Principle Component Analysis & Mosquitos

Eve198

Week 8

Maddie Armstrong & Rachael Bay

Climate and Urbanization Drive Mosquito Preference for Humans

Highlights

- African populations of *Ae. aegypti* vary in preference for human versus animal odor

Authors

Noah H. Rose, Massamba Sylla, Athanase Badolo, ..., Bradley J. White, Jacob E. Crawford, Carolyn S. McBride

Collected mosquito eggs with ovitraps
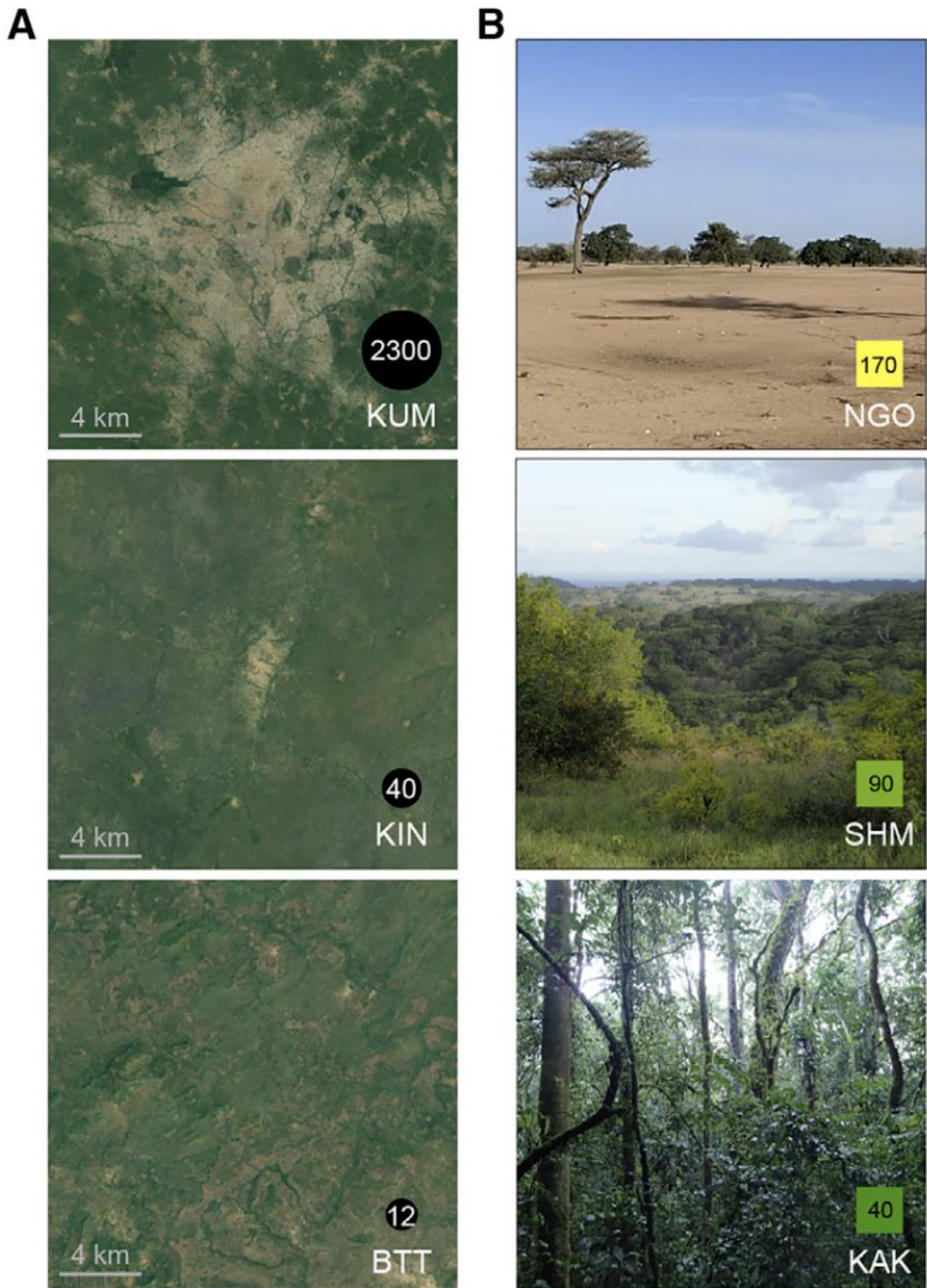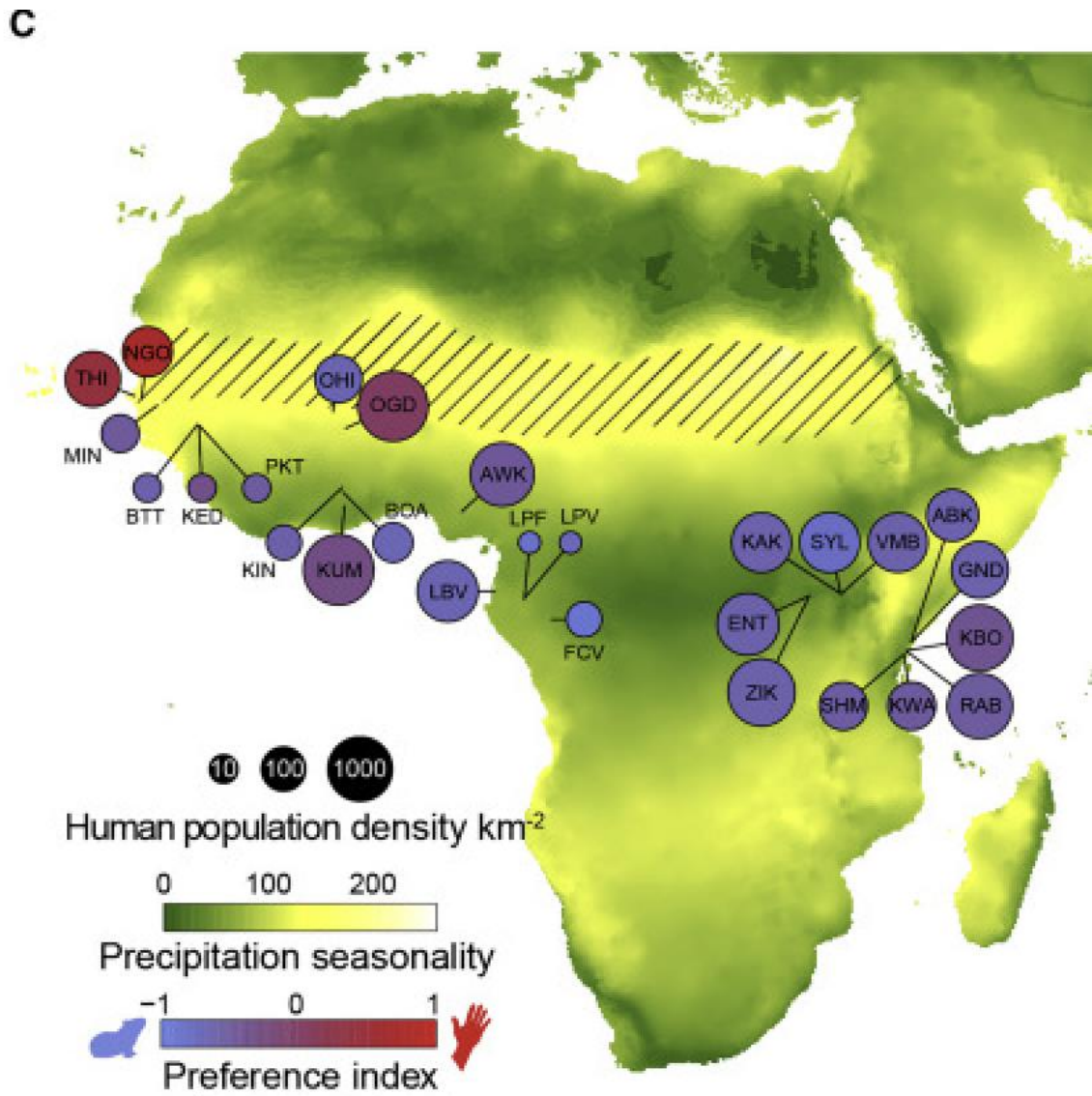Hatched eggs in lab for human preference experiments

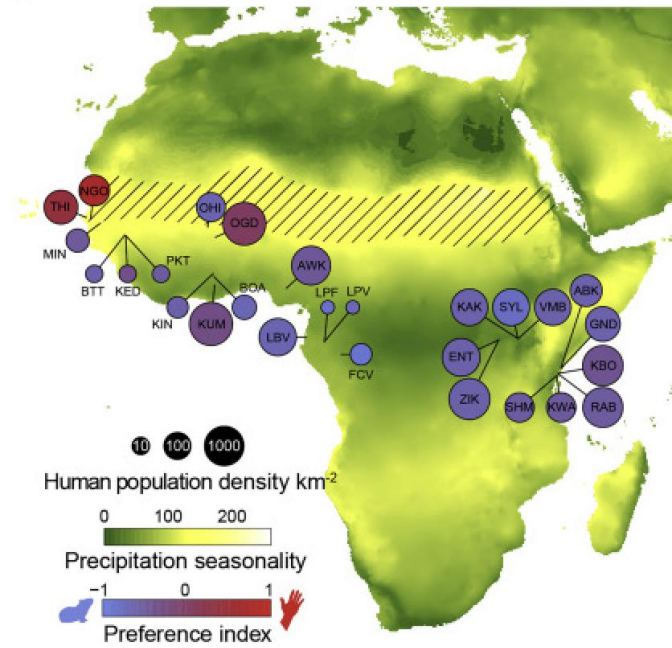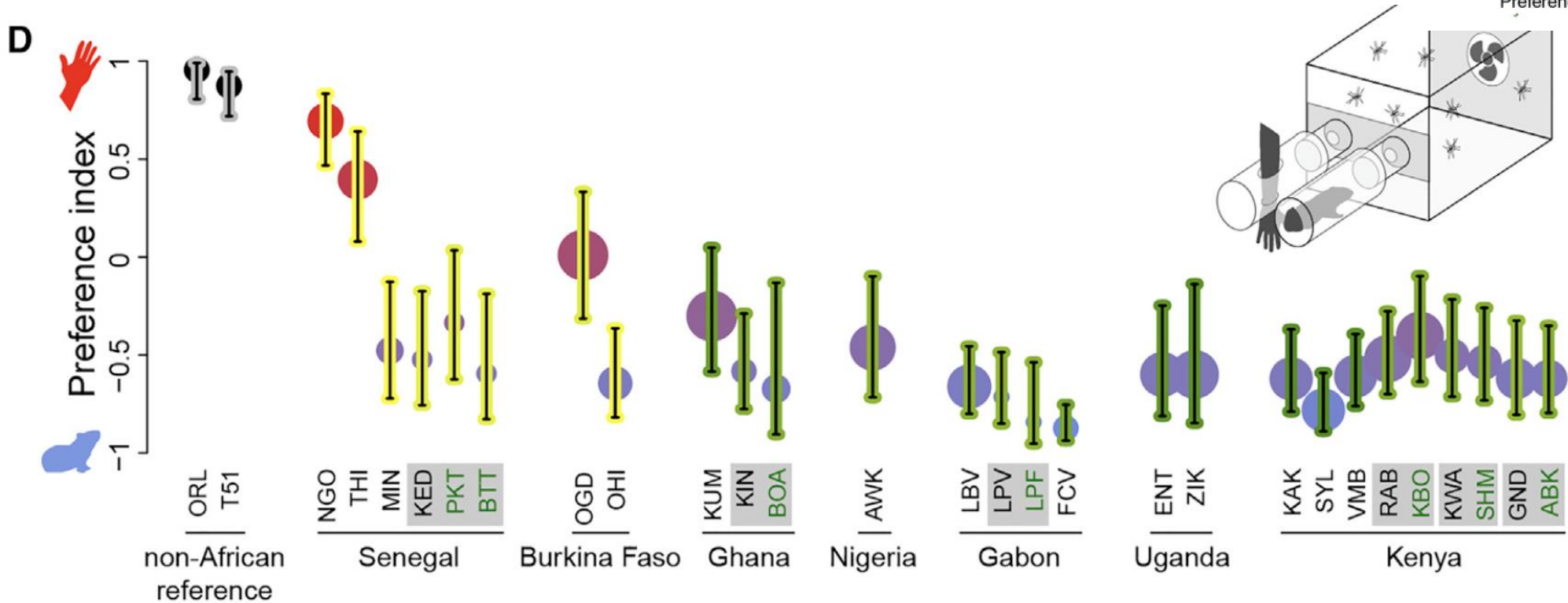Fig 1: Preference for human odor varies widely in Aedes aegypti mosquitos across Africa

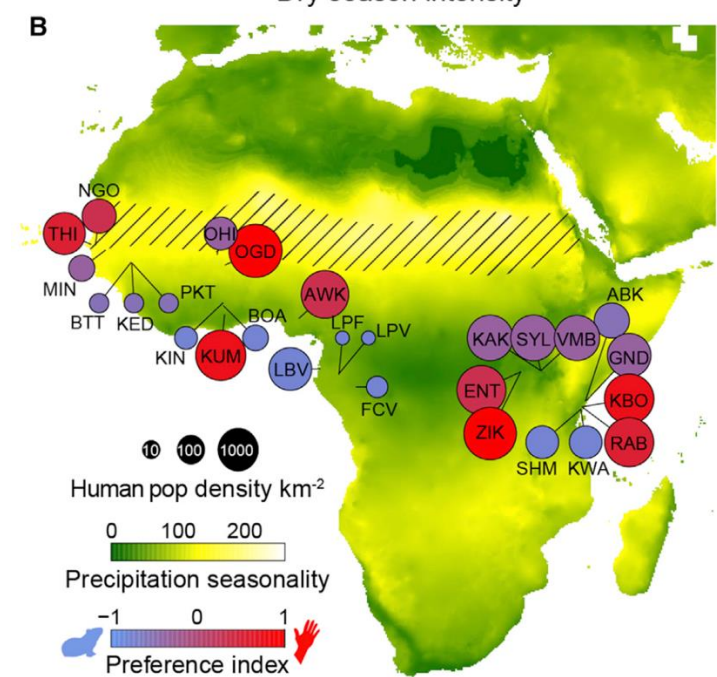# Fig 2: Human Population Density + Dry Season Explain Variation in Preference
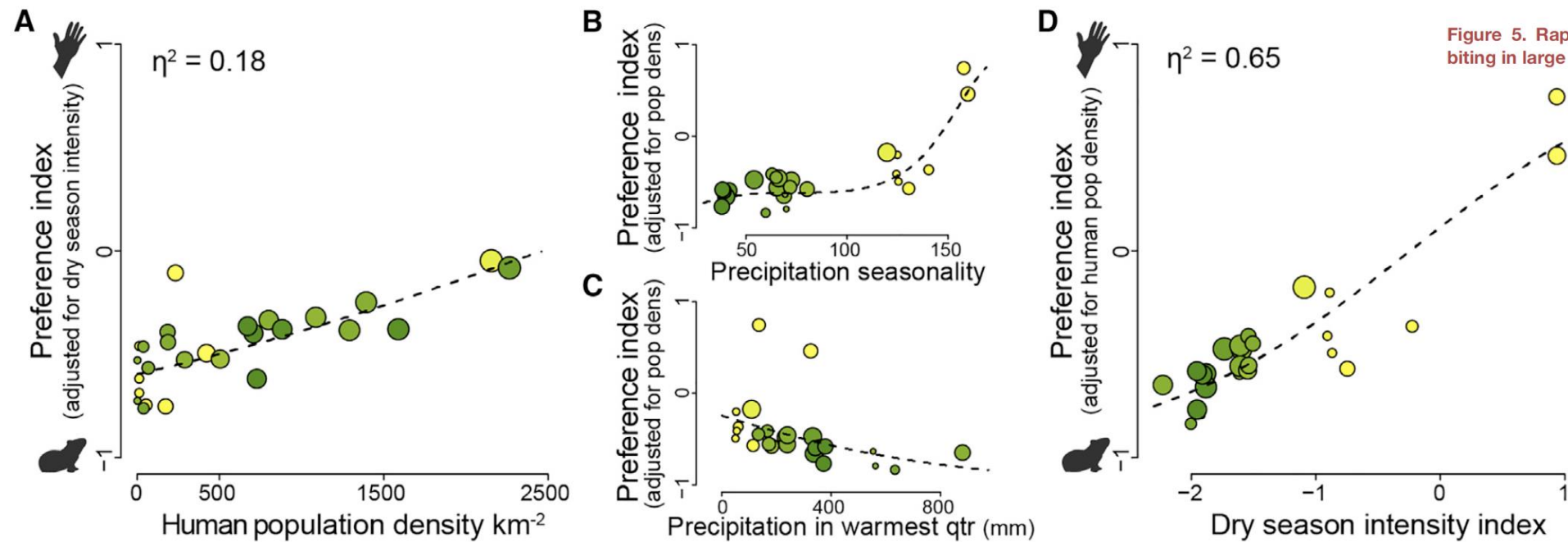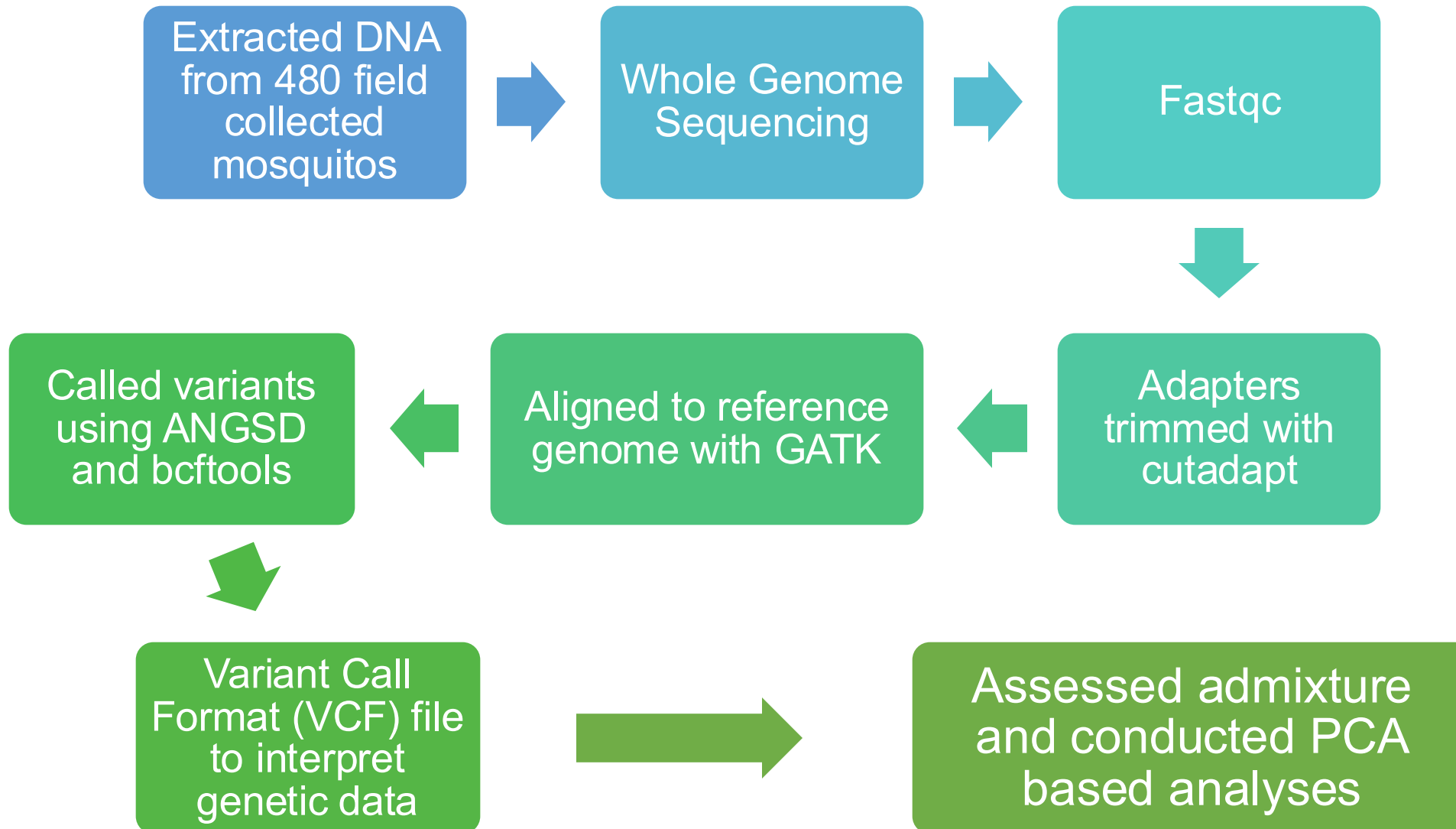

Figure 5. Rapid urbanization could favor a shift toward human-biting in large African cities by 2050

Bioinformatic Pipeline:

© Bronzit_poet via iNaturalist

Extracted DNA from 480 field collected mosquitos → Whole Genome Sequencing → Fastqc

↓

Called variants using ANGSD and bcftools ← Aligned to reference genome with GATK ← Adapters trimmed with cutadapt

↓

Variant Call Format (VCF) file to interpret genetic data → Assessed admixture and conducted PCA based analyses

Fig 3: Revisiting Population Structure:
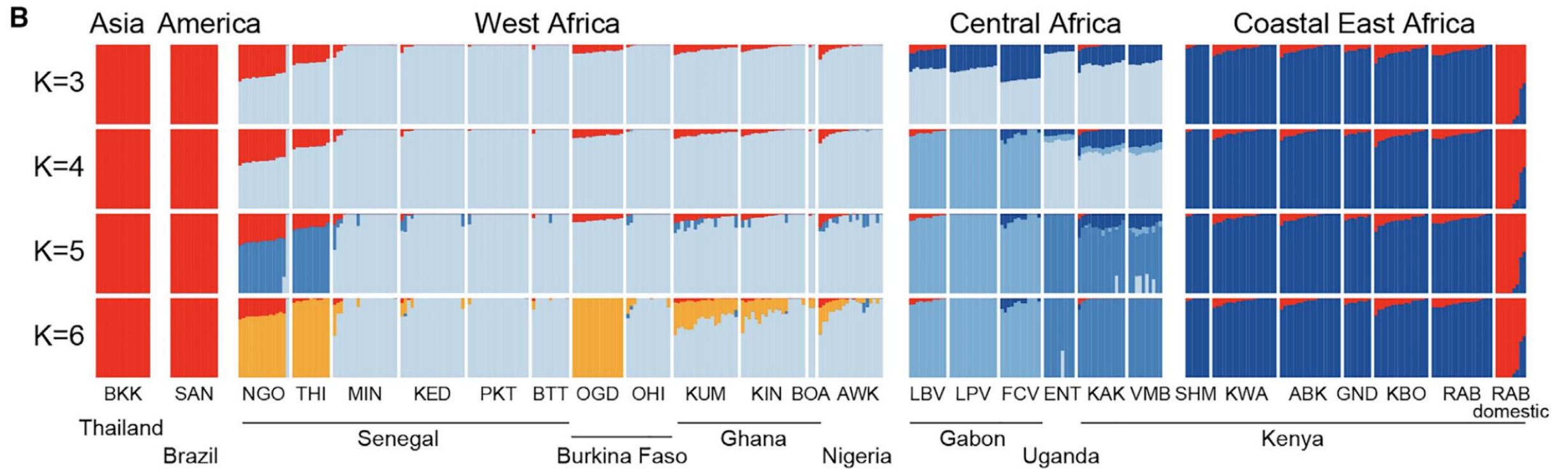
What is a Principle Component Analysis?

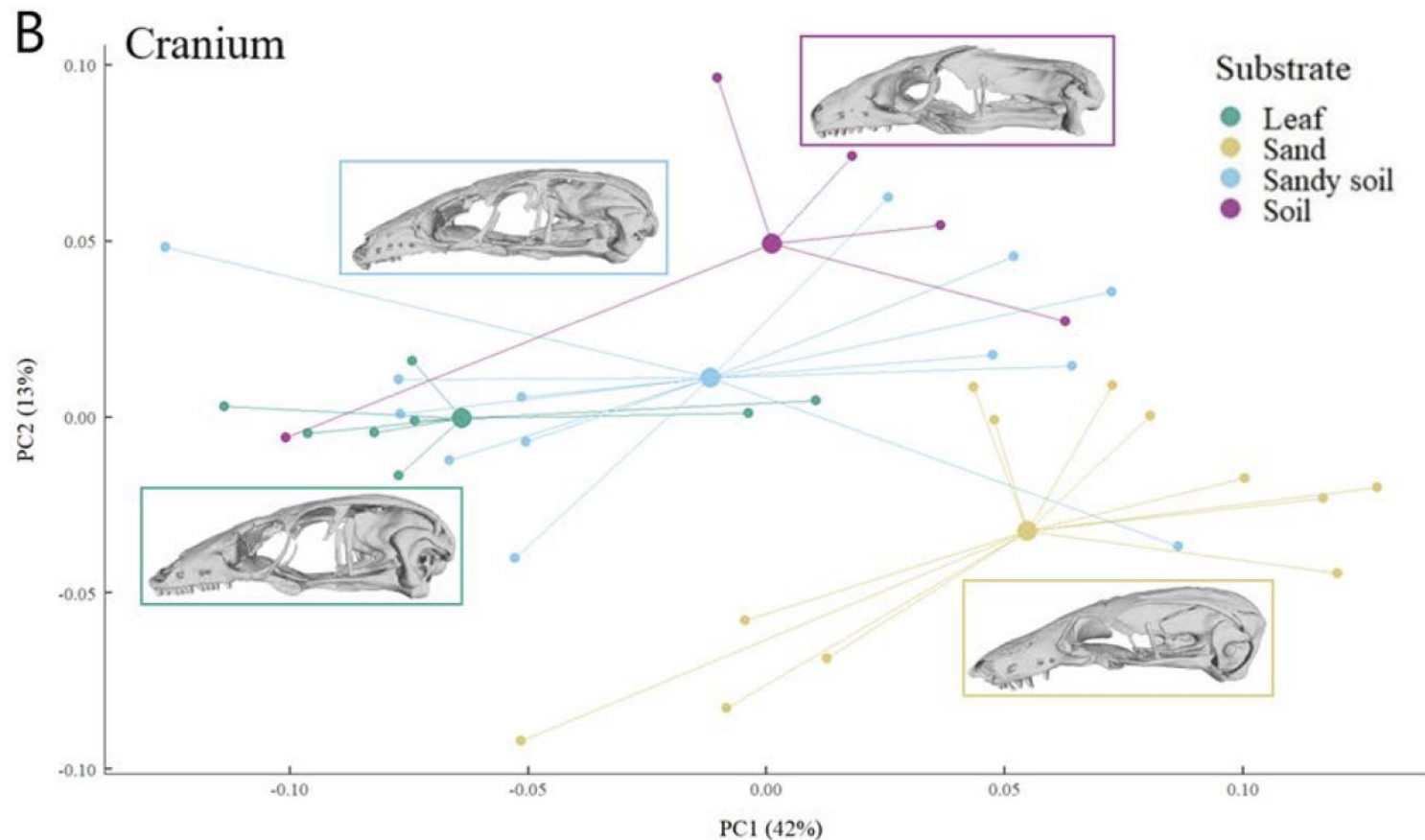- A method to reduce the dimensionality of large datasets

What is a Principle Component Analysis?

- A method to reduce the dimensionality of large datasets

(Stepanova & Bauer, 2021)

What is a Principle Component Analysis?

- A method to reduce the dimensionality of large datasets
  - Genomic Data= many SNPs for many individuals

© Nature Lover via iNaturalist
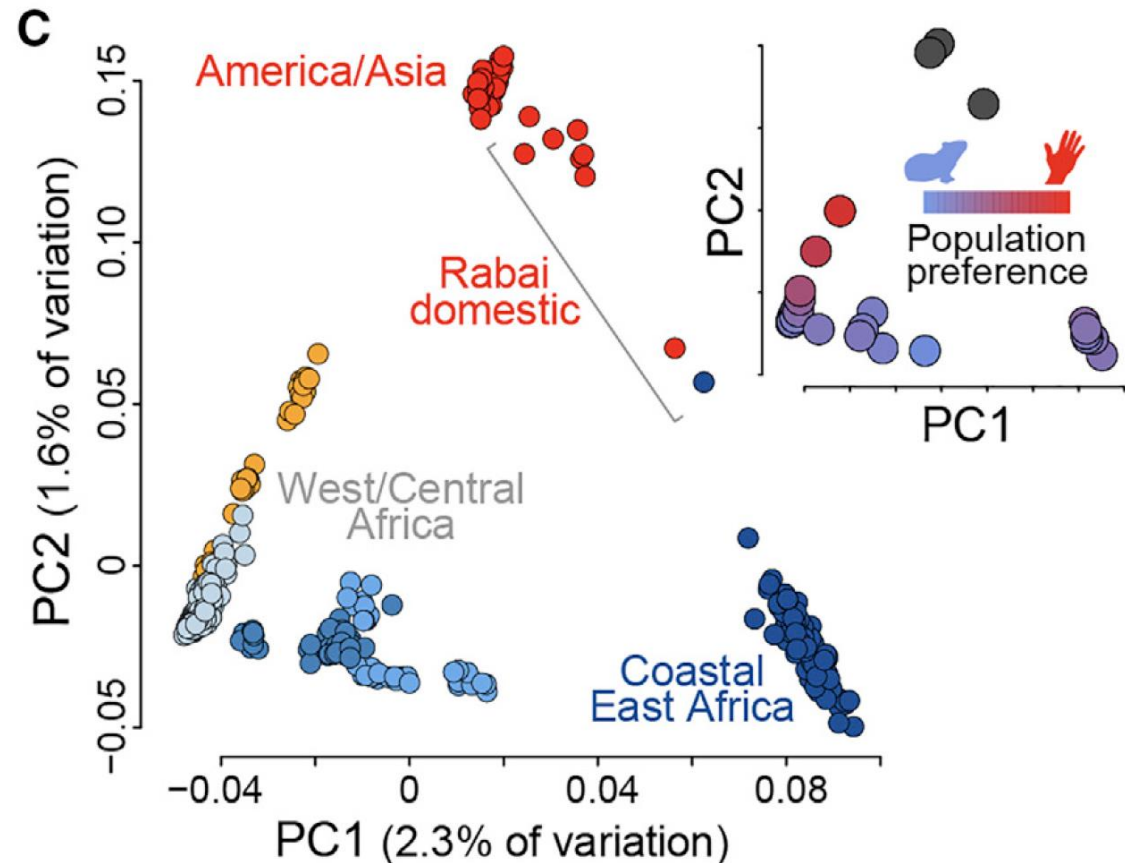
Steps involved in running a PCA

- Data standardization
  - PCA is sensitive to variances in the data
    - For example, if one site is much more variable than any other

  - Calculate a covariance matrix where (x, y, z) are individual samples

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

  - Compute the eigen vectors and eigen values of the covariance matrix
    - This is where we identify the principal components of the data!

Eigenvalues and eigenvectors

Basically:

An eigenvector points in the direction it is stretched by the transformation and the eigenvalue is how much it was stretched.

Vector transformation

Eigenvector demonstration

$$A = \frac{1}{2}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$
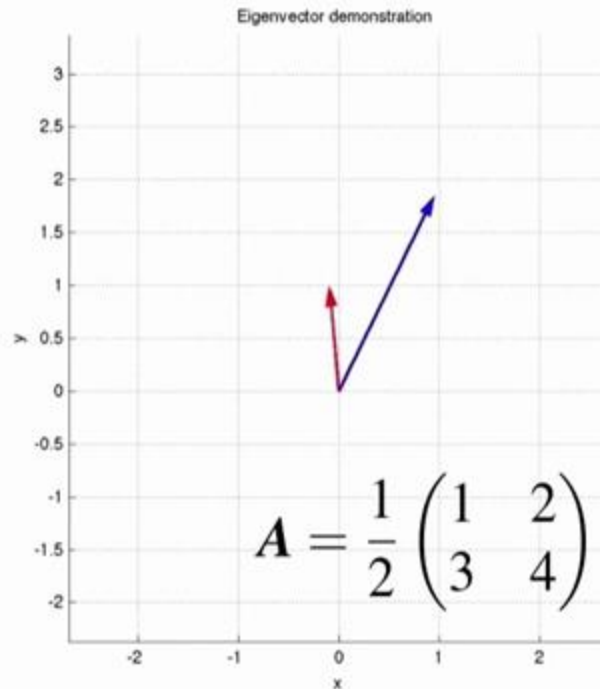
vector Nx1

matrix NxN

vector Nx1

$$v' = Av$$

- For every matrix A there are N vectors whose *direction is unchanged* by **A**
  - ↪ these are the **eigenvectors**
  - ↪ their length is scaled by the **eigenvalue**

What is a principal component?

- New variables constructed from a mixture of our initial variable.

- A PCA will return to you the same dimension of data you gave it.
  - But most of the information should be in the first few PCs
  - For 10 individuals genotyped at 100 loci you will get 10 principal components
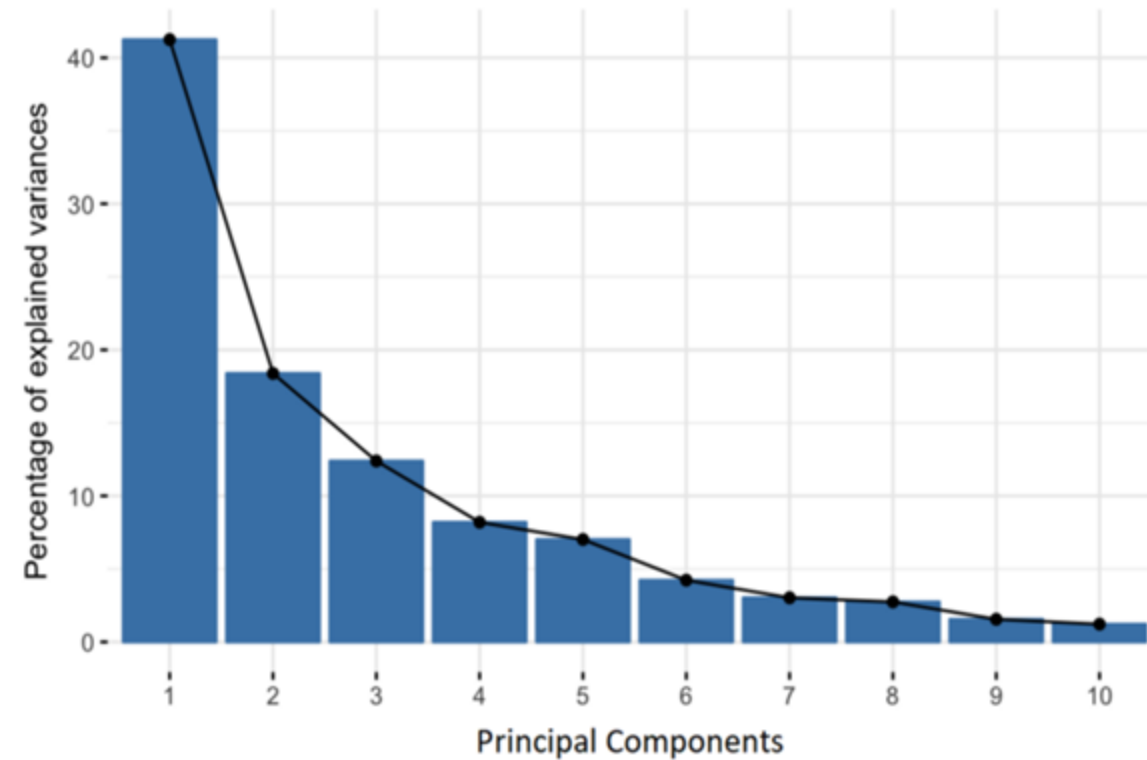
# Why do we care about genetic diversity?

- Tells us about population demography, gene flow

- Can inform conservation strategies

- Evolution requires variation in phenotypes and genotypes

© Marie Delport via iNaturalist

Working with a subset of data today:

Our methods differ from those used in this paper, primarily for convenience and time (ours).

- Just a section of chromosome 1

© Rentokil

1p   1q   *   3q

*

2p

3p

2q

*
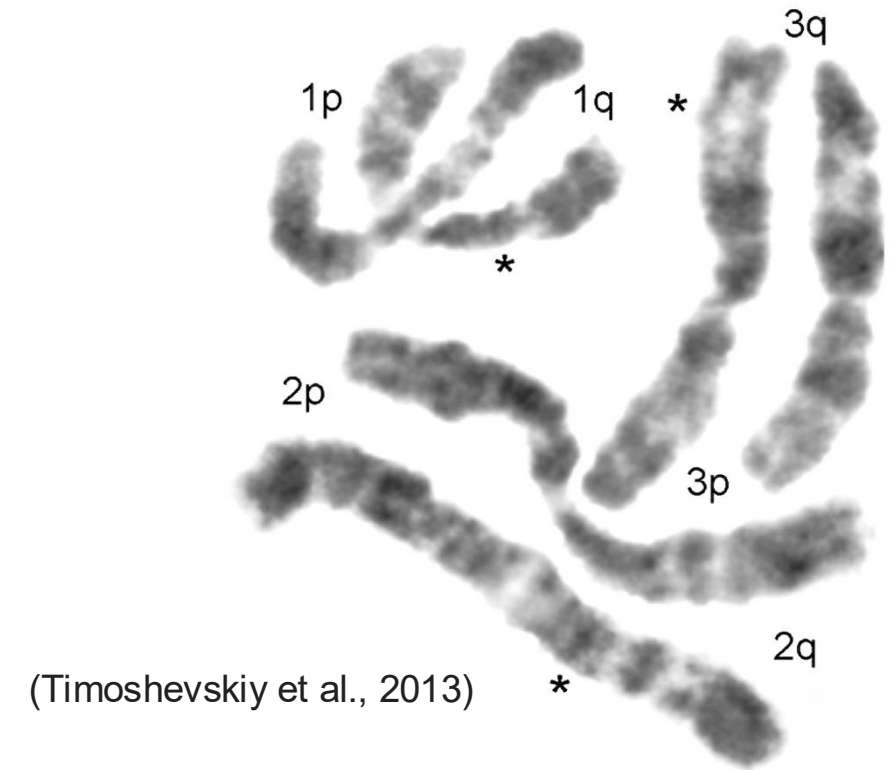
(Timoshevskiy et al., 2013)

An Integrated Linkage, Chromosome, and Genome Map for the Yellow Fever Mosquito *Aedes aegypti*

Vladimir A. Timoshevskiy, David W. Severson, Becky S. deBruyn, William C. Black, Igor V. Sharakhov, Maria V. Sharakhova ✉

Published: February 14, 2013 • https://doi.org/10.1371/journal.pntd.0002052