



**An R package for clustering and
visualizing ancient dna signatures**

By Hussein Al-Asadi & Kushal Dey

MapDamage

- MapDamage
- Show mapdamage results for anna gosling's data.
- 6 controls, 6 ancients, and 6 moderns

- C->T profiles of all 6 for each ancient, modern, and control

Problem with MapDamage

- Samples may have different levels of damage than others. Some are younger than others, different preservation conditions.
- Hard to visualize everything in one plot.
- Hard to some comparative analysis. Between sample comparison
- We only C->T patterns because that is dom... could be missing on other patterns

High-level intro to aRchiac

- Gets signatures from BAM files. Mention the flexibility here.
- Clusters and visualize samples based on signatures.
 - Allows you to visualize everything in one plots.
 - Allows you to do a between sample comparison.

- 1. Get signatures from data
- Show the signature.

Clustering and visualization

- Mixed membership. 50% modern and 50% ancient (hopefully). A contaminated sample could be 80% modern...
- First mention STRUCTURE. A person is like 80% African and 20% Europe.

Model

- Multinomial.
- What is ω , and what is θ ?
- ω is like admixture proportions and θ is like the allele frequency. Signature frequency.

Data 1

- Anna Gosling
- Anna Gosling + 1000g
- There is contamination

Scaling

Data 2

- Sards + Sherpa
- Sards + Sherpa + moderns.
- There is contamination

Data 3

- John Lindo.
- T->AAA. New signature. Put the plot for 1000g

Data 4

- IO + RISE + GOSLING + MODERN
- Batch effect or substructure. Be careful with interpretation

- C->T is most important. Filter out C->T and red position
- Show PCA plot.

- Fix two clusters, one representative of ancient and one of moderns.
- Get an idea of contamination.
- Show them plot for Anna Gosling's and Sard's data.