Variance partition reveals contrasting random effect contributions on the density and species composition of malaria-transmitting mosquitoes

Authors and affiliations

Tin-Yu J. Hui[1], Patric Stephane Epopa[2], Abdoul Azize Millogo[2,3], Franck A. Yao[2], Dao Koulmaga[2], Austin Burt[1], Abdoulaye Diabate[2]

1. Silwood Park Campus, Department of Life Sciences, Imperial College London, Ascot, United Kingdom
2. Institut de Recherche en Sciences de la Santé / Centre Muraz, Bobo-Dioulasso, Burkina Faso
3. Institut des Sciences des Sociétés, 03 BP 7047 Ouaga 03. Ouagadougou,Burkina Faso

Corresponding author: Tin-Yu Hui

Abstract

Spatial-temporal variation exists in the density and species composition of malaria-carrying mosquitoes, which will in turn influence that of disease transmission. While there has been extensive research on the season effect and other main drivers on the vectors' populations, the heterogeneity induced by random effects are less studied. To investigate the relative contributions of the between-house, between-village, and between-year variations, as well as other house-level covariates such as inhabitant size and bed net usage, intensive Pyrethroid Spray Catches (PSC) sampling was conducted across a 60-month period between 2012-2019 from four villages in the Sudano-Sahelian region of Burkina Faso.

For mosquito density, measured by the female *Anopheles gambiae s.l.* counts, our modelling showed that the between-house variation was the largest component, followed by the between-year then between-village variations, after accounting for seasonality and other covariates. Density increased with the number of habitants within a household but was unaffected by bed net usage. A subset of female mosquitoes from the species complex was genotyped for species identification. The composition of *An. coluzzii* and *An. gambiae*, the two dominant vectors in the region, varied hugely across villages without a clear seasonal tread. The between-village variance contributed up to 76% of the total random variation, followed by the between-year variance. The between-house variation was estimated to be insignificant. Neither household size nor bed net usage had any impact on their proportion.

In short, the relative importance of the random components in mosquito density was in the reverse order from species composition. The variance estimates and relative strengths help parameterise potential field trials for novel vector control programmes and monitoring.

Introduction

Despite the continued effort and investment on malaria control programmes, regions in sub-Saharan Africa remain disproportionally affected by the vector-borne disease. Harrowing figures of 263 million cases with over half a million deaths are still being reported on the continent, accounting for about 95% of the global malaria burden (WHO, 2024). Recent modelling predicted that the current intensity and range of interventions (such as long-lasting insecticidal nets and indoor residual spraying) can at best maintain the status quo but are unable to drive malaria towards eradication (Griffin et al., 2016). In fact, the recent increase in malaria incidence coincides with the vectors' resistance to insecticides (Kientega et al., 2024; Purugini et al., 2024), urging the need of developing complementary efficient, cost-effective and sustaining novel technologies for malaria control. Genetic modification (GM) of the vectors, such as by engineering a transgene into the wild populations to reduce vector competence, is a promising tool. This has been previously demonstrated by the release of sterile *Aedes aegypti* males for dengue control (Carvalho et al., 2015) and of many insect species that are deemed harmful (Bourtzis and Vreysen, 2021). Beyond the sterile insect technique, there has been extensive theoretical and laboratory-based research on developing gene drive systems for *Anopheles gambiae s.l.*, a species complex whose members contain some of the primary malaria vectors in sub-Saharan Africa. Gene drive allows the transgene to be passed onto their offspring and beyond at a higher fraction via super-Mendelian inheritance to achieve longer-term population suppression or replacement (Kyrou et al., 2018; Marshall and Akbari, 2016).

For contagious or vector-borne diseases, the basic reproductive number R_0 is the most important metric to quantify their transmission, with $R_0 < 1$ being a requirement for eradication. In malaria, R_0 is generally defined as the expected number of persons to be infected one generation after the malaria parasite was introduced, or equivalently the number of secondary cases (Smith et al., 2007; Griffin, 2016). As the life cycle of the malaria parasite consists of both mosquito and human phases, characterising the host-parasite-vector interaction can be a complex exercise (Mandal et al., 2011). The

entomological inoculation rate (EIR), for example, connects the sporozoite rate, the prevalence of the parasite in the salivary glands of female mosquitoes, and human biting rate (HBR) per night, which has the mosquito density per person embedded (Brady et al., 2016). Under the classical models R_0 and vectorial capacity are functions of these entomological rates at equilibrium. To reduce the intensity of transmission, most vector control programmes aim to supress factors concerning the vectors' density or competence (Smith et al., 2007). Malaria transmission is never uniform across space and time. Previous estimates of R_0 varied hugely from ~1 to >1000 across Africa (Smith et al., 2007), and mosquito density fluctuated across seasons often by orders of magnitude (Dao et al., 2014). In fact heterogeneity exists on almost all levels: from mosquito density and species distribution across countries, exposure among households within the same village, down to the susceptibility to malaria of individuals.

While many earlier works have been focusing on identifying the main drivers, such as environmental and climatic factors, associated with the trends in malaria vectors and disease transmission (Bationo et al., 2023), it is equally important to quantify the unexplained portion of the variation. For instance, how does mosquito exposure vary among neighbouring households given they are exposed to the same conditions? What is size of the variation across villages compared to other factors? Through our extensive mosquito collection survey (see below) we have the following aims: 1) We would like to estimate the random or unexplained variance components concerning mosquito density and species composition at different spatial-temporal levels, in particular, the between-village, between-house, and between-year variation, after accounting for seasonality and other fixed covariates. 2) We will discuss how these random components and their relative strengths impact vector density and malaria transmission. 3) With a potential clustered randomised controlled trial (cRCT) in mind, we will provide recommendations on the trial design and sample size considerations under various sampling and intervention scenarios.

<u>Sampling</u>

Mosquitoes were sampled in the houses from four villages in Burkina Faso's Sudano-Sahelian: Bana Market, Bana Village, Souroukoudinga, and Pala (Figure 1), where morphologically identical *An. gambiae (s.s.)*, *An. coluzzii*, and *An. arabiensis* are the primary vectors. The first three villages locate to the west of Bobo-Dioulasso, where our research base situates, while Pala is on the opposite direction from the rest (Epopa et al., 2018). Bana's two collection sites, Bana Market and Bana Village, are separated by a small semi-permanent river with different human activities (Epopa et al., 2017). Two phases of collection were conducted: from July 2012 – June 2014 and throughout 2017-2019, covering a total of 60 months. Monthly collection was scheduled aiming to sample 20 houses per village per month. Houses were chosen randomly from a pool of known houses based on local information, from which half of the houses were repeatedly visited over time (fixed houses), while the remaining half were further randomised each month (random houses). For each village, we attempted to sample all houses in the same or consecutive days to minimise other nuisance abiotic factors. For Pala and Souroukoudinga, the sampling interval changed to once every two months in the second phase (Figure 2). Sampling took place after sunrise via indoor Pyrethroid Spray Catches (PSC), using a well-known and locally-market available insecticide spray Kaltox® Paalga (Sphyto company, Burkina Faso). All mosquitoes were morphologically identified in the field (Holstein, 1949). The targeted organism *An. gambiae s.l.* was preserved in 80% ethanol for further analysis. During the second phase of collection PCR was performed on a subset of the preserved samples for species identification, such that the breakdown of the species (within the species complex) per visit was also known. The genotyping was based on the detection of SINE 200x locus (Santolamazza et al. 2008).

Methods

Each house visit generated a line of record, including the numbers of female and male *An. gambiae s.l.*, date of collection, and a unique house ID. Other potential covariates concerning the visit were also documented, such as the number of residents, and whether bed net was in use. The number of female *An. gambiae s.l.* collected per visit was the main response to proxy indoor mosquito density. To model seasonality we used

a six-level factor variable, with two consecutive months being pooled as one (e.g. January and February as one level, and so on) for a more balanced design. A Poisson Generalised linear mixed model (GLMM) was fitted. Seasonality, household size, and the installation of bed net (yes/no) were treated as fixed effects. Observational random effect was included as a measure of overdispersion (Harrison, 2014). Villages, year of sampling, and unique house IDs were considered as random effects. The interaction between village and seasonality was also included to allow additional variability. For species composition, a similar binomial GLMM was fitted to the proportion between the female counts of the two most dominant species in the region. Analyses were conducted in R-4.3.1 (R Core Team, 2024) with packages lme4 and emmeans (Lenth, 2025).

Results

We endeavoured to follow closely the intended sampling plan with 3212 visits made (Figure 2). After removing entries with missing information, 3133 visits from 383 distinct houses were retained over the said sampling period. Mosquito density as measured by the number of female *An. gambiae s.l.* is summarised as follows: A total of 28919 female *An. gambiae s.l.* were collected, with an average of 9.2 per visit. The high mean count however came with huge variation across visits, ranging from 0 to 333 per visit. The mean household size was just over 3 persons – the same value was subsequently used to classify household size into categories (none, small with 1-3 persons, large with >3 persons). Bed nets were found in about 90% of the visits.

A Poisson GLMM was fitted to the counts (Table 1A). Seasonality was well defined within a calendar year with a period of low and high counts. Lower counts were observed between November and February, then increased when transitioning into the next season towards the peak at the months of September and October. There was about 30-fold difference between the lowest and highest mean counts. As we grouped seasonality into six two-monthly factor levels, we performed contrasts to test whether their mean counts were different. All pairwise comparisons were significant, apart from

the three levels of months between May and October where they shared the same mean. The installation of bed net reduced mosquito count by 14%, but the effect was not statistically significant (p=0.099). Count also increased with household size: compared to an empty house a 3-fold (p=0.013) increase was observed if there were ≤3 occupants, then a further 53% increase (p<0.001) for a bigger household with >3 occupants. For the random components (Table 1B), observation-level overdispersion had the highest contribution, accounting for 65% of the total variance. The second largest term was the between-house variation (14%). The between-village variance was found to be of the lowest at 4.6%. The village-seasonality interaction allowing for additional variation on top of the two random components was estimated to be of moderate strength (8%). All five variance components were significantly larger than zero (p<0.05).

A subset of female *An. gambiae s.l.* from the second phase of collection (2017-2019) was genotyped for species identification. After genotyping, 3626 were identified as *An. coluzzii* and 1102 as *An. gambiae*, yielding an overall ratio of approximately 3:1 from 985 house visits (Figures 5-6). Apart from the two dominant species, a small proportion of *An. arabiensis* (~5%) was also found but excluded from subsequent analyses. The species proportions in our four villages were very different (Figure 4). In Bana Market and Bana Village *An. coluzzii* was the dominant species making up over 90% of the samples, and similarly 60% in Souroukoudinga, while in Pala it became the minority (24%). Although species composition varied throughout a year, the four villages did not seem to follow the same dynamics nor a pre-defined pattern. A binomial GLMM was then fitted to model the relative proportion of the two species. For the fixed effects, neither the use of bed nets (p=0.77) nor household size (p=0.89) had influence on the proportion (Table 2A). For the random components (Table 2B), the between-village variation was estimated to be the largest explaining 77% of the total random variance. This was followed by overdispersion (11.4%) and the village-seasonality interaction (3.9%). The between-house variation was not significant (1.5%, p=0.26). In other words, putting overdispersion aside, the relative importance of the random components was the exact opposite of that for mosquito density.

Discussion

Several baseline collections have been conducted in Burkina Faso and beyond to evaluate the vectors' density over space and time (Lehmann et al., 2014; Epopa et al., 2019; Muleba et al., 2023). Our results agree with the existing knowledge that mosquito density (as measured by female counts in this study) exhibits strong seasonal trends. The study region is characterised by a tropical climate marked by two main seasons, with average annual temperatures between 25-30 degrees Celsius and rainfall of 800-1100mm (Millogo et al., 2023). The rainy season is usually defined as the months from June to September, with May and October being in transition (Epopa et al., 2017), the same period at which we consistently observed higher counts. The peak at the late rainy season near October was also noticed by Dao et al. (2014) in a Sahelian village in the neighbouring country of Mali. After reaching the peak a period of low counts immediately followed from November into February. The clean switch between the two seasons is a signature of the Sahelian region as the surface water disappears, drastically reduces the availability of larval sites (Lehmann et al., 2017; Biasutti, 2019). The huge difference between the rainy and dry season mean counts we observed also fell within the same order of magnitude as existing census and genetic estimates (Dao et al., 2014; Mwima et al., 2024).

While many existing collections aimed to correlate seasonality and other deterministic drivers with mosquito dynamics through monthly summaries, curating individual PSC records gives opportunity to analyse covariates on finer scales such as on observational and household level. Bed net use, a factor considered as a fixed effect in this study, only reduced mosquito density by 14% (but not significant) despite its high coverage of over 90%. A separate study discovered the increase in frequency of insecticide resistance alleles during the same period which potentially explains its relatively low efficacy (Kientega et al., 2024). We also found a strong relationship between density and household size, a factor which was previously thought to be unimportant (Lehmann et al., 2014). Another dimension of this work is to estimate the random effects, which

induce uncontrolled heterogeneity that have not been explained by our main effects (Smith et al., 2014). We intentionally fitted a 6-level seasonality variable with 24 additional random village-seasonality interaction terms to our GLMMs to saturate the seasonality effect, allowing us to focus on these random components. The sizes and relative importance of the random components are of interest rather than how and why they differ. Several random components on various levels were considered: Overdispersion, assigned to each visit, refers to the intrinsic variation among PSC sampling thus little can be done to further reduce it. Overdispersion was also the largest component on mosquito density. By revisiting some households, the between-house variance was estimated to be the second largest component, hinting some houses were more susceptible to mosquitoes even though they were exposed to similar conditions as their neighbours. The difference could be due to building material, usage, and proximity to larval sites. Further modelling, including the use of GIS data, is welcome to further dissect this variance component (Lehmann et al., 2014). The between-village and between-year variances were minor with respect to density, each contributing to only a few percent of the total random variation.

There is no surprise that the two primary vectors *An. coluzzii* and *An. gambiae* were abundant in the region. Our binomial GLMM concluded that the proportion between the two species was unaffected by bed net usage or household size. The four sampled villages tended to have their own underlying proportions and seasonal variations: Souroukoudinga and the two sites from Bana were predominantly *An. coluzzii*, while Pala was the opposite. There was no observable switch in the proportions across seasons. The large between-village and village-seasonality interaction illustrated how diverse species composition can be across landscape. It is known that the two species coexist but with rather contrasting preferences. *An. coluzzii* prefers sites with permanent freshwater while *An. gambiae* can tolerate intermittent access to water (Lehmann et al., 2017). Bana and Souroukoudinga have access to water year-round due to agricultural activities thus explains the strong presence of *An. coluzzii*. Their respective persistence mechanisms may also affect the relative abundance in the dry season and transition months. It is generally thought that *An. coluzzii* aestivates locally

during the dry season thus an immediate population expansion is anticipated when the aestivators reemerge, whereas *An. gambiae* takes days to respond to the arrival of the rainy season as immigrants from further afield recolonise the focal population (Mwima et al., 2023). In short, the large between-village and lack of between-house variance were the key signatures of species composition. In fact, the relative importance of the random components in species composition is the exact opposite of that of density (Table 3), which itself is a very interesting finding. The design of population monitoring programmes and control strategies will therefore be influenced by the choice of response (see below). Future works should include the monitoring of *An. arabiensis*, which was also present in our study, especially in urban settings where it appears to be colonising (Longo-Pendy et al., 2021).

Both mosquito density and malaria transmission are subject to strong spatial-temporal variation. Besides our findings on density, a previous survey in the same region estimated a 25-fold difference in EIR between the rainy and dry season, mainly driven by the sharp rise in human landing rate (Epopa et al., 2019). Both malaria incidence and mosquito density found their peaks towards the late rainy season in Burkina Faso (Sangare et al., 2022), meaning seasonal strategies may be required (Grant et al., 2023), further straining the healthcare system. For GM vector control, modelling results demonstrated that elimination is more likely in regions with less pronounced seasonality (North et al., 2020). . The timing of the release of GM mosquitoes is also key to the effectiveness as maximum efficacy is achieved when the peak of transgene frequency synchronises with the trough of vector population size (Leung et al., 2022). While the heterogeneity in transmission across seasons or climatic zones have been anticipated, the unexplained variance components impose additional challenges on the road to the elimination of malaria. Persistent transmission can be maintained even under continued and effective malaria control (Andolina et al., 2021). It only requires a minority of human individuals with dry season exposure to constitute an infectious reservoir (Selvaraj et al., 2018). The higher the random variances in mosquito density the higher chance these small pockets of space are created for permanent transmission. Classical model showed that with heterogeneous biting, a factor of $(1 +$

α) is applied to R_0 where α is the squared coefficient of variation (CV) in biting rate. In other words, R_0 increases linearly with the variance in biting rate if all else being equal. While biting rates are best estimated from human landing catches, our PSC collection and between-house variance estimate in density could be carefully extrapolated to inform the degree of heterogeneity linking to indoor exposure and biting. With the many layers of random variations, the compounding effect on R_0 could potentially be huge even if the mean estimate suggests otherwise. Biting can also occur outdoor, thus measuring the outdoor mosquito density and species composition with other trapping methods will complement this study (Degefa et al., 2024). In terms of reporting, increased spatial and temporal heterogeneity in malaria cases in low transmission settings reduces the usefulness of national or regional level trends in incidence or prevalence (Routledge, 2018).

Any new intervention needs to undergo phases of rigorous testing and trial. A cRCT is the gold-standard study design for evaluating public health programmes, where interventions are addressed to some clusters then compared to those without (Hayes and Moulton, 2017; Hemming et al., 2018). Not only do this rich dataset and parameter estimates contribute towards the entomological baseline collection for a potential upcoming field trial, they also shed light on some of the key considerations concerning the trial design. . In many mosquito intervention programmes treatments must be addressed to a cluster of households (e.g. a village) as a whole hence it is the basic unit for randomisation. If the proposed intervention targets the entire species complex to suppress mosquito density by a certain percentage, then its effect can be evaluated from the difference in counts between the control and treatment arms via a Poisson GLMM like the one we fitted but with an additional variable indicating the treatment arm. We suggest sampling should prioritise in months with high abundance to maximise the chance of detecting the intervention effect. There are two reasons for this: The large baseline counts in the rainy seasons provide headroom for population suppression, while in the drier months the monthly aggregated counts are low with numerous zero readings from individual houses. The second reason is the ability to detect, that estimate is more likely to accompany with a lower CV (larger effect size). Observation-

level covariates such as household size were shown to be confounding with the baseline counts. For the same reason, the recruitment of households should target those with habitants. These covariates are essential when the cRCT is analysed from observation-level data (Campbell et al., 2000), and remain useful under the use of cluster-level summaries, where they can inform the pairing of clusters in restricted randomisation, or even a matched-pair cRCT design to further improve balance and statistical power (Imai et al., 2009).

The random effect estimates will too influence the design of the cRCT. The current sampling of 20 houses per month per village aimed to illustrate seasonality by averaging out the overly dispersed count data. One can find the optimal sampling strategy by altering the number of clusters and houses per month per cluster while keeping the total sampling effort constant. Given the relatively high importance of the between-house variation, sampling the same set of houses too intensively may risk inducing systematic bias. Small but non-zero between-year variation in density was detected across the 6-year period. Yearly variation is less of a problem for the classical parallel cRCT design as comparisons are made mostly within the same time period (vertical) between the two arms. A stepped-wedged design, however, contains both vertical and within-cluster (horizontal) comparisons, may be prone to this effect. In this design, clusters are randomised into several streams. All clusters are in control condition initially and will have received the intervention at the end, with intervention being gradually rolled out at different time points of the trial. As all clusters now include pre- and post-intervention phases, a good year (with respect to mosquitoes' survival) alone may result in a sudden increase in mosquito density even in clusters receiving the intervention. On a positive note, the between-year variance was estimated to be small with limited impact, contributing to only a few percent of the total variation in both mosquito counts and species proportions. While our four villages were chosen based on logistical reasons, they were considered as a random effect in the GLMMs as this would be the case in a cRCT. A non-zero between-village variance was estimated despite them being in the vicinity (~15km radius) of Bobo-Dioulasso. A full-scale cRCT would likely be conducted on a regional or even national basis with many more clusters,

thus this variance is expected to expand with geographical coverage. Note that these parameter estimates (Tables 1A, 1B) are reported in the log scale, which means they should be compared to the intervention strength of log(1-G) if the intervention suppresses the density by G. In trial designs, the between-village variance informs the intra-cluster correlation coefficient (ICC), which measures relative similarity of the observations from the same cluster compared to those from the others (cite). Since individual variances (i.e. overdispersion and between-house variation) were estimated to be much larger than between-village variance, the ICC is expected to be reasonably small, which helps reduce the number of clusters required.

Some novel vector control strategies can be tailored to target one or a subset of species while keeping the rest intact. The results from such a cRCT can be analysed with a similar Poisson GLMM on the counts of the targeted species often with the non-targeted species as covariates. Another way is to analyse the relative proportion of the two as a pair of joint response with a binomial GLMM, whose intervention effect is log(1-G) on the logit scale. Between *An. coluzzii* and *An. gambiae*, the ICC in proportion is thought to be much higher than in their combined density. In other words, most variation comes from clusters rather than the individual houses, thus the effective sample size approaches the number of clusters. Under high ICC the stepped-wedged design may be preferred as the efficiency is improved by the inclusion of horizontal comparisons (Hemming and Taljaard, 2016). The lack of between-house variation among houses has the potential to simplify the recruitment process, as one or a few households can represent the entire cluster. To conclude, our analyses provided crucial parameter estimates to be fed into trial design, power and sample size determination. With the many combinations of variables and scenarios, quantitative studies are necessary to fully explore their complex interplay, often aided by computer simulations.

References

1. Venkatesan P. WHO world malaria report 2024. *The Lancet Microbe* 2025.

2. Griffin JT, Bhatt S, Sinka ME, Gething PW, Lynch M, Patouillard E, et al. Potential for reduction of burden and local elimination of malaria by reducing Plasmodium falciparum malaria transmission: a mathematical modelling study. *The Lancet Infectious Diseases* 2016;16(4):465–472.

3. Kientega M, Morianou I, Traoré N, Kranjc N, Kaboré H, Zongo ON, et al. Genomic analyses revealed low genetic variation in the intron-exon boundary of the doublesex gene within the natural populations of An. gambiae sl in Burkina Faso. *BMC Genomics* 2024;25(1):1207.

4. Perugini E, Pichler V, Guelbeogo WM, Micocci M, Poggi C, Manzi S, et al. Longitudinal survey of insecticide resistance in a village of central region of Burkina Faso reveals co-occurrence of 1014F, 1014S and 402L mutations in Anopheles coluzzii and Anopheles arabiensis. *Malaria Journal* 2024;23(1):250.

5. Carvalho DO, McKemey AR, Garziera L, Lacroix R, Donnelly CA, Alphey L, et al. Suppression of a field population of Aedes aegypti in Brazil by sustained release of transgenic male mosquitoes. *PLoS neglected tropical diseases* 2015;9(7):e0003864.

6. Bourtzis K, Vreysen MJ. Sterile insect technique (SIT) and its applications. *Insects* 2021;12(7):638.

7. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A, Beaghton AK, et al. A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat Biotechnol* 2018;36(11):1062–1066.

8. Marshall JM, Akbari OS. Gene drive strategies for population replacement. *Genetic control of malaria and dengue*: Elsevier; 2016. p. 169–200.

9. Smith DL, McKenzie FE, Snow RW, Hay SI. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS biology* 2007;5(3):e42.

10. Griffin JT. Is a reproduction number of one a threshold for Plasmodium falciparum malaria elimination? *Malaria journal* 2016;15:1–12.

11. Brady OJ, Godfray HCJ, Tatem AJ, Gething PW, Cohen JM, McKenzie FE, et al. Vectorial capacity and vector control: reconsidering sensitivity to parameters for malaria elimination. *Trans R Soc Trop Med Hyg* 2016;110(2):107–117.

12. Dao A, Yaro AS, Diallo M, Timbiné S, Huestis DL, Kassogué Y, et al. Signatures of aestivation and migration in Sahelian malaria mosquito populations. *Nature* 2014;516(7531):387–390.

13. Bationo C, Cissoko M, Katilé A, Sylla B, Ouédraogo A, Ouedraogo JB, et al. Malaria in Burkina Faso: A comprehensive analysis of spatiotemporal distribution of incidence and environmental drivers, and implications for control strategies. *Plos one* 2023;18(9):e0290233.

14. Epopa PS, Maiga H, Hien DFdS, Dabire RK, Lees RS, Giles J, et al. Assessment of the developmental success of Anopheles coluzzii larvae under different nutrient regimes: effects of diet quality, food amount and larval density. *Malaria Journal* 2018;17:1–10.

15. Epopa PS, Millogo AA, Collins CM, North A, Tripet F, Benedict MQ, et al. The use of sequential mark-release-recapture experiments to estimate population size, survival and dispersal of male mosquitoes of the Anopheles gambiae complex in Bana, a west African humid savannah village. *Parasites & vectors* 2017;10:1–15.

16. Holstein M. Practical Guide to the Anophelines of French West Africa. 1949.

17. Santolamazza F, Mancini E, Simard F, Qi Y, Tu Z, della Torre A. Insertion polymorphisms of SINE200 retrotransposons within speciation islands of Anopheles gambiae molecular forms. *Malaria journal* 2008;7:1–10.

18. Harrison XA. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* 2014;2:e616.

19. R Core Team. 2024.

20. Lehmann T, Dao A, Yaro AS, Diallo M, Timbiné S, Huestis DL, et al. Seasonal variation in spatial distributions of Anopheles gambiae in a Sahelian village: evidence for aestivation. *J Med Entomol* 2014;51(1):27–38.

21. Millogo AA, Yaméogo L, Kassié D, de Charles Ouédraogo F, Guissou C, Diabaté A. Spatial modelling of malaria prevalence associated with geographical factors in Houet province of Burkina Faso, West Africa. *GeoJournal* 2023;88(2):1769–1783.

22. Biasutti M. Rainfall trends in the African Sahel: Characteristics, processes, and causes. *Wiley Interdisciplinary Reviews: Climate Change* 2019;10(4):e591.

23. Mwima R, Hui TJ, Kayondo JK, Burt A. The population genetics of partial diapause, with applications to the aestivating malaria mosquito Anopheles coluzzii. *Molecular Ecology Resources* 2024;24(4):e13949.

24. Mwima R, Hui TJ, Nanteza A, Burt A, Kayondo JK. Potential persistence mechanisms of the major Anopheles gambiae species complex malaria vectors in sub-Saharan Africa: a narrative review. *Malaria Journal* 2023;22(1):336.

25. Longo-Pendy NM, Tene-Fossog B, Tawedi RE, Akone-Ella O, Toty C, Rahola N, et al. Ecological plasticity to ions concentration determines genetic response and dominance of Anopheles coluzzii larvae in urban coastal habitats of Central Africa. *Scientific Reports* 2021;11(1):15781.

26. Sangaré I, Ouattara CA, Soma DD, Soma D, Assogba BS, Namountougou M, et al. Spatial-temporal pattern of malaria in Burkina Faso from 2013 to 2020. *Parasite epidemiology and control* 2022;18:e00261.

27. Grant J, Diawara H, Traore S, Koita F, Myers J, Sagara I, et al. Delivery strategies for malaria vaccination in areas with seasonal malaria transmission. BMJ global health 2023;8(5):e011838.

28. North AR, Burt A, Godfray HCJ. Modelling the suppression of a malaria vector using a CRISPR-Cas9 gene drive to reduce female fertility. *BMC biology* 2020;18:1–14.

29. Leung S, Windbichler N, Wenger EA, Bever CA, Selvaraj P. Population replacement gene drive characteristics for malaria elimination in a range of seasonal transmission settings: a modelling study. *Malaria Journal* 2022;21(1):226.

30. Andolina C, Rek JC, Briggs J, Okoth J, Musiime A, Ramjith J, et al. Sources of persistent malaria transmission in a setting with effective malaria control in eastern Uganda: a longitudinal, observational cohort study. *The Lancet Infectious Diseases* 2021;21(11):1568–1578.

31. Selvaraj P, Wenger EA, Gerardin J. Seasonality and heterogeneity of malaria transmission determine success of interventions in high-endemic settings: a modeling study. *BMC Infectious Diseases* 2018;18:1–14.

32. Degefa T, Yewhalaw D, Yan G. Methods of sampling malaria vectors and their reliability in estimating entomological indices in Africa. *J Med Entomol* 2024;61(3):573–583.

33. Routledge I, Chevéz JER, Cucunubá ZM, Rodriguez MG, Guinovart C, Gustafson KB, et al. Estimating spatiotemporally varying malaria reproduction numbers in a near elimination setting. *Nature communications* 2018;9(1):2476.

34. Hayes RJ, Moulton LH. *Cluster randomised trials*. : Chapman and Hall/CRC; 2017.

35. Hemming K, Taljaard M, Forbes G, Eldridge SM, Weijer C. Ethical implications of excessive cluster sizes in cluster randomised trials. *BMJ Quality & Safety* 2018;27(8):664–670.

36. Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. *Fam Pract* 2000;17(2):192–196.

37. Imai K, King G, Nall C. *The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation*. 2009.

38. Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. *J Clin Epidemiol* 2016;69:137–146.

Acknowledgements

We thank Samantha O'Loughlin, Silke Fuchs, and Camilla Beech for their feedback on an earlier draft. Authors receive funding through Target Malaria, which is supported by a grant from the Bill & Melinda Gates Foundation and from the Open Philanthropy Project Fund, an advised fund of the Silicon Valley Community Foundation. This work is supported by Wellcome Trust (224487).

Author contributions

P.S.E., A.A.M., F.A.Y. led the field collection. P.S.E., D.K. conducted the genotyping and prepared the raw data. P.S.E., A.A.M., T-Y.J.H. processed the datasets. T-Y.J.H. conducted the statistical analyses. T-Y.J.H wrote the manuscript with inputs from the other authors.

Data availability

Computing scripts and data are available upon request.

Conflict of interest

The authors declare no conflicts of interest.

Fixed effect	df	Comparison(s)	Estimate (s.e.)	p-value
Seasonality	5			0.000***
		(See Table 1C)		
Bed net	1			
		No vs Yes	-0.15072 (0.091)	0.099
Household size	2			0.000***
		None vs Small	1.110 (0.446)	0.013*
		Small vs Large	0.427 (0.066)	<0.001***
		None vs Large	1.537 (0.449)	0.000***

Table 1A. Fixed effect estimates from the Poisson GLMM on the counts of female *An. gambiae s.l.*. The significance of the fixed effects and p-values were based on Likelihood Ratio Tests (LRTs), while Wald p-values were computed for pairwise comparisons. Effect estimates are on log-scale. Asterisks are given to indicate statistical significance: (***) p<0.001, (**) p<0.01, (*) p<0.05.

Random effect	Number of levels	Variance estimate	Percentage	p-value
Overdispersion	3133	1.4413	64.88%	0.000***
Between-house	383	0.3127	14.07%	0.000***
Between-year	6	0.1881	8.47%	0.000***
Between-village x Seasonality	24	0.1768	7.96%	0.000***
Between-village	4	0.1025	4.61%	0.048*

Table 1B. Random effect estimates and percentage contributions from the same Poisson GLMM, ranking from the largest. LRTs were performed to obtain the p-values.

	0102	0304	0506	0708	0910	1112
0102		0.000***	0.000***	0.000***	0.000***	0.000***
0304	2.151		<0.001***	0.003**	0.000***	0.000***
0506	3.212	1.061		0.638	0.285	0.000***
0708	3.066	0.915	-0.146		0.130	0.000***
0910	3.544	1.394	0.332	0.479		0.000***
1112	0.798	-1.353	-2.414	-2.268	-2.746	

Table 1C. Pairwise contrast for the seasonality effect in the Poisson GLMM (Table 1A). Seasonality was categorised into 6 factor levels. For example, “0102” refers to the counts of Jan and Feb combined. Contrasts (lower-triangular) and the unadjusted p-values (upper-triangular, shaded in grey) were computed via the R package “emmeans”.

Fixed effect	df	p-value
Seasonality	5	0.028*
Bed net	1	0.771
Household size	2	0.886

Table 2A. Fixed effect estimates from the binomial GLMM on the proportion between female *An. coluzzii* and *An. gambiae*, with p-values obtained from LRTs. Effects are reported on logit scale.

Random effect	Number of levels	Variance estimate	Percentage	p-value
Between-village	4	3.476	76.47%	0.000***
Overdispersion	985	0.520	11.43%	0.000***
Between-village x Seasonality	24	0.304	6.68%	0.000***
Between-year	3	0.179	3.92%	0.000***
Between-house	200	0.0678	1.49%	0.263

Table 2B. Random effect estimates and percentage contributions from the binomial GLMM on species proportion, ranking from the largest. LRTs were performed to obtain the p-values.

Factor	Density	Species composition
Seasonality	Clear seasonal pattern with a rainy and dry season per year	No clear rainy and dry season patterns. Villages tend to have their own seasonality
Bed net installation	Not significant	Not significant
Household size	Density increased with household size	Not significant
Between-house variance	Large	Not significant
Between-year variance	Small but non-zero	Small but non-zero
Village-seasonality interaction variance	Moderate	Moderate
Between-village variance	Small but non-zero	Huge

Table 3. Descriptive summary of the fixed and random factors affecting mosquito density and species composition, based on the numerical results from the two GLMMs reported in Tables 1-2. For the random effects, the relative importance was reported.

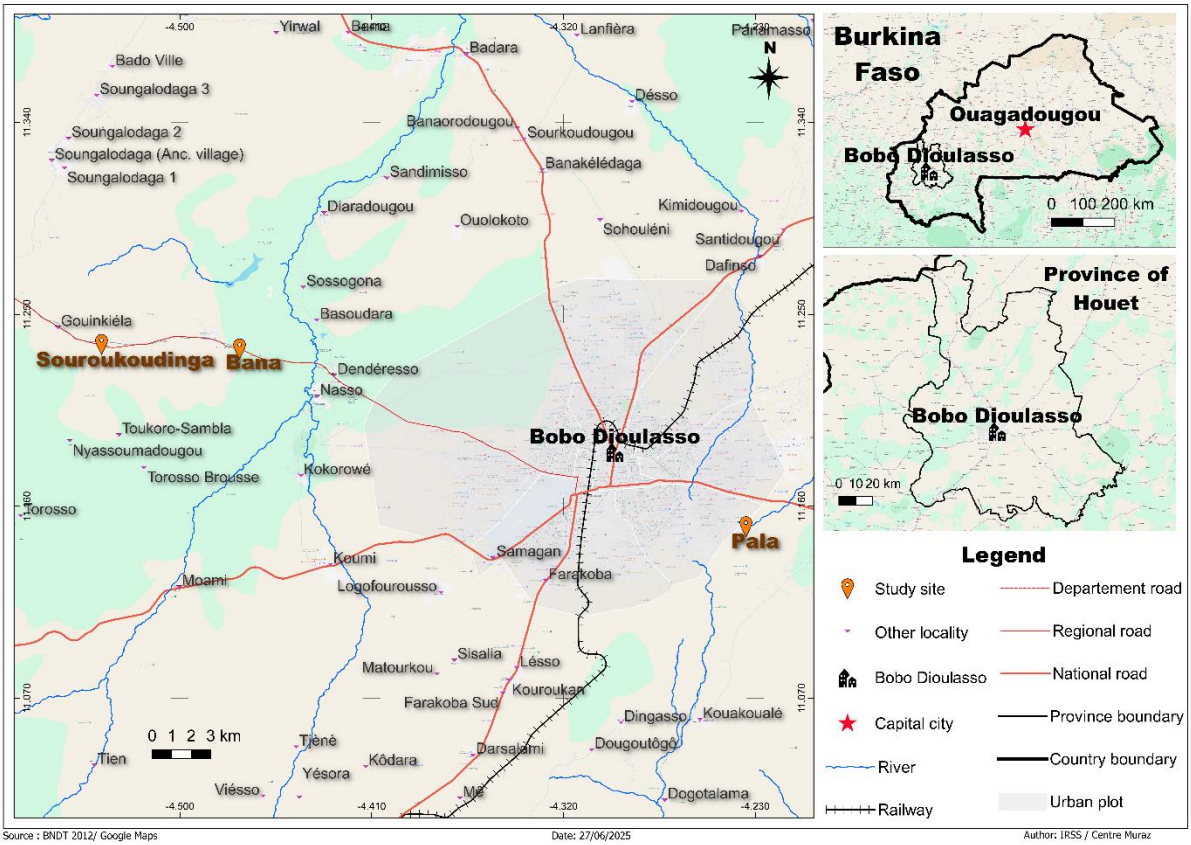


Figure 1. Location of our sampling locations in the Houet Province of western Burkina Faso. The coordinates of the sampling locations are: Bana [11.233, -4.472], Souroukoudinga [11.235, -4.537], Pala [11.151, -4.234]. In the centre of the map is Bobo-Dioulasso, the second largest city in Burkina Faso, where our research base locates. There are two sampling locations from Bana: Bana Village and Bana Market, which are about 23km from Bobo-Dioulasso. Souroukoudinga is about 7km further away in the same direction. Pala is on the opposite direction approximately 6km from Bobo-Dioulasso.

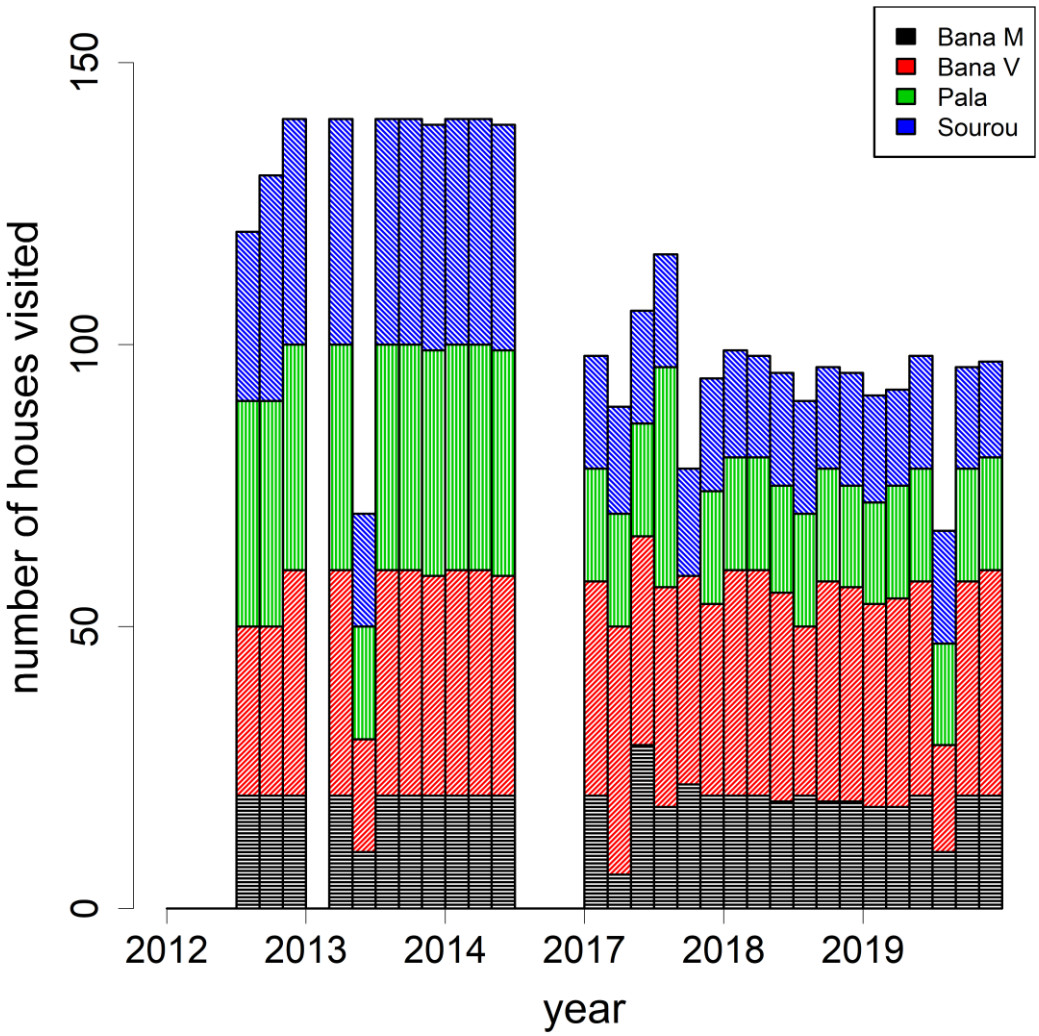
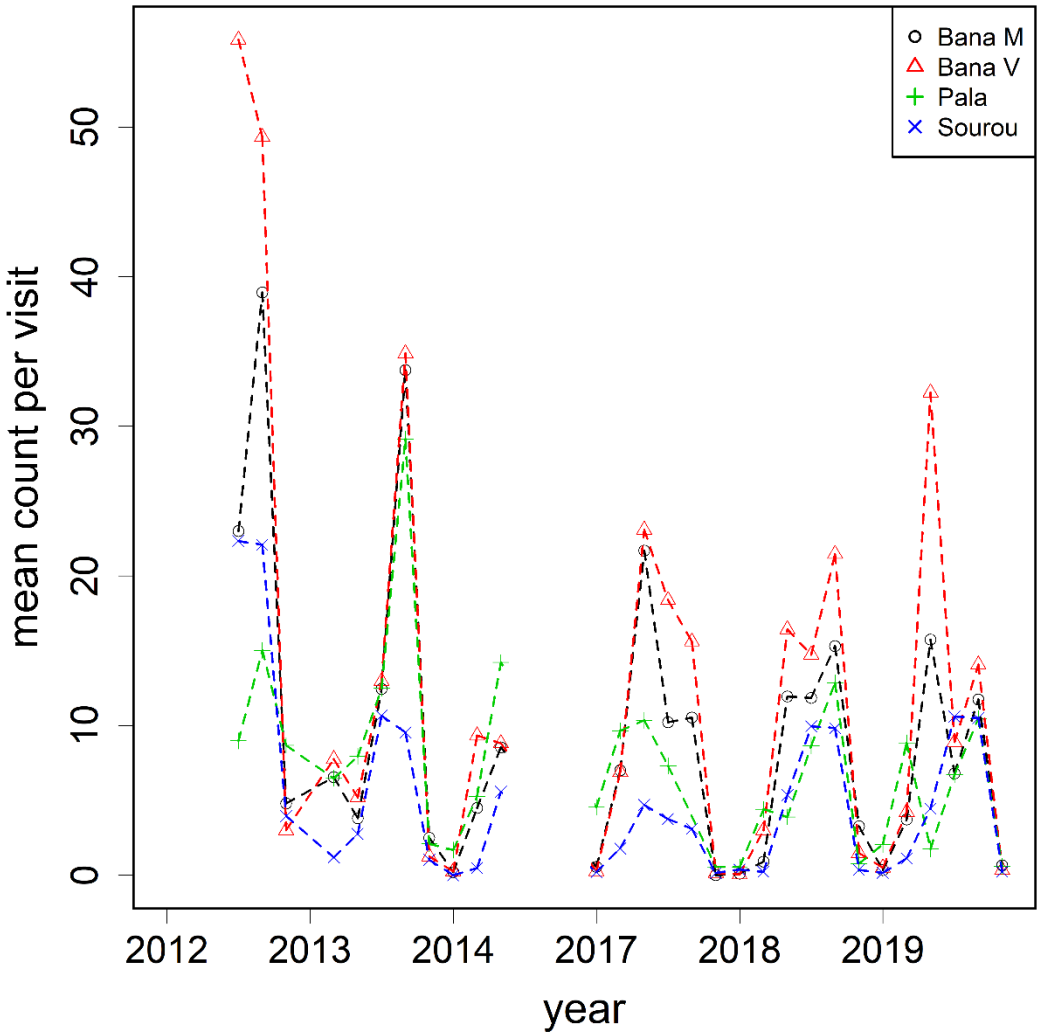


Figure 2. Sampling effort in our four villages over time: Bana Market (Bana M), Bana Village (Bana V), Pala, and Souroukoudinga (Sourou). Each bar represents the number of houses visited for PSC collection per two months. Sampling was paused between July 2015 and Dec 2016.



588

589 Figure 3. Mean female *An. gambiae s.l.* collected per visit by village over time. Note that

590 sampling was paused after June 2014 then resumed at the beginning of 2017.

591

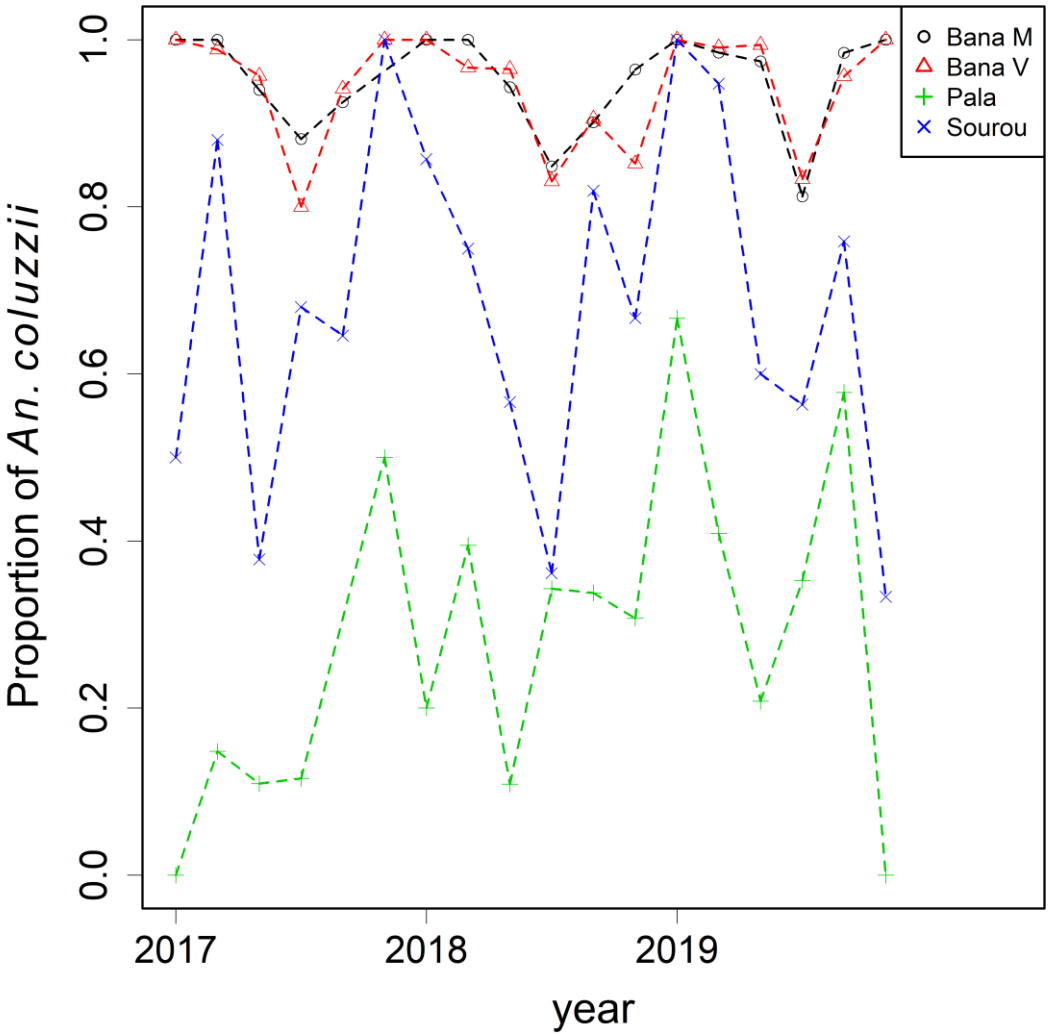


Figure 4. The relative proportion of *An. coluzzii* and *An. gambiae* by village over time. Species identification was performed to a subset of mosquitoes collected during 2017-2019.

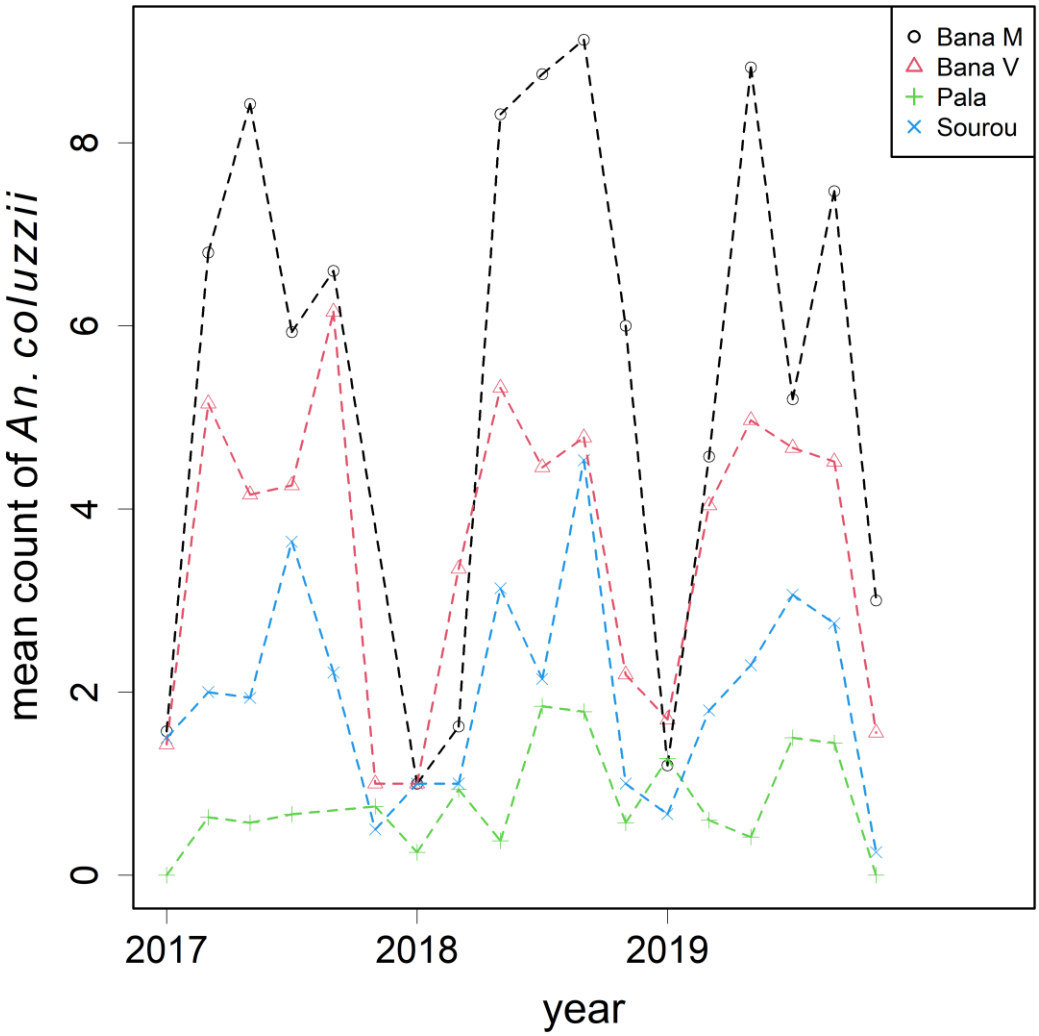
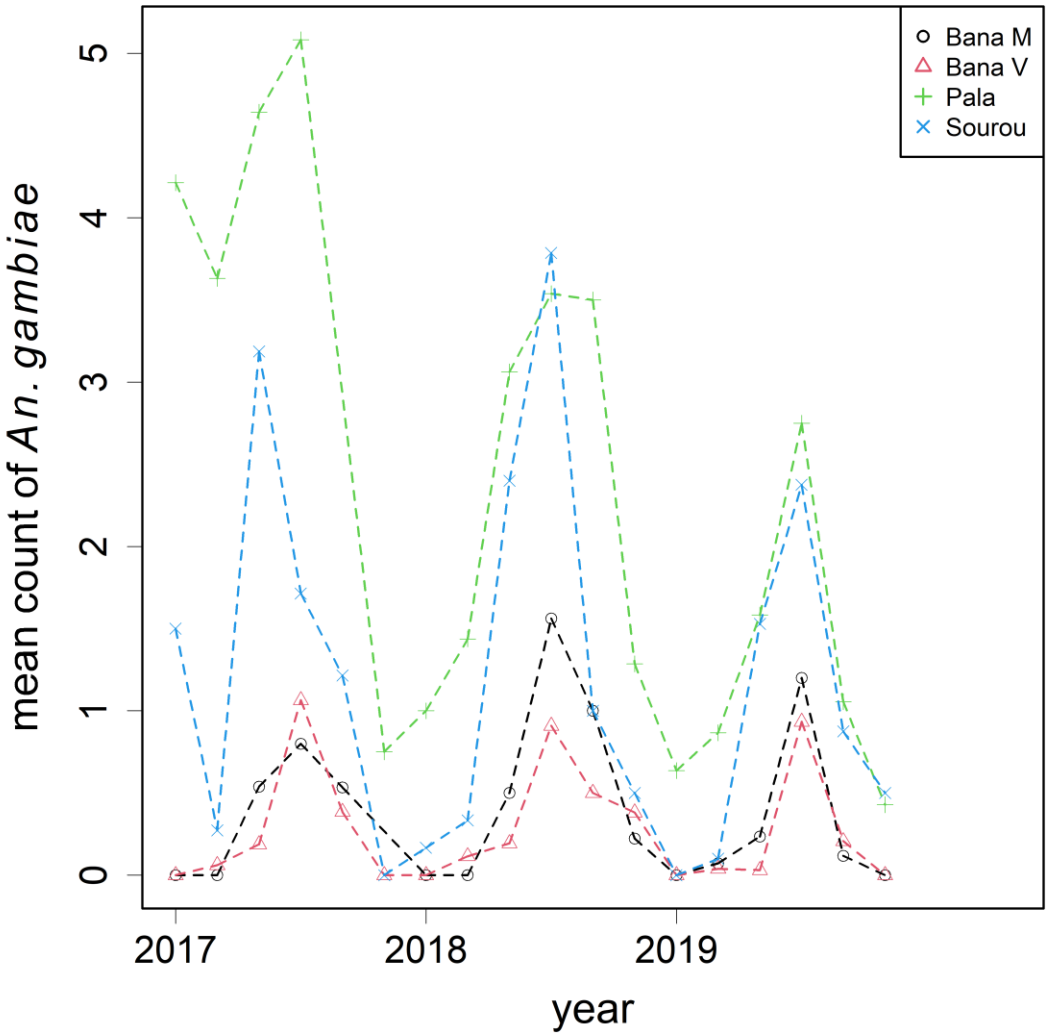


Figure 5. Mean number of female *An. coluzzii* per house visit by village over time.



601

602 Figure 6. Mean number of female *An. gambiae* per house visit by village over time.