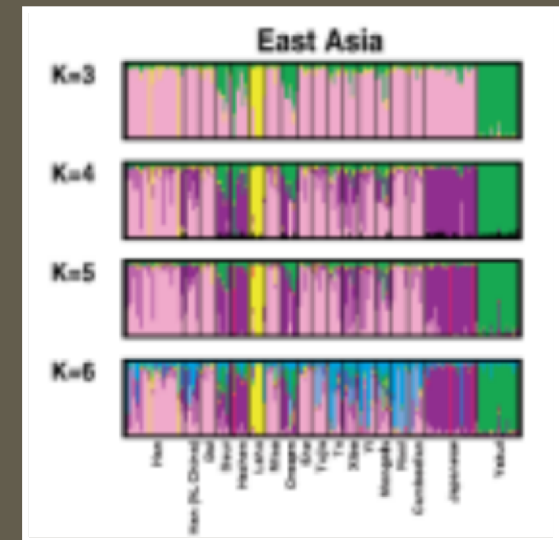
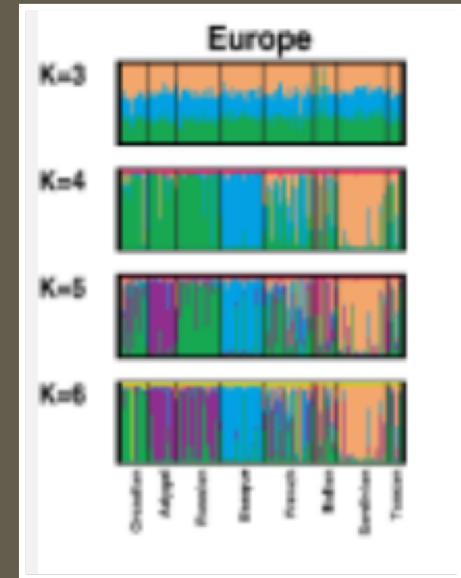


# REIMPLEMENTING STRUCTURE ALGORITHM WITH LINKED LOCI

Michael Li

## PROJECT GOALS

- **STRUCTURE** is an algorithm for multi locus genotype data that assigns individuals to distinct populations. It can be used further for studying hybrid zones, identifying admixed individuals, and estimating population allele frequencies in the case where there is a significant number of migrant or admixed individuals.



# PROJECT PLAN

- Find original paper and developments of STRUCUTRE
- Understand the basic algorithm and plan/design code
- Acquire original data used
- Write code to parse data
- Implement most basic version
- Test on original data and more if able to find
- Improve basic algorithm by adding Linked Loci extension

## IDEAS/METHODS USED IN STRUCTURE

- STRUCUTRE uses the idea of MCMC convergence with Gibbs Sampling.
- Sampling is from the Dirichlet Distribution
- Considering the possibility of linkage between loci and the correlations with admixed populations improves the model:
  - Improved population inference
  - Detection of admixture events farther in past
  - More accurate estimates of uncertainty

# DATA

- Was able to find the data from the original STRUCTURE paper and more from later studies, but the data was from microsatellites.

```

LABEL=1, POPDATA=1, NUMLOCI=7, MISSING=-9, ONEROWPERIND=0. Population Codes: 1=C, 2=M, 3=N, 4=Y
1 2 1 1 6 1 2 2 5
1 2 4 2 6 1 2 4 5
10 2 4 2 1 1 2 2 5
10 2 4 3 6 1 2 10 8
1219 2 4 3 6 1 3 2 1
1219 2 13 5 6 2 -9 2 4
1223 2 7 3 4 1 2 1 1
1223 2 12 3 6 3 2 2 6
1225 2 -9 2 1 1 2 2 5
1225 2 2 2 2 2 2 2 5

```

- Found another paper on STRUCUTRE using allele data that is equivalent to biallelic SNPs

[illegible]