MODELING THE ASSOCITATION GENOTYPE AND PHENOTYPE in ANOPHELES MOSQUITOES IN AFRICA

## I- Data inputs

• WHO discriminating concentration bioassay results from the Vector Atlas database,incorporating data from IR Mapper and WHO Malaria Threats Map. • Vector Atlas data results about genotype frequency • Covariate data : - Modelled use of ITNs 2000-2022 (MAP, see Bertozzi-villa et al. 2021) - Reported coverage of IRS 1997-2022 (WHO/MAP) - Population density 2000-2030 (Worldpop/MAP, adjusted to UN forecasts 2021-2030)

## II - Observation models

## 1- Bioassay data

Given data on mosquitoes tested and deaths observed, we assume these counts follow a Beta–Binomial distribution: and , the observed numbers of mosquitoes tested, and that died,we assume these follow a beta-binomial distribution:

Where is the modelled population-level susceptibility of the total vector population at the location and time from which this sample of mosquitoes was collected. The overdispersion parameter is specific to the insecticide class of insectcide , capturing the non-independence of individuals mosquitoes within the same bioassay sample. This accounts for the extra variability in mortality observed across different samples, reflecting the inherent heterogeneity in susceptibility measurements.

### Genotype test results for multilocus:

Let denote the number of mosquitoes tested and the number positive for genotype in a given record. Here, represents the vector of genotype frequencies in the population at the location and time of sampling. When multiple genotype categories are considered, we model these counts using a Dirichlet–Multinomial distribution, which accounts for both the multinomial sampling of genotypes and extra variability (overdispersion) in the population frequencies.

### Genotype test results for single locus:

Let denote the number of mosquitoes positive for a particular genotype at locus , and let be the total number of mosquitoes tested. Assuming Hardy–Weinberg equilibrium (HWE) with resistant-allele frequency (so

and

for the susceptible allele), the expected genotype frequencies in the population are given by

And the over-dispersion via a Dirichlet–Multinomial:

### Allele frequency data:

We consider is the number of mosquitoes tested at a gien time point , and let be the allele frequency of a specific allele at locus . We define as the number of carrying allele in the sample. These countes are modeled using a Betabinomial likelihood to account for sampling variability and the overdispersion :

## II - Process models

### 1- Computing phenotype frequencies:

The model computes the mortality probability for each insecticide at time as a weighted mixture of genotype-specific utilities. For each genotype at locus , the locus-specific susceptibility contribution is calculated based on allelic configuration and dominance :

where: and : are indicators for the allelic states of genotype at locus (SS and RR, respectively), is the locus-specific resistance effect (how much the locus reduces susceptibility to insecticide , is the dominance coefficient, representing the fraction of the resistant allele effect expressed in heterozygotes. The overall susceptibility of genotype to insecticide is obtained by multiplying across loci and scaling by the baseline mortality of the wild-type genotype :

At each time step , the population-level fraction of susceptible mosquitoes is then computed as the weighted average over genotype frequencies (which sum to 1):

This formulation captures how allelic states at multiple loci, locus-specific resistance effects, and dominance combine to determine both genotype-specific mortality and the population-level susceptibility to each insecticide.

### 2- Computing the probability of having a specific genotype

We construct the multilocus genotype frequency as the product of per-locus, genotype probabilities across the nloci:

Let denote the frequency of allele resistant at locus lo at time t (so is the frequency of allele S). For each multilocus genotype g and locus lo, define left and right allele indicator dummy matrices L and R (S =1, R=0). The per-locus probability that matches the genotype of g is

### Computing the frequency of the genotype at the next time series The genotype will be updated by normalization. At each time step we form an unnormalized genotype mass by multiplying the genotype prior (from allele frequencies) with its multilocus fitness. Concretely, with

genotypes over loci, so

The prior for genotype gat time tis (product of locus-wise left/right allele probabilities, accounting for the duplicate via ); the genotype fitness is

) The unnormalized next-step mass is

We then normalize across genotypes,

to obtain a valid probability distribution that sums to 1.

### Update the multilocus allele frequency

For each locus , we compute the next-step resistant-allele frequency as the posterior-weighted average number of resistant alleles per diploid, divided by 2. Concretely, is the normalized posterior probability of genotype at time (so ), and , encode the left and right alleles at that locus with and . Thus, is an indicator that the left allele is resistant, and indicates the right allele is resistant. The quantity . Averaging this count across genotypes with weights gives the expected number of resistant alleles per individual at that locus; dividing by 2 converts this expected count (out of two alleles) into a frequency:

### Selection pressure and relative fitness

We model locus- and insecticide-specific selection pressures as positively-constrained selection coefficients , representing selection towards resistance only; if no resistance driver is present, the selection is zero. These coefficients are expressed as a linear combination of observed covariates , which quantify the intensity of potential resistance drivers at each locus and time (scaled from 0 to 1, e.g., from 0 for 0% LLIN coverage to 1 for 100% coverage), and the positive selection effects for insecticide associated with driver :

and the relative fitness,

The selection effects are constrained to be positive via

, and the coefficients are modeled hierarchically: first sharing information between insecticides within the same class , and then across classes,

which allows the model to learn about common drivers of resistance for insecticides with sparse assay data and captures pleiotropic effects of resistance-conferring genotypes across multiple insecticides.