Adult (t_1), larva (t_2)	$P_{FSAL}(t_1, t_2)$	S1 Text §2.3
	Pupa (t_1), pupa (t_2)	$P_{FSPP}(t_1, t_2)$	S1 Text §2.3
	Pupa (t_1), larva (t_2)	$P_{FSPL}(t_1, t_2)$	S1 Text §2.3
	Larva (t_1), pupa (t_2)	$P_{FSLP}(t_1, t_2)$	S1 Text §2.3
	Pupa (t_1), adult (t_2)	$P_{FSPA}(t_1, t_2)$	S1 Text §2.3
	Adult (t_1), pupa (t_2)	$P_{FSAP}(t_1, t_2)$	S1 Text §2.3
Half-siblings	Larva (t_1), larva (t_2)	$P_{HSLL}(t_1, t_2)$	S1 Text §2.4
	Adult (t_1), adult (t_2)	$P_{HSAA}(t_1, t_2)$	§2.2.4
	Larva (t_1), adult (t_2)	$P_{HSLA}(t_1, t_2)$	S1 Text §2.4
	Adult (t_1), larva (t_2)	$P_{HSAL}(t_1, t_2)$	S1 Text §2.4
	Pupa (t_1), pupa (t_2)	$P_{HSPP}(t_1, t_2)$	S1 Text §2.4
	Pupa (t_1), larva (t_2)	$P_{HSPL}(t_1, t_2)$	S1 Text §2.4
	Larva (t_1), pupa (t_2)	$P_{HSLP}(t_1, t_2)$	S1 Text §2.4
	Pupa (t_1), adult (t_2)	$P_{HSPA}(t_1, t_2)$	S1 Text §2.4
	Adult (t_1), pupa (t_2)	$P_{HSAP}(t_1, t_2)$	S1 Text §2.4

2.2.1. Mother-offspring

Let us begin with the simplest possible kinship probability, $P_{MOL}(t_1, t_2)$, which represents the probability that, given an adult female sampled on day t_1, a larva sampled on day t_2 is her offspring. This can be expressed as the relative larval reproductive output on day t_2 of an adult female sampled on day t_1:

$$P_{MOL}(t_1, t_2) = \frac{\mathbb{E}\left[\text{Larval offspring at time } t_2 \text{ from an adult female sampled at time } t_1\right]}{\mathbb{E}\left[\text{Larval offspring at time } t_2 \text{ from all adult females at consistent times}\right]} = \frac{E_{MOL}(t_1, t_2)}{E_L}. \quad (1)$$

Here, $E_{MOL}(t_1, t_2)$ represents the expected number of surviving larval offspring on day t_2 from an adult female sampled on day t_1, and E_L represents the expected number of surviving larval offspring from all adult females in the population at times consistent

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with the time of larval sampling. Note that, since we are assuming a constant population size, E_L is independent of time and is given by:

$$E_L = \sum_{y_2=0-T_E-T_L}^{0-T_E} N_F \times \beta \times (1-\mu_E)^{T_E} \times (1-\mu_L)^{(0-y_2-T_E)}. \quad (2)$$

Here, N_F represents the equilibrium adult female population size (which is equal to half the equilibrium adult population size, $N_A/2$), and y_2 represents the day of egg laying. Considering day 0 as the reference day (in place of t_2), the egg must have been laid between days $(0-T_E-T_L)$ and $(0-T_E)$ (Figure 2, panel B). Equation 2 therefore represents the expected number of offspring laid by all adult females in the population that survive the egg and larva stages up to the time of sampling (day 0).

$E_{MOL}(t_1,t_2)$, on the other hand, is specific to the sampled adult female and the day of larval sampling, t_2. This is given by:

$$E_{MOL}(t_1,t_2) = \sum_{y_2=t_2-T_E-T_L}^{t_2-T_E} (1-\mu_A)^{(t_1-y_2)} \times \left(\mathbb{I}[(t_1-T_A) \leq y_2 \leq t_1] \times \beta \times (1-\mu_E)^{T_E} \times (1-\mu_L)^{(t_2-y_2-T_E)}\right). \quad (3)$$

Here, the day of egg-laying, y_2, is summed over days $(t_2-T_E-T_L)$ through (t_2-T_E), for consistency with the larva being present on the day of sampling (Figure 2, panel A). The first term in the summation represents the probability that the adult female sampled on day t_1 is alive on the day of egg-laying, and the second term (in larger brackets) represents the expected surviving larval output of this adult female on day t_2. This latter term is equal to their daily egg production, β, multiplied by the proportion of eggs that survive the egg and larva stages from the day they were laid up to the day of sampling. An indicator function is included to limit consideration to cases where the day of egg-laying lies within the adult female's possible lifetime - i.e., between days t_1 and (t_1-T_A), where T_A represents the maximum possible age of an adult mosquito. Although adult lifetime is exponentially-distributed, a value of T_A may be chosen that captures most of this distribution and leads to accurate parameter inference.

Next, we adapt the mother-offspring kinship probability for adult offspring to obtain $P_{MOA}(t_1,t_2)$, the probability that, given an adult female sampled on day t_1, an adult sampled on day t_2 is her offspring:

$$P_{MOA}(t_1,t_2) = \frac{\mathbb{E}\left[\text{Adult offspring at time } t_2 \text{ from an adult female sampled at time } t_1\right]}{\mathbb{E}\left[\text{Adult offspring at time } t_2 \text{ from all adult females at consistent times}\right]} = \frac{E_{MOA}(t_1,t_2)}{E_A}. \quad (4)$$

Here, $E_{MOA}(t_1,t_2)$ represents the expected number of surviving adult offspring on day t_2 from an adult female sampled on day t_1, and E_A represents the expected number of surviving adult offspring from all adult females at times consistent with the time of adult offspring sampling. Assuming a population at equilibrium, E_A is independent of time and is given by:

$$E_A = \sum_{y_2=0-T_E-T_L-T_P-T_A}^{0-T_E-T_L-T_P} N_F \times \beta \times (1-\mu_E)^{T_E} \times (1-\mu_L)^{T_L} \times (1-\mu_P)^{T_P} \times (1-\mu_A)^{(0-y_2-T_E-T_L-T_P)}. \quad (5)$$

Here, considering day 0 as the reference day (in place of t_2), the day of egg-laying, y_2, is summed over days $(0-T_E-T_L-T_P-T_A)$ through $(0-T_E-T_L-T_P)$, for consistency with the adult offspring being present on the day of sampling (Figure 2,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

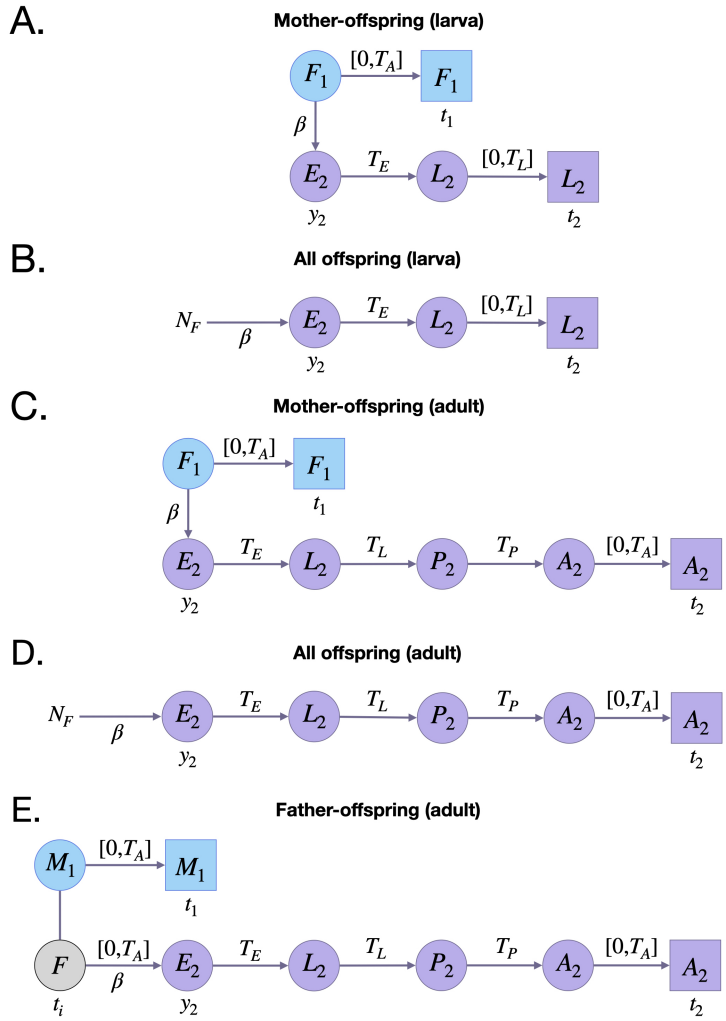

Fig 2. Consistent event histories for parent-offspring kinship probabilities. Parameters and state variables are as defined in Table 1 and §2.1. Subscript 1 refers to the parent, and subscript 2 refers to the offspring (the perspective from which probabilities are calculated). Circles represent living individuals and squares represent sampled individuals. Parents are sampled on day t_1, eggs are laid on day y_2, and offspring are sampled on day t_2. Offspring kinship probabilities are the ratio of the expected number of surviving offspring from a given adult on day t_2, and the expected number of surviving offspring from all adult females on this day. For mother-offspring pairs, this requires considering days of egg-laying, y_2, consistent with maternal ages at egg-laying in the range $[0, T_A]$, and with larval offspring ages at sampling in the range $[0, T_L]$ **(A)**, or adult offspring ages at sampling in the range $[0, T_A]$ **(C)**. The expected number of surviving offspring from all adult females requires considering days of egg-laying consistent with larval ages at sampling in the range $[0, T_L]$ (for larval offspring) **(B)**, or adult ages at sampling in the range $[0, T_A]$ (for adult offspring) **(D)**. Calculating the expected number of surviving adult offspring on day t_2 from an adult male requires considering days of mating, t_i, and egg-laying, y_2, consistent with maternal ages at egg-laying, and paternal and adult offspring ages at sampling in the range $[0, T_A]$ **(E)**.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

panel D). Equation 5 therefore represents the expected number of offspring laid by all adult females in the population that survive the egg, larva, pupa and adult stages up to the time of sampling (day 0). $E_{MOA}(t_1, t_2)$ is then given by:

$$E_{MOA}(t_1,t_2) = \sum_{y_2=t_2-T_E-T_L-T_P-T_A}^{t_2-T_E-T_L-T_P} (1-\mu_A)^{(t_1-y_2)} \times \left(\begin{array}{r} \mathbb{I}[(t_1-T_A) \leq y_2 \leq t_1] \times \beta \times (1-\mu_E)^{T_E} \times (1-\mu_L)^{T_L} \\ \times (1-\mu_P)^{T_P} \times (1-\mu_A)^{(t_2-y_2-T_E-T_L-T_P)} \end{array} \right). \quad (6)$$

Here, the day of egg-laying, y_2, is summed over days $(t_2 - T_E - T_L - T_P - T_A)$ through $(t_2 - T_E - T_L - T_P)$, for consistency with the adult offspring being present on the day of sampling (Figure 2, panel C). The terms within the summation are the same as for the mother-larval offspring case, with the exception that daily egg production is multiplied by the proportion of eggs that survive the egg, larva, pupa and adult stages from the day they were laid up to the day of sampling, t_2, which reflects the additional time elapsed for adult sampling.

Extending the mother-offspring kinship probability for pupal offspring is straightforward, involving similar adaptations as for the case of mother-adult offspring pairs. These extensions are provided in S1 Text §2.1.

2.2.2. Father-offspring

Next, we consider the father-adult offspring kinship probability, $P_{FOA}(t_1, t_2)$, which represents the probability that, given an adult male sampled on day t_1, an adult sampled on day t_2 is his offspring. This can be expressed as the relative adult reproductive output on day t_2 of adult females that mated with an adult male sampled on day t_1:

$$P_{FOA}(t_1,t_2) = \frac{\mathbb{E}\left[\text{Adult offspring at time } t_2 \text{ from an adult male sampled at time } t_1\right]}{\mathbb{E}\left[\text{Adult offspring at time } t_2 \text{ from all adult females at consistent times}\right]} = \frac{E_{FOA}(t_1,t_2)}{E_A}. \quad (7)$$

Here, $E_{FOA}(t_1, t_2)$ represents the expected number of surviving adult offspring on day t_2 of an adult male sampled on day t_1, and E_A is given in Equation 5. Each adult female mates once upon emergence and, since there are equal numbers of adult females and males in the population, each adult male mates on average once in their lifetime too. The day of this mating event, t_i, is unknown and so, in calculating $E_{FOA}(t_1, t_2)$, we treat this as a latent variable and take an expectation over all possible values it can take:

$$E_{FOA}(t_1,t_2) = \sum_{t_i=t_1-T_A}^{t_1} p_A(t_1-t_i) \times E_{FOA}(t_1,t_2|t_i). \quad (8)$$

Here, the expectation over the day of mating, t_i, is taken over days $(t_1 - T_A)$ through t_1, for consistency with the day of adult male sampling (Figure 2, panel E). The term $E_{FOA}(t_1, t_2|t_i)$ represents the expected number of adult offspring on day t_2 of an adult male sampled on day t_1, conditional upon the day of mating being t_i, and $p_A(t_1 - t_i)$ represents the probability that the mating event occurs on day $(t_1 - t_i)$. In general, $p_A(t)$ represents the probability that a given adult in the population has age t which, following from the daily survival probability, $(1 - \mu_A)$, is given by:

$$p_A(t) = (1-\mu_A)^t \Big/ \sum_{t_j=0}^{T_A} (1-\mu_A)^{t_j}. \quad (9)$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$E_{FOA}(t_1, t_2|t_i)$ is then given by:

$$E_{FOA}(t_1, t_2|t_i) = \sum_{y_2=t_i}^{t_i+T_A} (1-\mu_A)^{(y_2-t_i)} \times \begin{pmatrix} \mathbb{I}[(y_2+T_E+T_L+T_P) \leq t_2 \leq (y_2+T_E+T_L+T_P+T_A)] \\ \times \beta \times (1-\mu_E)^{T_E} \times (1-\mu_L)^{T_L} \\ \times (1-\mu_P)^{T_P} \times (1-\mu_A)^{(t_2-y_2-T_E-T_L-T_P)} \end{pmatrix}. \quad (10)$$

Here, the day of egg-laying, y_2, is summed over days t_i through $(t_i + T_A)$, for consistency with the mother's potential lifespan (Figure 2, panel E). The first term in the summation represents the probability that the mother is alive on the day of egg-laying, and the second term (in larger brackets) represents the expected surviving adult output of this adult female on day t_2. This latter term is the same as for the mother-adult offspring case, with the exception that the indicator function limits consideration to cases where the day of adult sampling, t_2, lies within the adult offspring's possible lifetime - i.e. between days $(y_2 + T_E + T_L + T_P)$ and $(y_2 + T_E + T_L + T_P + T_A)$.

Extending the father-offspring kinship probability for larval and pupal offspring is straightforward, involving similar adaptations as per the case of mother-offspring pairs. These extensions are provided in S1 Text §2.2.

2.2.3. Full-siblings

Next, we consider the full-sibling kinship probability for larva-larva pairs, $P_{FSLL}(t_1, t_2)$, which represents the probability that, given a larva sampled on day t_1, a larva sampled on day t_2 is their full-sibling. This can be expressed as the relative larval reproductive output on day t_2 of the mother of a larva sampled on day t_1:

$$P_{FSLL}(t_1, t_2) = \frac{\mathbb{E}\left[\text{Larvae at time } t_2 \text{ that are full-siblings of a larva sampled at time } t_1\right]}{\mathbb{E}\left[\text{Larval offspring at time } t_2 \text{ from all adult females at consistent times}\right]} = \frac{E_{FSLL}(t_1, t_2)}{E_L}. \quad (11)$$

Here, $E_{FSLL}(t_1, t_2)$ represents the expected number of surviving larvae on day t_2 that are full-siblings of a larva sampled on day t_1, and E_L is given in Equation 2. For convenience, let us refer to the larva sampled on day t_1 as individual 1. To calculate $E_{FSLL}(t_1, t_2)$, there are two unknown event times that we treat as latent variables and take an expectation over - i) the day that egg 1 is laid, y_1, and ii) the day that individual 1's mother emerges as an adult, t_i:

$$E_{FSLL}(t_1, t_2) = \sum_{y_1=t_1-T_E-T_L}^{t_1-T_E} p_L(t_1 - y_1 - T_E) \times \sum_{t_i=y_1-T_A}^{y_1} p_A(y_1 - t_i) \times E_{FSLL}(t_1, t_2|y_1, t_i). \quad (12)$$

Here, the expectation over the day that egg 1 is laid, y_1, is taken over days $(t_1 - T_E - T_L)$ through $(t_1 - T_E)$, for consistency with the day that larva 1 is sampled, and the expectation over the day of their mother's emergence, t_i, is taken over days $(y_1 - T_A)$ through y_1, so that egg 1 may be laid during their mother's potential lifetime (Figure 3, panel A). The term $E_{FSLL}(t_1, t_2|y_1, t_i)$ represents the expected number of surviving larvae on day t_2 that are full-siblings of larva 1, conditional upon egg 1 being laid on day y_1, and their mother emerging as an adult on day t_i. Additionally, $p_L(t_1 - y_1 - T_E)$ represents the probability that egg 1 is laid on day $(t_1 - y_1 - T_E)$, and $p_A(y_1 - t_i)$ represents the probability that their mother emerges on day $(y_1 - t_i)$. In general, $p_A(t)$ is given in Equation 9, and $p_L(t)$ represents the probability that a given

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

larva in the population has age t which, following from the daily larval survival probability, $(1-\mu_L)$, is given by:

$$p_L(t) = (1-\mu_L)^t \Big/ \sum_{t_j=0}^{T_L} (1-\mu_L)^{t_j}. \tag{13}$$

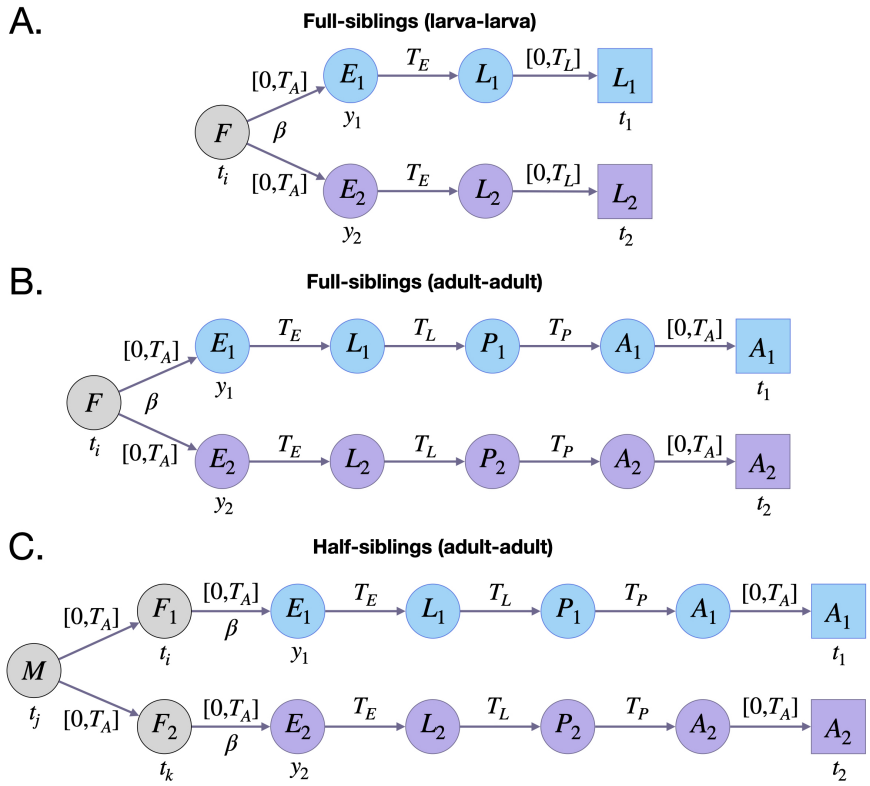

Fig 3. Consistent event histories for sibling kinship probabilities. Parameters and state variables are as defined in Table 1 and §2.1. Subscript 1 refers to the reference sibling, and subscript 2 refers to the sibling from whose perspective the probabilities are calculated. Circles represent living individuals and squares represent sampled individuals. The reference sibling is sampled on day t_1 and laid on day y_1. Sibling 2 is sampled on day t_2 and laid on day y_2. Sibling kinship probabilities are the ratio of the expected number of surviving siblings of a given individual on day t_2, and the expected number of surviving offspring from all adult females on this day. Calculating the expected number of surviving full-siblings requires considering days of their mother emerging as an adult, t_i, and of egg-laying, y_1 and y_2, that are consistent with maternal ages at egg-laying in the range $[0, T_A]$, and with larval ages at sampling in the range $[0, T_L]$ (for larval siblings) **(A)**, or adult ages at sampling in the range $[0, T_A]$ (for adult siblings) **(B)**. Calculating the expected number of surviving adult half-siblings of an adult requires considering days of their father emerging as an adult, t_j, their mothers emerging as adults, t_i and t_k, and of egg-laying, y_1 and y_2, that are consistent with paternal ages at mating in the range $[0, T_A]$, maternal ages at egg-laying in the range $[0, T_A]$, and adult ages at sampling in the range $[0, T_A]$ **(C)**.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$E_{FSLL}(t_1, t_2|y_1, t_i)$ is then given by:

$$E_{FSLL}(t_1, t_2|y_1, t_i) = \sum_{y_2=t_i}^{t_i+T_A} (1-\mu_A)^{(y_2-t_i)} \times \begin{pmatrix} \mathbb{I}[(t_2 - T_E - T_L) \leq y_2 \leq (t_2 - T_E)] \\ \times \beta \times (1-\mu_E)^{T_E} \times (1-\mu_L)^{(t_2-y_2-T_E)} \end{pmatrix}. \quad (14)$$

Here, the day of sibling egg-laying, y_2, is summed over days t_i through $(t_i + T_A)$, for consistency with the mother's potential lifespan (Figure 3, panel A). The first term in the summation represents the probability that the mother is alive on the day of sibling egg-laying, and the second term (in larger brackets) represents the expected larval output of the mother on day t_2. This latter term is the same as for the mother-larval offspring case, with the exception that the indicator function limits consideration to cases where the day of sibling egg-laying, y_2, is between days $(t_2 - T_E - T_L)$ and $(t_2 - T_E)$, for consistency with a larval sibling being sampled on day t_2.

Extending the full-sibling kinship probability to other life stage pairs is straightforward. We consider the case of adult-adult full-sibling pairs here, and provide the remaining cases in S1 Text §2.3. For adult-adult pairs, the full-sibling kinship probability is denoted by $P_{FSAA}(t_1, t_2)$ and represents the probability that, given an adult sampled on day t_1, an adult sampled on day t_2 is their full-sibling. This can be expressed as:

$$P_{FSAA}(t_1, t_2) = \frac{\mathbb{E}\left[\text{Adults at time } t_2 \text{ that are full-siblings of an adult sampled at time } t_1\right]}{\mathbb{E}\left[\text{Adult offspring at time } t_2 \text{ from all adult females at consistent times}\right]} = \frac{E_{FSAA}(t_1, t_2)}{E_A}. \quad (15)$$

Here, $E_{FSAA}(t_1, t_2)$ represents the expected number of surviving adults on day t_2 that are full-siblings of an adult sampled on day t_1, and E_A is given in Equation 5. For convenience, let us refer to the adult sampled on day t_1 as individual 1. To calculate $E_{FSAA}(t_1, t_2)$, there are two unknown event times that we treat as latent variables and take an expectation over - i) the day that egg 1 is laid, y_1, and ii) the day that individual 1's mother emerges as an adult, t_i:

$$E_{FSAA}(t_1, t_2) = \sum_{y_1=t_1-T_E-T_L-T_P-T_A}^{t_1-T_E-T_L-T_P} p_A(t_1 - y_1 - T_E - T_L - T_P) \times \sum_{t_i=y_1-T_A}^{y_1} p_A(y_1 - t_i) \times E_{FSAA}(t_1, t_2|y_1, t_i). \quad (16)$$

This is the same equation as for the larva-larva case with two exceptions: i) the expectation over the day that individual 1 is laid is taken over days $(t_1 - T_E - T_L - T_P - T_A)$ through $(t_1 - T_E - T_L - T_P)$ to account for the additional time elapsed between the larva and adult life stages, and ii) the probability that egg 1 is laid on day $(t_1 - y_1 - T_E - T_L - T_P)$, $p_A(t_1 - y_1 - T_E - T_L - T_P)$, reflects the adult age probability distribution in Equation 9 as this is the relevant life stage (Figure 3, panel B). The term $E_{FSAA}(t_1, t_2|y_1, t_i)$ represents the expected number of surviving adults on day t_2 that are full-siblings of adult 1, conditional upon egg 1 being laid on day y_1, and their mother emerging as an adult on day t_i. This is given by:

$$E_{FSAA}(t_1, t_2|y_1, t_i) = \sum_{y_2=t_i}^{t_i+T_A} (1-\mu_A)^{(y_2-t_i)} \times \begin{pmatrix} \mathbb{I}[(t_2 - T_E - T_L - T_P - T_A) \leq y_2 \leq (t_2 - T_E - T_L - T_P)] \\ \times \beta \times (1-\mu_E)^{T_E} \times (1-\mu_L)^{T_L} \\ \times (1-\mu_P)^{T_P} \times (1-\mu_A)^{(t_2-y_2-T_E-T_L-T_P)} \end{pmatrix}. \quad (17)$$

This is the same equation as for the larva-larva case with two exceptions: i) daily egg production is multiplied by the proportion of eggs that survive the egg, larva, pupa

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and adult stages up to the day of sampling to reflect the fact that adults rather than larvae are being sampled, and ii) the indicator function limits consideration to cases where the day of sibling egg-laying, y_2, is between days $(t_2 - T_E - T_L - T_P - T_A)$ and $(t_2 - T_E - T_L - T_P)$, again accounting for the additional time elapsed between the larva and adult life stages.

2.2.4. Half-siblings

Next, we consider the half-sibling kinship probability for adult-adult pairs, $P_{HSAA}(t_1, t_2)$, which represents the probability that, given an adult sampled on day t_1, an adult sampled on day t_2 is their half-sibling. This can be expressed as the relative adult reproductive output on day t_2 of adult females that mate with the father of an adult sampled on day t_1:

$$P_{HSAA}(t_1, t_2) = \frac{\mathbb{E}\left[\text{Adults at time } t_2 \text{ that are half-siblings of an adult sampled at time } t_1\right]}{\mathbb{E}\left[\text{Adult offspring at time } t_2 \text{ from all adult females at consistent times}\right]} = \frac{E_{HSAA}(t_1, t_2)}{E_A}. \quad (18)$$

Here, $E_{HSAA}(t_1, t_2)$ represents the expected number of surviving adults on day t_2 that are half-siblings of an adult sampled on day t_1, and E_A is given in Equation 5. For convenience, let us refer to the adult sampled on day t_1 as individual 1. To calculate $E_{HSAA}(t_1, t_2)$, there are three unknown event times that we treat as latent variables and take an expectation over - i) the day that egg 1 is laid, y_1, ii) the day of the mating event between individual 1's mother and father, t_i, and iii) the day that individual 1's father emerges as an adult, t_j:

$$\begin{aligned} E_{HSAA}(t_1, t_2) = \sum_{y_1 = t_1 - T_E - T_L - T_P - T_A}^{t_1 - T_E - T_L - T_P} & p_A(t_1 - y_1 - T_E - T_L - T_P) \times \sum_{t_i = y_1 - T_A}^{y_1} p_A(y_1 - t_i) \\ & \times \sum_{t_j = t_i - T_A}^{t_i} p_A(t_i - t_j) \times E_{HSAA}(t_1, t_2 | y_1, t_i, t_j). \end{aligned} \quad (19)$$

Here, i) the expectation over the day that egg 1 is laid, y_1, is taken over days $(t_1 - T_E - T_L - T_P - T_A)$ through $(t_1 - T_E - T_L - T_P)$, for consistency with the day that adult 1 is sampled, ii) the expectation over the day of the mating event, t_i, is taken over days $(y_1 - T_A)$ through y_1, for consistency with egg 1 being laid during their mother's potential lifetime, and iii) the expectation over the day that their father emerges, t_j, is taken over days $(t_i - T_A)$ through t_i, so that the mating event overlaps with their father's potential lifetime (Figure 3, panel C). The term $E_{HSAA}(t_1, t_2 | y_1, t_i, t_j)$ represents the expected number of surviving adults on day t_2 that are half-siblings of adult 1, conditional upon egg 1 being laid on day y_1, their mother and father mating on day t_i, and their father emerging as an adult on day t_j. Additionally, $p_A(t_1 - y_1 - T_E - T_L - T_P)$ represents the probability that egg 1 is laid on day $(t_1 - y_1 - T_E - T_L - T_P)$, $p_A(y_1 - t_i)$ represents the probability that the mating event happens on day $(y_1 - t_i)$, and $p_A(t_i - t_j)$ represents the probability that their father emerges on day $(t_i - t_j)$, where $p_A(t)$ is given by Equation 9. $E_{HSAA}(t_1, t_2 | y_1, t_i, t_j)$ is then given by:

$$E_{HSAA}(t_1, t_2 | y_1, t_i, t_j) = \sum_{t_k = t_j}^{t_j + T_A} (1 - \mu_A)^{(t_k - t_j)} \times \mu_A \times \sum_{y_2 = t_k}^{t_k + T_A} (1 - \mu_A)^{(y_2 - t_k)} \times \begin{pmatrix} \mathbb{I}[(t_2 - T_E - T_L - T_P - T_A) \leq y_2 \\ \leq (t_2 - T_E - T_L - T_P)] \\ \times \beta \times (1 - \mu_E)^{T_E} \\ \times (1 - \mu_L)^{T_L} \times (1 - \mu_P)^{T_P} \\ \times (1 - \mu_A)^{(t_2 - y_2 - T_E - T_L - T_P)} \end{pmatrix}. \quad (20)$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In order to produce a half-sibling, adult 1's father must mate with another adult female and that adult female must produce an offspring. Here, the day of the second mating event, t_k, is summed over days t_j through $(t_j + T_A)$, for consistency with the father's potential lifespan, and the day of sibling egg-laying, y_2, is summed over days t_k through $(t_k + T_A)$, for consistency with the mother's potential lifespan (Figure 3, panel C). The terms in the first summation represent: i) the probability that the father survives days t_k through t_j and therefore is alive on the day of the second mating event, and ii) the probability that the father mates on this day. This latter probability is equal to the adult mortality rate, μ_A, since, for a population at equilibrium, the adult emergence and mortality rates are the same, and the mating rate is equal to the emergence rate since females are assumed to mate upon emergence. Finally, the terms in the second summation represent the probability that the mother is alive on the day of sibling egg-laying, and the expected adult output of the mother on day t_2. This latter term (in big brackets) is the same as that for the full-sibling adult-adult case. We provide half-sibling kinship probabilities for other life stage pairs in S1 Text §2.4.

2.3. Likelihood calculation

The goal of this mosquito CKMR analysis is to make inferences about demographic and life history parameters given data on the frequency and timing of observed close-kin pairs. Here, we calculate the likelihood of parent-offspring and sibling pairs in a manner that takes advantage of the nature of the kinship probabilities and the sampling process. The kinship probabilities for each pair of individuals are assumed to be independent of each other, even though they are not. For this reason the combined likelihood is referred to as a "pseudo-likelihood" [1]. The pseudo-likelihood approach has been shown to produce accurate parameter and variance estimates provided the size of each sampling event is sufficiently low relative to the total population size [2, 3].

2.3.1. Parent-offspring pairs

Let us begin by considering the mother-adult offspring kinship probability, $p_{MOA}(t_1, t_2)$, which represents the probability that, given an adult female sampled on day t_1, an adult sampled on day t_2 is her offspring. Now consider $n_F(t_1)$ adult females sampled on day t_1. The probability that a given adult has a mother amongst the $n_F(t_1)$ sampled adult females, $p_{MOA}(t_1, t_2)$, is equal to one minus the probability that none of the $n_F(t_1)$ sampled adult females are the adult's mother, i.e.:

$$p_{MOA}(t_1, t_2) = 1 - (1 - P_{MOA}(t_1, t_2))^{n_F(t_1)}. \tag{21}$$

Here, $P_{MOA}(t_1, t_2)$ is as defined in Equation 4. Now consider $n_A(t_2)$ adults sampled on day t_2, and let $k_{MOA}(t_1, t_2)$ be the number of adults sampled on day t_2 that have a mother amongst the adult females sampled on day t_1. The pseudo-likelihood that $k_{MOA}(t_1, t_2)$ of the $n_A(t_2)$ adults sampled on day t_2 have a mother amongst the adult females sampled on day t_1 follows from the binomial distribution:

$$L(k_{MOA}(t_1, t_2)) = \binom{n_A(t_2)}{k_{MOA}(t_1, t_2)} \times p_{MOA}(t_1, t_2)^{k_{MOA}(t_1,t_2)} \times (1 - p_{MOA}(t_1, t_2))^{n_A(t_2) - k_{MOA}(t_1,t_2)}. \tag{22}$$

The full log-pseudo-likelihood for mother-adult offspring pairs, Λ_{MOA}, follows from summing the log-pseudo-likelihood over all adult female sampling days, t_1, and over

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

consistent adult offspring sampling days, t_2:

$$\Lambda_{MOA} = \sum_{t_1} \sum_{t_2=t_1+T_E+T_L+T_P-T_A}^{t_1+T_E+T_L+T_P+T_A} k_{MOA}(t_1,t_2) \log p_{MOA}(t_1,t_2) + (n_A(t_2) - k_{MOA}(t_1,t_2)) \log (1 - p_{MOA}(t_1,t_2)). \quad (23)$$

Note that, for the purpose of parameter interference, we can drop the first term in the pseudo-likelihood equation, and for the purpose of efficient computation, we consider consistent adult sampling days from $(t_1 + T_E + T_L + T_P - T_A)$ through $(t_1 + T_E + T_L + T_P + T_A)$. The earliest adult sampling day (relative to t_1) corresponds to the case where the mother laid the offspring at the beginning of her life, was sampled at the end of her life, and the adult offspring was sampled at the beginning of its life. The latest adult sampling day (relative to t_1) corresponds to the case where the mother was sampled on the day they laid their offspring, and the adult offspring was sampled at the end of its life. For cases where $t_1 = t_2$, the number of sampled adults, $n_A(t_2)$, is reduced by one to account for the fact that an adult cannot be its own parent.

Parent-offspring pseudo-likelihood equations for other sampled sexes and life stages follow an equivalent formulation. The main point to note is that consistent offspring sampling days are specific to the kinship and sampled life stages being considered (these can be deduced from event history diagrams like those in Figure 2). The joint log-pseudo-likelihood for parent-offspring pairs is then given by:

$$\Lambda_{PO} = \Lambda_{MOL} + \Lambda_{MOP} + \Lambda_{MOA} + \Lambda_{FOL} + \Lambda_{FOP} + \Lambda_{FOA}. \quad (24)$$

Here, Λ_{MOL}, Λ_{MOP}, Λ_{FOL}, Λ_{FOP} and Λ_{FOA} denote the log-pseudo-likelihoods for mother-larval offspring pairs, mother-pupal offspring pairs, father-larval offspring pairs, father-pupal offspring pairs and father-adult offspring pairs, respectively.

2.3.2. Full-sibling pairs

For siblings, we begin with the adult-adult full-sibling kinship probability, $P_{FSAA}(t_1,t_2)$, defined in Equation 15, which represents the probability that, given an adult sampled on day t_1, an adult sampled on day t_2 is their full-sibling. We consider a given adult, indexed by i and sampled on day $t_1(i)$, and $n_A(t_2)$ adults sampled on day t_2. Let $k_{FSAA}(i,t_2)$ be the number of adults sampled on day t_2 that are full-siblings of adult i. The pseudo-likelihood that $k_{FSAA}(i,t_2)$ of the $n_A(t_2)$ sampled adults on day t_2 are full-siblings of adult i follows from the binomial distribution:

$$L(k_{FSAA}(i,t_2)) = \binom{n_A(t_2)}{k_{FSAA}(i,t_2)} \times P_{FSAA}(t_1(i),t_2)^{k_{FSAA}(i,t_2)} \times (1 - P_{FSAA}(t_1(i),t_2))^{n_A(t_2)-k_{FSAA}(i,t_2)}. \quad (25)$$

Note that, for cases where $t_1(i) = t_2$, the number of sampled adults on day t_2, $n_A(t_2)$, is reduced by one to account for the fact that an adult cannot be its own sibling. Additionally, when counting siblings, we only consider siblings with indices $> i$ to avoid double-counting. The full log-pseudo-likelihood for adult-adult full-sibling pairs, Λ_{FSAA}, follows from summing the log-pseudo-likelihood over all sampled adults, i, and over consistent adult sampling days, t_2:

$$\Lambda_{FSAA} = \sum_{i=1}^{n_A-1} \sum_{t_2=t_1(i)-2T_A}^{t_1(i)+2T_A} k_{FSAA}(i,t_2) \log P_{FSAA}(t_1(i),t_2) + (n_A(t_2) - k_{FSAA}(i,t_2)) \log (1 - P_{FSAA}(t_1(i),t_2)). \quad (26)$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consistent adult sampling days for this case are from $(t_1(i) - 2T_A)$ through $(t_1(i) + 2T_A)$. The earliest adult sampling day (relative to $t_1(i)$) corresponds to the case where the mother laid individual 2 at the beginning of her life and individual 1 at the end of her life, adult 1 was sampled at the end of its life, and adult 2 was sampled soon after emergence. The latest adult sampling day (relative to $t_1(i)$) corresponds to the reverse case. Full-sibling pseudo-likelihood equations for other life stage pairs follow an equivalent formulation, with consistent sampling days specific to the kinship and sampled life stages being considered (these can be deduced from event history diagrams like those in Figure 3). The joint log-pseudo-likelihood for full-sibling pairs is then given by:

$$\Lambda_{FS} = \Lambda_{FSLL} + \Lambda_{FSLP} + \Lambda_{FSLA} + \Lambda_{FSPL} + \Lambda_{FSPP} + \Lambda_{FSPA} + \Lambda_{FSAL} + \Lambda_{FSAP} + \Lambda_{FSAA}. \quad (27)$$

Here, Λ_{FSLL}, Λ_{FSLP}, Λ_{FSLA}, Λ_{FSPL}, Λ_{FSPP}, Λ_{FSPA}, Λ_{FSAL} and Λ_{FSAP} denote the log-pseudo-likelihoods for larva-larva, larva-pupa, larva-adult, pupa-larva, pupa-pupa, pupa-adult, adult-larva and adult-pupa full-sibling pairs, respectively.

2.3.3. All sibling pairs

In moving from full-siblings to both full and half-siblings, we adopt a multinomial approach in which each pair of individuals can either be full-siblings, half-siblings or neither. Consider again a given adult, indexed by i and sampled on day $t_1(i)$, and $n_A(t_2)$ adults sampled on day t_2, and let $k_{HSAA}(i, t_2)$ be the number of adults sampled on day t_2 that are half-siblings of adult i. The pseudo-likelihood that $k_{FSAA}(i, t_2)$ and $k_{HSAA}(i, t_2)$ of the $n_A(t_2)$ sampled adults on day t_2 are full-siblings and half-siblings of adult i, respectively, follows from the multinomial distribution:

$$L(k_{FSAA}(i,t_2), k_{HSAA}(i,t_2)) \propto \begin{pmatrix} P_{FSAA}(t_1(i),t_2)^{k_{FSAA}(i,t_2)} \times P_{HSAA}(t_1(i),t_2)^{k_{HSAA}(i,t_2)} \\ \times (1 - P_{FSAA}(t_1(i),t_2) - P_{HSAA}(t_1(i),t_2))^{n_A(t_2) - k_{FSAA}(i,t_2) - k_{HSAA}(i,t_2)} \end{pmatrix}. \quad (28)$$

Here, $P_{FSAA}(t_1(i), t_2)$ is as defined in Equation 15, and $P_{HSAA}(t_1(i), t_2)$ represents the probability that, given an adult sampled on day $t_1(i)$, an adult sampled on day t_2 is their half-sibling, as defined in Equation 18. The full log-pseudo-likelihood for all adult-adult sibling pairs, Λ_{SAA}, follows from summing the log-pseudo-likelihood over all sampled adults, i, and over consistent adult sampling days, t_2:

$$\Lambda_{SAA} = \sum_{i=1}^{n_A - 1} \sum_{t_2 = t_1(i) - 3T_A}^{t_1(i) + 3T_A} \begin{pmatrix} k_{FSAA}(i,t_2) \log P_{FSAA}(t_1(i),t_2) + k_{HSAA}(i,t_2) \log P_{HSAA}(t_1(i),t_2) \\ + (n_A(t_2) - k_{FSAA}(i,t_2) - k_{HSAA}(i,t_2)) \\ \times \log (1 - P_{FSAA}(t_1(i),t_2) - P_{HSAA}(t_1(i),t_2)) \end{pmatrix}. \quad (29)$$

The range of consistent adult sampling days is larger when half-siblings are included, due to the additional event histories involved. The full and half-sibling pseudo-likelihood equations for other life stage pairs follow an equivalent formulation, and the joint log-pseudo-likelihood is given by:

$$\Lambda_S = \Lambda_{SLL} + \Lambda_{SLP} + \Lambda_{SLA} + \Lambda_{SPL} + \Lambda_{SPP} + \Lambda_{SPA} + \Lambda_{SAL} + \Lambda_{SAP} + \Lambda_{SAA}. \quad (30)$$

Here, Λ_{SLL}, Λ_{SLP}, Λ_{SLA}, Λ_{SPL}, Λ_{SPP}, Λ_{SPA}, Λ_{SAL} and Λ_{SAP} denote the log-pseudo-likelihoods for larva-larva, larva-pupa, larva-adult, pupa-larva, pupa-pupa, pupa-adult, adult-larva and adult-pupa full and half-sibling pairs, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.3.4. Parameter inference

Despite parent-offspring and sibling kinship probabilities not being independent, the pseudo-likelihood approach enables us to combine these likelihoods, provided the size of each sampling event is sufficiently low relative to the total population size [1]. As we will see later, our simulation studies suggest this to be the case. We therefore combine these log-pseudo-likelihoods to obtain a log-pseudo-likelihood for the entire data set:

$$\Lambda = \Lambda_{PO} + \Lambda_S. \tag{31}$$

Parameter inference can then proceed by varying a subset of the demographic and life history parameters in Table 1 in order to minimize $-\Lambda$. We used the **nlminb** function implemented in the **optimx** function in R [25] to perform our optimizations. This function implements a Newton-type algorithm and performed the best, in terms of speed and accuracy, among the 13 algorithms available through the **optimx** function.

2.4. Individual-based simulation model

We developed an individual-based simulation model of mosquito life history to test the effectiveness of the CKMR approach at estimating mosquito demographic and bionomic parameters. The model is an individual-based adaptation of our previous model, **MGDrivE** [26], which is a genetic and spatial extension of the lumped age-class model applied to mosquitoes by Hancock and Godfray [17] and Deredec *et al.* [18] (Figure 1). The simulation time-step is one day. Functionality is included to account for spatial population structure, with mosquitoes being distributed across populations in a metapopulation [26], and each population having an equilibrium adult population size, N_A^*, and exchanging migrants with the other populations; however, in the present analysis, a single panmictic population is modeled. Each population is partitioned according to discrete life stages - egg, larva, pupa and adult - with sub-adult stages having fixed durations as defined earlier. Daily mortality rates are as defined earlier, and implemented according to a Bernoulli distribution for each individual. Density-independent juvenile mortality rates are calculated for consistency with observed population growth rates for *Ae. aegypti* (Table 1). Additional density-dependent mortality occurs at the larval stage and regulates population size (see S1 Text §1 for formulae and derivations). Sex is modeled at the adult stage - half of pupae emerge as females, and the other half as males, implemented according to a Bernoulli distribution with probability 0.5. Females mate once upon emergence, with the male mate being chosen at random. Males mate throughout their lifespan, and independently of previous mating events. Females lay eggs at a daily fecundity rate, β, for the duration of their lifespan with daily egg production of each adult female following a Poisson distribution.

Sampling is lethal, and is implemented as specified, with collection days, locations and sample sizes for each life stage defined by the user. To enable close-kin relationships to be inferred for sampled individuals, each individual is labeled with a unique IN, and parental INs are stored as attributes. Output CSV (comma-separated value) files are produced for each sampled life stage (larva, pupa, adult female and adult male, as appropriate), and include the time (day) and location (patch) of collection, as well as the individual's age at the time of sampling, their IN, and maternal and paternal INs. Inference of mother-offspring, father-offspring, full-sibling and (paternal) half-sibling pairs from this data is straightforward. Age information was not used in this analysis; but may be useful in the future as new technologies emerge to estimate the age of wild-caught adults [27].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Results

We used simulated data from the individual-based mosquito model to explore the feasibility of CKMR methods to infer demographic and bionomic parameters for *Ae. aegypti*. Our simulated population consisted of 3,000 adults with bionomic parameters listed in Table 1. Open questions concern the suitability of CKMR methods for *Ae. aegypti*, and the range of demographic and bionomic parameters that can be accurately estimated using them. To address these questions, we explored logistically-feasible sampling schemes to accurately infer adult and juvenile parameters by varying: i) sampled life stage (larva, pupa or adult), sex (adult female or male), sampling frequency (daily, biweekly, weekly or fortnightly), sampling duration (1-4 months), and total sample size (500-5,000 sequenced individuals). For adults, we focused on adult population size, N_A, and mortality rate, μ_A, and for juvenile life stages, we focused on larval mortality rate, μ_L, and the duration of the larval stage, T_L. By default, our likelihood calculations were based on parent-offspring and full-sibling pairs. Half-sibling pairs were only included for optimal sampling schemes, due to the computational burden that half-siblings present by requiring integration over six latent event times (Figure 3, panel C). We also considered subsets of likelihood components in our analyses, in the event that these may provide increased accuracy or precision.

3.1. Optimal sampling schemes to estimate adult parameters

To estimate adult parameters, our default sampling scheme consisted of a total of 1,000 sequenced individuals sampled daily over a three-month period (i.e., ca. 11-12 individuals sampled each day, for a total of 1,000 individuals after three months of sampling). We first explored the optimal distribution of sampled life stage and sex to estimate N_A and μ_A. Sampled larval, adult female and adult male life stage proportions were varied in 25% increments and limited to scenarios where the number of sampled adult females was greater than or equal to the number of sampled adult males (this reflects the case in the field due to the relative difficulty of sampling adult males). We also considered the case where only pupae were sampled, as pupae are often used as indicators of adult population size in entomological field studies [28]. Results of 100 simulation-and-analysis replicates for each of ten sampling scenarios are depicted in Figure 4 (panels A and B). The key result from this analysis is that the most accurate estimates of N_A and μ_A - in terms of both accuracy of the median and tightness of the interquartile range (IQR) - are obtained when only adult females are sampled. This is an intuitive result, as N_A and μ_A both describe the adult population, and adult females provide the most direct information on kinship - i.e., calculating the kinship probability for father-offspring pairs as compared to mother-offspring pairs involves integrating over an additional latent event time (Figure 2). Other key messages from this analysis are that IQRs of inferred parameters are wider for samples dominated by larvae (75% or higher) or pupae (100%), and there is a bias towards higher estimates of population size and lower estimates of adult mortality in all cases except the optimal case of adult female sampling. Given these results, we focused on adult female sampling while refining other details of the sampling schemes for estimating adult parameters.

Next, we explored the most efficient sampling frequency to estimate N_A and μ_A. While we consider daily sampling a theoretical gold standard, mosquito collections in the field tend to be at most biweekly [29], with weekly collections being more common [7]. For completeness, we also considered collections occurring every two weeks, with results of 100 replicates for each of the four sampling scenarios depicted in Figure 4 (panels C and D). The key result from this analysis is that CKMR estimates of N_A and μ_A are robust for daily, biweekly, weekly, and even fortnightly collections, which is reassuring for the logistical feasibility of the method. In the field, the decision on sampling

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

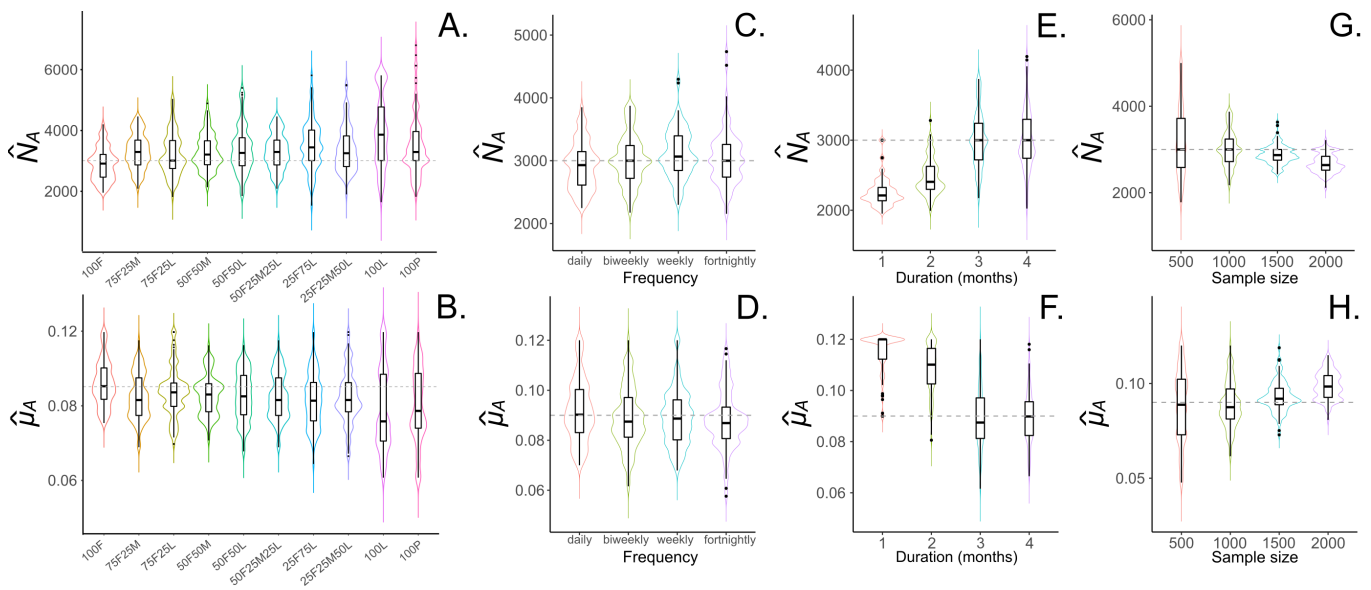

Fig 4. Sampling schemes to estimate N_A and μ_A for *Ae. aegypti*. Violin plots depict estimates of N_A and μ_A for sampling scenarios described in §3.1. The simulated population consists of 3,000 adult *Ae. aegypti* with bionomic parameters listed in Table 1. Boxes depict median and interquartile ranges of 100 simulation-and-analysis replicates for each scenario, thin lines represent 5% and 95% quantiles, points represent outliers, and kernel density plots are superimposed. The default sampling scheme consists of 1,000 individuals sampled as ca. 11-12 individuals a day over three months. In panels **(A-B)**, sampled larval, adult female and adult male life stage proportions are varied in 25% increments and limited to scenarios where the number of sampled adult females exceeds the number of sampled adult males. The case 100% sampled pupae is also included. In panels **(C-D)**, all sampled individuals are adult females, and four sampling frequencies are considered - daily, biweekly, weekly and fortnightly. In panels **(E-F)**, biweekly sampling is adopted, and sampling durations of 1-4 months are explored. In panels **(G-H)**, a sampling duration of three months is adopted, and total sample sizes of 500, 1,000, 1,500 and 2,000 adult females are explored. The optimal sampling scheme consists of 1,000 adult females collected biweekly-to-fortnightly over a three month period.

frequency will be based on the required total sample size, and the sampling frequency required to achieve it. We decided to focus on biweekly sampling henceforth, given its precedent in the field, and considering it allows more mosquitoes to be collected than weekly sampling. CKMR methods rely on the day of sampling to be known, and so mosquitoes must be collected within a single day of trapping, unlike regular mosquito surveillance efforts in which they are pooled over the days between collections.

Following this, we explored the most efficient sampling duration to estimate N_A and μ_A. We explored durations of 1-4 months as, given the short generation time of mosquitoes [14], parent-offspring pairs could potentially be collected within a month, and given the seasonality of mosquito populations, a maximum sampling period of four months corresponds to a season when the constant population assumption may approximately apply. 100 replicates for each of four sampling scenarios are depicted in Figure 4 (panels E and F). These results suggest that sampling durations of 3-4 months provide unbiased estimates of N_A and μ_A, while sampling durations of 1-2 months lead to adult mortality being overestimated, and adult population size being underestimated. Interestingly, N_A is also underestimated for sampling durations of 1-2 months when it is

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the only parameter being estimated (results not shown). Given these results, we retained a three-month sampling period as the most efficient option.

Next, we explored the optimal sample size to estimate N_A and μ_A for *Ae. aegypti*. We performed 100 simulation-and-analysis replicates for each of four total sample sizes - 500, 1,000, 1,500 and 2,000 adult females - depicted in Figure 4 (panels G and H). Results suggest that, while estimates of N_A and μ_A become more precise for larger sample sizes (as measured by the IQR), adult mortality is overestimated for total sample sizes of 1,500 or higher, and adult population size is correspondingly underestimated. These are likely a reflection of lethal sampling removing individuals from the population and hence increasing adult mortality and reducing adult population size. We therefore converged on an optimal sample size of 1,000 adult females, collected biweekly-to-fortnightly over a three month period, as providing accurate and unbiased estimates of N_A and μ_A.

Given this optimal sampling scheme to estimate N_A and μ_A, we next explored the likelihood components used in these analyses. Curiously, we found that including half-siblings in our analyses introduced significant biases in our parameter estimates, leading to an underestimate of μ_A and an overestimate of N_A (Figure 5). This could potentially be due to the half-sibling likelihood component requiring a sampling period longer than three months to produce accurate parameter estimates, as half-sibling kinship probabilities require integrating over several more latent event times than full-sibling and parent-offspring kinship probabilities. Interestingly, adult parameter estimates inferred from combined parent-offspring and full-sibling likelihood components are more accurate and precise (as measured by the median and IQR of replicate parameter estimates, respectively) compared to those inferred from either likelihood component in isolation (Figure 5). This confirms that the parameter estimates from the optimal sampling scheme in Figure 4 are indeed optimal - namely an adult population size estimate of 3,000 (IQR: 2,719-3,241), and an adult mortality rate estimate of 0.087 per day (IQR: 0.081-0.089 per day).

3.2. Optimal sampling schemes to estimate juvenile parameters

Preliminary explorations of sampling schemes to estimate juvenile parameters suggested this was not possible when including all likelihood components. We therefore tested likelihood components on a component-by-component basis to see whether some were more informative of juvenile parameters than others. We found that mother-larval offspring pairs provided accurate estimates of larval mortality, μ_L, and that mother-adult offspring pairs provided accurate estimates of the duration of the larval stage, T_L. We were not able to estimate pupal parameters (μ_P or T_P), likely due to the brevity of this life stage. Preliminary explorations suggested a sample of 1,000 adult females satisfied the adult requirement for larval parameter estimates, and had already been recommended for estimation of N_A and μ_A. We therefore focused our systematic exploration on the supplemental larval sampling requirement to estimate μ_L and T_L. We estimated these parameters simultaneously using a grid search, varying T_L discretely in the range $[1, 10]$, inferring the value of μ_L that minimized $-\Lambda$ for each value of T_L, and determining the values of μ_L and T_L that minimized $-\Lambda$ overall.

Our default sampling scheme consisted of a total of 1,000 sequenced adult females and an additional number of larvae sampled daily over a three month period. We first explored the optimal larval sample size to estimate μ_L and T_L. We performed 100 simulation-and-analysis replicates for each of four total larval sample sizes - 500, 1,000, 2,000 and 4,000 - depicted in Figure 6 (panels A and B). Results suggest that estimates of μ_L and T_L are unbiased for the larval sample sizes of 1,000 or more, but precision of the estimates, particularly of μ_L (as measured by the IQR), improves as larval sample size is increased. E.g., for a larval sample size of 1,000, the IQR for μ_L is 0.454-0.580

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

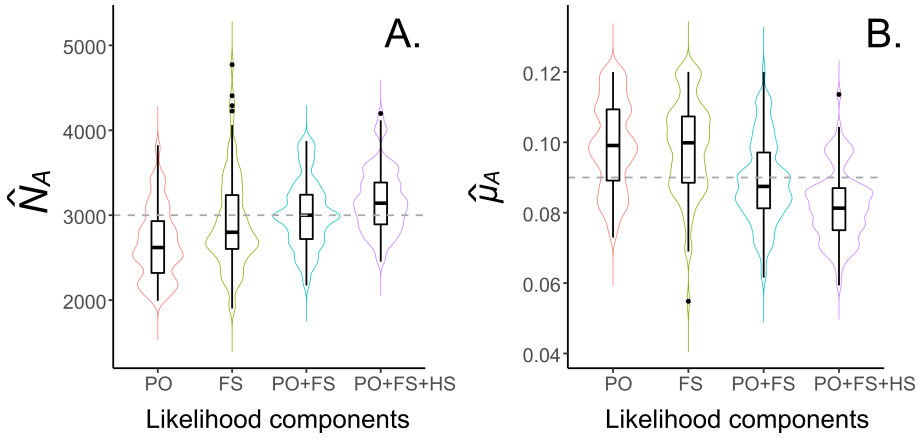

Fig 5. Likelihood components to estimate N_A and μ_A for *Ae. aegypti*. Violin plots depict estimates of N_A **(A)** and μ_A **(B)** for the optimal sampling scheme determined in Figure 4 (1,000 adult females collected biweekly over a three month period) and various included likelihood components. The simulated population consists of 3,000 adult *Ae. aegypti* with bionomic parameters listed in Table 1. Boxes depict median and interquartile ranges of 100 simulation-and-analysis replicates for each scenario, thin lines represent 5% and 95% quantiles, points represent outliers, and kernel density plots are superimposed. Adult parameter estimates inferred from combined parent-offspring and full-sibling likelihood components are more accurate than those inferred from either likelihood component in isolation and more accurate than those inferred by inclusion of half-sibling pairs.

per day, while for a larval sample size of 4,000, the IQR is 0.499-0.573 per day (the true value is 0.554 per day, Table 1). We therefore proceeded with a sample size of 4,000 larvae in addition to the 1,000 adult females previously recommended, although note that a larval sample of 1,000 or 2,000 is also adequate for daily sampling.

Next, we explored the most efficient sampling frequency to estimate μ_L and T_L. As for the adult parameter case, we considered four sampling frequencies - daily, biweekly, weekly and fortnightly - with results of 100 replicates for each scenario depicted in Figure 6 (panels C and D). The key result from this analysis is that CKMR estimates of μ_L and T_L are accurate and unbiased for daily and biweekly collections; but that weekly and fortnightly collections are inadequate for estimating μ_L and less reliable for estimating T_L. While this is a more frequent sampling requirement than that for estimating adult parameters, there is a precedent for biweekly collections in the field [29]. Biweekly collections were also our default recommendation for adult collections due to their field precedent, and because they allow a greater number of individuals to be collected over time.

Finally, we explored the most efficient sampling duration to estimate μ_L and T_L. As for the adult parameter case, we explored durations of 1-4 months, with results of 100 replicates for each scenario depicted in Figure 6 (panels E and F). These results suggest that sampling durations of 3-4 months provide accurate estimates of μ_L and T_L, while sampling durations of 1-2 months lead to larval mortality being underestimated, and estimates of T_L being less accurate. We therefore converged on an optimal sample size of 4,000 larvae supplementing the 1,000 adult females recommended earlier, collected biweekly over a three month period, as providing accurate and accurate and unbiased estimates of μ_L and T_L. This produces parameter estimates for μ_L of 0.534 (IQR: 0.499-0.573), and for T_L of 5 days (IQR: 4-6 days).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

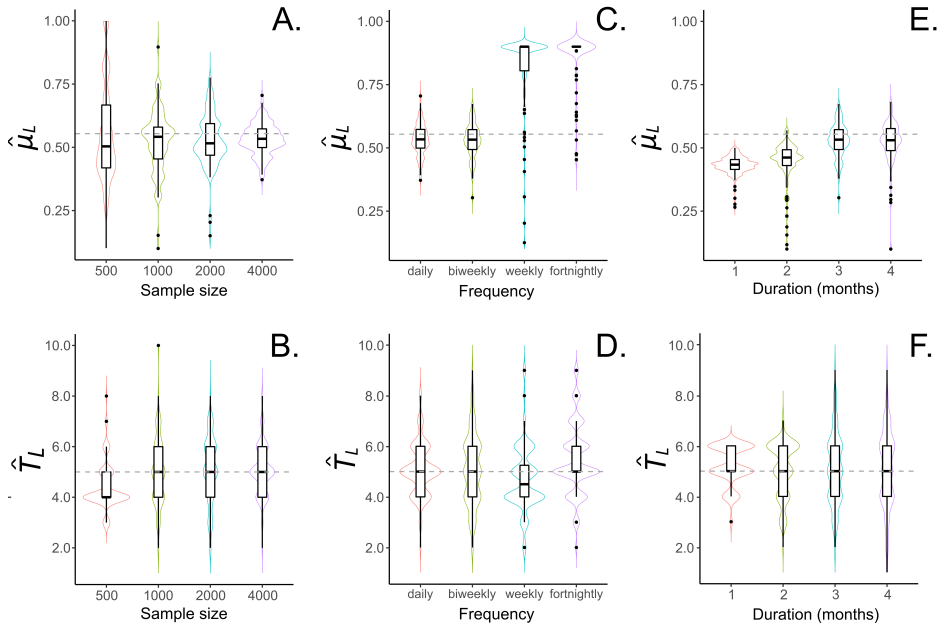

Fig 6. Sampling schemes to estimate μ_L and T_L for *Ae. aegypti*. Violin plots depict estimates of μ_L and T_L for sampling scenarios described in §3.2. The simulated population consists of 3,000 adult *Ae. aegypti* with bionomic parameters listed in Table 1. Boxes depict median and interquartile ranges of 100 simulation-and-analysis replicates for each scenario, thin lines represent 5% and 95% quantiles, points represent outliers, and kernel density plots are superimposed. The default sampling scheme consists of 1,000 adult females and supplemental larvae sampled daily over a three month period. In panels **(A-B)**, total larval sample sizes of 500, 1,000, 2,000 and 4,000 are explored. In panels **(C-D)**, a larval sample size of 4,000 is adopted, and four sampling frequencies are considered - daily, biweekly, weekly and fortnightly. In panels **(E-F)**, biweekly sampling is adopted, and sampling durations of 1-4 months are explored. The optimal sampling scheme consists of 4,000 larvae and 1,000 adult females collected biweekly over a three month period.

4. Discussion

We have demonstrated the application of the CKMR formalism described by Bravington *et al.* [1] to estimate demographic parameters for mosquitoes with *Ae. aegypti*, a major vector of dengue, Zika, chikungunya and yellow fever, as a case study. Using an individual-based simulation based on the lumped age-class model [15, 16] applied to mosquitoes [17], we have shown that these methods accurately estimate adult population size, N_A, adult mortality rate, μ_A, larval mortality rate, μ_L, and larval life stage duration, T_L, for logistically feasible sampling schemes when model assumptions are satisfied. Encouragingly, the optimal sampling scheme inferred from this analysis is consistent with *Ae. aegypti* ecology and field studies. Estimating adult parameters will likely be of most interest, and in this case, only adult females need to be sampled. Conveniently, adult females are preferentially attracted to most commercial traps through cues that mimic potential blood-meals, while adult males are more difficult to trap as they do not blood-feed [29]. Estimating larval parameters requires larval collections, and although larval breeding sites need to be actively sought out, larvae are an abundant life stage that can easily be collected with a cup or pipette [30].

Other details of the CKMR-optimal sampling scheme are also consistent with *Ae.*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

aegypti ecology. The sampling duration required for accurate estimates of both adult and larval parameters is three months, which is consistent with the length of a season, during which time the constant population size assumption in this analysis approximately holds. For estimating adult parameters, the total sample size of 1,000 adult females collected over three months is reasonable, and sequencing these 1,000 mosquitoes to the extent required to accurately infer close-kin relationships should fall within the budget of current mosquito surveillance programs [7]. For estimating larval parameters, the sample size of 4,000 larvae collected over three months is achievable, given the abundance of this life stage, although currently the sequencing expense would be burdensome. That said; as sequencing continues to become cheaper, and as more scalable methods become available to estimate relatedness, large-scale larval sequencing may also fall within the budget of surveillance programs.

Finally, the sampling frequency requirement of these CKMR methods is consistent with mosquito field studies, with biweekly sampling being adequate for accurate estimation of both adult and larval parameters. This is commonplace among mosquito surveillance programs [29]. If estimates of only adult parameters are desired, sampling frequency can be less frequent (e.g., fortnightly), although achieving the total required sample size may be a barrier to less frequent sampling. For CKMR methods, temporal information is of utmost importance, and so the day of collection should be known. This means that samples from a mosquito trap should represent collections for a single day, rather than the accumulation of mosquitoes over several days, as is the case for regular mosquito surveillance programs. A total sample size of 1,000 adult females over three months corresponds to biweekly collections of ca. 40 mosquitoes or weekly collections of ca. 80 mosquitoes. With these numbers in mind, the expected daily mosquito yield of a given location can inform the required sampling frequency.

As a preliminary exploration of the application of CKMR methods to mosquitoes, and as a modeling exercise, this study has several limitations. Firstly, the same life history model Figure 1 was used as a basis for both the population simulations and the CKMR analysis. Additionally, other than the parameters being estimated, the same parameters were used in both simulations and analysis. This represents an overly generous scenario as compared to the field, where true life history is varied and complex, and where life history parameters are only approximately known. That said; this is an appropriate starting point to verify the utility of the method for mosquitoes - it first needs to be shown to infer the true value of a parameter given the true model. Subsequent analyses should explore the robustness of parameter inference when other parameters in the model are dynamic or misspecified, or when kinship data are generated from a more detailed model - e.g., the CIMSiM model of *Ae. aegypti* population dynamics [22]. Another modest model variation would be to increase the variance in the fecundity parameter, β. Presently, the daily number of offspring generated by each adult female is Poisson-distributed and distributing this according to an overdispersed negative Binomial distribution would reduce the effective population size, N_e, while maintaining the census adult population size, N_A [13], the impact of which would be interesting to explore.

A second limitation of the application of our methods is that we have assumed perfect kinship inference throughout. A variety of molecular methods for kinship inference are available [31–33], the accuracy of which should be assessed for *Ae. aegypti* and other species of interest. Incorporating kinship uncertainty into the CKMR likelihood equations is theoretically possible [34], although this has produced little improvement in parameter inference at large computational cost when applied to data from fish species [2]. Likely, the best approach would be to introduce errors in kinship assignment at the simulation phase, and to test the robustness of the methods to this. Here, there is an important distinction between type I (false positive) and II (false

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

negative) error rates. Studies in fish species suggest that kinship inference must have an especially low type I error rate in order for CKMR parameter inference to be informative [1]. Kinship inference methods should be calibrated accordingly. On a related note, there is a debate over the conditions for inclusion of half-siblings in CKMR analyses. Half-sibling relationships are indistinguishable from avuncular (e.g., aunt-niece) and grandparent-grandchild relationships, introducing kinship assignment errors into likelihood calculations. Possible solutions have been proposed - e.g., restricting the time window of recording half-sibling pairs to include mostly same-cohort captures [1] - however this is a moot point for the present analysis, given that inclusion of half-siblings reduces the accuracy of parameter estimates even when precisely known.

A third limitation of the current analysis is that it ignores spatial structure. The population of 3,000 adults in the *Ae. aegypti* simulation was based on studies that suggest this to be a reasonable estimate for the number of *Ae. aegypti* adults within a characteristic dispersal radius in a variety of settings [19–21]; however, *Ae. aegypti* adults tend to be relatively sessile, often remaining within the same household unit for the duration of their lifetime [11]. With this in mind, a more accurate model might be *Ae. aegypti* populations distributed across households with migration between them [35]. Areas of future research would be to test the robustness of single-population CKMR methods to data from spatially-structured simulations [36], and to incorporate spatial structure into the CKMR analyses themselves, opening the potential to estimate dispersal parameters using these methods. The theoretical underpinnings of this latter approach have been outlined by Bravington *et al.* [1], and an analogous approach limited to discrete generations and parentage data has been used to estimate dispersal parameters for coral trout [37]. Alternative close-kin methods have also been used to characterize dispersal distances for *Ae. aegypti* [6, 7], and it will be interesting to see whether a spatially-structured CKMR approach can infer complementary information.

The application of CKMR methods beyond fish species has been contemplated since their inception [1], and extending their application to the egg-larva-pupa-adult life history of *Ae. aegypti* mosquitoes is promising for their application to insect species with comparable life histories. A species of particular interest is *Anopheles gambiae*, the main African malaria vector, which has a similar life history, increased dispersal [11] and larger population sizes than *Ae. aegypti* [38, 39]. Age-grading methods are also available for this species, based on ovariole measurements and emerging biochemical and spectroscopic techniques [27]. Incorporating approximate age-at-capture information with kinship data should greatly enhance the precision of CKMR parameter inference, as has been seen for applications to southern bluefin tuna [2] and sharks [3]. The larger size of *An. gambiae* populations also means that smaller population proportions need to be sampled in order to obtain accurate parameter estimates [13]. Although the total required sample size will be higher, lethal sampling is less likely to bias the mortality rate estimate upwards and the population size estimate downwards (as seen for *Ae. aegypti* in Figure 4). Several species of insect agricultural crop pests should also be suited to these CKMR methods, including the medfly and spotted wing *Drosophila*; although theoretical assessments will first be needed, especially for more long-lived pest species such as the pink bollworm.

5. Conclusions

We have theoretically demonstrated the application of CKMR methods to estimate adult and larval parameters for mosquitoes, with *Ae. aegypti* as a case study. CKMR methods have advantages over traditional mark-release methods, as the mark is genetic, removing the need for physical marking and recapturing. Particularly encouraging is the fact that the inferred optimal sampling scheme is consistent with *Ae. aegypti* ecology

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and field studies, meaning that the requisite samples may be obtained with only minor adjustments to current mosquito surveillance programs. Sequencing requirements are significant, particularly for estimating larval parameters; however, as sequencing becomes cheaper and more efficient, this will become less burdensome and perhaps even routine. Work remains to test the robustness of these methods under a range of scenarios in which model components and parameters vary, and in which kinship inference is imperfect; however this study represents an important first demonstration that parameter inference is accurate when the underlying model is known. Application to other insects of epidemiological and agricultural significance is promising, particularly for *An. gambiae*, a major malaria vector for which age-grading methods are available.

Supporting information

S1 Text. Supplemental model equations. Additional equations describing the lumped age-class model of mosquito population dynamics, and kinship probabilities for parent-offspring and sibling pairs that, for brevity, were not included in the manuscript.

Acknowledgments

We thank Dr. Igor Filipović for help with parallelizing code and running simulation replicates, Dr. Eileen Jeffrey Gutiérrez and Dr. Tomás León for discussions regarding mosquito sampling and life history, and Yi Li for help with formulating the kinship probabilities.

Author contributions

Conceptualization: John M. Marshall, Gordana Rašić
Data curation: Yogita Sharma, Jared B. Bennett
Formal analysis: John M. Marshall, Yogita Sharma, Jared B. Bennett
Funding acquisition: John M. Marshall, Gordana Rašić
Investigation: John M. Marshall, Yogita Sharma
Methodology: John M. Marshall, Yogita Sharma, Jared B. Bennett
Project administration: John M. Marshall
Resources: John M. Marshall, Gordana Rašić
Software: Jared B. Bennett, John M. Marshall
Supervision: John M. Marshall
Validation: John M. Marshall
Visualization: John M. Marshall, Gordana Rašić
Writing - original draft preparation: John M. Marshall
Writing - review & editing: Yogita Sharma, Jared B. Bennett, Gordana Rašić, John M. Marshall

Funding statement

This work was supported by a National Institutes of Health R01 Grant (1R01AI143698-01A1) awarded to J.M.M. and G.R. and a DARPA Safe Genes Program Grant (HR0011-17-2-0047) awarded to J.M.M.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Competing interests

The authors have declared that no competing interests exist.

Data availability

The source code for the individual-based mosquito simulation model is available at https://github.com/GilChrist19/mPlex. Documentation, including vignettes, are included for all simulation functions. The source code for inferring parameters based on the likelihood of the kinship data is available at https://github.com/MarshallLab/CKMR. Both sets of code are available under the GPL3 License and are free for other groups to modify and extend as needed.

References

1. Bravington MV, Skaug HJ, Anderson EC. Close-kin mark-recapture. Statistical Science. 2016; 31:259-274.

2. Bravington MV, Grewe PM, Davies CR. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. Nature Communications. 2016; 7:1-8.

3. Hillary RM, Bravington MV, Patterson TA, Grewe P, Bradford R, Feutry P, Gunasekera R, Peddemors V, Werry J, Francis MP, Duffy CA. Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. Scientific Reports. 2018; 8:1-9.

4. Ruzzante DE, McCracken GR, Førland B, MacMillan J, Notte D, Buhariwalla C, Mills Flemming J, Skaug H. Validation of close-kin mark-recapture (CKMR) methods for estimating population abundance. Methods in Ecology and Evolution. 2019; 10:1445-1453.

5. Wacker S, Skaug HJ, Forseth T, Solem Ø, Ulvan EM, Fiske P, Karlsson S. Considering sampling bias in close-kin mark-recapture abundance estimates of Atlantic salmon. Ecology and Evolution. 2021; 11:3917-3932.

6. Jasper M, Schmidt TL, Ahmad NW, Sinkins SP, Hoffmann AA. A genomic approach to inferring kinship reveals limited intergenerational dispersal in the yellow fever mosquito. Molecular Ecology Resources. 2019; 19:1254-1264.

7. Filipović I, Hapuarachchi HC, Tien WP, Razak MA, Lee C, Tan CH, Devine GJ, Rašić G. Using spatial genetics to quantify mosquito dispersal for control programs. BMC Biology. 2020; 18:1-5.

8. Nazni WA, Hoffmann AA, NoorAfizah A, Cheong YL, Mancini MV, Golding N, Kamarul GM, Arif MA, Thohir H, NurSyamimi H, ZatilAqmar MZ. Establishment of *Wolbachia* strain wAlbB in Malaysian populations of *Aedes aegypti* for dengue control. Current Biology. 2019; 29:4241-4248.

9. The Project *Wolbachia* – Singapore Consortium, Ng LC. *Wolbachia*-mediated sterility suppresses *Aedes aegypti* populations in the urban tropics. medRxiv. 2021; doi:10.1101/2021.06.16.21257922.

10. Perkins TA, Scott TW, Le Menach A, Smith DL. Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. PLoS Computational Biology. 2013; 9:e1003327.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11. Guerra CA, Reiner RC, Perkins TA, Lindsay SW, Midega JT, Brady OJ, Barker CM, Reisen WK, Harrington LC, Takken W, Kitron U. A global assembly of adult female mosquito mark-release-recapture data to inform the control of mosquito-borne pathogens. Parasites & Vectors. 2014; 7:1-5.

12. Anderson, EC. CKMRpop: Forward-in-time simulation and tabulation of pairwise kin relationships in age-structured populations. Molecular Ecology Resources. 2021; doi:10.1111/1755-0998.13513.

13. Waples RS, Feutry P. Close-kin methods to estimate census size and effective population size. Fish and Fisheries. 2022; 23:273-293.

14. Eisen L, Monaghan AJ, Lozano-Fuentes S, Steinhoff DF, Hayden MH, Bieringer PE. The impact of temperature on the bionomics of *Aedes* (Stegomyia) *aegypti*, with special reference to the cool geographic range margins. Journal of Medical Entomology. 2014; 51:496-516.

15. Gurney WS, Nisbet RM. The systematic formulation of delay-differential models of age or size structured populations. In Freedman HI, Strobeck C (Eds.), Population Biology (pp. 163-172). 1983; Springer, Berlin.

16. Gurney WS, Nisbet RM, Lawton JH. The systematic formulation of tractable single-species population models incorporating age structure. Journal of Animal Ecology. 1983; 52:479-495.

17. Hancock PA, Godfray HC. Application of the lumped age-class technique to studying the dynamics of malaria-mosquito-human interactions. Malaria Journal. 2007; 6:98.

18. Deredec A, Godfray HC, Burt A. Requirements for effective malaria control with homing endonuclease genes. Proceedings of the National Academy of Sciences. 2011; 108:E874-880.

19. Carvalho DO, McKemey AR, Garziera L, Lacroix R, Donnelly CA, Alphey L, Malavasi A, Capurro ML. Suppression of a field population of *Aedes aegypti* in Brazil by sustained release of transgenic male mosquitoes. PLoS Negl Trop Dis. 2015; 9:e0003864.

20. Lounibos LP. Genetic-control trials and the ecology of *Aedes aegypti* at the Kenya coast. In Takken W, Scott TW (Eds.), Ecological Aspects for Application of Genetically Modified Mosquitoes (pp. 33-46). 2003; Springer, Wageningen.

21. Sheppard PM, Macdonald WW, Tonn RJ, Grab B. The dynamics of an adult population of *Aedes aegypti* in relation to dengue haemorrhagic fever in Bangkok. Journal of Animal Ecology. 1969; 38:661-702.

22. Focks DA, Haile DG, Daniels E, Mount GA. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. Journal of Medical Entomology. 1993; 30:1003-1017.

23. Otero M, Solari HG, Schweigmann N. A stochastic population dynamics model for *Aedes aegypti*: formulation and application to a city with temperate climate. Bulletin of Mathematical Biology. 2006; 68:1945-1974.

24. Simoy MI, Simoy MV, Canziani GA. The effect of temperature on the population dynamics of *Aedes aegypti*. Ecological Modelling. 2015; 314:100-110.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.19.481126; this version posted February 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. Nash JC, Varadhan R. Unifying optimization algorithms to aid software system users: optimx for R. Journal of Statistical Software. 2011; doi:10.18637/jss.v043.i09.

26. Sánchez C. HM, Wu SL, Bennett JB, Marshall JM (2018) MGDrivE: A modular simulation framework for the spread of gene drives through spatially-explicit mosquito populations. Methods in Ecology and Evolution. 2020; 11:229–239.

27. Johnson BJ, Hugo LE, Churcher TS, Ong OTW, Devine GJ. Mosquito age grading and vector-control programmes. Trends in Parasitology. 2020; 36:39-51.

28. Ha T, León TM, Lalangui K, Ponce P, Marshall JM, Cevallos V. Household-level risk factors for *Aedes aegypti* pupal density in Guayaquil, Ecuador. Parasites Vectors. 2021; 14:458.

29. Ndiaye EH, Diallo D, Fall G, Ba Y, Faye O, Dia I, Diallo M. Arboviruses isolated from the Barkedji mosquito-based surveillance system, 2012-2013. BMC Infectious Diseases. 2018; 18:642.

30. Ferede G, Tiruneh M, Abate E, Kassa WJ, Wondimeneh Y, Damtie D, Tessema B. Distribution and larval breeding habitats of *Aedes* mosquito species in residential areas of northwest Ethiopia. Epidemiol Health. 2018; 40:e2018015.

31. Christie MR. Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. Molecular Ecology Resources. 2010; 10:115-128.

32. Wang J. Estimating pairwise relatedness in a small sample of individuals. Heredity. 2017; 118:302.

33. Waples RK, Albrechtsen A, Moltke I. Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. Molecular Ecology. 2018; 28:35-48.

34. Skaug HJ. Allele-sharing methods for estimation of population size. Biometrics. 2001; 57:750-756.

35. Sánchez C. HM, Bennett JB, Wu SL, Rašić G, Akbari OS, Marshall JM. Confinement and reversibility of threshold-dependent gene drive systems in spatially-explicit *Aedes aegypti* populations. BMC Biology. 2020; 18:50.

36. Conn PB, Bravington MV, Baylis S, Ver Hoef JM. Robustness of close-kin mark-recapture estimators to dispersal limitation and spatially varying sampling probabilities. Ecology and Evolution. 2020; 10:5558-5569.

37. Bode M, Williamson DH, Harrison HB, Outram N, Jones GP. Estimating dispersal kernels using genetic parentage data. Methods in Ecology and Evolution. 2017; 9:490-501.

38. Taylor CE, Touré YT, Carnahan J, Norris NE, Dolo G, Traoré SF, Edillo FE, Lanzaro GC. Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. Genetics. 2001; 157:743-750.

39. Touré YT, Dolo G, Petrarca V, Traoré SF, Bouaré M, Dao A, Carnahan J, Taylor CE. Mark-release-recapture experiments with *Anopheles gambiae* s.l. in Banambani village, Mali, to determine population size and structure. Medical and Veterinary Entomology. 1998; 12:74-83.