Selection of Sites for Field Trials of Genetically Engineered Mosquitoes with Gene Drive.

Lanzaro G.C.[1,*], Campos M.[1], Crepeau M.[1], Cornel A.[1], Estrada A.[1], Gripkey H.[1], Haddad Z.[2], Kormos A.[1], Palomares S.[1] and Sharpee W.[1]

[1]Vector Genetics Laboratory, Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California, Davis, CA

[2]California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA

[*]Corresponding author

Keywords: malaria, *Anopheles gambiae*, genetic control, islands, population modification,

Abstract: Novel malaria control strategies using genetically engineered mosquitoes (GEMs) are on the horizon. Population modification is one approach wherein mosquitoes are engineered with genes rendering them refractory to the malaria parasite coupled with a low-threshold, Cas9-based gene drive. When released into a wild vector population, GEMs preferentially transmit these beneficial genes to their offspring, ultimately modifying a vector population into a non-vector one. Deploying this technology awaits evaluation including ecologically contained field trials. Here, we consider a process for site selection, the first critical step in designing a trial. Our goal is

to identify a site that maximizes prospects for success, minimizes risk, and serves as a fair, valid, and convincing test of efficacy and impacts of a GEM product intended for large-scale deployment in Africa. We base site selection on geographical, geological, and biological, rather than social or legal, criteria. We recognize the latter as critically important but not preeminent. We propose physical islands as being the best candidates for a GEM field trial and present an evaluation of 22 African islands. We consider geographic and genetic isolation, biological complexity, island size, topography, and identify two island groups that satisfy key criteria for ideal GEM field trial sites.

1. Introduction

We present a framework employed by the University of California Irvine Malaria Initiative (UCIMI) for the selection of sites to conduct field trials of a genetically engineered mosquito (GEM) with gene drive. These GEMs are designed to offer safe, cost-effective, and sustainable malaria control in sub-Saharan Africa. This will be achieved using a population modification strategy [1] wherein parasite blocking effector genes are engineered into vector mosquitoes, rendering them incapable of transmitting the parasite [2]. An essential GEM component is an efficient gene drive [3] which serves two critical purposes: to establish the effector genes at high frequency in the mosquito population at the immediate release site and to facilitate its spread into

neighboring populations via normal mosquito dispersal and gene flow. This GEM is designed to eliminate the malaria parasite without eliminating the mosquito.

Achieving malaria control on a large spatial scale, requires a so-called low-threshold gene drive; meaning one with a maximum capability for spreading across the environment (invasiveness). Henceforth, when we refer to a GEM, we mean a mosquito engineered with anti-*Plasmodium* effector genes and a low threshold, highly invasive gene drive. This is the GEM that UCIMI aims to evaluate in a field trial.

Our primary goal for a site selection process is identification of a site that maximizes the prospects for success, minimizes risk, and serves as a fair, valid, and convincing test of the efficacy and impacts of a GEM product intended for large-scale deployment in sub-Saharan Africa. The purpose of the field trial itself is to describe the behavior of a GEM when introduced into a natural population of a target species, in this case *Anopheles gambiae* and/or its sister species *Anopheles coluzzii*.

A multi-phase pathway for the development and evaluation of GEMs has been proposed by the World Health Organization (WHO) [4]. This protocol has been widely endorsed [5, 6] and serves as the foundation for the framework described here. PHASE 1 of the WHO pathway includes design and construction of the GEM product and initial evaluation of its efficacy. This evaluation assesses the phenotype generated by the transgenes, transgene inheritance (especially as it relates to the efficiency of the gene drive component), the stability of the construct over time, and a rudimentary

61 evaluation of overall fitness [3, 7]. GEM products that show promise then move into
62 PHASE 2 field trials with a strong emphasis on containment.

63 Early guidelines recommended that initial tests be conducted in large, artificially
64 contained greenhouse-like cages designed to simulate natural conditions [8-11]. Data
65 generated in such caged environments are limited in several important ways: they do
66 not allow analysis of community and ecosystem-level interactions in any meaningful
67 sense, they cannot replicate food web structure, and they do not permit examination
68 of ecological phenomenon (e.g., dispersal) across spatial scales [12, 13]. Critically,
69 experiments conducted in artificial environments often yield highly replicable, but
70 spurious results [14]. These limitations were recognized in later guidelines and the use of
71 artificially contained environments is now suggested as optional, unless required by
72 regulatory authorities [15, 16].

73 A different strategy that has been proposed for dealing with containment is to
74 conduct field trials in a stepwise fashion with early trials using high-threshold drives, such
75 as split-drive systems which have limited invasiveness and are therefore self-contained
76 [17, 18]. Threshold-dependent drives have their place in controlling vectors on a small
77 spatial scale, such as in urban settings [19]; however, deploying a high threshold drive
78 to achieve malaria control at the scale of continental Africa is not feasible [15].

79 From our perspective, conducting trials in large cages or with high-threshold drives
80 does not satisfy our goal that field tests be valid and convincing. Therefore, we propose

to use ecologically confined PHASE 2 field trials in their place. The issue of containment can be mitigated by selecting the appropriate site.

The first consideration in the selection of a GEM field site should be based on defining biological and physical characteristics that would make a site ideal, or as near to ideal, as possible [8, 11, 20]. Ethical, social, and legal issues are critically important, and no field test can be undertaken before these are addressed [21-23]. However, valuable resources, relationships, and infrastructure are best developed at a site that has been determined to be scientifically suitable. Here we describe a set of criteria that may be applied to a thoughtful consideration and assessment of potential field trial sites. When completed, this framework should provide a cogent justification for why a particular site was selected for GEM testing.

Ecologically confined field sites offer geographical, environmental, and/or biological confinement [4]. Physical islands have been suggested as ideal for conducting GEM field trials [6, 11]. In addition to containment, islands have numerous characteristics that favor their use as GEM field trial sites, including relatively small size, distinct boundaries, simplified biotas, and relative geological youth. These features led to the development of the "Dynamic Equilibrium Theory of Island Biogeography" [24-27], which we rely on to inform our assessment of the advantages of island over mainland sites for the evaluation of GEM.

2. Materials and methods

(a) Selection of candidate island sites

Site selection was initiated with the identification of all potential island sites, which we define broadly as any island associated with the continent of Africa (Figure 1).

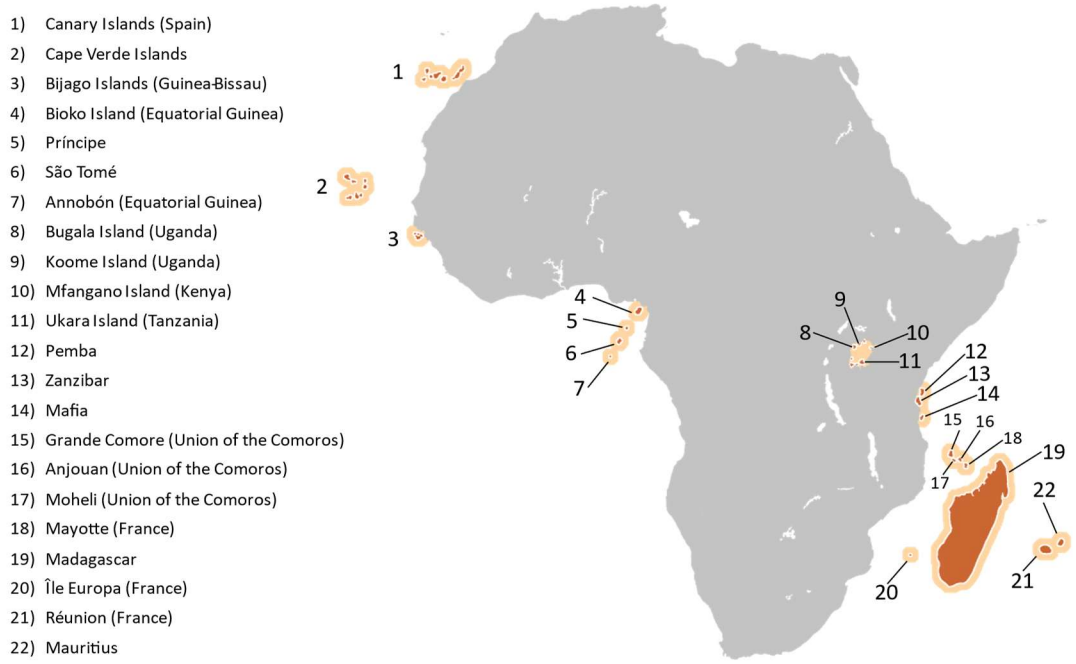

Figure 1. African islands and island groups considered potential field sites for genetically engineered mosquitoes for malaria eradication.

Data for each site was obtained from published sources except for some genetic data which was generated *de novo* by us. These data were used to inform the suitability of potential sites by determining if they meet the set of criteria listed in Box 1. This information includes descriptions of entomological, genetic, geographic, and geophysical features of the sites and mosquito populations therein.

BOX 1. Criteria for the selection of field sites.

Primary Criteria	**Rationale**
1. Presence of *Anopheles gambiae*/*A. coluzzii*	Major vector; target of GEM production
2. Geographic isolation	Containment of GEMs
3. Genetic isolation	Containment of transgene constructs
4. Genetic diversity	Potential detriment to GEM function
5. Island size	Feasibility vs. validity
6. Topography	Evaluation of GEM dispersal capacity
7. Anopheline species richness	Logistics; HGT*; confound epi. endpoints
Other considerations	**Rationale**
1. Insecticide susceptibility/resistance	Match GEM to indigenous mosquitoes
2. *Plasmodium* prevalence	Estimation of epidemiological impact
3. Presence of endangered species	Potential for negative GEM interactions
4. Travel feasibility	Operational logistics and cost

*HGT=horizontal gene transfer

(b) Measuring Island Geographic Isolation

Geographic isolation for each island was defined using three methods, all reported in Table 1. The first is simply the geographic distance to the nearest mainland. Distances for individual islands were calculated as the shortest great circular distance between an island's mass centroid and the mainland coast. For archipelagos, distances from the nearest island to the mainland were used [28]. Distance to the mainland for each Lake Victoria island and for Annobón was determined using Google Earth's distance and area measuring tool. The two closest points on the mainland and island shores were used as measuring points. The significance of distance to mainland is that the nearest mainland is assumed to be the richest gene pool and the source of populations on the islands [28, 29].

Table 1. Bioclimatic and isolation index values used for the evaluation of potential island field sites. DD = decimal degrees; UNEP = United Nations Environment Programme; SLMP = Surrounding Landmass Proportion; GMMC = Glacial Maximum Mainland Connection a proxy variable for island geological history which indicates whether an island was connected to the mainland during the Last Glacial Maximum (LGM) (1 = true and 0 = false); - refers to missing or incomplete data. Additional data sources: Bugala:[30-33]; Mfango [34]; Ukara: [35-37]; Koome [38-42]

Island	Archipelago	Island type	Area (km²)	Distance to mainland (km)	UNEP Isolation Index	SLMP	GMMC	Elevation max (m)
Canary Islands	Canary Islands	Oceanic	7509.66	116.63	30.4	0.812	0	3705
Cape Verde	Cape Verde Islands	Oceanic	4088.52	586.53	55	0.466	0	2813
Bijagos	Bijagos Islands	Continental	1944.72	0.83	10.8	1.123	1	59
Bioko	Cameroon Line	Continental	1950.46	73.03	17	1.148	1	3011
Annobón	Cameroon Line	Oceanic	15.7	350	45	-	0	587
São Tomé	Cameroon Line	Oceanic	854.8	283.63	39	0.753	0	1977
Príncipe	Cameroon Line	Oceanic	143.16	221.72	39	0.86	0	934
Bugala	Lake Victoria	Lacustrine	296	3.7	5.487	-	1	160
Koome	Lake Victoria	Lacustrine	100	14.3	10.688	-	1	180
Mfangano	Lake Victoria	Lacustrine	66	7.4	10.414	-	1	551
Ukara	Lake Victoria	Lacustrine	80	22.8	15.623	-	1	162
Pemba	Zanzibar	Continental	987.08	68.6	31.231	1.178	0	149
Zanzibar	Zanzibar	Continental	1591.5	50.9	17	1.337	1	133
Mafia	Mafia	Continental	443.24	36.06	29.432	1.194	1	66
Grand Comore	Comoros	Oceanic	1021.61	307.45	49	0.736	0	2368
Moheli	Comoros	Oceanic	212.09	340.61	49	0.754	0	793
Anjouan	Comoros	Oceanic	432.08	417.78	49	0.706	0	1591
Mayotte	Comoros	Oceanic	371.42	490.16	47	0.669	0	636
Madagascar	Madagascar	Oceanic	590547.4	780.51	58	0.46	0	2876
Ile Europa	French Territory	Oceanic	32.64	492.95	67.941	0.736	0	20
Réunion	Mascarene Islands	Oceanic	2512.65	1699.32	73	0.467	0	3066
Mauritius	Mascarene Islands	Oceanic	1868.44	1874.49	87	0.399	0	816

A second metric is the United Nations Environment Programme (UNEP) Isolation Index, which is calculated as "the sum of the square roots of the distances to the nearest equivalent or larger island, the nearest group or archipelago, and the nearest continent [43]." The higher the value, the more geographically isolated the island is.

The third isolation index is Surrounding Land Mass Proportion (SLMP) where the isolation of the focal island is proportional to the area of the surrounding landmass [28]. SLMP is calculated as the sum of the proportions of landmass within buffer distances of 100, 1000, and 10,000 km around the island perimeter. SLMP accounts for the coastline shape of large landmasses by considering only regions that extend into the measured buffers. SLMP values for the Canary Islands, Cape Verde Islands, and Bijagós Islands were represented as the average of all islands in their respective archipelagos [28]. SLMP is a preferred index for analysis of species variation on a focal island. The equilibrium theory of island biogeography supports this index as individual islands may act as stepping-stones for species dispersal and establishment, which this index accounts for by shortening the distance between an island and potential source populations [26]. A larger SLMP value indicates that an island is surrounded by more landmass. For this study, we are focusing on islands with a lower SLMP value since these islands will have less surrounding landmass which could facilitate mosquito dispersal into or out of the target island.

(c) Island size and topography

Island size information, presented in Table 1 as area, is taken from the publication by Weigelt et al. [28]. They describe island size by using the Database of Global Administrative Areas (GADM) to obtain high-resolution island polygons. Area was calculated for each GADM polygon in a cylindrical equal area projection. Areas for archipelagos (Canary Islands, Bijagos, Cape Verde) were reported here as the sum of all islands in each archipelago [28]. The area for Annobón was obtained from the

United Nations Environmental Programme [43]. The areas for the Lake Victoria islands (excluding Koome) were taken from the literature [33-35] and the area for Koome Island was approximated using Google Earth's distance and area measuring tool.

Elevation maximum and minimum of each island were obtained from the AW3D30 Global Digital Surface Model of the Japan Aerospace Exploration Agency [44]. GeoTIFF files were downloaded, and the highest elevation of each island/archipelago was identified. Island topography was further described using the United States National Aeronautics and Space Administration (NASA) 90-m resolution elevation data from the Shuttle Radar Topography Mission (SRTM) 90m Digital Elevation Model database [45]. In this case, altitude and magnitude of steepest gradient measurements were used to generate heat maps as graphic descriptors of topography.

(d) Population genomics analyses

We conducted a comparative genomics analysis of mainland and island populations of the two target species. The locations and sample sizes per site are provided in Figure 2. In total, 420 individual *Anopheles gambiae* and *A. coluzzii* genome sequences were analyzed in this study. The UC Davis Vector Genetics Laboratory (VGL) generated 167 genomes (Supplemental Table 1). In addition, 196 genomes were obtained from the *Anopheles gambiae* 1,000 Genome Project phase 2 [46] and 57 were taken from a published Lake Victoria islands study [47].

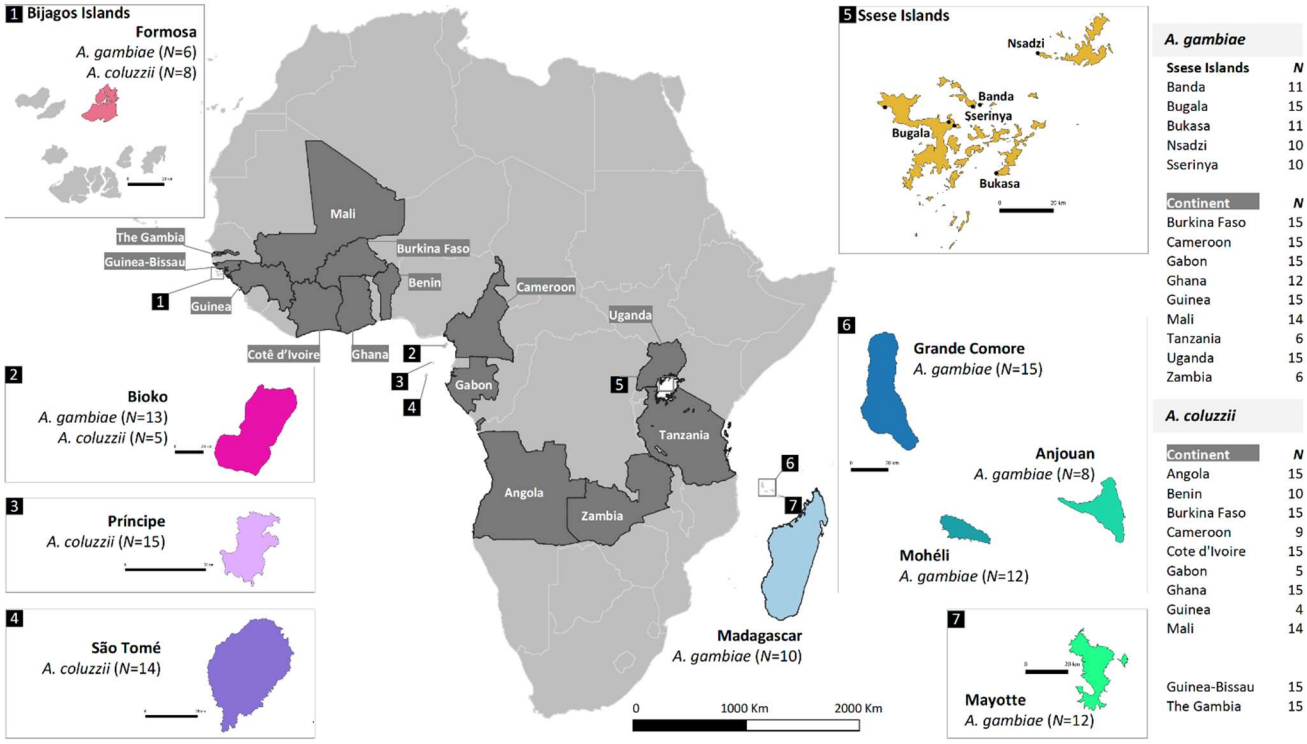

Figure 2. Study sampling locations. Samples of *A. gambiae* and *A. coluzzii* were from 12 countries (dark shade) in continental Africa: Angola, Benin, Burkina Faso, Cameroon, Cotê d'Ivoire, Gabon, Ghana, Guinea, Tanzania, Uganda and Zambia. Populations from Guinea-Bissau and The Gambia (dark shade) were included with no species assignment. Table on the right displays the number of samples for each of the mainland populations. The insert maps show African islands sampled in this study: 1) Formosa islands within Bijagos archipelago; 2) Bioko, 3) Príncipe and São Tomé, 4) islands in the Gulf of Guinea, 5) five islands in Ssese islands in Lake Victoria in Uganda, 6) in the Comoros (Anjouan, Mohéli and Grande Comore), and 7) Mayotte. Madagascar island is shown in the main map. The number of samples included for each island is shown in parenthesis. Insert maps contain a scale of 20km length.

Individual mosquito DNAs from the VGL samples were extracted using a Qiagen Biosprint (Qiagen, Hilden, Germany) following our established protocol [48]. 10 ng of genomic DNA was used for individual libraries using a KAPA HyperPlus Kit (Roche Sequencing Solutions, Indianapolis, Indiana, USA), as described in Yamasaki et.al [49].

Sequencing was performed on an Illumina HiSeq 4000 instrument (Illumina, San Diego, California, USA) at the UC Davis DNA Technologies Core facility. Methods used for genome sequencing of individuals from other sources are described elsewhere [47, 50].

Demultiplexed raw reads of VGL samples were filtered and trimmed using Trimmomatic v0.36 [51] and saved as FastQ files. Sequences from Ag1000G and Lake Victoria study [47] were downloaded and converted to FastQ files using BEDTools v.2.2 [52]. All specimens were mapped to the reference AgamP4 [53, 54] using BWA-MEM v0.7.15 [55] with default settings. Duplicate reads were removed using Sambamba markdup [56]. Freebayes v1.2.0 [57] was used for variant calling, with standard filters and the "-no-population-priors", "theta = 0.01", and "max-comple-gap = 0" options. Variants were normalized with *vt normalize* v0.5 [58].

SNPs were filtered out when they did not pass the accessibility mask from Ag1000G, missingness >10%, a minimum depth of 8 and minor allele frequency (MAF) < 1%. In addition, population structure analysis was based on chromosome 3 SNPs only. This was done to avoid confounding signals from polymorphic inversions on chromosomes 2 and X [53]. Heterochromatic regions on chromosome 3R (3R:38,988,757-41,860,198; 3R:52,161,877-53,200,684) and 3L (3L:1-1,815,119; 3L:4,264,713-5,031,692) were also filtered out [53].

Description of population structure was performed by Principal component analysis (PCA) after pruning for LD using scikit-allel v1.2.0 [59]. Hudson's estimator [60, 61] was used for pairwise fixation indices F_{ST} calculation implemented in scikit-allel v1.2.0.

Nucleotide diversity (π) was calculated in nonoverlapping windows of 10 kb on euchromatic regions of chromosome 3 using VCFtools [62]. The results were grouped by population and significance tests performed between the islands and mainland populations using a Wilcoxon rank-sum test in R.

(e) Anthropogenic sources of dispersal

The prospects for a GEM emigrating out of a field trial site into a non-target site or vice versa by air or ship transport was assessed by determining the frequency of departures from select mainland and island sites. Airline flight data including the annual (Jan. 1-Dec. 31, 2019) number of international departures from airports within a specific country were obtained from CIRIUM, an aviation data analytics provider [63]. Similarly, shipping data for the annual (Jan.1-Dec.31, 2020) number of commercial ship departures was obtained from the [64].

(f) Anopheline species richness

Published compilations of Afrotropical *Anopheles* species distributions [65, 66] were used to assemble the information for mainland and island countries. The first criterion for field site selection is the presence of the target species, which is in our case *Anopheles gambiae sensu stricto* and/or its sister species *Anopheles coluzzii*. Species are designated as primary or secondary vectors or as "other" if they are non-vectors or their status as vectors is not clear. Species that commonly had sporozoite infection rates above 1%, as determined by salivary gland dissections, CSP ELISA or PCR of head and thorax were listed as primary vectors. Species with infection rates of <1% were listed as

239 secondary vectors. Our knowledge of the population structure and biology of almost all
240 the secondary vectors is limited and their role in malaria transmission varies from
241 location to location.

242

243 **3. Results and discussion**

244 **(a) Identification of potential field sites**

245 We evaluated 22 potential field sites, including 5 individual islands, multiple islands
246 within 7 archipelagos and 4 islands within Lake Victoria (Figure 1). The sites identified
247 include three island types: continental, oceanic, and lacustrine. Each type possesses
248 features that impact its utility as a GEM trial site. Continental or land bridge islands are
249 unsubmerged portions of the continental shelf and were, at one time, connected to
250 the mainland. Oceanic islands arise from the ocean floor and were never connected
251 to the mainland. Lacustrine islands are islands within lakes and are typically formed by
252 deposits of sedimentary rock, as are the Lake Victoria islands. For comparison, our
253 analyses include mainland sites closest to the islands and those in which GEM field trials
254 are currently under consideration (e.g., Burkina Faso, Mali, Uganda). We then proceed
255 by defining and justifying a prioritized set of criteria (Box 1) on which to base
256 evaluations.

257 **(b) Geographic isolation**

258 Geographic isolation is among the most significant features favoring islands as GEM
259 field trial sites. Although some mosquito species are known to disperse on prevailing

winds over long distances [67, 68], there are, to our knowledge, no reliable reports of open-ocean wind dispersal of malaria vector species over the distances (hundreds of kilometers) separating some of the oceanic islands under consideration here. Emigration of GEMs out of the field trial site into neighboring, non-target sites, either on nearby islands or the mainland, pose a problem, especially as it relates to risk and regulatory concerns. Equally important is immigration of wild type individuals from neighboring sites into the trial site. Immigration, in this case, will confound efforts to measure GEM invasiveness and could potentially render the gene drive inefficient or even ineffective. Island biogeography theory predicts that choosing a remote island as an initial field trial site greatly reduces the potential for gene flow between vector populations both into and out of the island site. This is further supported by the results of our population genomics assessment, as discussed below.

We evaluated geographic isolation for all candidate islands using distance to mainland, UNEP Isolation Index, and SLMP (Table 1). We excluded any island with a UNEP Isolation Index of less than 15 and we used a Surrounding Landmass Proportion (SLMP) value of 1 as a cutoff, so islands with an SLMP value >1 were considered unacceptable. Sites considered unacceptable based on these criteria include the Bijago Islands, Bugala, Koome, Mfangano, Pemba, Zanzibar, and Mafia.

(c) Island size and topography

There are no well-defined criteria to guide decisions with respect to an appropriately sized area for a GEM field trial. One important consideration is mosquito

281 flight range. To evaluate the dispersal capacity of a GEM, the site should exceed the
282 flight range of the target species. For our considerations we assumed a maximal daily
283 flight range of 10 km for *A. gambiae* [69]. Generally, we aimed to identify sites small
284 enough to be manageable, but large enough to be convincing, keeping the following
285 considerations as a guide.

286 Area (km²) is an important parameter influencing the biology of populations residing
287 on an island. Large island areas typically include more habitat types and can support
288 larger populations. This characteristic can increase the rate of speciation and lower
289 extinction rates over time [27]. Using island size as a criterion we exclude the islands of
290 Annobón and Île Europa for being too small and Madagascar for being too large.

291 Evaluating the dispersal capabilities of a GEM is a critical outcome from a field trial.
292 This capacity is best evaluated at a site that possess topographical features that may
293 pose a challenge to dispersal, as would be encountered in continental Africa.
294 Elevation was used as a measure of topographic complexity and as a proxy for
295 environmental heterogeneity. The difference between the elevation maximum and
296 minimum of each island measured from sea level is reported in the "Elevation" column
297 in Table 1. Elevation relates to the number of available habitats because of differences
298 between windward and leeward sites, temperature decrease with altitude, and high
299 precipitation regimes at certain altitudes [28].

300 Altitude and magnitude of steepest gradient were used to generate a graphic
301 representation of topography for each island. A representative sample of these

302 analyses for the islands of Grande Comore and São Tomé are presented in
303 Supplemental Figure 1A and B to illustrate sites having a complex topography and for
304 the islands of Zanzibar and Mafia in Supplemental Figure 1C and D to illustrate a lack of
305 topographic complexity. Sites lacking topographic complexity were excluded from
306 consideration, these included the Bijago Islands, the islands in Lake Victoria, Zanzibar,
307 Pemba, Mafia and Ile Europa.

308 **(d) Genetic isolation**

309 Genetic isolation relates to the level of gene flow between populations and may be
310 inferred by measuring the degree of genetic divergence between populations under
311 the assumption that gene flow reduces genetic divergence.

312 Single nucleotide polymorphism (SNP) data were analyzed to reveal genetic
313 relationships among populations and results were visualized using principal components
314 analysis (PCA). The position of individuals in the space defined by the principal
315 components can be interpreted as revealing levels of genetic similarity/dissimilarity
316 among the populations from which those individuals were sampled. Populations
317 occupying the same space are presumed to be very similar genetically and those
318 widely separated, very different.

319 Results of the PCA for *A. gambiae* populations are illustrated in Figure 3A. This
320 analysis reveals a high degree of genetic similarity between mainland and both
321 lacustrine and continental islands. Conversely, oceanic islands (Comoros archipelago

and Madagascar) form discrete individual clusters, indicating that they are genetically distinct both from the mainland and from each other.

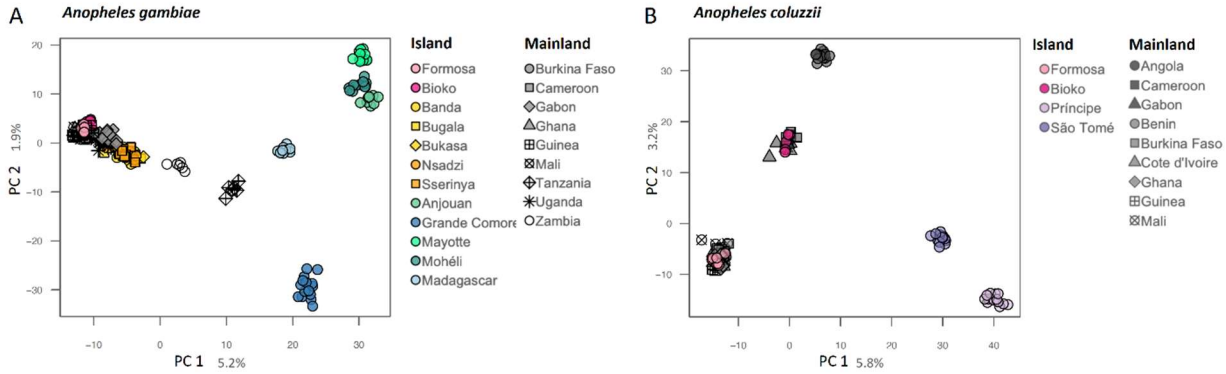

Figure 3. Population structure analysis by PCA. 2D-plot of *A. gambiae* (A) and *A. coluzzii* (B) from islands and mainland populations across Africa. Analyses were based on 50,000 biallelic SNPs from euchromatic regions on chromosome 3. Each marker represents one individual mosquito. Geographic location for each site and numbers of genome analysed per site are provided in Figure 2.

Results of the PCA for *A. coluzzii* (Figure 3B) confirm that populations on continental islands form tight clusters that include mainland populations. Oceanic islands form discrete clusters indicating genetic divergence from mainland populations and from each other. These results indicate high levels of genetic isolation for oceanic island populations of both *A. coluzzii* and *A. gambiae*.

The extent to which individuals move (migrate) between two populations can be approximated by measuring the level of genetic divergence between those populations. Migration (*m*) can be thought of as including the genotypes of the individuals doing the moving and, in this context, migration results in gene flow. Genetic divergence can be described using the statistic F_{ST}, which is the genetic variance in a

subpopulation (S) relative to the total variance (T). F_{ST} values range between 0 and 1 and are higher when populations are considerably diverged. The relationship between F_{ST} and *m* is complex, but excluding the effects of drift and selection, the more gene flow between two populations the lower the F_{ST} value. All pairwise F_{ST} values for the populations of *A. gambiae* and *A. coluzzii* analyzed in this study (Figure 2) are presented in Supplemental Figure 2.

Pairwise F_{ST} values for 14 populations of *A. gambiae* spanning its range across sub-Saharan Africa are provided in Supplemental Figure 2A. Results are consistent with the PCA (Figure 3A). West-central populations, including the island of Formosa in the Bijago archipelago, are very similar (F_{ST}=0.000-0.009). Divergence between Bioko island and the nearest mainland in Cameroon is higher (F_{ST}=0.036). The pattern is quite different in east Africa, wherein mainland populations are far more diverged (F_{ST}=0.033-0.090). Divergence between the islands in Lake Victoria and the nearest mainland in Uganda are lower (F_{ST}=0.003-0.029). Considerably higher divergence is observed between the Comoros islands and the nearest mainland populations in Tanzania (F_{ST} = 0.130-0.169) and between the Comoros and Madagascar (F_{ST}=0.126-0.196).

The F_{ST} values for populations of *A. coluzzii* are likewise consistent with the PCA (Figure 3B). Divergence between the continental island of Formosa and the nearest mainland populations in Guinea-Bissau and between the island of Bioko and nearest sites in Cameroon are low (F_{ST}=0.015 and 0.022 respectively). Populations of *A. coluzzii* on the oceanic islands of São Tomé and Príncipe were, by far, the most genetically isolated from mainland populations (F_{ST}=0.144 and 0.199 respectively). In addition, the

two islands were highly diverged from each other (F_{ST}=0.130). The islands within Lake Victoria were excluded from consideration because the *A. gambiae* populations residing on them lacked the high level of divergence that would indicate genetic isolation. No genetic data was available for the Canary Islands, Cape Verde, Zanzibar, Pemba, Mafia, Mauritius, Réunion or Ile Europa.

Taken together, the data summarized in Figures 4 and Supplemental Figure 2 reveal a high degree of genetic isolation among oceanic islands compared with either continental or lacustrine islands. These results suggest limited dispersal (gene flow) between islands and nearest landmasses and are consistent with expectations based on island biogeography theory as described above and reinforce the benefits of selecting a contained island site for conducting GEM field trials. Genetic data is not currently available for several potential island sites, including the Canary Islands, Cape Verde, Île Europa, Zanzibar, Pemba, and Mafia. Genetic isolation, as measured here, was deemed inadequate for the Lake Victoria islands (Bugala, Koome, Mfangano, and Ukara).

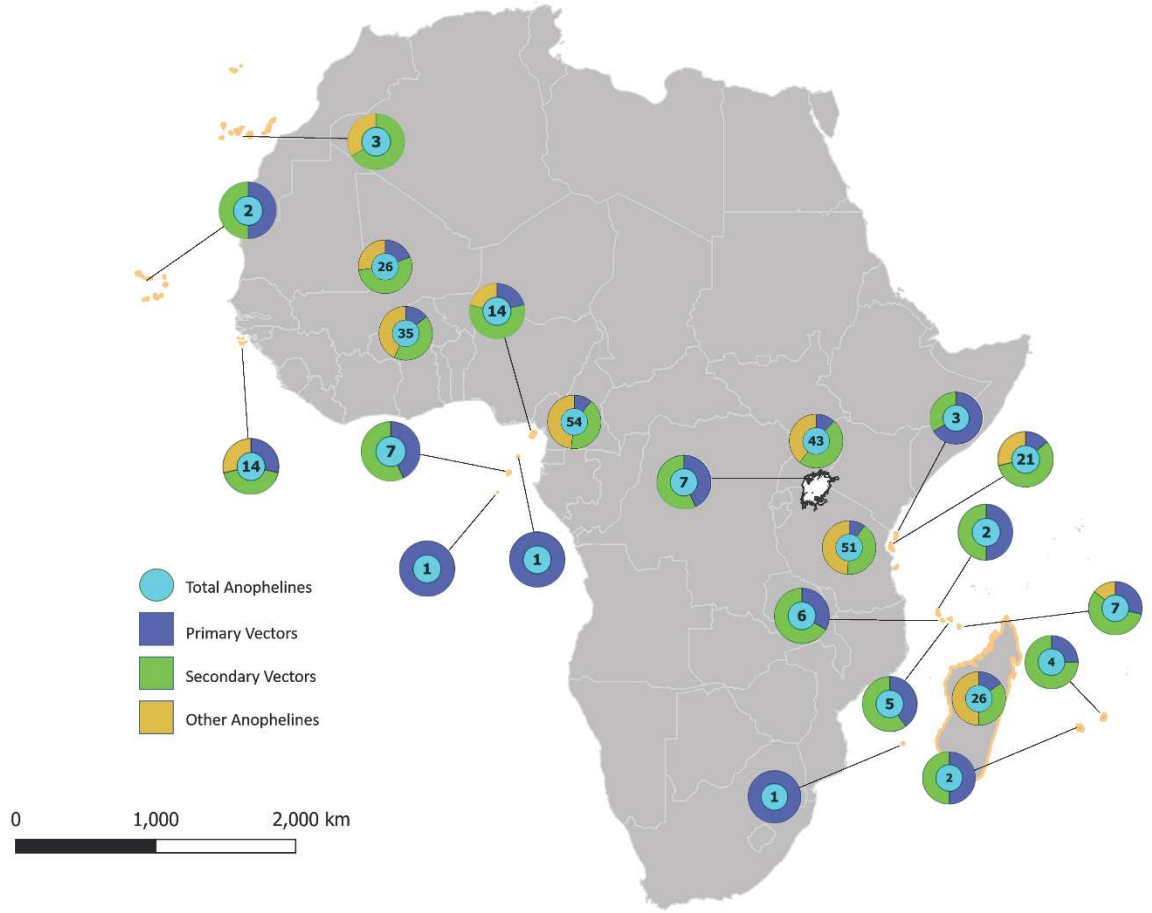

Figure 4. *Anopheles* species complexity in Africa including island and select mainland sites. Map locations and summary of data presented in Table 2. Cyan circle = total number of *Anopheles* spp.; blue proportion of primary vector species; green = proportion of secondary vectors; yellow = proportion of species identified as non-vector or for which vector status unknown.

(e) Anthropogenic dispersal

Anthropogenic dispersal of mosquitoes from inside the release site into nontarget populations may occur and should be considered in selecting a field trial site. The level of genetic divergence between island and mainland populations of *A. coluzzii* and *A.*

389 *gambiae* is generally high suggesting that dispersal off the islands is low. Nonetheless,
390 dispersal that may occur is most likely to rely on anthropogenic conveyance [68, 70].

391 The most significant source of passive anthropogenic dispersal of mosquitoes is by
392 rail and road. This poses significant risk for mainland field sites, where extensive in-
393 country and trans-boundary connections exist [71-73]. Risk by this mode of mosquito
394 dispersal is reduced to zero for oceanic island test sites.

395 Frequency of air and sea departure to interim and final destinations for a sample of
396 mainland and island populations is presented in Figure 5 and Supplemental Tables 3
397 and 4. Islands, due to their smaller human populations and geographic areas generally
398 originate less trans-boundary air and sea traffic compared with the continent (Figure
399 5A). This results in remote islands having inherently lower risk levels for these modes of
400 anthropogenic dispersal. A notable exception is the Cape Verde archipelago which
401 has relatively high ship travel due to its location as a major refueling site (Figure 5B).
402 Traffic has increased with the completion of two new ports and upgrades to existing
403 ports in 1997. Airline and shipping traffic data were only obtained for the locations
404 shown in Figure 5, therefore assessment of the potential for anthropogenic dispersal for
405 the majority of island sites was not assessed. Results for São Tomé and Príncipe and for
406 the Comoros suggest that the likelihood of mosquitoes migrating into or out of these
407 islands is minimal.

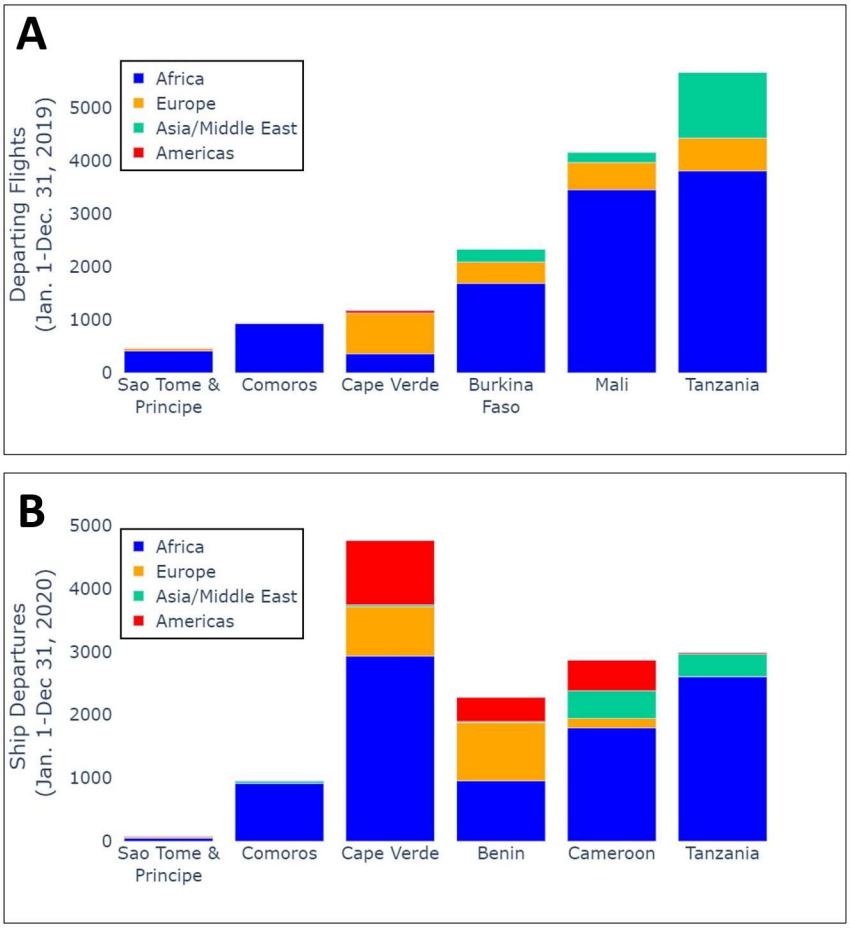

Figure 5. Annual departures by air (A) and sea (B) from representative island and mainland locations in Africa. Colors indicate destinations, grouped by geographic region. Air traffic data provided by Cirium*. (**This information has been extracted from a Cirium product. Cirium has not seen or reviewed any conclusions, recommendations or other views that may appear in this document. Cirium makes no warrantees, express or implied, as to the accuracy, adequacy, timeliness, or completeness of its data or its fitness for any particular purpose. Cirium disclaims any and all liability relating to or arising out of use of its data and other content or to the fullest extent permissible by law.*) Sea traffic data provided by the MarineTraffic Global Ship Tracking Intelligence database.

(f) Anopheline species richness

The number of primary, secondary and other (malaria vector status unclear) species present in island sites and select locations on the mainland are illustrated in Figure 4

422 (and Supplemental Table 2). It is generally agreed that potential field sites with the
423 fewest number of non-target *Anopheles* species are desirable [16, 74]. If multiple sister
424 species are present, there exists the possibility that the transgene will move between
425 species via natural hybridization [75, 76] which could add an additional level of
426 complexity to post-release assessments. Although the movement of transgene elements
427 between malaria vector species may be considered desirable, it raises the specter of
428 horizontal transfer, which is generally identified as a risk to this technology [77].

429 Assessment of entomological endpoints following a GEM release requires repeated
430 mosquito collections to quantify changes in the ratio of GEM to wild type mosquitoes.
431 This necessitates sorting large numbers of individual field-collected mosquito samples to
432 separate target from non-target species. For members of sibling species complexes,
433 which are morphologically indistinguishable, this requires the application of PCR-based
434 diagnostics to each individual specimen. If collections include larvae, time-consuming
435 microscopic examination to identify species is required even for those species
436 distinguishable morphologically in the adult stage. Logistically these procedures are
437 greatly simplified where fewer non-target *Anopheles* species are present, positively
438 impacting the time and resources required for successful assessment.

439 Although entomological endpoints are the main consideration in evaluating the
440 outcome of a PHASE 2 trial, epidemiological impacts should be considered where
441 feasible. If epidemiological endpoints are to be assessed, the presence of multiple
442 primary and secondary vectors are problematic as they can lengthen the season of
443 malaria transmission [78]. Therefore, their presence can mask the effects that GEMs

might have on transmission at a field site by maintaining the rate of transmission, even if the parasite is not present in the target mosquito species. If a site is selected in which very few malaria vector species occur, it becomes more likely that the GEM release will have a measurable impact on the level of malaria transmission.

The number of anopheline species present in the mainland sites included here ranged from 26-54. Continental sites (Bijago Islands, Bioko and Zanzibar) had between 14 and 21 species. As expected, oceanic islands contained far fewer, ranging from 1 to 7 species. These results favor the selection of the oceanic islands, São Tomé and Príncipe, Annobón and the Comoros, for field trials. The oceanic islands including the Canary Islands, Cape Verde, Mauritius and Reunion likewise had low numbers of anopheline species, but these were excluded because our target species *A. coluzzii* and *A. gambiae* are absent from these islands.

(g) Genetic complexity

Genetic complexity was measured using the nucleotide diversity statistic (π), defined as the average number of pairwise nucleotide differences per nucleotide site. Mean nucleotide diversities (π) on oceanic islands (*A. gambiae*- 0.80%, *A. coluzzii*- 0.88%) were significantly lower ($p < 0.0001$) than mainland population means (*A. gambiae*- 1.17%, *A. coluzzii*- 1.11%) for both species (Figure 6A and B).

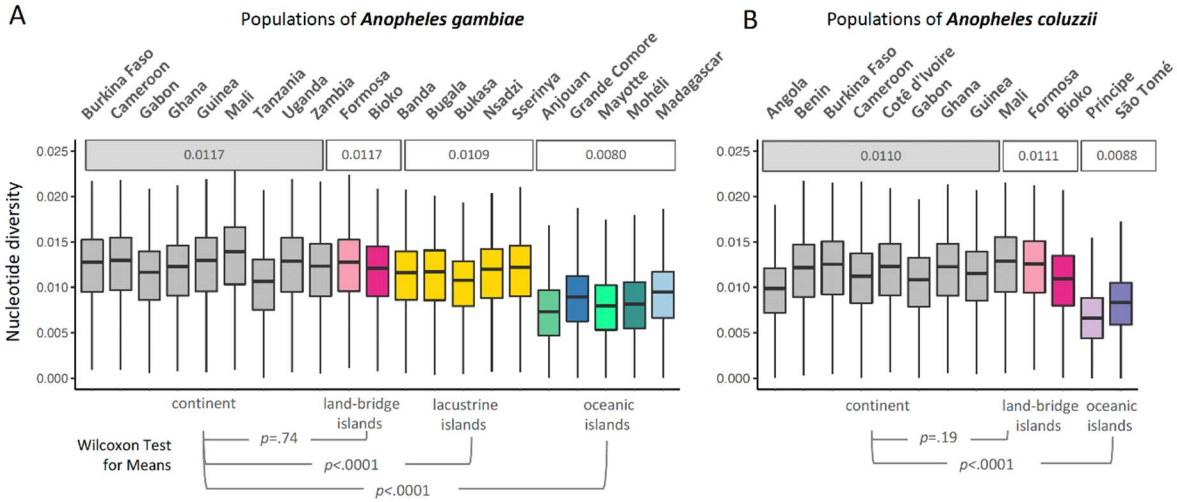

Figure 6. Population diversity. Metric is grouped by sampling locations of (A) *A. gambiae* and (B) *A. coluzzii* populations from island and mainland (grey boxplots). Boxplot of nucleotide diversity (π) performed in 10 kb windows of euchromatic regions of chromosome 3. The midline in all boxplots represents the median, with upper (75th percentile) and lower (25th percentile) limits, whiskers show maximum and minimum values, and outliers are not shown. Mean nucleotide diversity for set of populations are shown above the boxplots; *A. gambiae* populations were divided into four groups: mainland continental (grey), land-bridge (pink), lacustrine (yellow) and oceanic (green/blue) islands; *A. coluzzii* in three: mainland continental (grey), land-bridge (pink) and oceanic (blue) islands. P-value for testing of means between islands and mainland are shown below. Geographic location for each site and numbers of genome analysed per site are provided in Figure 2.

Comparisons among island types yielded results that were consistent with island biogeography theory. Nucleotide diversity in continental island populations did not differ from mainland populations, and lacustrine islands had only slightly lower, but statistically significant, values for π. These observations are expected given the geologic history and proximity of continental and lacustrine islands to the coast. Anjouan island populations presented the lowest (0.73%) nucleotide diversity (π) for *A. gambiae* and

Príncipe island for *A. coluzzii* (0.66%), likely due their small size and high degree of isolation.

In general, the lower biocomplexity on isolated islands includes reduced genetic variation [24]. Our results are concordant with this observation (Figure 6). Selecting field sites with populations containing the lowest levels of variation should decrease the potential for transgene/genome interactions that might negatively impact GEM performance. These include São Tomé and Príncipe and the Comoros.

(h) Selection of candidate field sites

Each potential site was evaluated based on the criteria listed in Box 1. Evaluations were based on information available from the literature or calculated by us as summarized in the narrative above. Sites that fail to meet all primary criteria were eliminated from further consideration. Those sites that met all primary criteria were raised from potential status to candidate status. Some criteria require further analysis or site visits before a final evaluation can be completed. Sites visits are recommended for candidate sites only. Evaluation of insecticide resistance should be conducted during site visits. Security at candidate sites is dynamic and should be evaluated regularly before and during trials. Evaluation of potential impacts on endangered species requires knowledge about the extent to which these overlap ecologically with *A. coluzzii* and/or *A. gambiae* which can only be thoroughly evaluated by mosquito collections made during early site visits.

Overall evaluations are presented in Box 2.

BOX 2. Overall summary of evaluation of potential island sites. ✓ =site meets criterion; **x**=site fails to meet criterion; - = data not available; * = to be determined when site is visited.

	Primary criteria								Other considerations			
ISLAND	A. coluzzii/A. gambiae present	Geographic isolation	Genetic Isolation	Genetic diversity	Size (area)	Topography	Anopheline species richness		Insecticide resistance	Plasmodium prevalence	Endangered species	Travel feasibility
OCEANIC ISLANDS												
Canary Islands	x	✓	-	-	✓	✓	✓		*	x	*	✓
Cape Verde	x	✓	-	-	✓	✓	✓		*	x	*	✓
Annobón	✓	✓	✓	-	x	✓	✓		*	✓	*	x
São Tomé	✓	✓	✓	✓	✓	✓	✓		*	✓	*	✓
Príncipe	✓	✓	✓	✓	✓	✓	✓		*	✓	*	✓
Grand Comore	✓	✓	✓	✓	✓	✓	✓		*	✓	*	✓
Moheli	✓	✓	✓	✓	✓	✓	✓		*	✓	*	✓
Anjouan	✓	✓	✓	✓	✓	✓	✓		*	✓	*	✓
Mayotte	✓	✓	✓	✓	✓	✓	✓		*	✓	*	✓
Ile Europa	-	✓	-	-	x	x	✓		*	-	*	x
Madagascar	✓	✓	✓	✓	x	✓	x		*	✓	*	✓
Mauritius	x	✓	-	-	✓	✓	✓		*	x	*	✓
Réunion	x	✓	-	-	✓	✓	✓		*	x	*	✓
CONTINENTAL ISLANDS												
Bijagos	✓	x	✓	x	✓	x	x		*	✓	*	✓
Bioko	✓	✓	✓	-	✓	✓	x		*	✓	*	✓
Zanzibar	✓	x	-	-	✓	x	x		*	✓	*	✓
Pemba	✓	x	-	-	✓	x	✓		*	✓	*	✓
Mafia	✓	x	-	-	✓	x	-		*	✓	*	✓
LACUSTRINE ISLANDS (Lake Victoria)												
Bugala	✓	x	x	x	✓	x	✓		*	✓	*	✓
Koome	✓	x	x	x	✓	x	✓		*	✓	*	✓
Mfangano	✓	x	x	x	✓	x	✓		*	✓	*	✓
Ukara	✓	x	x	x	✓	x	✓		*	✓	*	✓

503 Evaluation of all twenty-two potential field sites indicate that Bioko, São Tomé &
504 Príncipe, and the Comoros Islands (Anjouan, Grand Comore, Mayotte and Moheli) can
505 be elevated from "potential" to "candidate" GEM field trial sites. The Mascarene
506 (Mauritius and Réunion) and Cape Verde Islands fit many criteria, but *Anopheles*
507 *gambiae* does not occur in these islands. Annobón scores well based on several our
508 criteria but travel there was determined to be infeasible, and the island was deemed
509 too small to represent a trial which would provide compelling outcomes.

510 Therefore, we propose the following as the lead candidate sites for a PHASE 2 GEM
511 field trial: the Comoros Islands, São Tomé and Príncipe and Bioko.

512

513 **4. Conclusions**

514 Our early decision to consider physical islands as the ideal sites for a GEM field trial was
515 guided by contemporary island biogeography theory. This theory provides the basis for
516 certain expectations concerning species richness, in our case, anopheline species
517 richness and also features such as genetic isolation and variability. Our results confirm
518 the relationships between geographic isolation and both genetic isolation (pairwise F_{ST}
519 values) and genetic diversity (nucleotide diversity, pi) which are significantly correlated
520 (Supplemental Figure 3).

521 The framework described here has been applied by the University of California Irvine
522 Malaria Initiative (UCIMI) as they enter PHASE 2 of GEM research. It is our belief that this
523 comprehensive framework provides identification of site(s) that will maximize the

prospect for success, minimize risk, and will serve as a fair, valid, and convincing test of the efficacy and impacts of the UCIMI GEM product, meeting the goal of a PHASE 2 field trial. Furthermore, this process provides a well-reasoned, science-based justification for selecting these sites for GEM field trials, and a solid foundation on which to approach ethical, social, and legal considerations with field site stakeholders.

References

1. Carballar-Lejarazú, R., & James, A. A. (2017). Population modification of Anopheline species to control malaria transmission. *Pathog Glob Health, 111*(8), 424-435. doi:10.1080/20477724.2018.1427192
2. Isaacs, A. T., Jasinskiene, N., Tretiakov, M., Thiery, I., Zettor, A., Bourgouin, C., & James, A. A. (2012). Transgenic Anopheles stephensi coexpressing single-chain antibodies resist Plasmodium falciparum development. *Proc Natl Acad Sci U S A, 109*(28), E1922-1930. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22689959. doi:10.1073/pnas.1207738109
3. Carballar-Lejarazú, R., Ogaugwu, C., Tushar, T., Kelsey, A., Pham, T. B., Murphy, J., Schmidt, H., Lee, Y., et al. (2020). Next-generation gene drive for population modification of the malaria vector mosquito, Anopheles gambiae. *Proceedings of the National Academy of Sciences, 117*(37), 22805-22814. Retrieved from https://www.pnas.org/content/pnas/117/37/22805.full.pdf. doi:10.1073/pnas.2010214117
4. WHO/TDR and FNIH. (2014). *The Guidance Framework for testing genetically modified mosquitoes* [Framework]. Retrieved from https://www.who.int/tdr/publications/year/2014/Guidance_framework_mosquitoes.pdf
5. African Union and NEPAD. (2018). *Gene Drives for Malaria Control and Elimination in Africa*. Retrieved from https://www.nepad.org/publication/gene-drives-malaria-control-and-elimination-africa
6. National Academies of Sciences Engineering and Medicine. (2016). *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. Washington, DC: The National Academies Press.
7. Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., et al. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology, 34*(1), 78-83. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26641531. doi:10.1038/nbt.3439

8. Alphey, L., Beard, C. B., Billingsley, P., Coetzee, M., Crisanti, A., Curtis, C., Eggleston, P., Godfray, C., et al. (2002). Malaria control with genetically manipulated insect vectors. *Science, 298*(5591), 119-121. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12364786. doi:10.1126/science.1078278

9. Benedict, M., D'Abbs, P., Dobson, S., Gottlieb, M., Harrington, L., Higgs, S., James, A., James, S., et al. (2008). Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: recommendations of a scientific working group. *Vector Borne and Zoonotic Diseases, 8*(2), 127-166. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18452399. doi:10.1089/vbz.2007.0273

10. Facchinelli, L., Valerio, L., Ramsey, J. M., Gould, F., Walsh, R. K., Bond, G., Robert, M. A., Lloyd, A. L., et al. (2013). Field cage studies and progressive evaluation of genetically-engineered mosquitoes. *PLoS Negl Trop Dis, 7*(1), e2001. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23350003. doi:10.1371/journal.pntd.0002001

11. Scott, T. W., Takken, W., Knols, B. G., & Boete, C. (2002). The ecology of genetically modified mosquitoes. *Science, 298*(5591), 117-119. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12364785. doi:10.1126/science.298.5591.117

12. Carpenter, S. R. (1996). Microcosm Experiments have Limited Relevance for Community and Ecosystem Ecology. *Ecology, 77*(3), 677-680. Retrieved from http://www.jstor.org/stable/2265490. doi:10.2307/2265490

13. Wynn, G., & Paradise, C. J. (2001). Effects of microcosm scaling and food resources on growth and survival of larval Culex pipiens. *BMC Ecology, 1*, 3. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11527507. doi:10.1186/1472-6785-1-3

14. Schindler, D. W. (1998). Whole-Ecosystem Experiments: Replication Versus Realism: The Need for Ecosystem-Scale Experiments. *Ecosystems, 1*(4), 323-334. Retrieved from https://doi.org/10.1007/s100219900026. doi:10.1007/s100219900026

15. James, S., Collins, F. H., Welkhoff, P. A., Emerson, C., Godfray, H. C. J., Gottlieb, M., Greenwood, B., Lindsay, S. W., et al. (2018). Pathway to Deployment of Gene Drive Mosquitoes as a Potential Biocontrol Tool for Elimination of Malaria in Sub-Saharan Africa: Recommendations of a Scientific Working Group(dagger). *American Journal of Tropical Medicine and Hygiene, 98*(6_Suppl), 1-49. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29882508. doi:10.4269/ajtmh.18-0083

16. James, S. L., Marshall, J. M., Christophides, G. K., Okumu, F. O., & Nolan, T. (2020). Toward the Definition of Efficacy and Safety Criteria for Advancing Gene Drive-Modified Mosquitoes to Field Testing. *Vector Borne and Zoonotic Diseases, 20*(4), 237-251. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/32155390. doi:10.1089/vbz.2019.2606

17. Cisnetto, V., & Barlow, J. (2020). The development of complex and controversial innovations. Genetically modified mosquitoes for malaria eradication. *Research Policy, 49*(3), 103917.

Retrieved from https://www.sciencedirect.com/science/article/pii/S0048733319302355. doi:https://doi.org/10.1016/j.respol.2019.103917

18. Nash, A., Urdaneta, G. M., Beaghton, A. K., Hoermann, A., Papathanos, P. A., Christophides, G. K., & Windbichler, N. (2019). Integral gene drives for population replacement. *Biol Open, 8*(1). Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/30498016. doi:10.1242/bio.037762

19. Li, M., Yang, T., Kandul, N. P., Bui, M., Gamez, S., Raban, R., Bennett, J., Sanchez, C. H., et al. (2020). Development of a confinable gene drive system in the human disease vector Aedes aegypti. *Elife, 9*. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/31960794. doi:10.7554/eLife.51701

20. Knols, B., & Bossin, H. (2006). Identification and characterization of field sites for genetic control of disease vectors. In (pp. 203-209).

21. Kolopack, P. A., & Lavery, J. V. (2017). Informed consent in field trials of gene-drive mosquitoes. *Gates Open Res, 1*, 14. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29355214. doi:10.12688/gatesopenres.12771.1

22. Neuhaus, C. P. (2018). Community Engagement and Field Trials of Genetically Modified Insects and Animals. *Hastings Cent Rep, 48*(1), 25-36. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29457234. doi:10.1002/hast.808

23. Resnik, D. B. (2014). Ethical issues in field trials of genetically modified disease-resistant mosquitoes. *Dev World Bioeth, 14*(1), 37-46. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23279283. doi:10.1111/dewb.12011

24. Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity (Edinb), 78 (Pt 3)*, 311-327. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9119706. doi:10.1038/hdy.1997.46

25. Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature, 457*(7231), 830-836. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19212401. doi:10.1038/nature07893

26. MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton, N.J.,: Princeton University Press.

27. Santos, A. M. C., Field, R., & Ricklefs, R. E. (2016). New directions in island biogeography. *Global Ecology and Biogeography, 25*(7), 751-768. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.12477. doi:https://doi.org/10.1111/geb.12477

28. Weigelt, P., Jetz, W., & Kreft, H. (2013). Bioclimatic and physical characterization of the world's islands. *Proceedings of the National Academy of Sciences, 110*(38), 15307-15312. Retrieved from https://www.pnas.org/content/pnas/110/38/15307.full.pdf. doi:10.1073/pnas.1306309110

29. Itescu, Y., Foufopoulos, J., Pafilis, P., & Meiri, S. (2020). The diverse nature of island isolation and its effect on land bridge insular faunas. *Global Ecology and Biogeography, 29*(2), 262-280.

Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13024. doi:https://doi.org/10.1111/geb.13024

30. Kayondo, J. K., Mukwaya, L. G., Stump, A., Michel, A. P., Coulibaly, M. B., Besansky, N. J., & Collins, F. H. (2005). Genetic structure of Anopheles gambiae populations on islands in northwestern Lake Victoria, Uganda. *Malar J, 4*, 59. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16336684. doi:10.1186/1475-2875-4-59

31. Nambuya A, S. C., Nkwiine C, Wetala P. (2013, 20–25 October 2013). *Abundance and ecological functional categories of soil macrofauna as indicators of soil chemical properties status in oil palm plantations in Bugala island Kalangala District, Uganda.* Paper presented at the Joint Proceedings of 27th Soil Science Society of East Africa & 6th African Soil Science Society, Nakuru, Kenya.

32. Ssegawa, P., & Nkuutu, D. N. (2006). Diversity of vascular plants on Ssese islands in Lake Victoria, central Uganda. *African Journal of Ecology, 44*(1), 22-29. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2028.2006.00609.x. doi:https://doi.org/10.1111/j.1365-2028.2006.00609.x

33. Zeemeijer, I. M. (2012). *Who gets What, When and How? New Corporate Land Acquisitions and the Impact on Local Livelihoods in Uganda.* (Master's Thesis), Utrecht University,

34. Idris, Z. M., Chan, C. W., Kongere, J., Gitaka, J., Logedi, J., Omar, A., Obonyo, C., Machini, B. K., et al. (2016). High and Heterogeneous Prevalence of Asymptomatic and Sub-microscopic Malaria Infections on Islands in Lake Victoria, Kenya. *Sci Rep, 6*, 36958. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27841361. doi:10.1038/srep36958

35. Lounio, T. (2014). *Population Dynamics and Livelihood Change on Ukara Island, Lake Victoria.* (Master's Thesis), University of Helsinki,

36. Mugono, M., Konje, E., Kuhn, S., Mpogoro, F. J., Morona, D., & Mazigo, H. D. (2014). Intestinal schistosomiasis and geohelminths of Ukara Island, North-Western Tanzania: prevalence, intensity of infection and associated risk factors among school children. *Parasit Vectors, 7*, 612. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25533267. doi:10.1186/s13071-014-0612-5

37. Smith, A. (1955). The Transmission of Bancroftial Filariasis on Ukara Island, Tanganyika. I.—A Geographical and Ecological Description of the Island with an annotated List of Mosquitos and other Arthropods of medical Importance. *Bulletin of Entomological Research, 46*(2), 419-436. Retrieved from https://www.cambridge.org/core/article/transmission-of-bancroftial-filariasis-on-ukara-island-tanganyika-ia-geographical-and-ecological-description-of-the-island-with-an-annotated-list-of-mosquitos-and-other-arthropods-of-medical-importance/193FF09AACF861834482ACB9CF8AB3D6. doi:10.1017/S000748530003100X

38. BakamaNume, B. B. (2010). *A Contemporary Geography of Uganda*: Mkuki Na Nyota.

39. Google Earth (Cartographer). Map showing location of Kome Island. Retrieved from https://earth.google.com/web/search/Koome+Island,+Uganda/@-

0.08724243,32.74244361,1211.47630539a,48380.8541521d,35y,16.3547808h,11.41826421t,0r/data=CigiJgokCQxDZ9yeljVAEQtDZ9yeljXAGRf30zaOXUJAIW6hBeDnvlDA

40. Jackson, G., & Gartlan, J. S. (1965). The Flora and Fauna of Lolui Island, Lake Victoria: A Study of Vegetation, Men and Monkeys. *Journal of Ecology, 53*(3), 573-597. Retrieved from http://www.jstor.org/stable/2257622. doi:10.2307/2257622

41. Nampijja, M., Webb, E. L., Kaweesa, J., Kizindo, R., Namutebi, M., Nakazibwe, E., Oduru, G., Kabuubi, P., et al. (2015). The Lake Victoria Island Intervention Study on Worms and Allergy-related diseases (LaVIISWA): study protocol for a randomised controlled trial. *Trials, 16*, 187. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25902705. doi:10.1186/s13063-015-0702-5

42. Tuhebwe, D., Bagonza, J., Kiracho, E. E., Yeka, A., Elliott, A. M., & Nuwaha, F. (2015). Uptake of mass drug administration programme for schistosomiasis control in Koome Islands, Central Uganda. *PLoS One, 10*(4), e0123673. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25830917. doi:10.1371/journal.pone.0123673

43. Dahl, A. L. (1991). *Island directory* (Prelim. ed.). Nairobi, Kenya: UNEP.

44. Japan Aerospace Exploration Agency. (2020). *ALOS Global Digital Surface Model (DSM) ALOS World 3D-30m (AW3D30) Version 3.1.* Retrieved from: https://www.eorc.jaxa.jp/ALOS/en/aw3d30/aw3d30v31_product_e_a.pdf

45. Jarvis A., H. I. R., A. Nelson, E. Guevara. (2008). *Hole-filled seamless SRTM data V4.* I. C. f. T. A. (CIAT). Retrieved from: https://srtm.csi.cgiar.org/

46. Anopheles gambiae 1000 Genomes Consortium. (2020). Genome variation and population structure among 1142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii. *Genome Research, 30*, 1-14. doi:10.1101/gr.262790.120

47. Bergey, C. M., Lukindu, M., Wiltshire, R. M., Fontaine, M. C., Kayondo, J. K., & Besansky, N. J. (2020). Assessing connectivity despite high diversity in island populations of a malaria mosquito. *Evolutionary Applications, 13*(2), 417-431. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.12878. doi:https://doi.org/10.1111/eva.12878

48. Nieman, C. C., Yamasaki, Y., Collier, T. C., & Lee, Y. (2015). A DNA extraction protocol for improved DNA yield from individual mosquitoes. *F1000Res, 4*, 1314. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26937269. doi:10.12688/f1000research.7413.1

49. Yamasaki, Y. K., Nieman, C. C., Chang, A. N., Collier, T. C., Main, B. J., & Lee, Y. (2016). Improved tools for genomic DNA library construction of small insects [version 1; not peer reviewed]. *F1000Research, 5*, 211 (poster). Retrieved from https://doi.org/10.7490/f1000research.1111322.1.

50. Clarkson, C. S., Miles, A., Harding, N. J., O'Reilly, A. O., Weetman, D., Kwiatkowski, D., Donnelly, M. J., & Anopheles gambiae Genomes, C. (2021). The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors Anopheles gambiae and

Anopheles coluzzii. *Molecular Ecology*. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/33590926. doi:10.1111/mec.15845

51. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24695404. doi:10.1093/bioinformatics/btu170
52. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20110278. doi:10.1093/bioinformatics/btq033
53. Sharakhova, M. V., Hammond, M. P., Lobo, N. F., Krzywinski, J., Unger, M. F., Hillenmeyer, M. E., Bruggner, R. V., Birney, E., & Collins, F. H. (2007). Update of the Anopheles gambiae PEST genome assembly. *Genome Biol, 8*(1), R5. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17210077. doi:10.1186/gb-2007-8-1-r5
54. Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., et al. (2002). The genome sequence of the malaria mosquito Anopheles gambiae. *Science, 298*(5591), 129-149. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12364791. doi:10.1126/science.1076181
55. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv, 1303*.
56. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics, 31*(12), 2032-2034. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25697820. doi:10.1093/bioinformatics/btv098
57. Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv, 1207*.
58. Tan, A., Abecasis, G. R., & Kang, H. M. (2015). Unified representation of genetic variants. *Bioinformatics, 31*(13), 2202-2204. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25701572. doi:10.1093/bioinformatics/btv112
59. Miles, A., & Harding, N. (2017). *cggh/scikit-allel: v1.1.8 (Version v1.1.8).* Zenodo. Retrieved from: http://doi.org/10.5281/zenodo.822784
60. Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics, 132*(2), 583-589. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1427045.
61. Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting FST: the impact of rare variants. *Genome Research, 23*(9), 1514-1521. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23861382. doi:10.1101/gr.154831.113
62. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., et al. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21653522. doi:10.1093/bioinformatics/btr330

63. LNRS Data Services Inc. (2021). *Q-00223232 Historical Data Export.* Cirium.
64. Exmile Solutions Ltd. (2021). *Historical AIS data (port calls) for 2019-01-01 to 2019-12-31 for PORT = BJCOO, CMDLA, TZDAR, CVRAI, CVMIN, ZADUR, KMYVA, KMMUT, Fomboni, STTMS, STPCP.* Marine Traffic.
65. Irish, S. R., Kyalo, D., Snow, R. W., & Coetzee, M. (2020). Updated list of Anopheles species (Diptera: Culicidae) by country in the Afrotropical Region and associated islands. *Zootaxa, 4747*(3), zootaxa 4747 4743 4741. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/32230095. doi:10.11646/zootaxa.4747.3.1
66. Kyalo, D., Amratia, P., Mundia, C. W., Mbogo, C. M., Coetzee, M., & Snow, R. W. (2017). A geo-coded inventory of anophelines in the Afrotropical Region south of the Sahara: 1898-2016. *Wellcome Open Res, 2*, 57. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28884158. doi:10.12688/wellcomeopenres.12187.1
67. Huestis, D., Dao, A., Diallo, M., Sanogo, Z. L., Samake, D., Yaro, A., Ousman, Y., Linton, Y.-M., et al. (2019). Windborne long-distance migration of malaria mosquitoes in the Sahel. *Nature*.
68. Service, M. W. (1997). Mosquito (Diptera: Culicidae) dispersal--the long and short of it. *Journal of Medical Entomology, 34*(6), 579-588. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9439109. doi:10.1093/jmedent/34.6.579
69. Kaufmann, C., & Briegel, H. (2004). Flight performance of the malaria vectors Anopheles gambiae and Anopheles atroparvus. *J Vector Ecol, 29*(1), 140-153. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15266751.
70. Belkin, J. N. (1962). *The Mosquitoes of the South Pacific: (Diptera, Culicidae)*: University of California Press.
71. Campos, E. G., Trevino, H. A., & Strom, L. G. . (1961). The Dispersal of Mosquitoes by Railroad Trains involved in International Traffic. *Mosquito News, 21*(3), 190-192.
72. Eritja, R., Palmer, J., Roiz, D., Sanpera-Calbet, I., & Bartumeus, F. (2017). Direct Evidence of Adult Aedes albopictus Dispersal by Car. *Scientific Reports, 7*, 14399. doi:10.1038/s41598-017-12652-5
73. Frean, J., Brooke, B., Thomas, J., & Blumberg, L. (2014). Odyssean malaria outbreaks in Gauteng Province, South Africa, 2007 - 2013. *S Afr Med J, 104*(5), 335-338. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25212198. doi:10.7196/samj.7684
74. Brown, D. M., Alphey, L. S., McKemey, A., Beech, C., & James, A. A. (2014). Criteria for identifying and evaluating candidate sites for open-field trials of genetically engineered mosquitoes. *Vector Borne and Zoonotic Diseases, 14*(4), 291-299. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24689963. doi:10.1089/vbz.2013.1364
75. Lee, Y., Marsden, C. D., Norris, L. C., Collier, T. C., Main, B. J., Fofana, A., Cornel, A. J., & Lanzaro, G. C. (2013). Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, Anopheles gambiae. *Proc Natl Acad Sci U S A, 110*(49), 19854-19859. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24248386. doi:10.1073/pnas.1316851110

76. White, G. B. (1971). Chromosomal evidence for natural interspecific hybridization by mosquitoes of the Anopheles gambiae complex. *Nature, 231*(5299), 184-185. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/4930676. doi:10.1038/231184a0

77. Courtier-Orgogozo, V., Danchin, A., Gouyon, P. H., & Boete, C. (2020). Evaluating the probability of CRISPR-based gene drive contaminating another species. *Evol Appl, 13*(8), 1888-1905. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/32908593. doi:10.1111/eva.12939

78. Antonio-Nkondjio, C., Kerah, C. H., Simard, F., Awono-Ambene, P., Chouaibou, M., Tchuinkam, T., & Fontenille, D. (2006). Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. *Journal of Medical Entomology, 43*(6), 1215-1221. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17162956. doi:10.1603/0022-2585(2006)43[1215:cotmvs]2.0.co;2