Estimating *Anopheles spp.* population size

Isabelle Rajendiran, Tin-Yu Hui[1] and Vassiliki Koufopanou[1]
[1]Department of Life Sciences, Silwood Park, Imperial College London

Background

Anopheles spp, especially *A.gambiae* and *A.coluzzii*, are key **vectors of human malaria**.[1] Knowing the population size of these mosquitoes is critical for vector control strategies:

- **Monitoring pre- and post- intervention effectiveness** i.e., lower population size
- **Determining the necessary amount of intervention** e.g., enough gene drive is needed to prevent its loss in the population by genetic drift[1]

Here, **effective population size (*Ne*)** will be estimated - the size of an ideal (Wright-Fisher) population which loses heterozygosity at the same rate as the observed population.[2] Includes:

- **Historical *Ne***: *Ne* estimation over a longer period and wider geographical scale i.e., global measure[2]
  - Can be calculated using **nucleotide diversity (π)**
- **Contemporary *Ne***: *Ne* estimation over recent years[2] and local geographical scale[3]
  - Can be calculated using **genetic drift** (temporal change in allele frequencies)

Large *Ne*

Larger π
Smaller drift

Small *Ne*

Smaller π
Larger drift

WHY?
Smaller populations experience greater drift which also leads to loss of genetic variation over time (due to loss and fixation of alleles)[4]

Aims

1. Estimate historical and contemporary effective population sizes (*Ne*) of chosen *Anopheles spp.* populations
2. Investigate the appropriate data to use in the estimations i.e., chromosome, chromosome position and SNP site type

Research Plan

The project is conducted using Python on Jupyter Notebook.

Access the *Anopheles gambiae* 1000 Genomes (Ag1000G) Project phase 3 dataset from MalariaGEN[5]

↓

Explore database metadata and select appropriate samples

Criteria: Minimum of 50 samples taken within the same rough geographical area. **Diversity requires a minimum of 1 sample. Drift requires a minimum of 2 samples, taken between 2-5 years apart.**

↓

Filtering the SNPs required for diversity and drift calculation

Different subsets drawn for various objectives e.g., 3R vs X chromosome, and intergenic vs 4-fold degenerate coding sites.

↓

Genetic drift and nucleotide diversity calculation

↓

Estimate historical and contemporary *Ne*

Figure 1. The experimental workflow for estimating *Ne* from genomic data

Risk and Mitigation

- **Not enough samples in each population**: Samples within the same rough location were arbitrarily grouped as one population here but this could be re-conducted by using FST to genetically identify populations
- **Varying diversity across the chromosome:** Measure recombination levels across the chromosome and re-conduct analysis with new chromosome positions (removing area of low recombination)

Preliminary Results

Table 1. Nucleotide diversity results for *A.coluzzii* samples taken in 2013 and 2015 in Niono, Mali. Results compare between chromosomes (X and 3R) and neutral site types (intergenic and 4-fold degenerate coding sites). A sample size of 70 mosquitoes was used for all.

Sample Cohort	Genomic Region	Nucleotide Diversity (π)	
		Intergenic	4-CDS
ML-4_Niono_colu_2013	3R	0.0136	0.0240
	X	0.0080	0.0078
ML-4_Niono_colu_2015	3R	0.0136	0.0239
	X	0.0080	0.0078

1. **No difference in diversity is observed between 2013 and 2015** for all comparisons
2. **The observed 3R:X ratio was lower than the expected ratio of 1:0.75**, but consistent with previous findings from the Ag1000G project[5]
3. **Intergenic sites on chromosome 3R showed almost double the diversity compared to 4-CDS sites**, while no such difference was observed for chromosome X

Evaluation

Next steps

- Investigate the reason for the difference between the intergenic vs 4-CDS diversity levels
- Conduct analyses for other species and countries - potentially compare Africa-wide and local *Ne* estimations

References

1. Monawwer, S. A., Alzubaidi, A. O. I., Yasmin, F., Haimour, S. M. Q., Shah, S. M. I. & Ullah, I. (2022) Gene-drive mosquitoes: a prospect for future malaria control. The Pan African Medical Journal. 41 109. 10.11604/pamj.2022.41.109.31687.
2. Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K. & Allendorf, F. W. (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. Conservation Genetics. 11 355-373.
3. Hui, T. J., Brenas, J. H. & Burt, A. (2021) Contemporary Ne estimation using temporally spaced data with linked loci. Molecular Ecology Resources. 21 (7), 2221-2230. 10.1111/1755-0998.13412.
4. Hartl, D. L., & Clark, A. G. (2007). Principles of population genetics (4th ed.). Oxford University Press.
5. The Anopheles gambiae 1000 Genomes Consortium. (2017) Genetic diversity of the African malaria vector Anopheles gambiae. Nature. 552 (7683), 96-100. 10.1038/nature24995.

Acknowledgment

Jon Brenas for for addressing my python-related queries