Detecting recurrent extinction in a metapopulation of *Anopheles gambiae*: preliminary results using simulation

NICHOLAS C. MANOUKIS[†] YOOSOOK LEE[†] EDGAR E. VALLEJO[†§]

CHARLES E. TAYLOR[†]

Dept. Organismic Biology, Ecology and Evolution
University of California, Los Angeles[†]
621 S. Charles Young Dr. East, Los Angeles, CA 90095
USA

Computer Science Dept.
Tecnológico de Monterrey, Campus Estado de Mexico[§]
Carretera Lago de Guadalupe km 3.5,
Atizapán de Zaragoza, Estado de México
MÉXICO

manoukis@taylor0.biology.ucla.edu http://taylor0.biology.ucla.edu/malaria

Abstract

In this paper we present a method for studying recurrent extinction in a metapopulation of *Anopheles gambiae*. We model the metapopulation as a collection of populations, in which individuals are represented as haploid DNA sequences that evolve through mutation, genetic drift and clonal replication. Analysis of simulation results demonstrates the potential of repeated measurements of Watterson's θ for discriminating hypotheses of population structure. These results demonstrate that simulation is an appropriate methodology for studying population genetic questions in biology.

Key-Words : *Anopheles gambiae*, Mali, modeling, simulation, genetic diversity, metapopulation

1 Introduction

The increasing interest in the introduction of genetically modified vectors of disease has heightened the need for suitable population models. Of particular interest is the genetic modification of populations of malaria vectors, especially in areas where the disease is endemic [1]. Population models are especially needed for researchers to assess the potential of such a disease control strategy. The demographic and evolutionary history of biological populations can be inferred using population genetic analysis[2, 3].

Population genetics models have long been formulated as systems of differential equations[4]. While often helpful, these models are also subject to limitations. For example, they commonly assume a very large population size. This assumption misses important small population effects such as genetic drift or extinction[5]. In addition, these models usually employ summary statistics, but describing their behavior under varying demographic conditions in closed form is very difficult. Finally, the introduction of non linearities in rates of drift and other parameters typically makes these models intractable. Even the most promising developments in population genetics, such as coalescent theory[6], have yet to provide an unambiguous window to the past. Consequently, simulation is often helpful by providing an alternative way

to study population biology. The desire is to make the simulation model powerful enough to capture the essentials of population biology scenarios, and at the same time be amenable to manipulation, repeatability, analysis and to be understandable.

In this paper we describe a method to distinguish the genetic signatures of seasonal extinction versus population bottleneck using a well-defined and parametrized situation from the field. The effectiveness of the method is assessed using a realistic simulation of molecular evolution in a metapopulation. We hypothesize that Watterson's θ (θ_w)[7], a ratio of diversity to genealogical depth, will rise very quickly from a lower level with seasonal extinction compared to its change under seasonal bottlenecking.

Preliminary results indicate that θ_w is able to discern between different scenarios in a metapopulation of *Anopheles gambiae*. These results demonstrate that simulation is helpful for identifying and clarifying patterns of population structure.

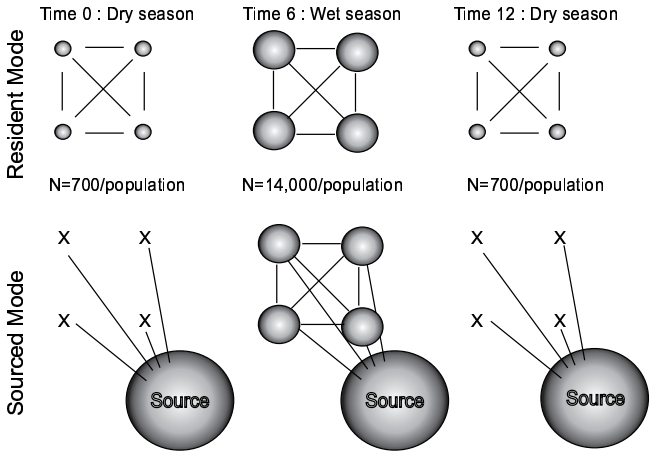

Figure 1: Two extreme hypothesis of *An. gambiae* metapopulation dynamics: resident mode, where peripheral populations bottleneck but don't go extinct and sourced mode where extinction occurs during the dry season followed by re-colonization.

2 Background

Malaria is one of the world's gravest health problems. There are 300 million acute cases of malaria in the world each year, resulting in over a million deaths. 90% of these deaths occur in Sub-Saharan Africa, where it is the leading cause of child mortality (20%)[8]. The costs in human suffering and lost GDP (US$ 12 billion) are staggering, and inflicted on the nations least able to afford them[8].

The mosquito *An. gambiae* is the principal vector of the malaria parasite *Plasmodium falciparum* in Africa. It has been suggested that genetic modification of this species might make it refractory to the malaria parasite, thereby interrupting transmission of the disease[9]. Beyond the creation or discovery of a gene that effectively causes resistance, a major obstacle to this effort is understanding how to introduce the modified organism into natural populations so that it becomes dominant or displaces the wild type[10]. Since temporal and geographic structure in a population can significantly affect its genetic diversity and evolution[11, 12], the population structure of *An. gambiae* is of particular concern.

Forms of population structure include geographic heterogeneity, non random mating or dispersal limits, all of which affect the movement of genes within the target species. In the case of *An. gambiae* several types of population structure have been identified[13].

The village of Banambani (near Bamako) has received especially thorough study[14]. Banambani is located in the Sudan Savanna region of south central Mali. This area is characterized by a long dry season from about November to May[14]. The village comprises 60–70 compounds, and is surrounded by a network of other nuclear villages of similar size, separated by a few kilometers. It is estimated that population sizes become very small during the dry season, especially during March of each year[13].

It is currently unknown if populations of *An. gambiae* in Banambani and surrounding villages go extinct during this time, and then are re-colonized at the outset of the next wet season. It may equally be that village populations are severely bottlenecked, but do not go extinct.

These two alternative scenarios represent hypothetical extremes of metapopulation dynamics. We use a broad definition of a metapopulation as a collection of populations with some degree of gene flow and some probability of extinction and recolonization[15]. In this study we simulated these scenarios, with the goal of identifying population ge-

netic signatures that might allow them to be distinguished.

We term the extreme dynamics *resident* and *sourced* modes (see Fig.1). The difference between them is the permanence of peripheral populations (villages) in face of large overall population size changes within a year. It is likely that some populations may go extinct and others not. Also, there may be some influence of a source despite persistence over the dry season. We explore the latter possibility, namely a *mixed* mode, in addition to the two extreme modes.

3 Methods

We simulated a collection of four peripheral populations plus a large *source* population. Each population is a group of individuals that inhabit a single geographic location (equivalent to a village) and is reproductively continuous in time. An individual is represented as a haploid, neutral sequence of DNA. As described below this should equally well represent a population that is diploid, but with half the population size. The four populations are subject to several types of events: mutation, migration, birth, and death.

We use exclusively point mutations with an infinite alleles model. The length of a sequence is set to be equivalent to the length of data we can easily generate with a single sequencing reaction using current technology. The mutation rate (μ) is comparable to fast-evolving neutral nuclear DNA [4]. The use of a neutral gene represents the absence of functional constraints on where a mutation can occur.

A migration event removes a randomly chosen individual from one population and appends it to another. The migration rate (m) is taken from estimates for *An. gambiae* in Banambani[16]. A small effect of inter-village distance on gene flow was found[16] for distances in the range of the villages we are simulating, so we use a fully connected network between the villages. Under this arrangement each destination is equiprobable for a dispersing individual. In reality migration is a bit more complicated.

Births are equivalent to clonal reproduction, where a new copy of a randomly selected sequence is made and appended to the population. The number of mutations (N_μ) is tied to the number of births (N_b):

$$N_\mu = \mu \times N_b \tag{1}$$

Thus while mutation may not occur exactly during reproduction they are not independent. Death is the removal of a randomly selected individual from a population.

The change in population size (ΔN) is a product of the number of births (N_b) minus the number of deaths (N_d):

$$\Delta N = N_b - N_d \tag{2}$$

Every village has the same predefined profile of population size change, based on inter and intra year size estimates from genetic and demographic surveys in Banambani[13] (See Table 1). The model explicitly includes turn over: a baseline number of births and deaths in addition to the excess of one or the other needed to achieve a particular ΔN. We set this to a value constrained by computational time and a reasonable amount of molecular evolution between time steps.

Each simulation starts with four populations and a source of genetically identical individuals. For every time step, the number of births, deaths, mutations and inter-village migrations are calculated. The events are executed in arbitrary order creating an asynchronous simulation. At the beginning of every time step, θ_w is calculated from a small sample for each village using equation 3,

$$\theta_w = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{i}} \tag{3}$$

Table 1: Peripheral population size(N_p) at each time step of a single year. Note that the source population size(N_s) is constant.

Time	Population size[a]	Time	Population size
00	700	06	13996
01	2916	07	11780
02	5132	08	9564
03	7348	09	7348
04	9564	10	5132
05	11780	11	2916

Table 2: Experimental Conditions: extinction probablity during dry season (e) and migrant origin (m_o: 0 = all peripheral)

	Mode		
Variable	Resident	Sourced	Mixed
e	0	1	0
m_o	0	0	0.5

Table 3: Mode independent parameters. Maximum and minimum peripheral sizes ($N_{p,max}$ and $N_{p,min}$), source size (N_s), mutation rate (μ) and migration rate (m), sample proportion of population(P_{sample}) to estimate θ_w.

Variable	Value	Variable	Value
t_{total}	10,000	μ[b]	3×10^{-6}
$N_{p,max}$	14,000	m	0.01
$N_{p,min}$	700	Turnover	0.4
N_s	30,000	P_{sample}	0.01

[b]The unit of μ is per nucleotide per update.

where S is the number of segregating sites and n is the number of alleles (or lineages) in the sample[7].

The experimental conditions (or modes) we wish to distinguish differ in two parameters (Table 2): Extinction during the dry season (e) and migrant origin (m_o). Resident and sourced mode have extinction probabilities of 0 and 1, respectively, for the time step when population size is the smallest. After extinction in sourced mode a founding pod of individuals is drawn from the source. All migrations (non-founding movements) in both these modes involve movement of individuals among peripheral populations. Under mixed mode there is no extinction but 50% of the migrants originate from the large source population (note that $m_o = 1$ means all migrants come from the source).

We parametrized the elements and other events in the model according to our best understanding of the situation in the field, as summarized in Table 3. Most of these parameters are described by Taylor *et. al.*[17].

θ_w is calculated from a small sample to make the estimates comparable to those we expect from the field, where we will capture only a small fraction of the total population. We conducted statistical analysis on the variation of θ_w for four continuous sample years at the end of the simulation. We assume mutation-drift equilibrium at this point in the simulation since t is greater than $4N_e$ generations [18, 4] by then. A brief algorithm of the simulation is given in Fig. 2.

```
for each village v_i
  create an initial population P_i
end for
if mode = source or mode = mixed
  create a source population S
end if
do until number of generations is met
  for each village v_i
    select a random subpopulation Q_i ⊂ P_i
    compute θ_w^i for Q_i
    for each different event e_j
      compute the number of events n_j for e_j
    end for
    create a list L of ∑_j n_j events l_k
    for each event l_k ∈ L
      case l_k
        birth:
          select a random individual p_x ∈ P_i
          create a copy p_y of individual p_x
          append p_y to P_i
        death:
          select a random individual p_x ∈ P_i
          remove p_x from P_i
        mutation:
          select a random individual p_x ∈ P_i
          select a random loci p_x^l in p_x
          substitute the base at p_x^l randomly
        migration:
          select a random village v_h ≠ v_i
          select a random individual p_x ∈ P_i
          remove p_x from P_i
          append p_x to P_h
      end case
    end for
  end for
end do
```

Figure 2: Algorithm for the simulations

4 Experimental Results

We executed each simulation for 10,000 time steps, equivalent to over 800 years. We analyzed mean θ_w for the four village populations over four years, between year 829 and 833 of the simulation. Typical results for individual villages under resident and sourced modes are shown in Fig. 3. Each mean θ_w can be considered the value for a metapopulation un-

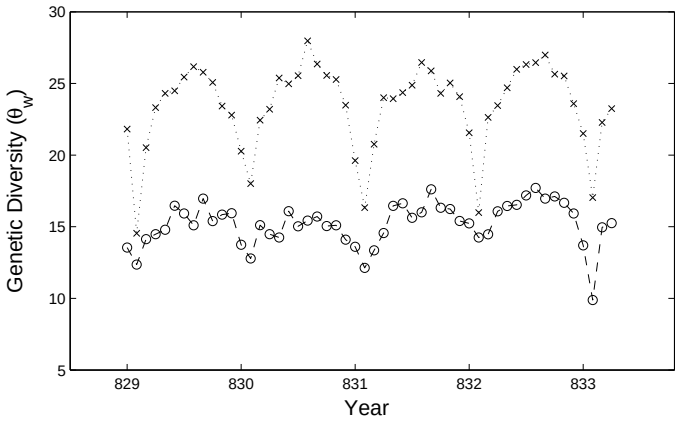

Figure 3: Changes in genetic diversity in two sample villages under resident (∘) and sourced (×) modes

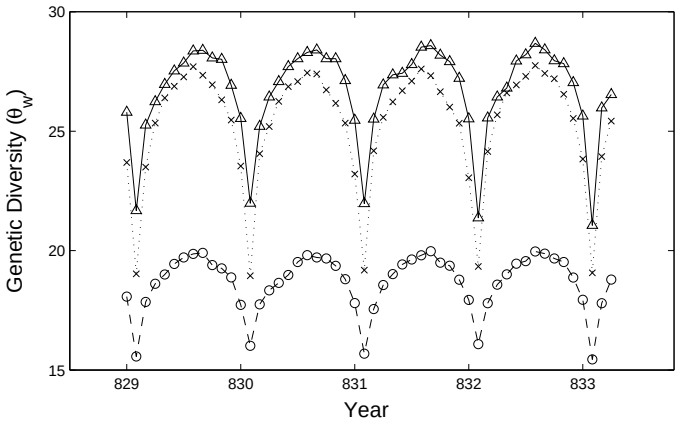

Figure 4: Average θ_w from four villages within six separate runs under each of the experimental conditions: resident (∘), sourced (×) and mixed (△). The amplitude of the curves are significantly different across conditions for these four years (see text).

Table 4: 2 way ANOVA for effect of year and mode on amplitude of genetic diversity fluctuation. Significance codes: $^{*}(p \leq 0.05)$, $^{**}(p \leq 0.01)$, $^{***}(p \leq 0.001)$

	df	SS	MS	F
year	3	1.09	0.36	0.2252
mode	2	328.80	164.40	101.7030***
mode*year	6	2.66	0.44	0.2744
total	60	96.99	1.62	

der the condition being investigated. We carried out 6 runs per mode. The mean θ_w of each mode are plotted in Fig. 4.

We measured the difference between the minimum and maximum θ_w for each of the metapopulations per year under each condition, which is the amplitude of the θ_w fluctuation. We found no effect of year on the amplitude of θ_w. There was, however, a statistically significant effect of mode on this measure (See Table 4). An analysis of only sourced and mixed modes using ANOVA also revealed statistically significant differences (result not shown).

5 Discussion

The experimental results show that the sourced mode metapopulation begins with low genetic diversity which rises quickly to high levels. These may be indistinguishable from the wet season peak values under resident mode depending on the rate of migration and the exact population size fluctuation (results not shown). This may seem counter intuitive because overall higher levels of diversity are expected in a large source population. The reason for the low diversity immediately following peripheral colonization is that the founding pod is low in diversity since it is a small fraction of the size of the source population. As soon as any migration occurs between villages, though, diversity increases rapidly.

Mixed mode, which explores the effect of migrant origin on genetic diversity fluctuation is also revealing: the levels of wet season diversity are higher with a source contributing migrants, as expected. However, the dry season dip is deeper than under resident mode despite the fact that neither experiences extinction. We are unsure why this is occurring but plan to explore it in further analysis, described below.

We expect these results generally are dependent on the size of the source population. A small source may cause a smaller amplitude of genetic diversity fluctuation. We plan to quantify the effect of the model parameters in a sensitivity analysis to explore this hypothesis and investigate the role of migrant origin fully.

It should be noted that *An. gambiae* is a diploid organism. If there is no dominance among haplotypes, the diploid model should provide the same results as the haploid model [19]. However, a diploid model will allow extensions of the mutation model to include recombination and the examination of other population

genetics parameters, such as heterozygocity. The development of such a model is currently in progress in our laboratory.

Overall, the simulations and analysis presented here show that it should be possible to distinguish complex dynamics with repeated measurement of a simple population genetic parameter. We have begun collections at Banambani that will allow us to determine what happens during the dry season. We expect that our observations will bear on the feasibility of genetic modification of *An. gambiae* for malaria control.

Acknowledgement We thank Travis C. Collier for helpful discussions and programming advice. This work is supported by a grant from NIH-NIAID to CT. EV is supported by a grant from UC MEXUS-CONACYT.

References

[1] Http://www.gatesfoundation.org.

[2] Taylor C. and Powell J. *Population structure of Drosophila: genetics and ecology.*, vol. 3D, chap. 28. Academic Press, 1983, pp. 26–60.

[3] Slatkin M. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, vol. 12(3), 1985, pp. 253–262.

[4] Hartl D.L. and Clark A.G. *Principles of Population Genetics.* Sinauer Associates, Sunderland, Massachusetts USA, 1997.

[5] Taylor C.E. and Jefferson D. Artificial Life as a Tool for Biological Inquiry. *Artificial Life*, vol. 1, 1994, pp. 1–13.

[6] Fu Y. and Li W. Coalescing into the 21st century: An overview and prospects of coalescent theory. *Theoretical Population Biology*, vol. 56, 1999, pp. 1–10.

[7] Watterson G. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, vol. 7(2), 1975, pp. 256–276.

[8] *RBM information sheets.* WHO Roll Back Malaria Initiative, 2002.

[9] James A.A. and Handler A.M. Control of disease transmission through genetic modification of mosquitoes. *Insect transgenesis: Methods and applications*, 2000, pp. 319–333.

[10] Boete C. and Koella J. Evolutionary ideas about genetically manipulated mosquitoes and malaria control. *Trends in Parasitology*, vol. 19, 2003, pp. 32–38.

[11] Slatkin M. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, vol. 12(3), 1977, pp. 253–262.

[12] Pannell J. Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution*, vol. 57, 2003, pp. 949–961.

[13] Taylor C. and Manoukis N. *Effective population size in relation to genetic modification of Anopheles gambiae sensu stricto*, chap. 10. Kluwer Academic, 2003, pp. 133–148.

[14] Touré Y., Dolo G., Petrarca V., Traore S., Bouare M., Dao A., Carnahan J., and Taylor C. Mark-release-recapture experiments with Anopheles gambiae sl in Banambani Village, Mali, to determine population size and structure. *Medical and Veterinary Entomology*, vol. 12, 1998, pp. 74–83.

[15] Hanski I. and Gilpin M. *Metapopulation biology: ecology, genetics and evolution.* Academic Press, 1997.

[16] Carnahan J., Zheng L., Taylor C., Toure Y., Norris D., Dolo G., Diuk-Wasser M., and Lanzaro G. Genetic differentiation of Anopheles gambiae s.s. populations in Mali, West Africa, using microsatellite loci. *Journal of Heredity*, vol. 93, 2002, pp. 249–253.

[17] Taylor C., Touré Y., Carnahan J., Norris D., Dolo G., Traor S., Edillo F., and Lanzaro G. Gene flow among populations of the malaria vector Anopheles gambiae, in Mali, West Africa. *Genetics*, vol. 157(2), 2001, pp. 743–750.

[18] Kimura M. and Crow J. The number of alleles that can be maintained in a finite population. *Genetics*, vol. 49, 1964, pp. 725–738.

[19] Calabretta R., Galbiati R., Nolfi S., and Parisi D. Investigating the role of diploidy in simulated populations of evolving individuals. In *Electronic Proceedings of the 1997 European Conference on Artificial Life*, 1997.