

Basic population genomics

Dr Johanna Rhodes

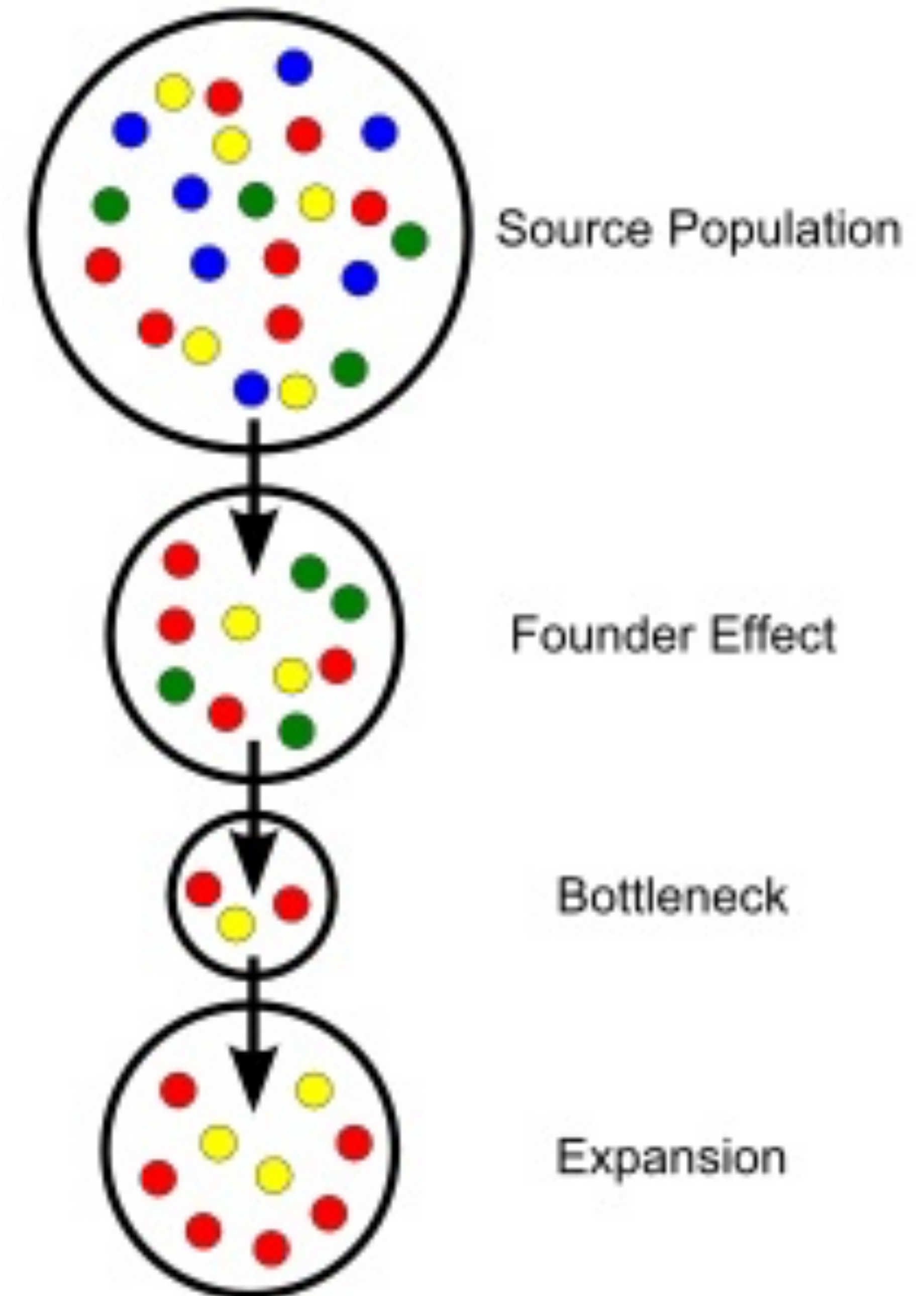
Learning outcomes

- **Basic population genetics**
 - theory
 - usage in fungal research
 - using SNP data

Population genetics

Using genomic data

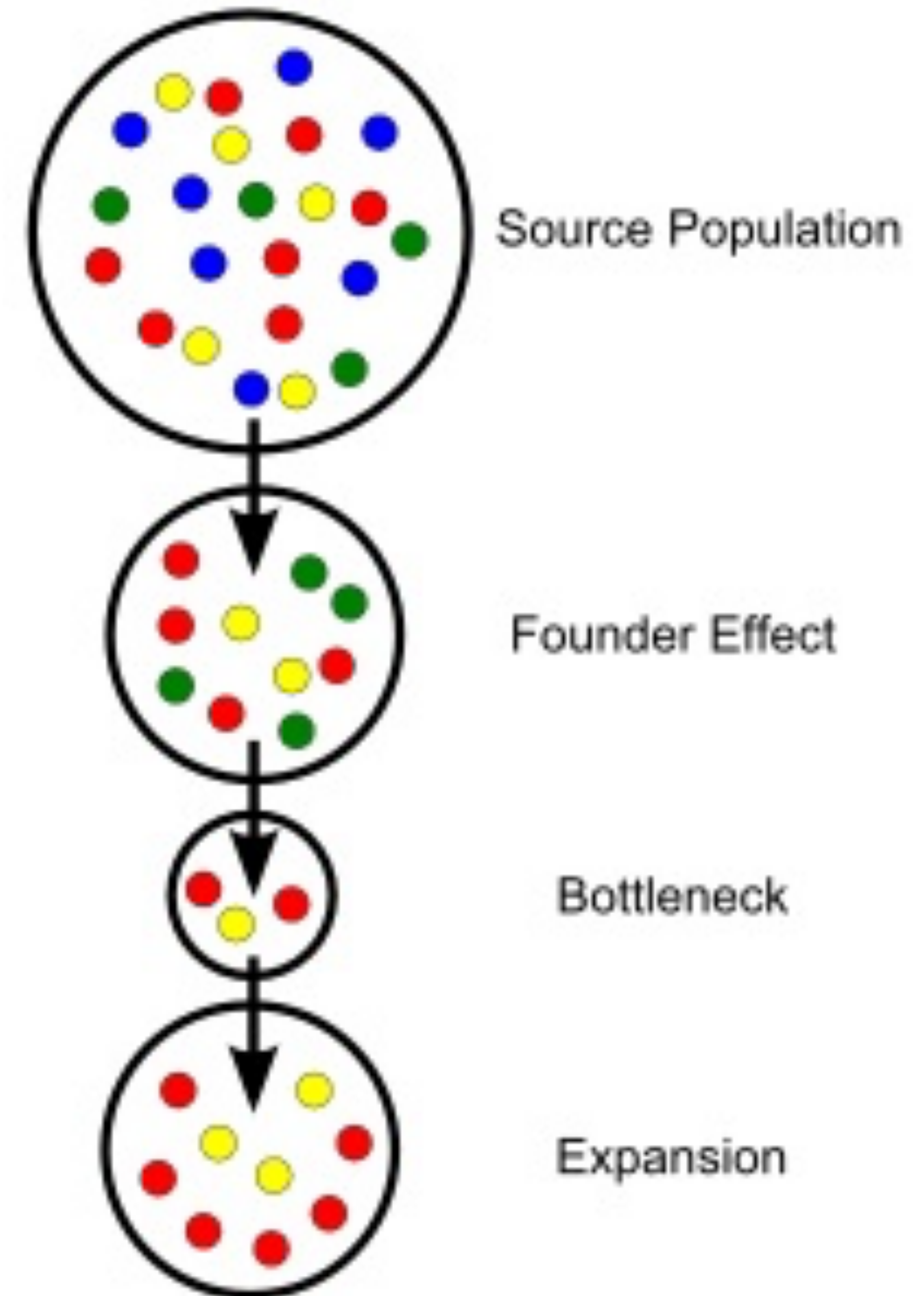
- Study of genetic variation within populations
- changes in frequencies of genes and alleles over space and time
- many genes in a population will be polymorphic - they will occur in a number of different forms



Population genetics

Using genomic data

- Principles in fungal populations include:
 - mutation
 - gene flow
 - transfer of genes among populations
 - genetic drift
 - natural selection



Population genetics

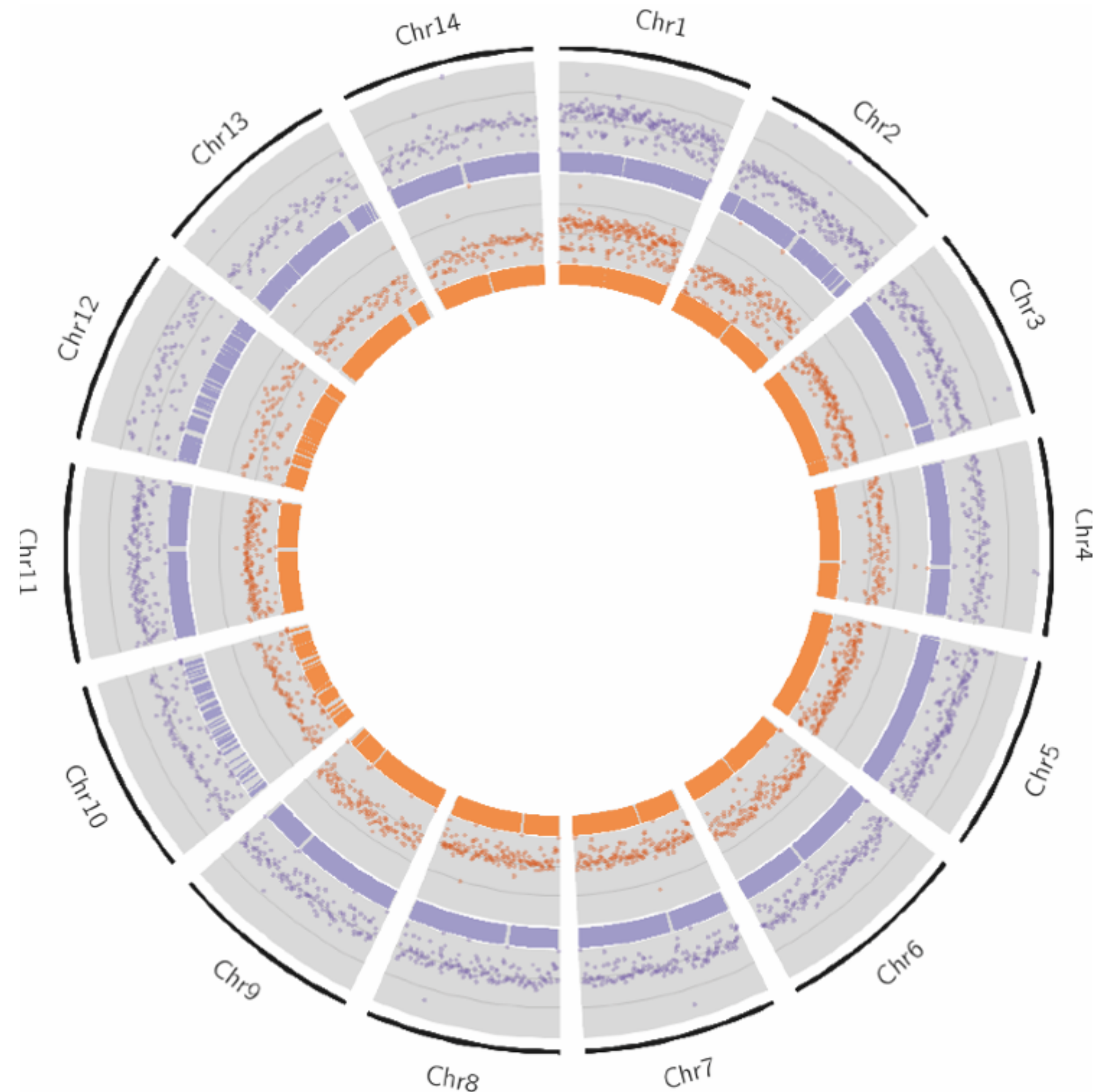
In fungi

- How do fungal pathogens evolve, spread, and interact with hosts and environments
- Tracking genetic diversity and evolution
 - hotspots for mutation and recombination
- Understand transmission patterns and epidemiological surveillance
- Host-pathogen interaction and adaptation
- Population structure and gene flow
 - are they highly clonal, or is there significant genetic exchange between populations?

Population genetics

In fungi

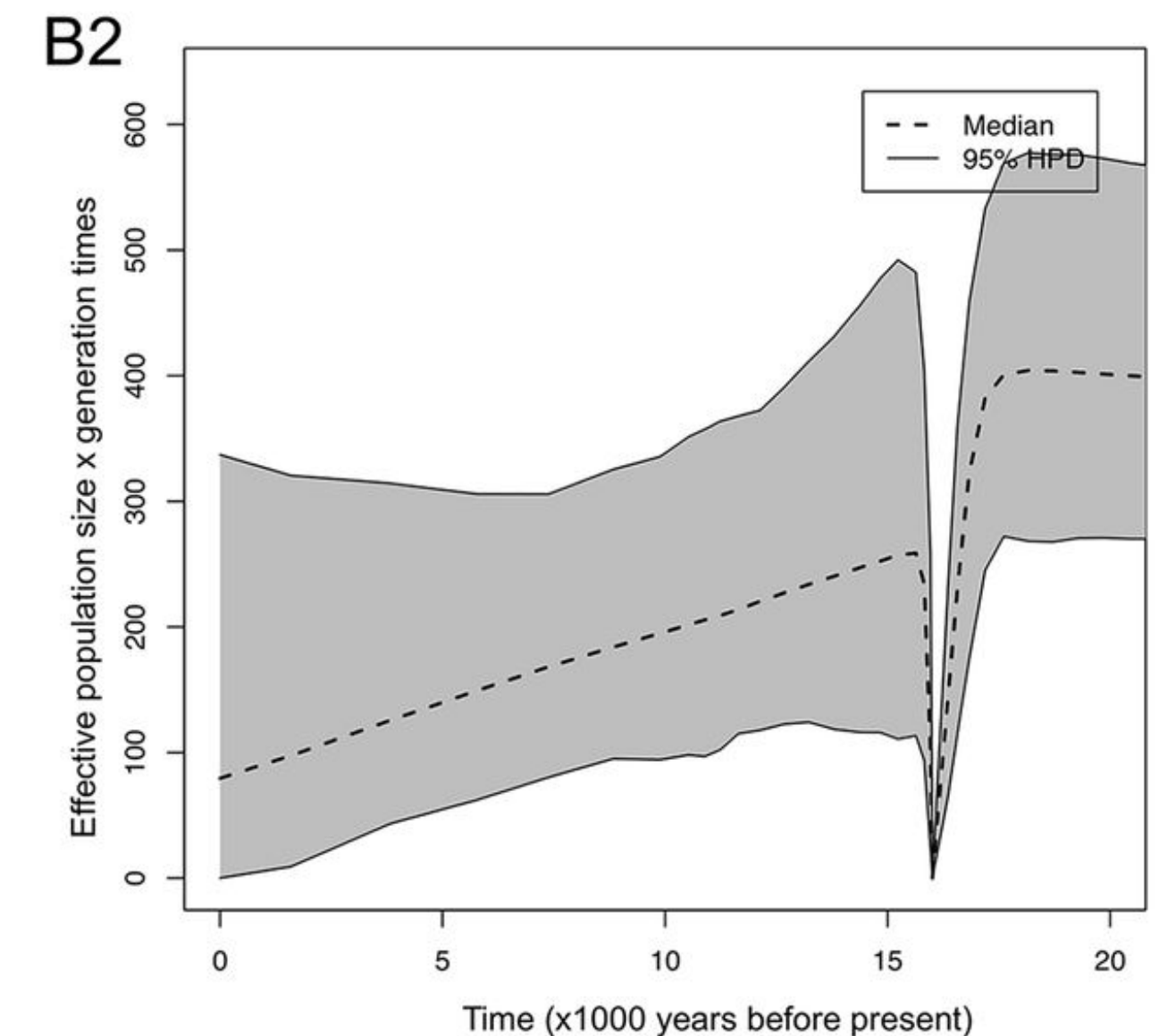
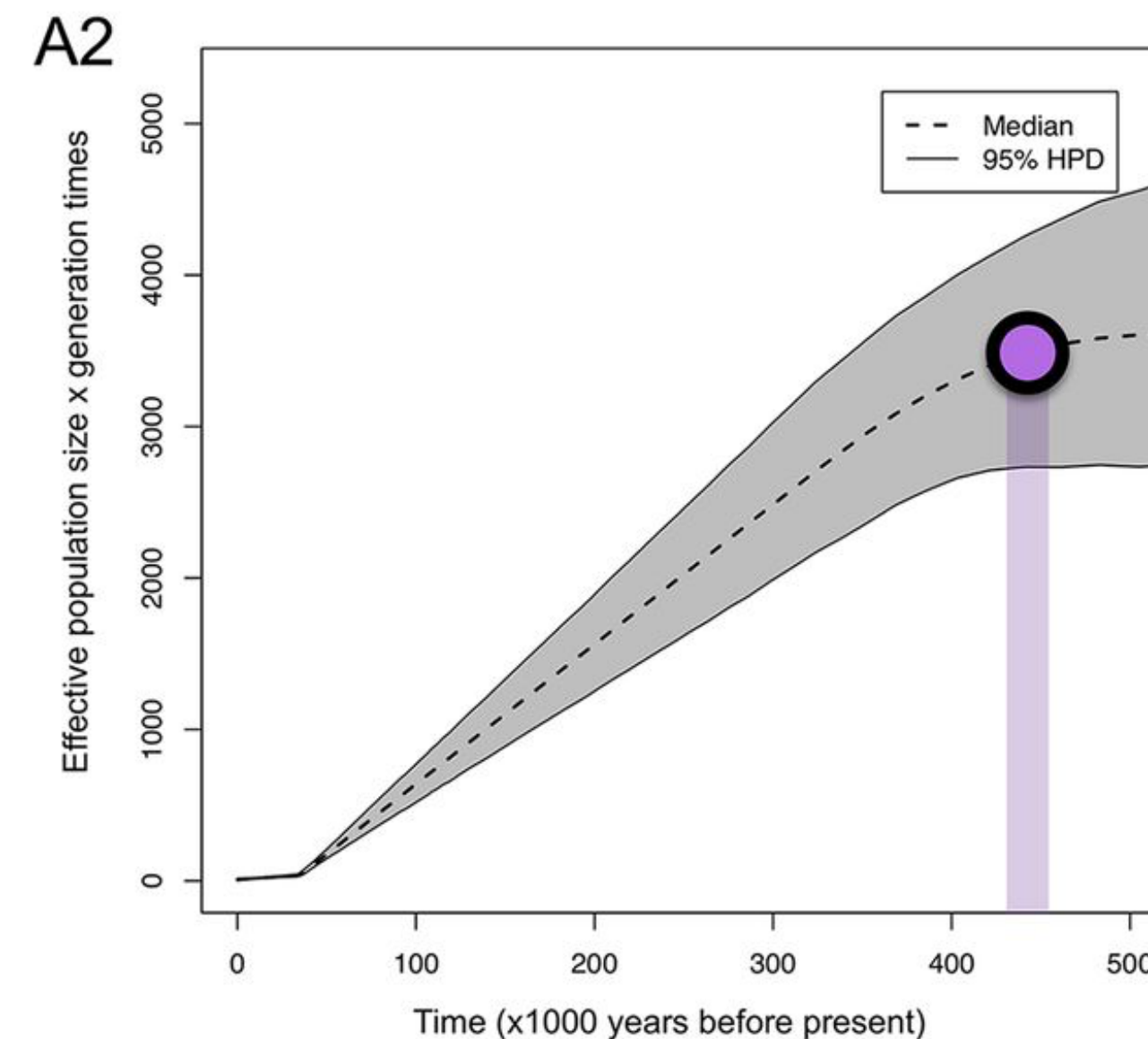
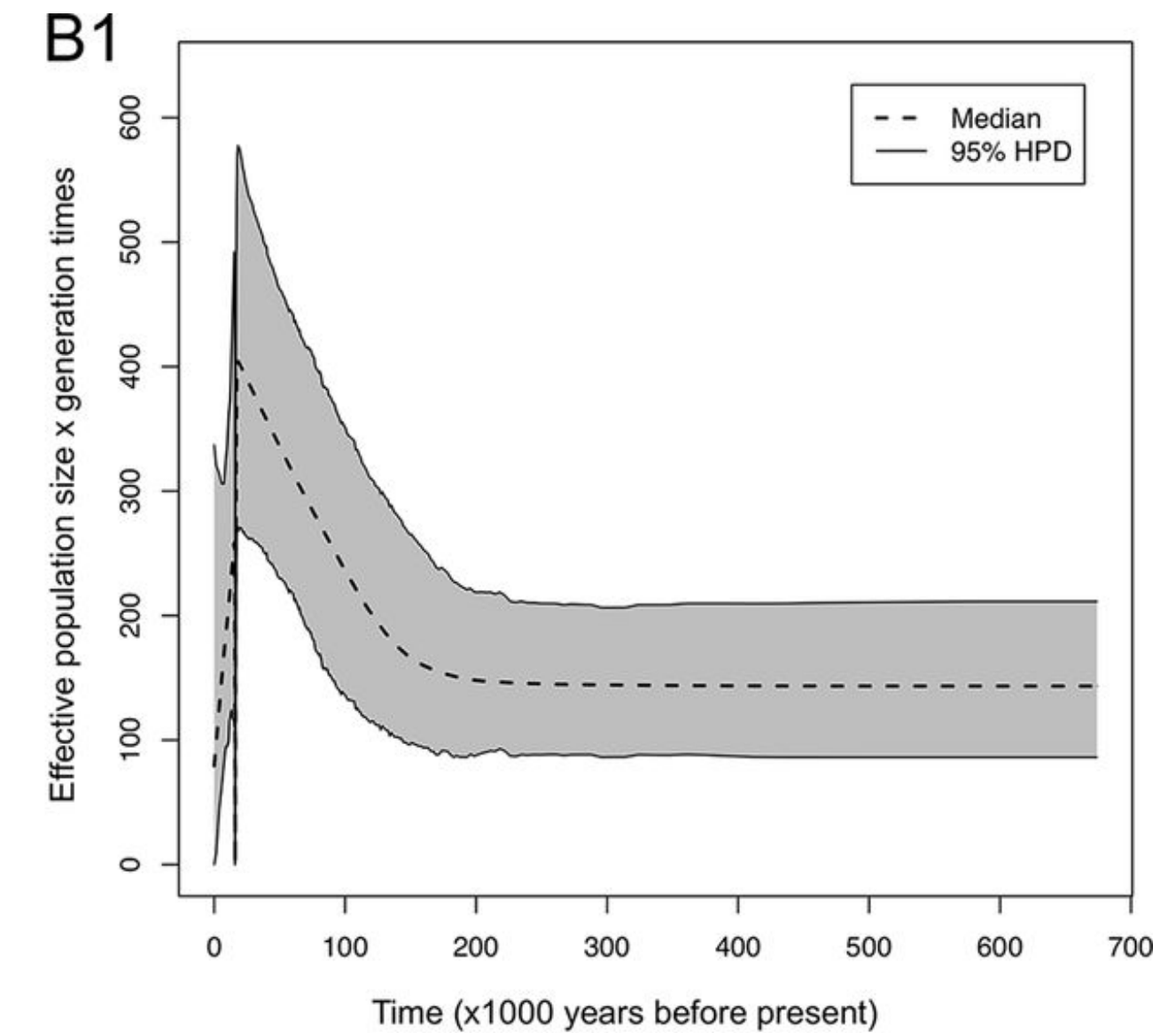
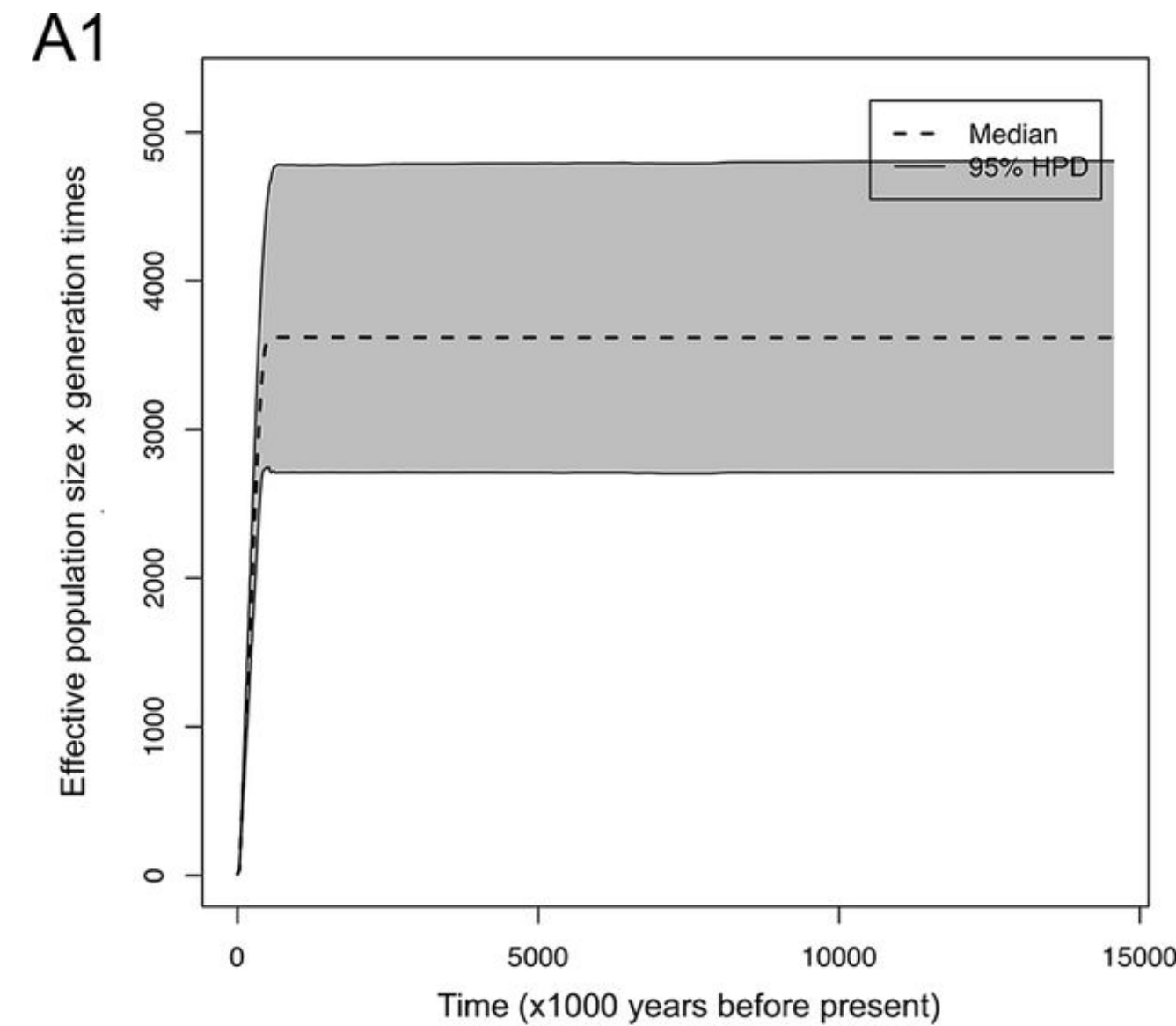
- Population genomics in *Cryptococcus neoformans* to understand in-host microevolution:
 - Assess whether patients are re-infected or relapse of original infection
 - Hypermutation as a rapid adaptation tool for drug resistance



Population genomics

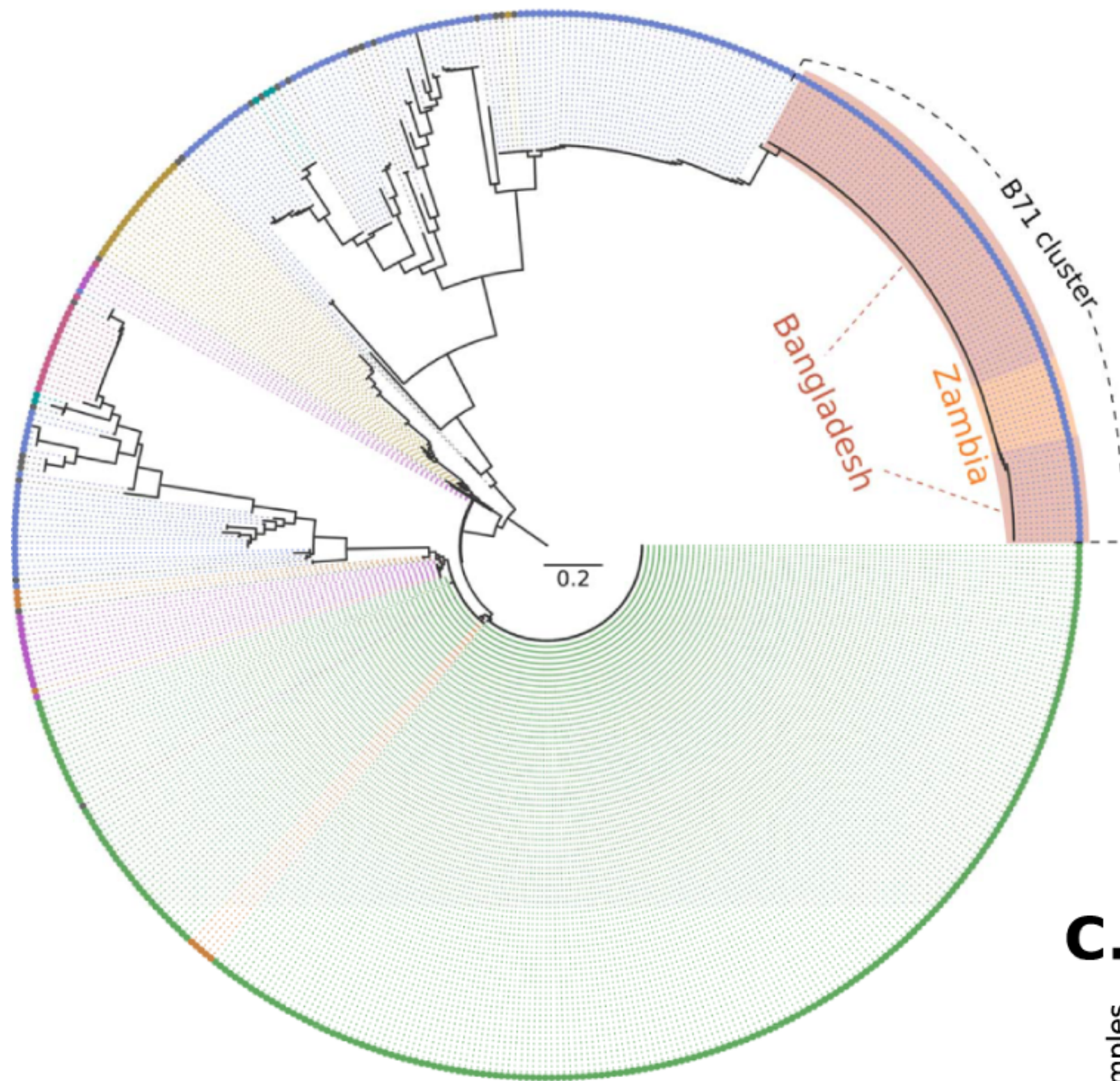
In fungi

- Investigate the demographic history of natural populations infecting various hosts e.g. *Pneumocystis*
- Detect recombination and variations shaping the global population structure
- Evolutionary rates and molecular clock estimates
 - host shift of *P. jirovecii*

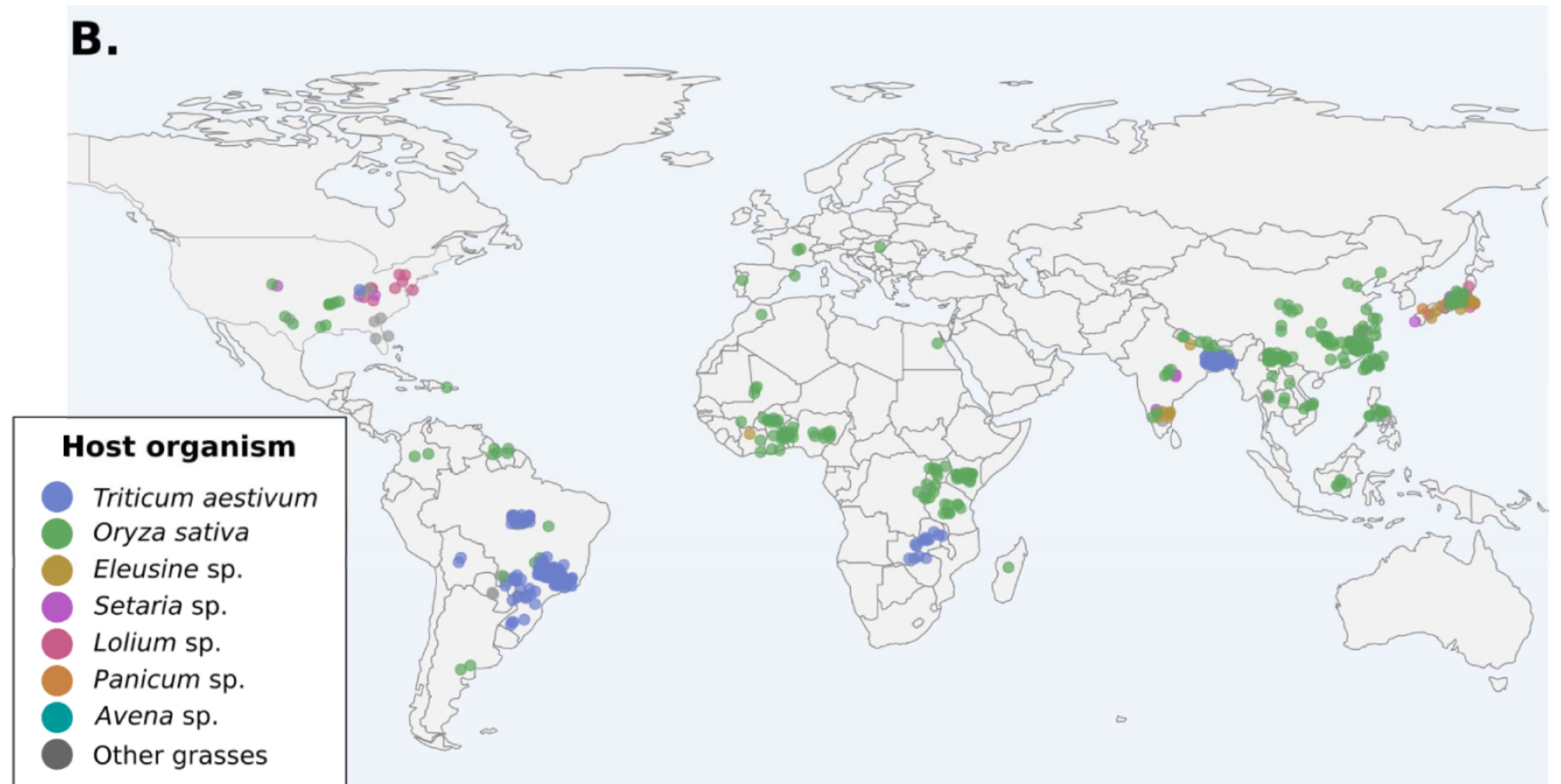


Population genomics

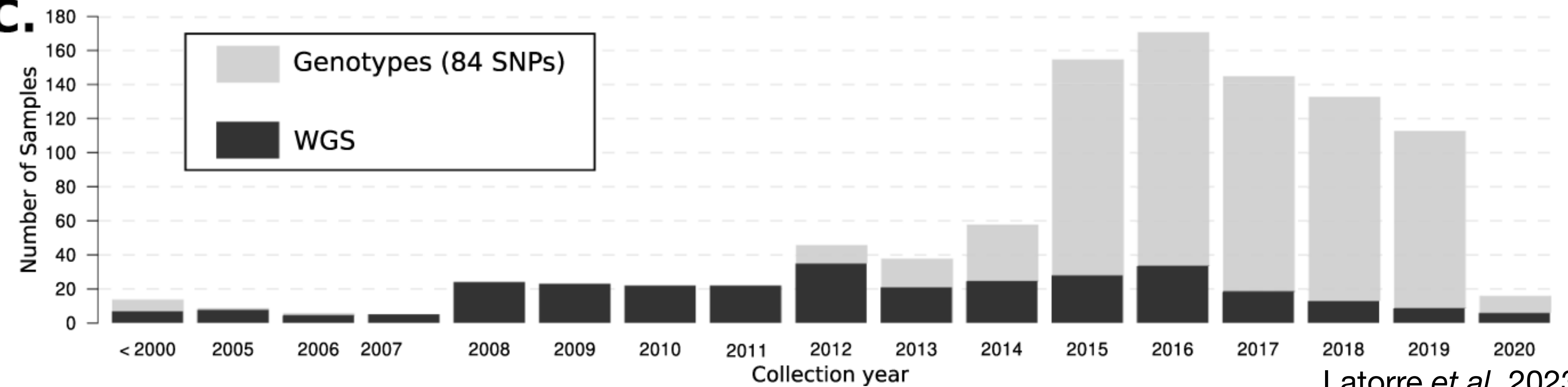
A.



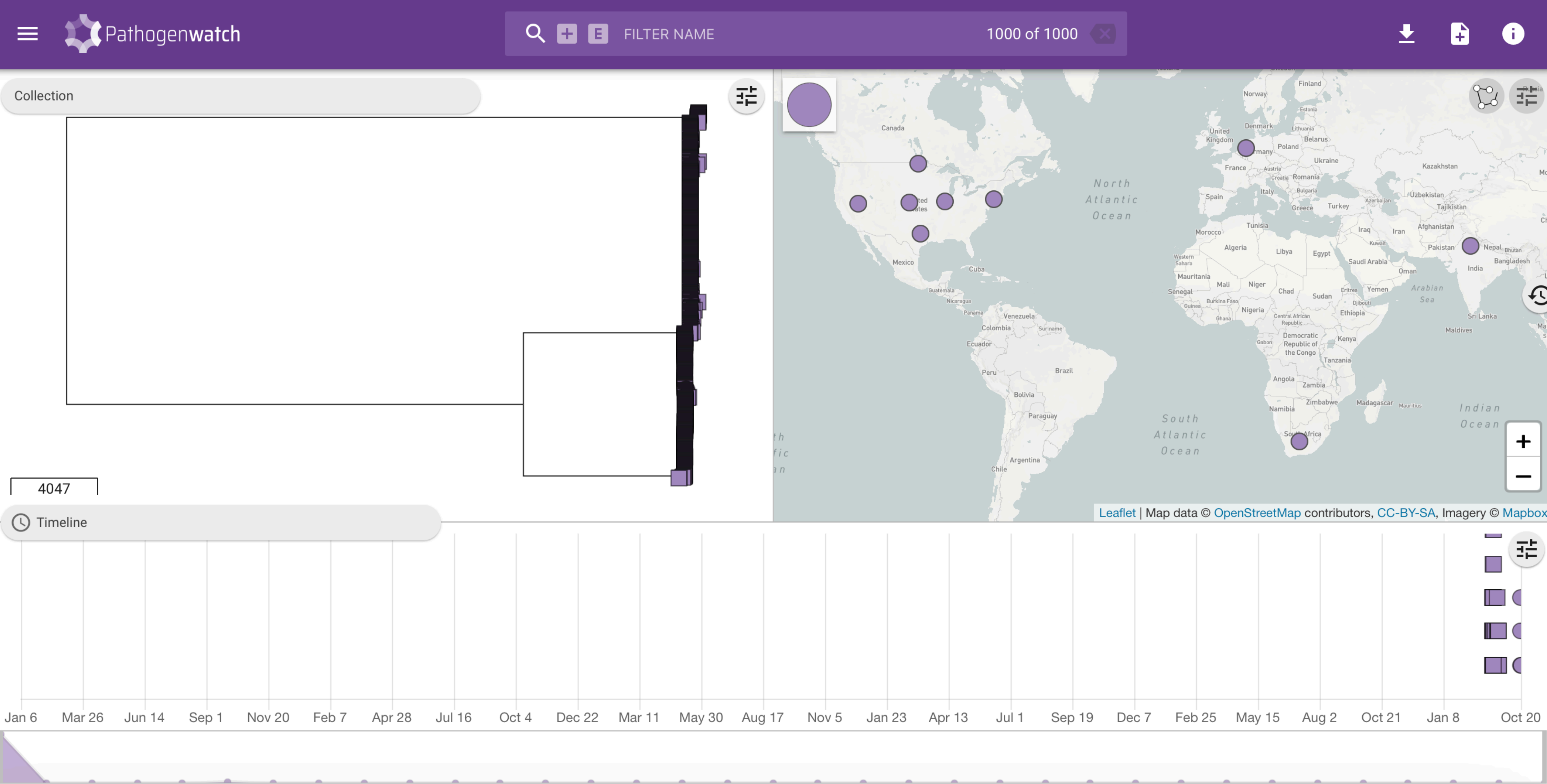
B.



C.



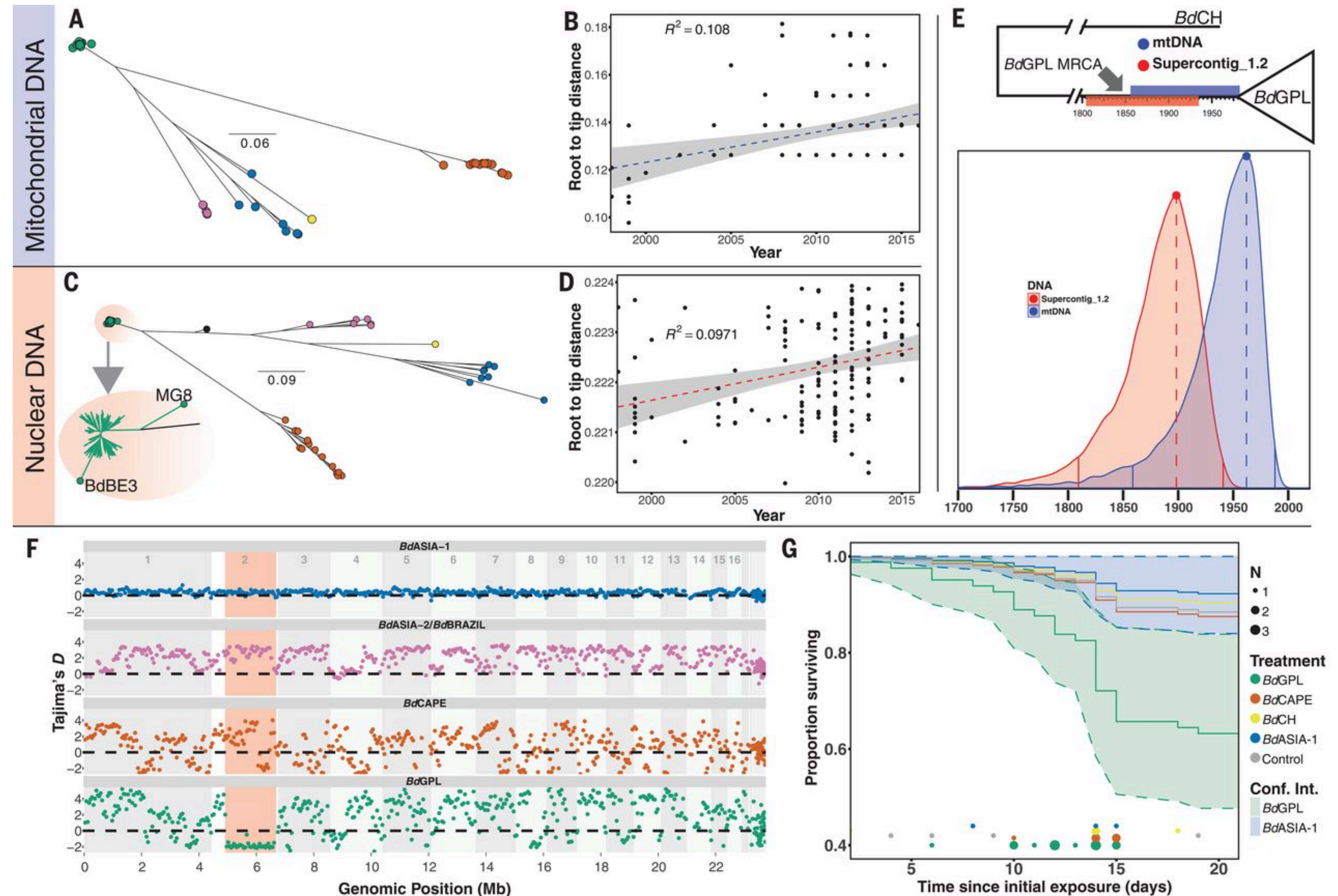
Population genomics



Population genomics

In fungi

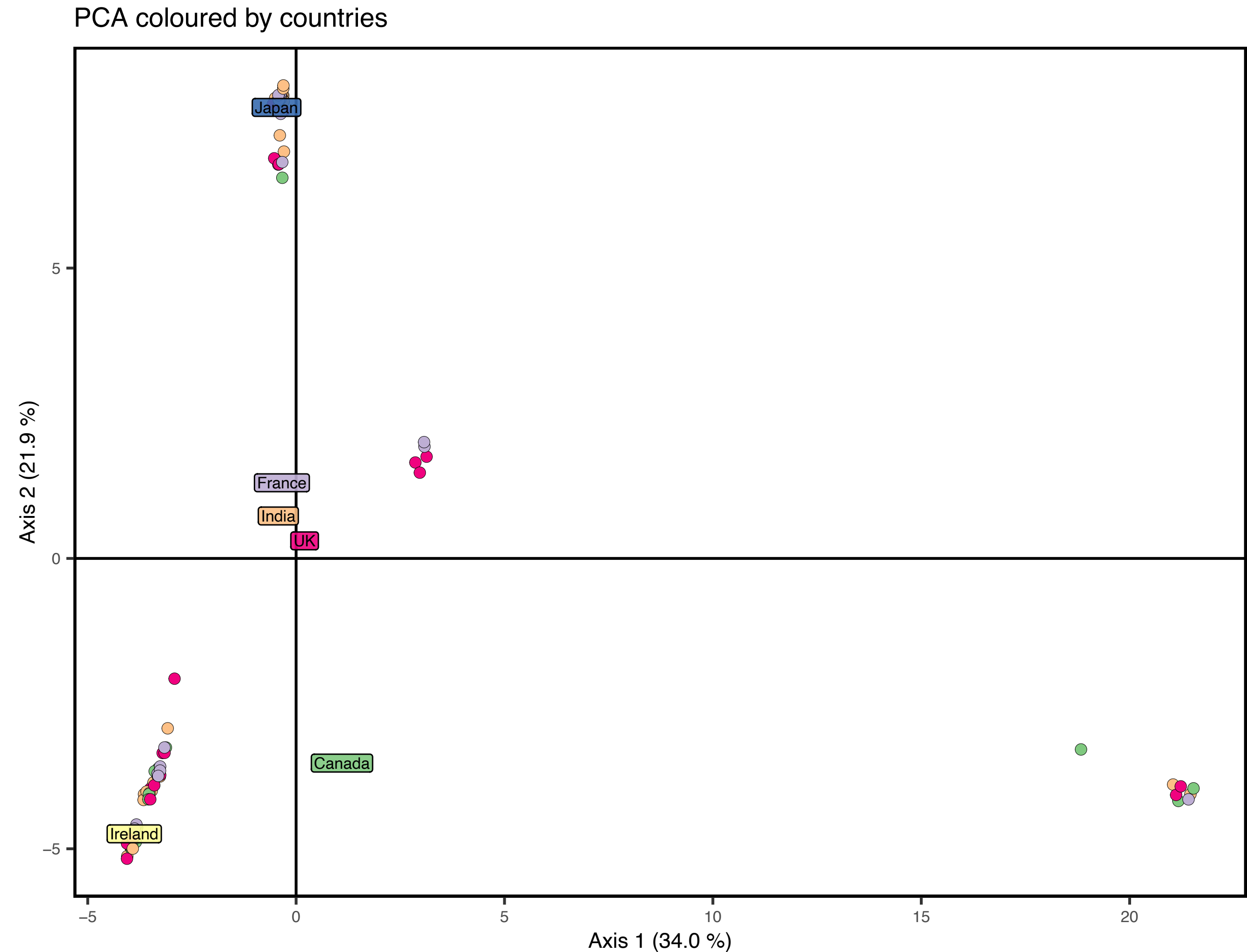
- Identify ancestral populations seeding global panzootic of *Bd*.
 - ‘Flat’ profile of *Bd*ASIA-1 indicates mutation-drift equilibrium reflects pathogen endemism to the region
- Population statistics (such as Tajima’s D) can also be used to identify recombination hot- and coldspots



What is a population?

Using wgSNPs

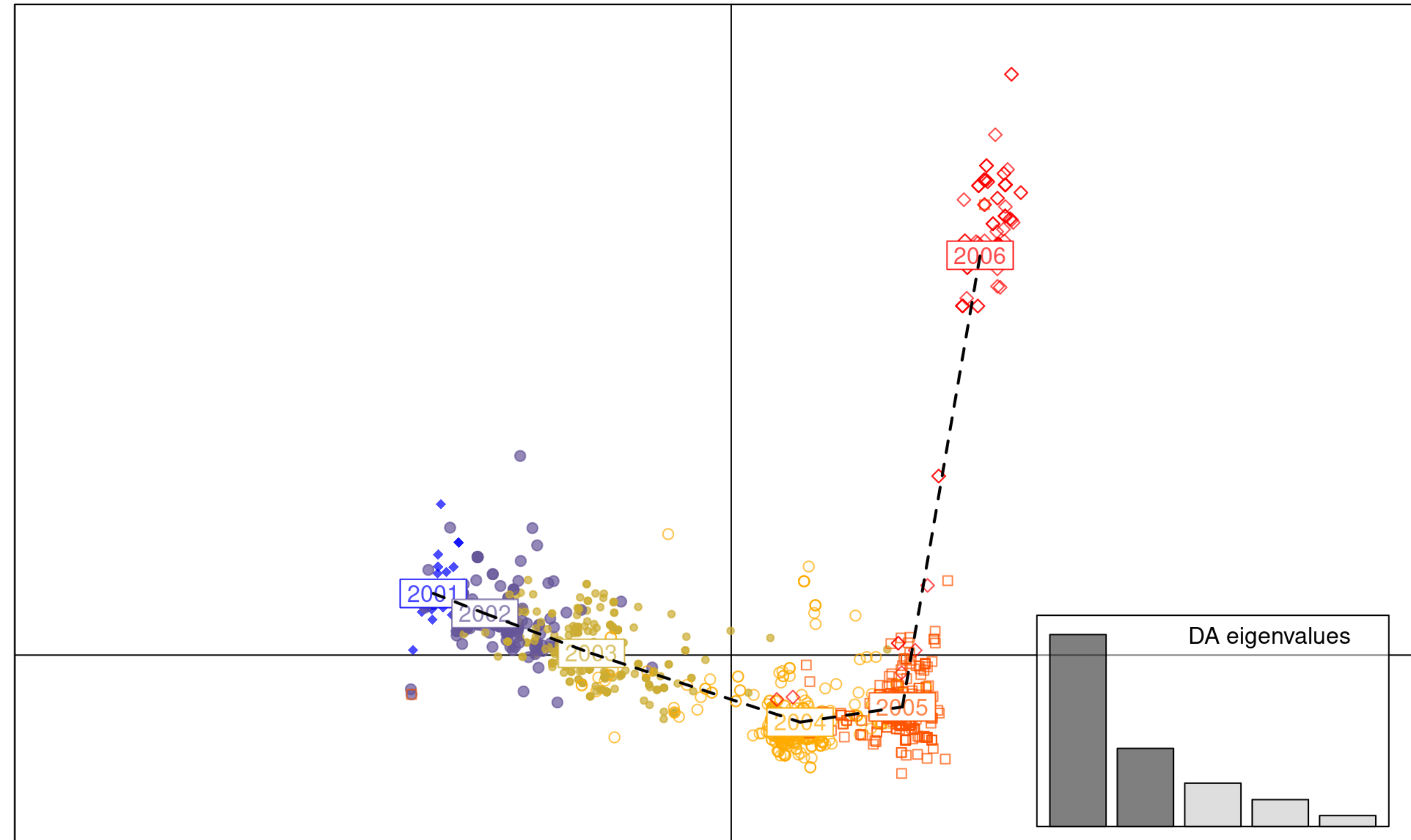
- Principal component analysis (PCA)
- reduces data to its essential features - principal components
- principal components explain the variance of the variables



What is a population?

Using wgSNPs

- Discriminant analysis of principal components (DAPC)
- PCA fails to discriminate groups accurately; DAPC maximises the separation between groups while minimising variation within groups



Principal component analysis (PCA)

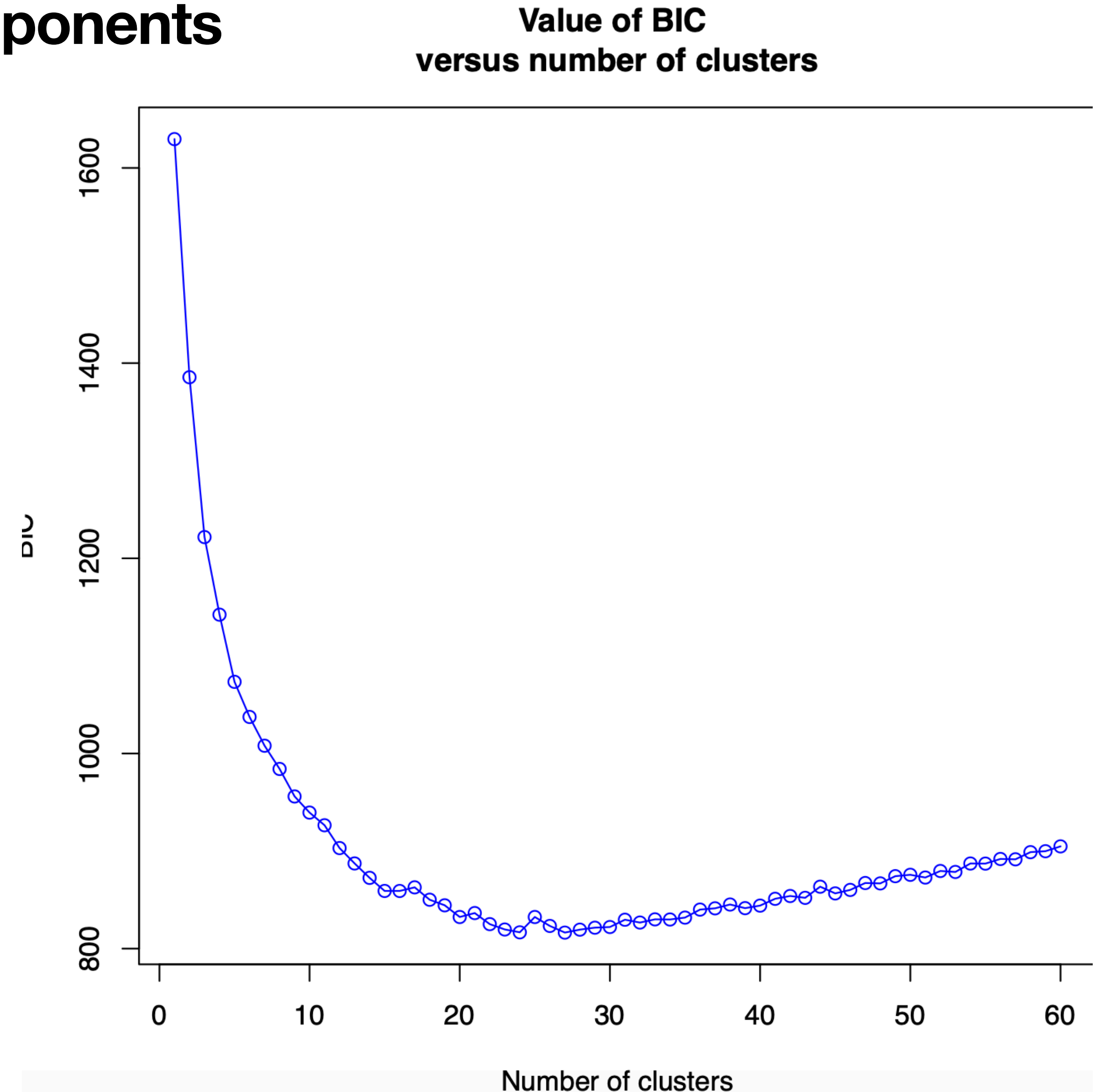
In practice

- In R/RStudio
- Need:
 - multiFASTA file
 - installation of packages: ape, adegenet, viridis
 - txt file with all isolates plus group assignment

Discriminant Analysis of Principal Components (DAPC)

In practice

- Same inputs as for PCA. How does it look?
- Do other groupings explain the variation better?
 - finding the true 'k' value (hypothetically)
- 'find.clusters(x)' in the dapc package



Summary

And suggested reading

- Assess genetic diversity and understand population structure
 - WGS
 - GWAS
 - Phylogenetic and clustering analysis
- Also consider reading literature in plant fungal pathogens
 - Daniel Croll group
 - Bruce Macdonald