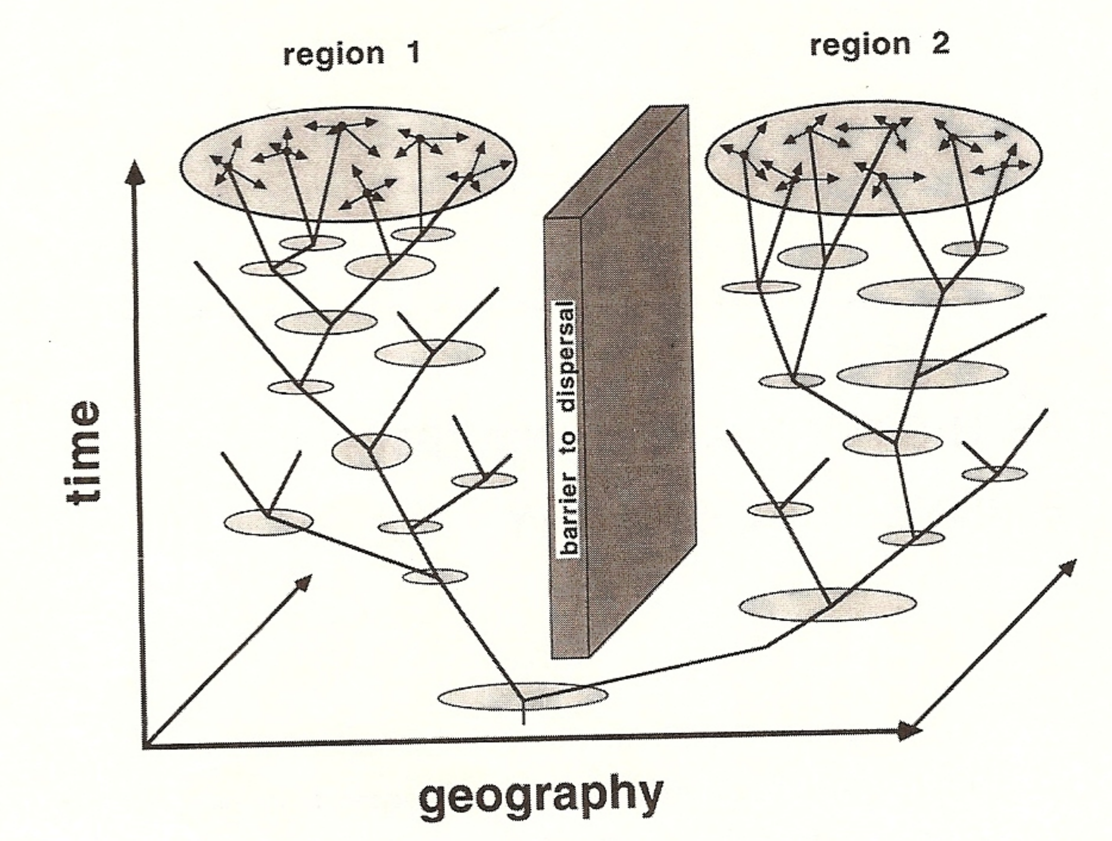
Phylogeography

What is Phylogeography?

- “A field of study concerned with the principles and processes governing the geographic distribution of genealogical lineages, especially those within and among closely related species.” (Avise 2000)
- Phylogenetics + Biogeography
- How is genetic variation distributed in time and space?
- What factors account for the spatial and temporal distribution of genetic lineages?

What is phylogeography?



- Gene genealogies of interest are mapped in space and time.
- Goal: understand the factors contributing to the formation of population genetic structure.

Avise 2000

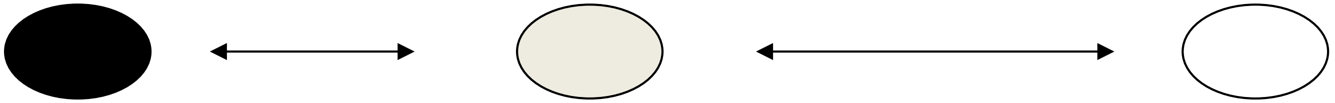
Why is phylogeography important?

- Geographic distribution of haplotypes can tell us something about migration
  - Geographic history
  - Dispersal routes
  - Taxonomy
- Deep versus shallow may indicate units of interest
  - Deep: Evolutionarily significant unit (ESU)
  - Shallow: Management significant unit (MSU)
- Multispecies concordance of discordance
  - Congruence: Important biogeographic barriers
  - Discordance: Biased gene flow

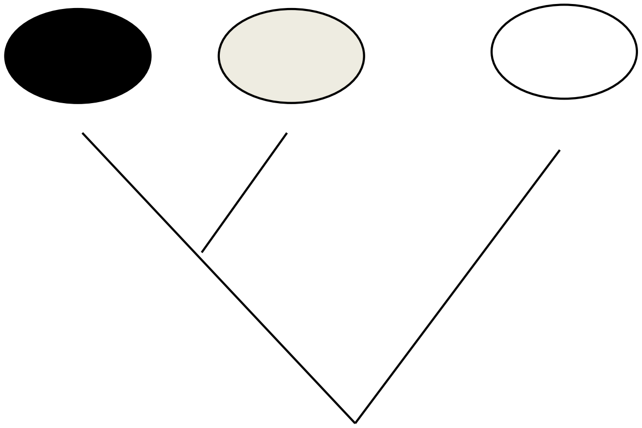
Determinants of genetic structure

Historical relationships and contemporary gene flow influence spatial patterns of genetic variation.

Similarity due to gene flow



Similarity due to historical relationships

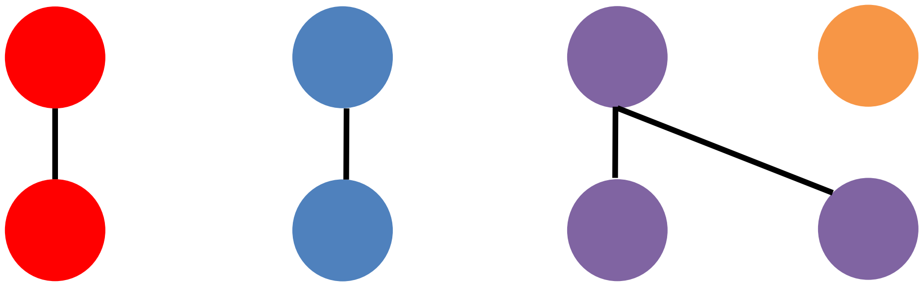


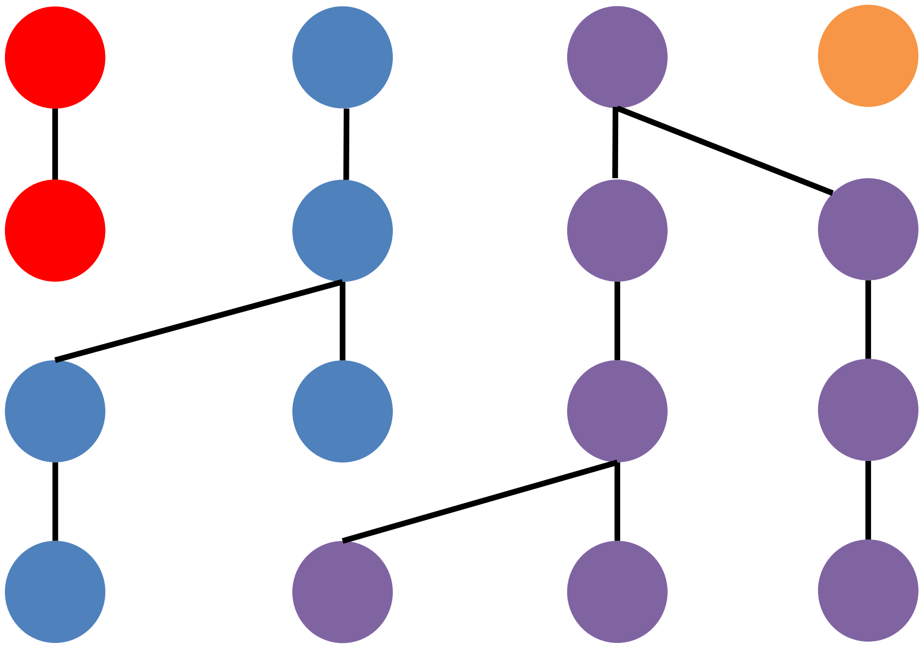
Determinants of genetic structure

- “**Phylogeographic methods** provide a means of examining the history of genetic exchange among populations, with the potential to distinguish biogeographic patterns of genetic variation caused by gene flow from those caused by common ancestry.” (Schaal, 1998)
- Tease apart *contemporary* forces of genetic exchange from *historical* relationships
  - Relies on population genetics and phylogenetics coupled with biogeographic data
  - Relies on gene genealogies

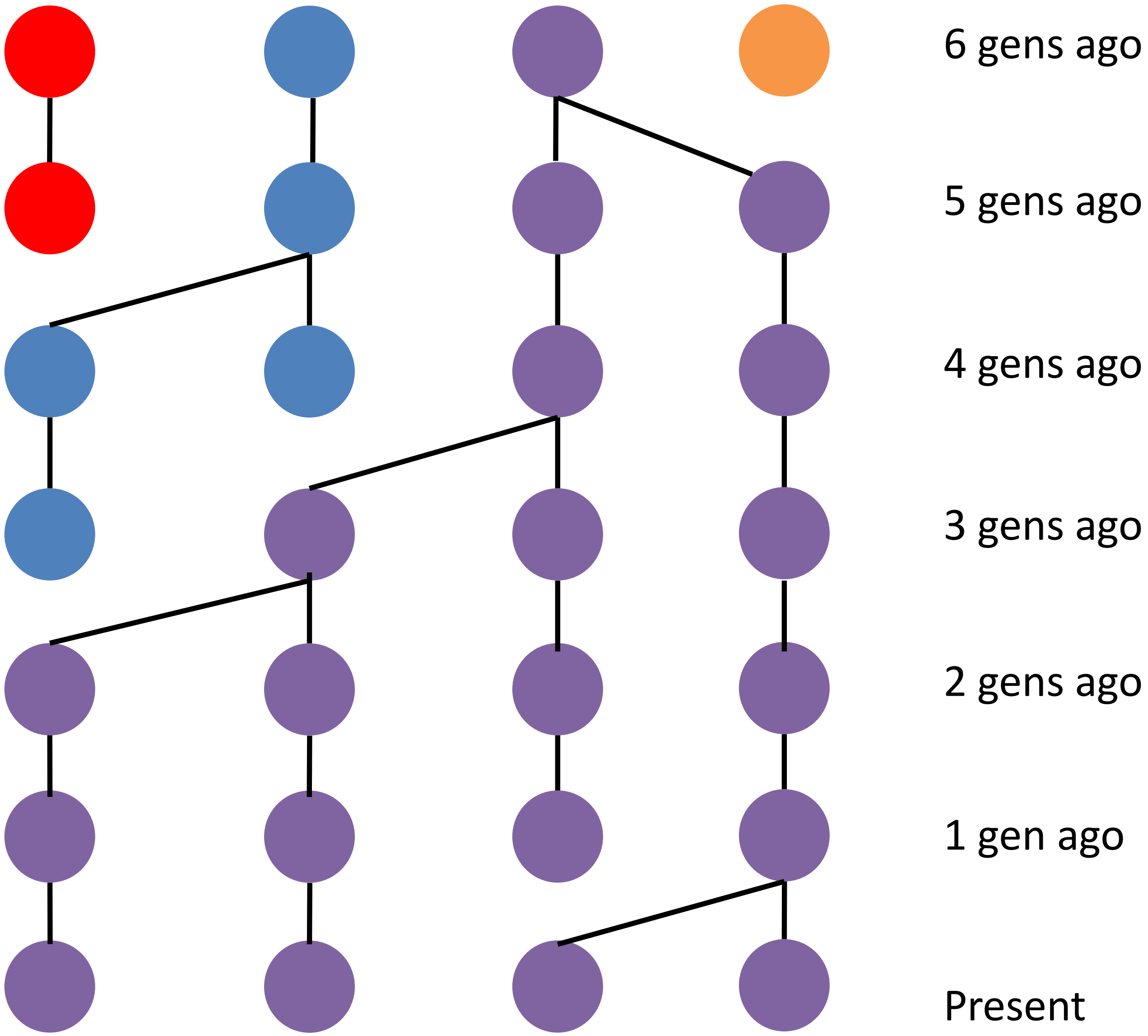
Genealogies

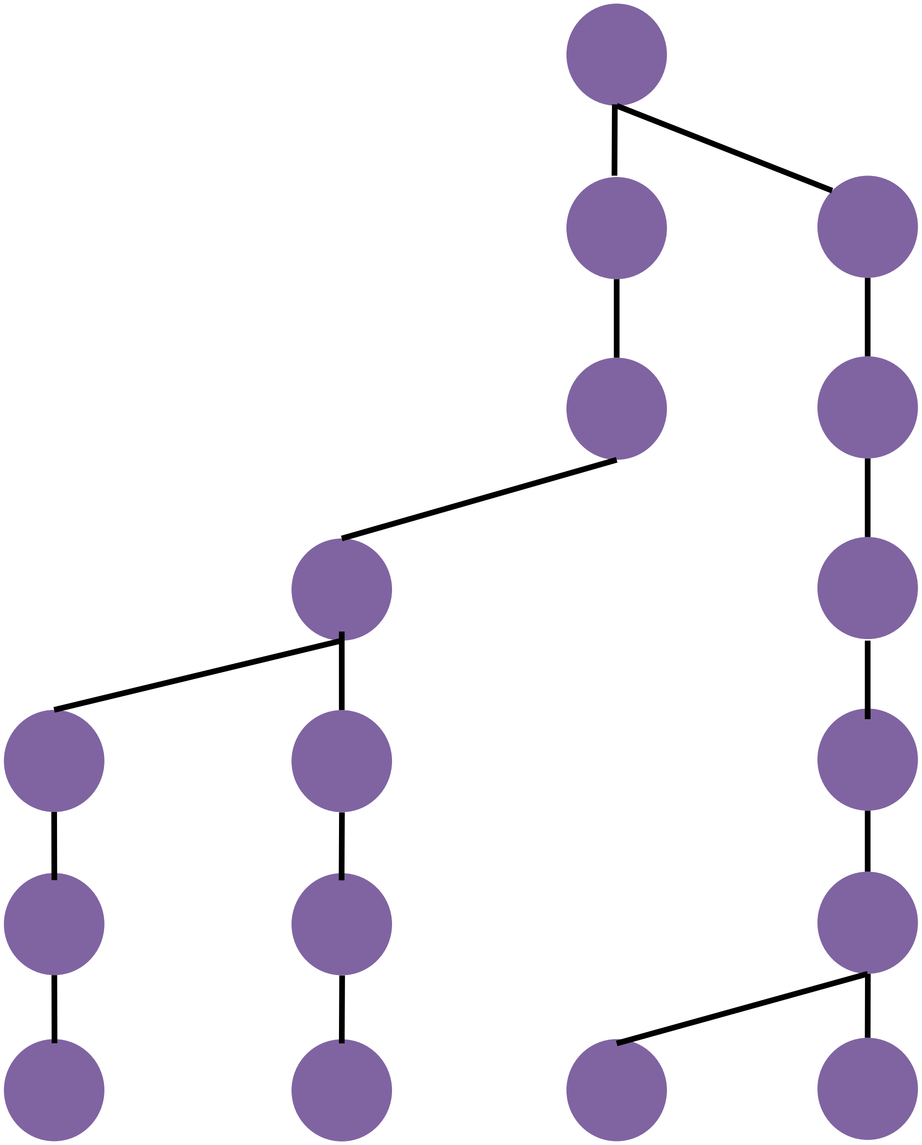
- Describe inheritance/relatedness relationships among alleles
- Similar to phylogenetic tree





MRCA = 6 gen ago





6 gens ago

5 gens ago

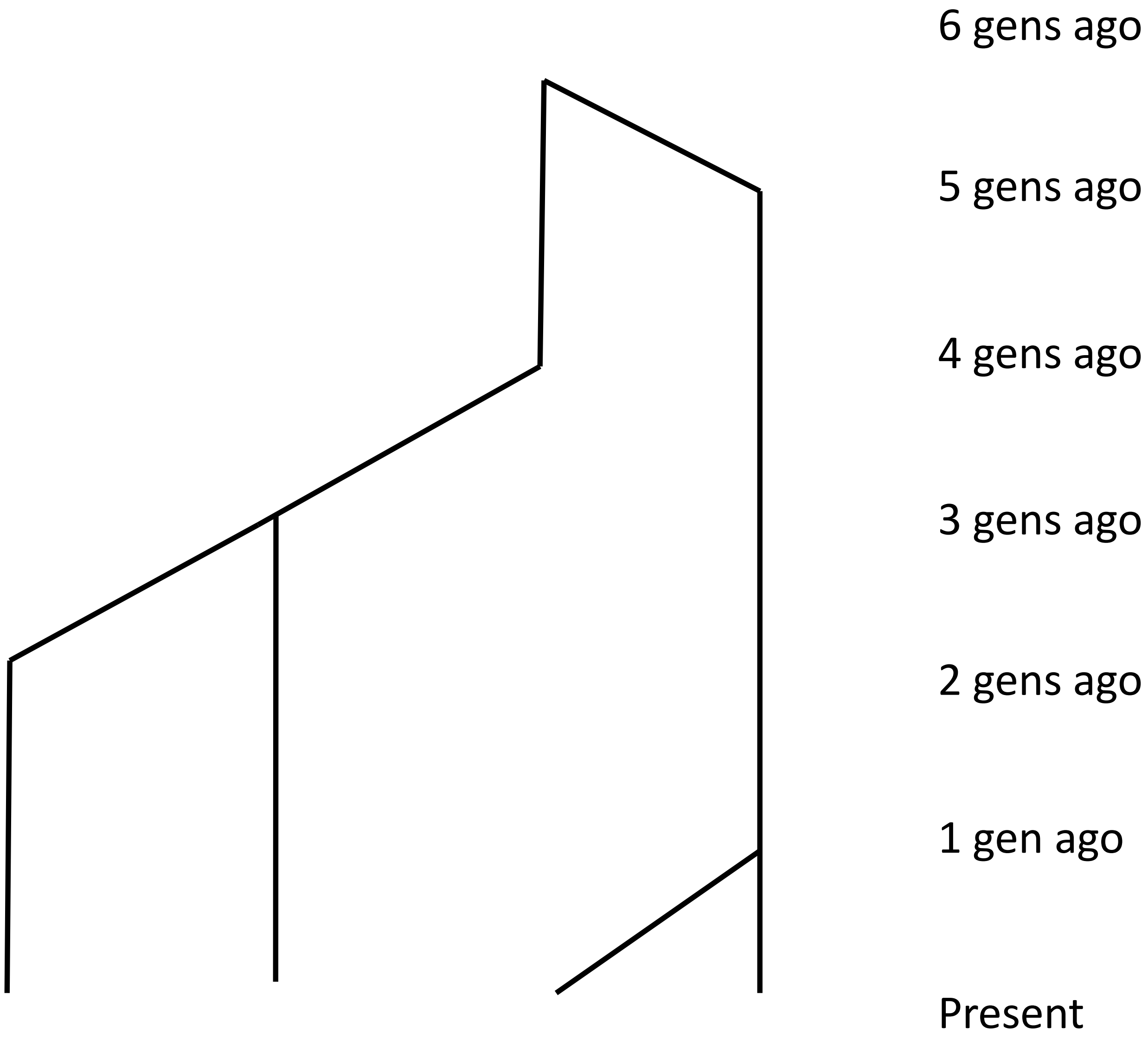
4 gens ago

3 gens ago

2 gens ago

1 gen ago

Present



6 gens ago

5 gens ago

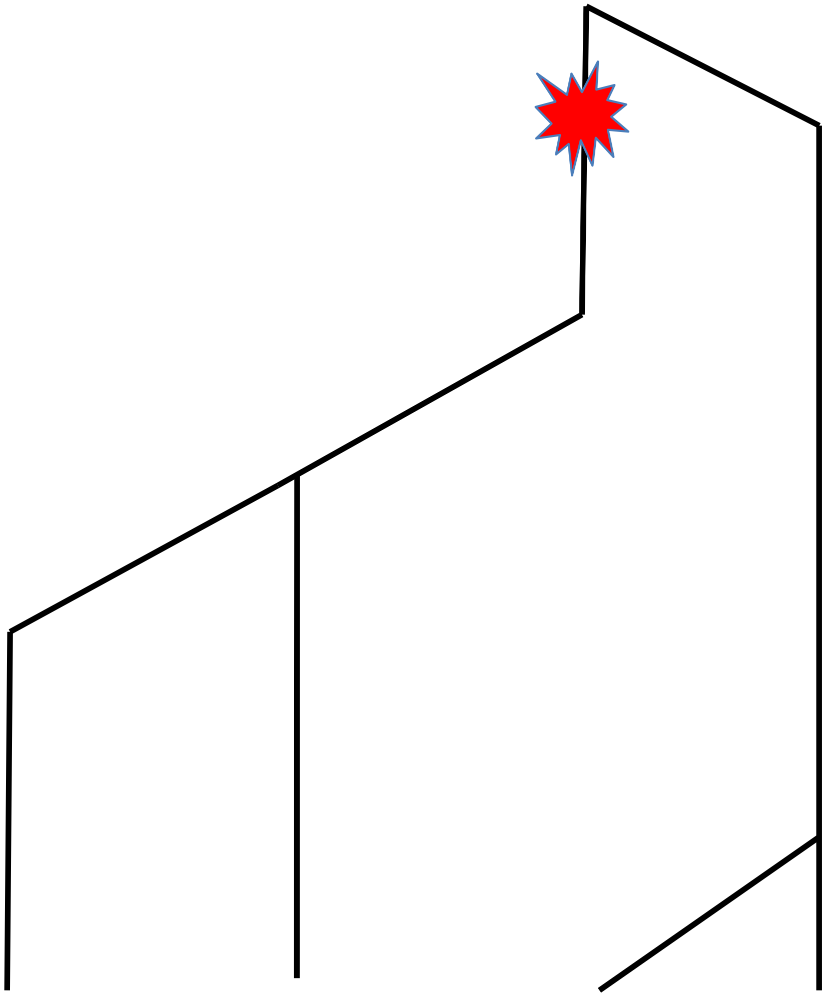
4 gens ago

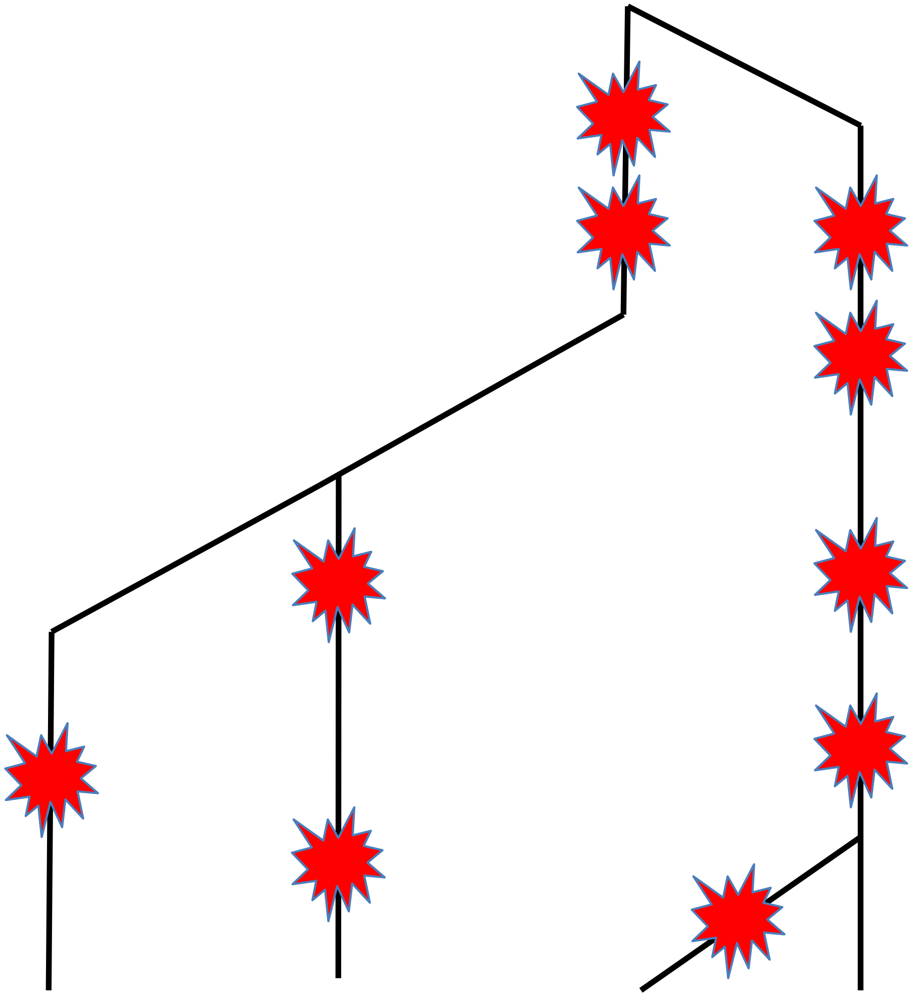
3 gens ago

2 gens ago

1 gen ago

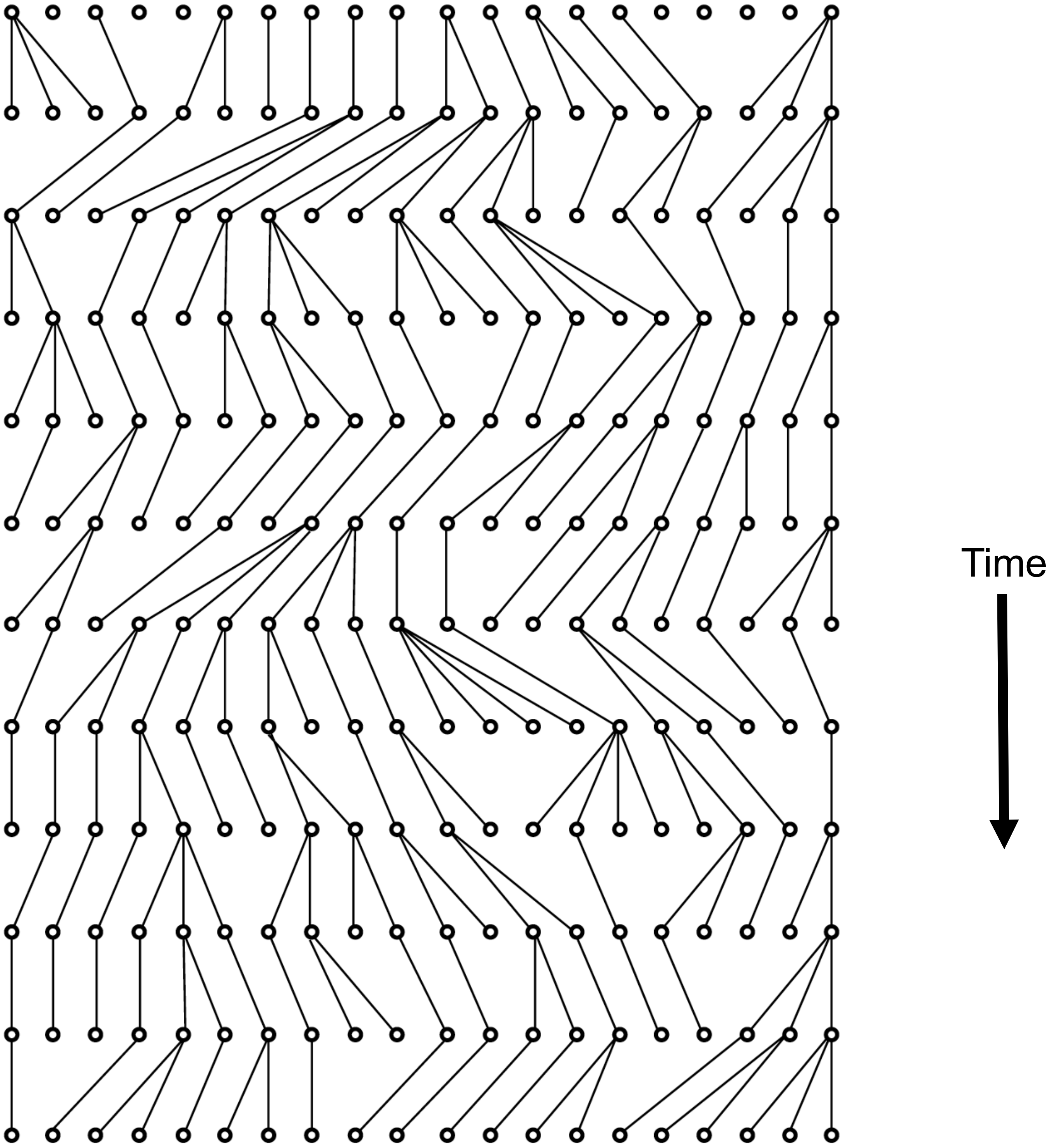
Present

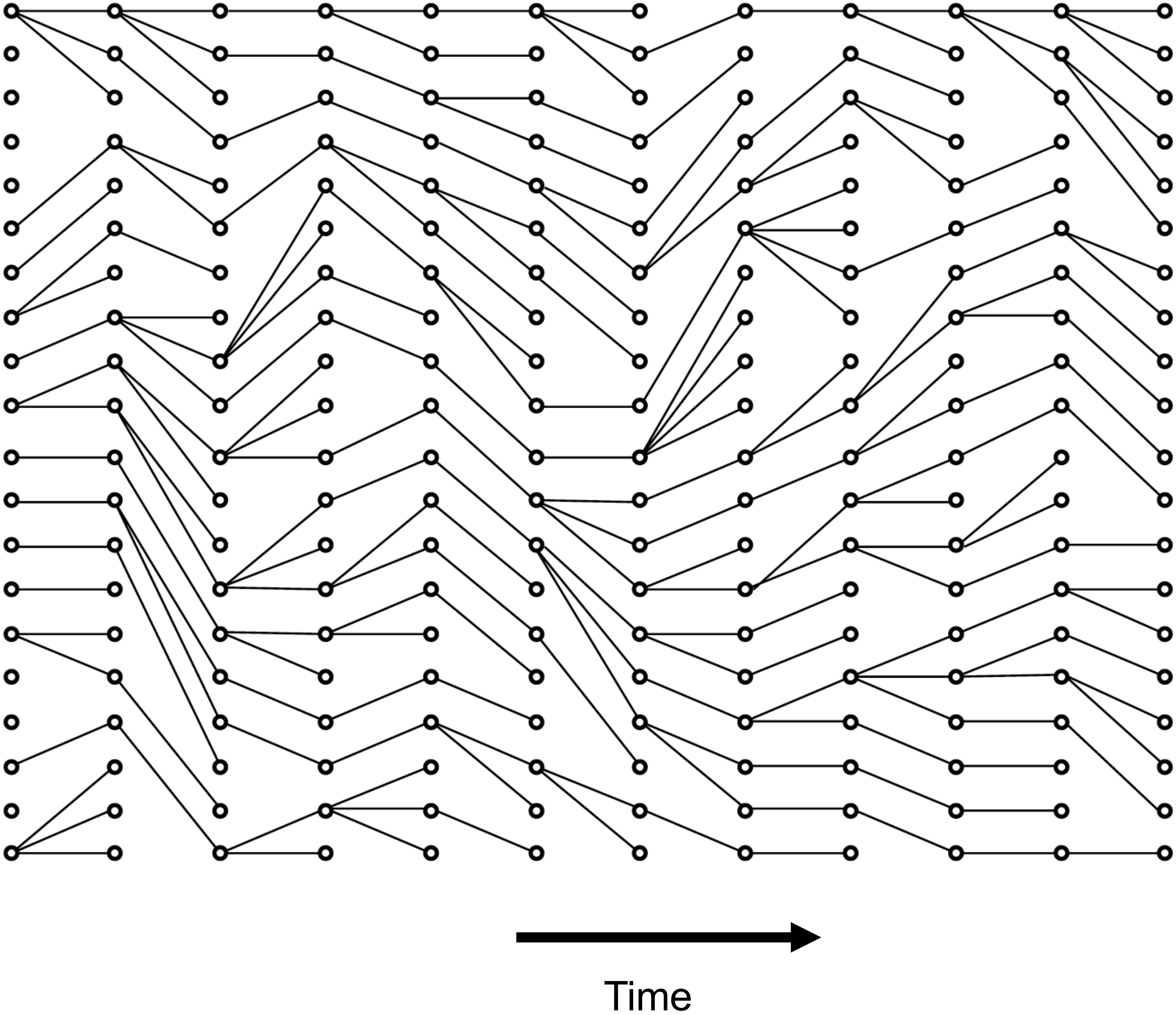


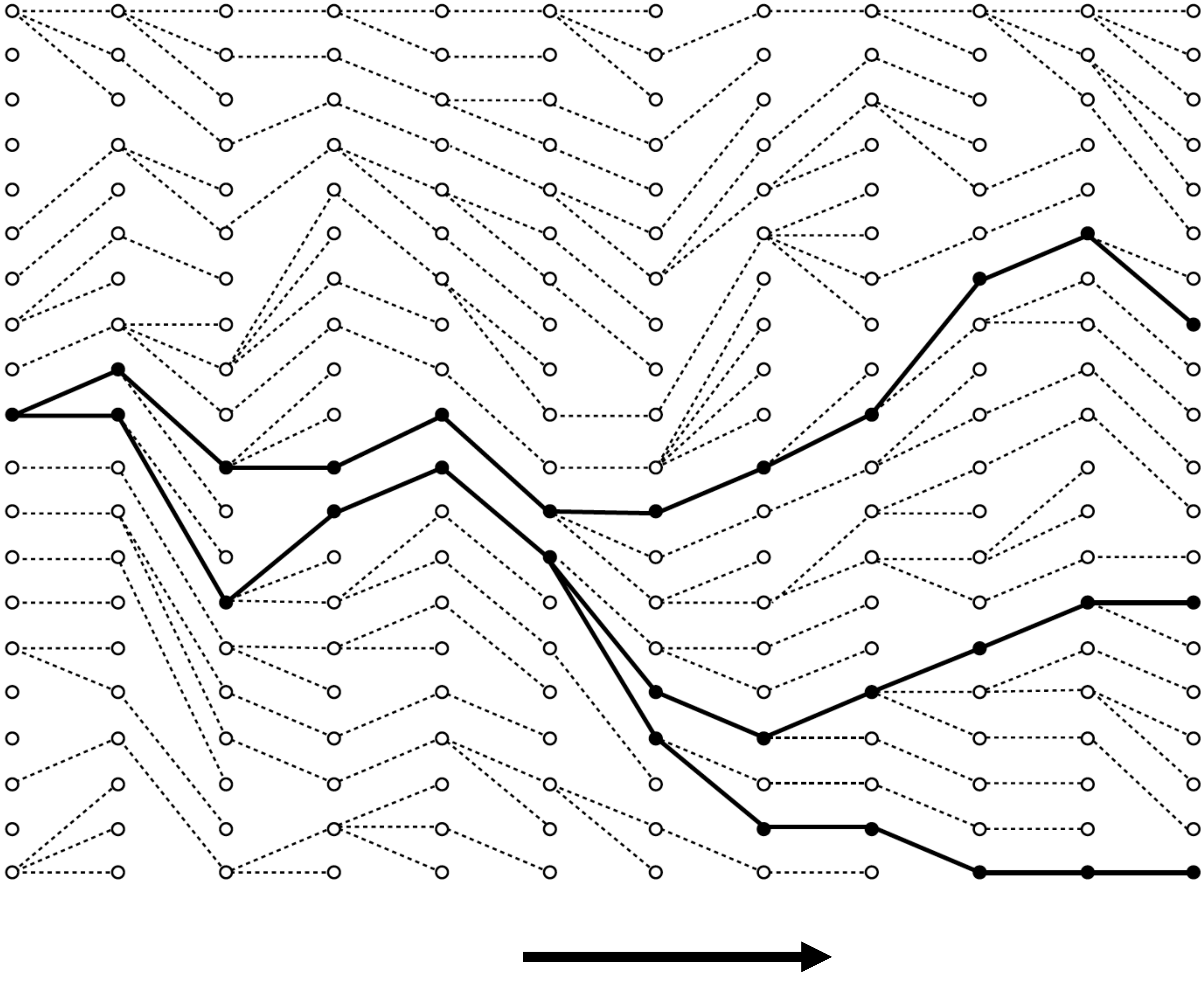


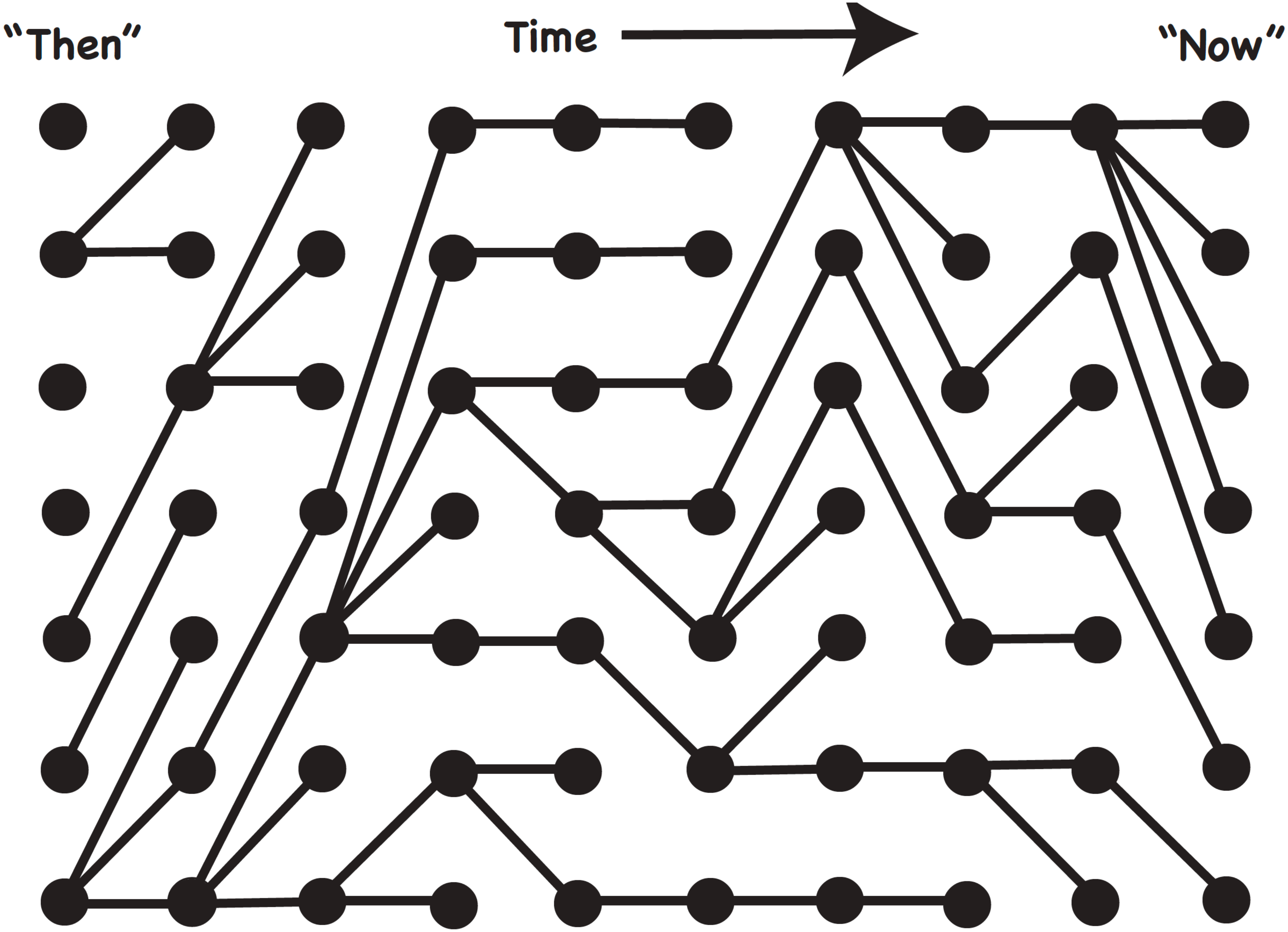
The Coalescent

- Wright-Fisher Model: Model of population genetics going forward in time
  - Simulation
  - Population genetic inference
- Coalescent Process: Model of population genetics going backward in time
  - Statistical inference

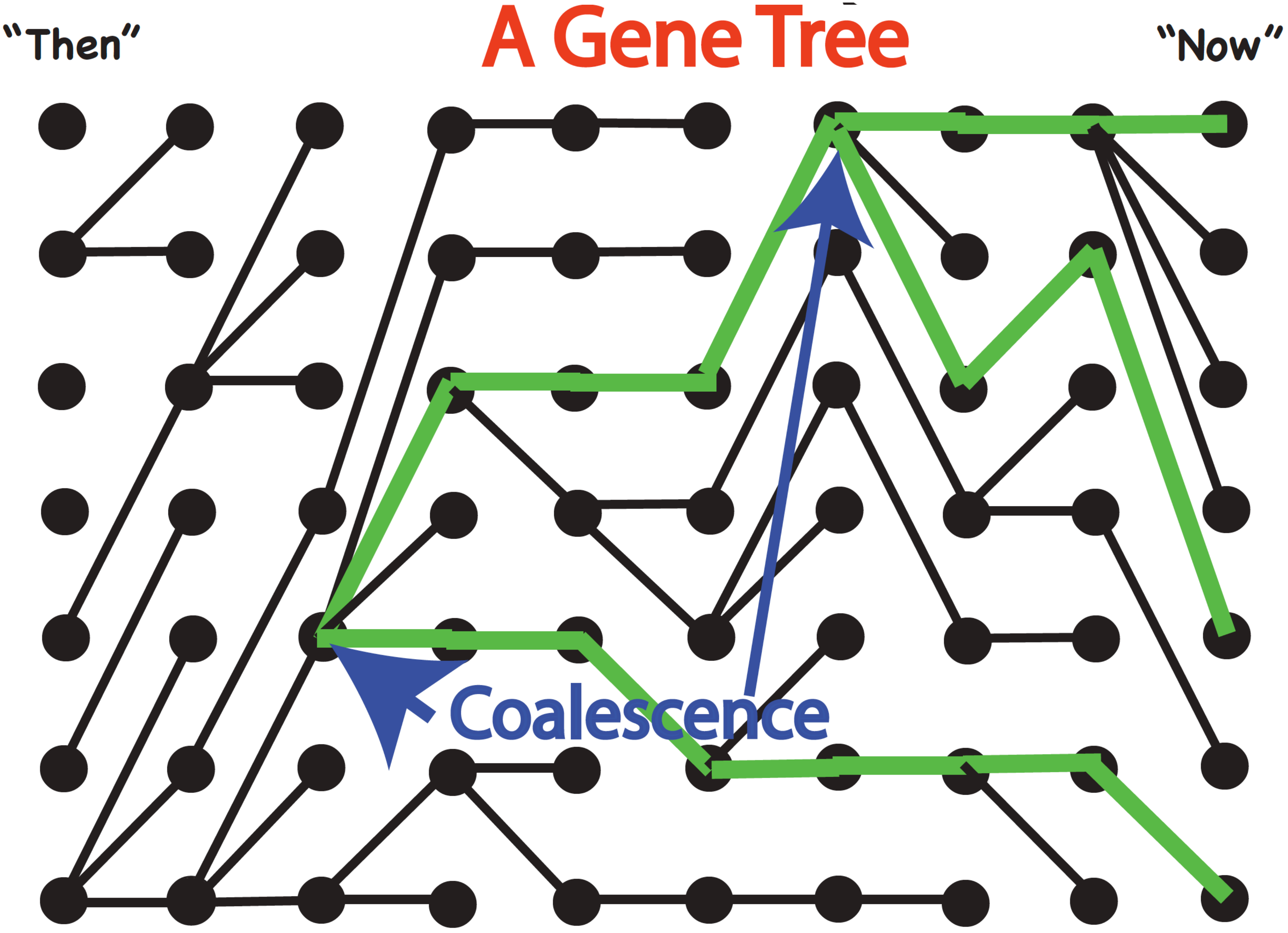

Time


Time


Time


"Then"
Time
"Now"

A Gene Tree



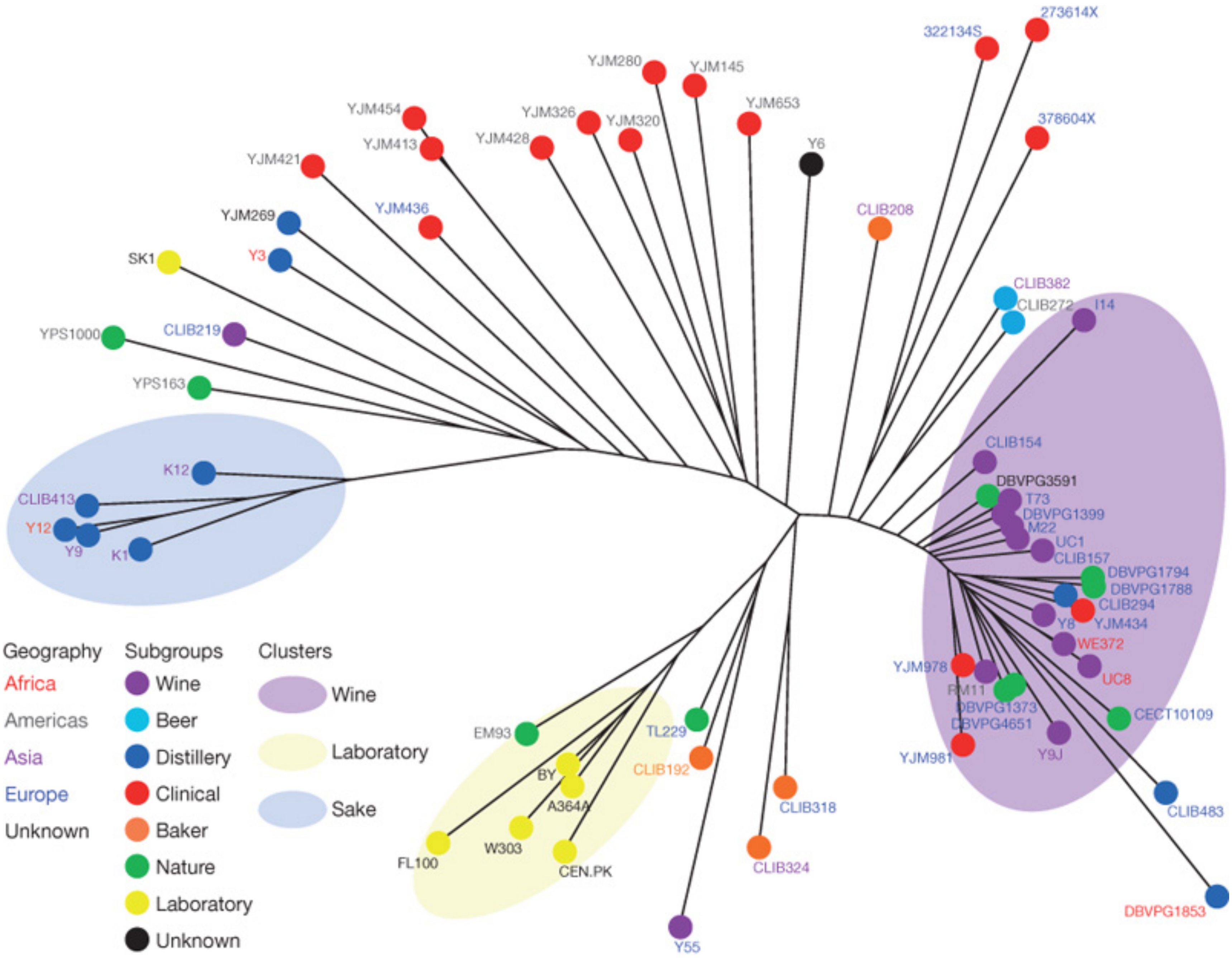
The coalescent

- A way to approximate probability of a specific gene tree when neutral evolution is assumed
  - Wright-Fisher
  - Moran (does not require discrete generations)
- Basic assumptions
  - Population size N is large
  - Number of gene copies in a sample is much smaller than 2N

The coalescent

- Number of generations since two randomly sampled gene copies have a common ancestor is distributed exponentially
- Diploid population of N individuals, mean of exponential distribution is 2N (generations)

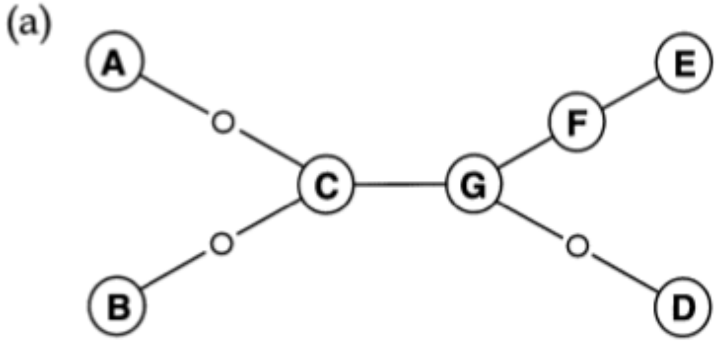
Neighbour-joining tree of 63 *S. cerevisiae* strains.



J Schacherer *et al. Nature* (2009) doi:10.1038/nature07670

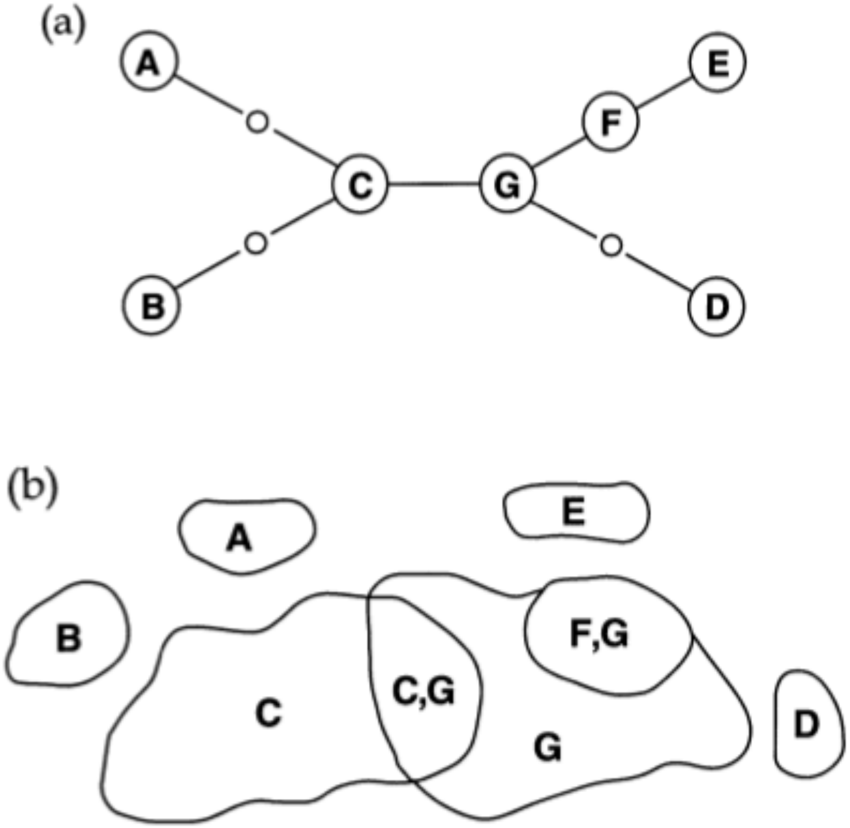
Phylogeographic inference

- Interpreting patterns of congruence/discordance between geographical distribution of haplotypes and their genealogical relationships



Phylogeographic inference

- Interpreting patterns of congruence/discordance between geographical distribution of haplotypes and their genealogical relationships



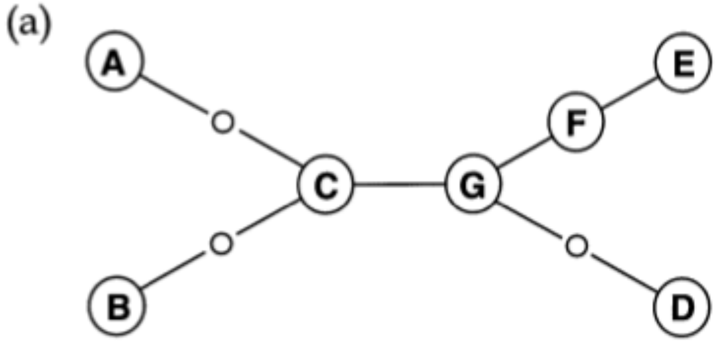
Clades of closely related haplotypes are geographically restricted and occur in proximity to each other.

CONGRUENCE

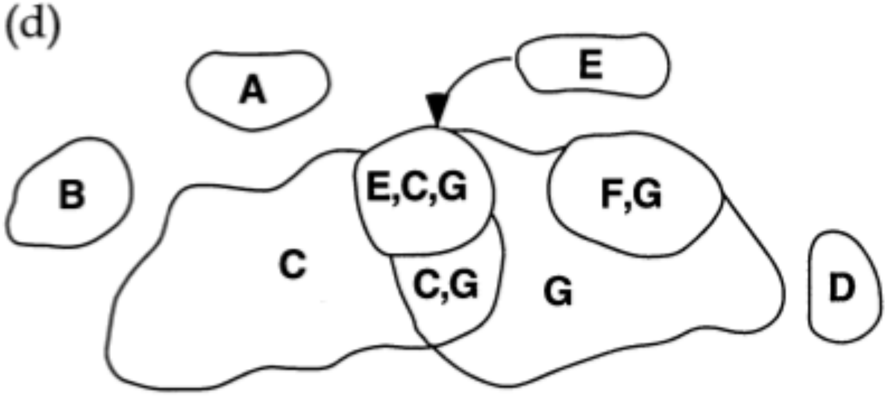
Long-standing pattern of highly restricted gene flow.

Phylogeographic inference

- Interpreting patterns of congruence/discordance between geographical distribution of haplotypes and their genealogical relationships



Clades of closely related haplotypes are NOT geographically restricted and do NOT occur in proximity to each other.

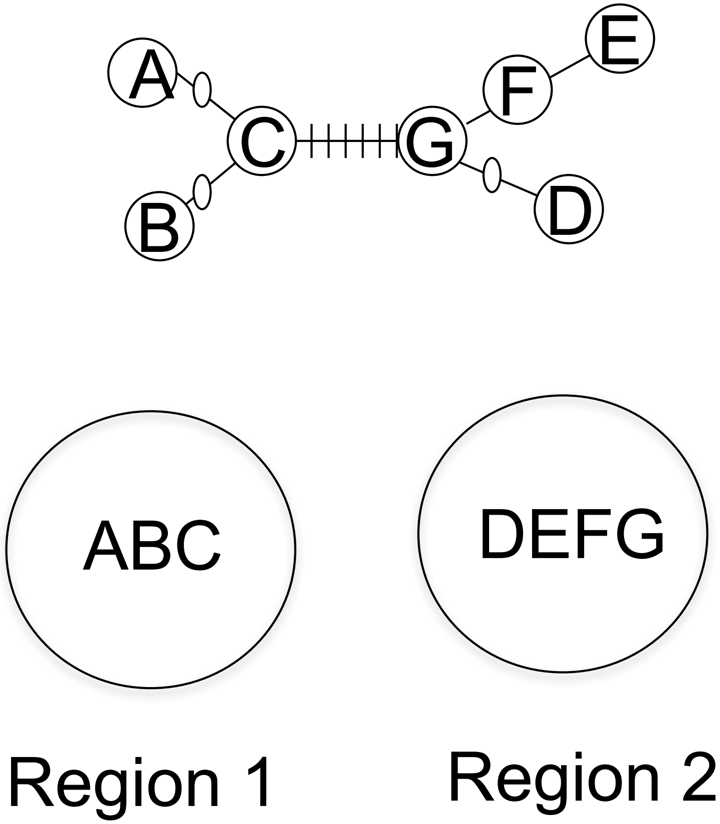


DISCORDANT

Inter-population gene flow

Phylogeographic inference

- Interpreting patterns of congruence/discordance between geographical distribution of haplotypes and their genealogical relationships



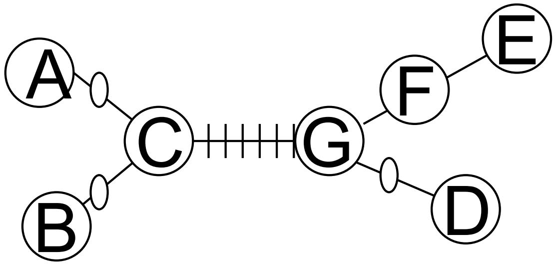
Deep gene tree

Major lineages **allopatric**

Long term barriers to gene flow

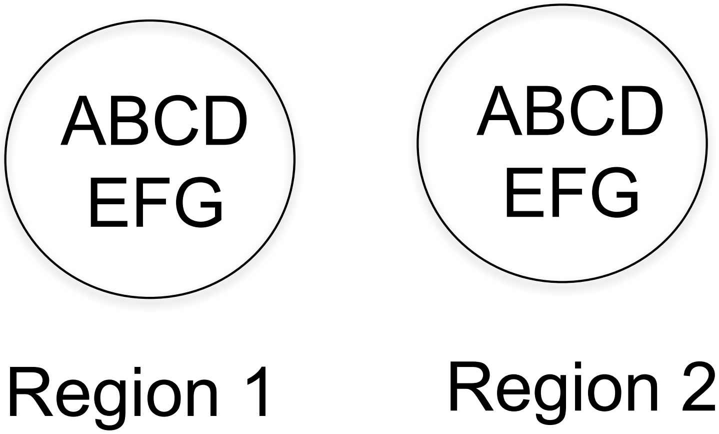
Phylogeographic inference

- Interpreting patterns of congruence/discordance between geographical distribution of haplotypes and their genealogical relationships



Deep gene tree

Major lineages **sympatric**



ABCD
EFG

Region 1

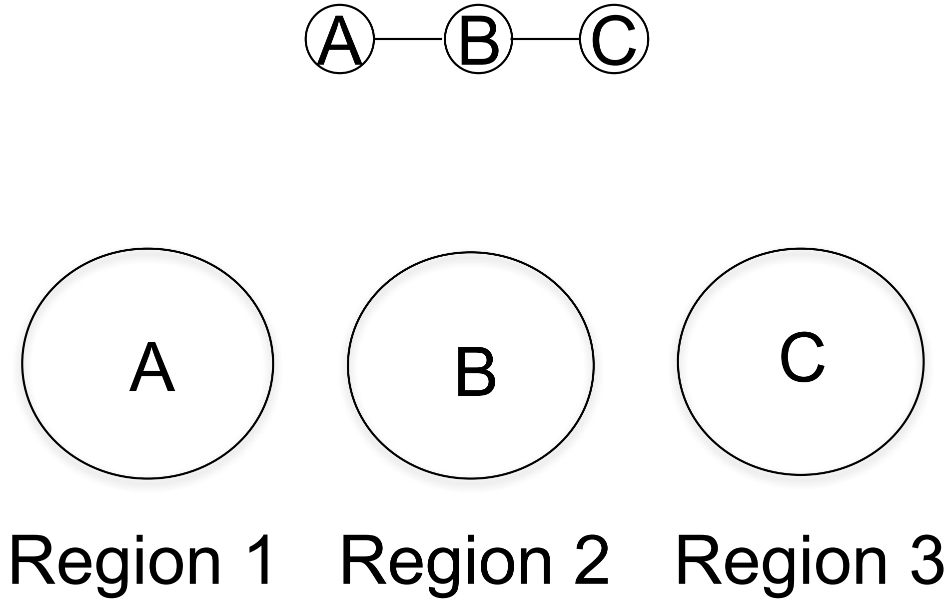
ABCD
EFG

Region 2

Long term barriers to gene flow but more recent admixture

Phylogeographic inference

- Interpreting patterns of congruence/discordance between geographical distribution of haplotypes and their genealogical relationships



Shallow gene tree

Major lineages **allopatric**

Recent colonization followed by restricted gene flow

Phylogeographic methods

- Nested Clade Phylogeographic Analysis (NPCA)
  - Templeton 1998, 2004
  - Couples estimates of gene genealogies with permutations of spatially-explicit summary statistics
  - Geographically contextualized gene genealogies
  - Infer demographic history of each taxon

Phylogeographic methods

- Statistical phylogeography
  - Based on coalescent models for parameter estimation and hypothesis testing
  - 1: Summary statistic calculated from simulated datasets under each model to obtain a distribution of the summary statistic under different models. Probabilities of each model are then evaluated with respect to the summary statistic estimated from empirical data
  - 2: Full Likelihood/Bayesian approach
    - Assume single model, estimate parameters using all of the data
  - 3: Approximate Bayesian Computation (ABC)
    - Bypasses the problem of calculating full likelihoods

Phylogeographic methods

- ABC
  - Approximate the posterior distribution of parameters
  - Based on simulating data from a coalescent model using parameter values randomly drawn from priors
  - Posterior approximated based on simulated data sets that most closely match the observed data (based on summary statistics)
  - Extremely flexible (as complex as a scenario as imaginable)
  - Can be used for parameter estimation AND for hypothesis testing.
  - Does not use all the data, requires choice of most informative summary statistics

Phylogeographic software

- LAMARC: Likelihood analysis with metropolis algorithm using random coalescence
- Kuhner 2006 (Bioinformatics)
- Markov chain Monte Carlo coalescent genealogy sampler
- Estimates population genetic parameters from genetic data
  - Population size
  - Population growth rate
  - Migration rate
  - Recombination rate

Phylogeographic software

- BEAST 2: Bayesian Evolutionary Analysis by Sampling Trees
- Bouckaert et al 2014 (PLOS Comp Bio)
- Platform for MCMC-based Bayesian inference and phylogenetics
- Implements structured coalescent model that allows inference of subpopulation sizes and migration rates together with (structured trees) from genetic data

Phylogeographic software

- STRUCTURE
- Pritchard et al. 2000 (Genetics)
- Model-based clustering method for using multilocus genotype data
- Applications:
  - Infer presence of distinct populations
  - Assign individuals to populations
  - Study hybrid zones
  - Identify migrants and admixed individuals

Phylogeographic software

- BAPS (Bayesian Analysis of Population Structure
- Corander and Tang 2007 (Mathematical Biosciences)
- Bayesian predictive classification model using multilocus genotype data
- Applications:
  - Infer presence of distinct populations
  - Assign individuals to populations
  - Spatial analysis of molecular variance (SAMOVA)

Phylogeographic software

- SPREAD: Spatial phylogenetic reconstruction of evolution dynamics
- Bielejec et al. 2011 (Bioinformatics)
- Platform to analyze and visualize Bayesian phylogeographic reconstructions incorporating spatial-temporal diffusion
- Extension of suite of Bayesian inference approaches for the joint reconstruction of evolutionary and geographic history
- Models that accommodate spatial diffusion in discrete and continuous space produce statistical distributions of temporal-spatial phylogenies

Phylogeographic software

- GenAlEx
- Population genetic analysis in the Microsoft Excel environment
- Frequency-based analyses
  - F-statistics, heterozgosity, HWE, population assignment, relatedness
- Distance-based analyses
  - AMOVA, PCoA, Mantel test, multivariate spatial autocorrelation analysis

Phylogeographic software

- Arlequin 3.5
- Excoffier and Lischer 2010 (Molecular Ecology Resources)
- Platform for population genetic and statistical tests to extract information on genetic and demographic features of a collection of population samples
- Implements
  - AMOVA
  - Fst
  - Exact test of population differentiation
  - Mantel test

Phylogeography: Applications

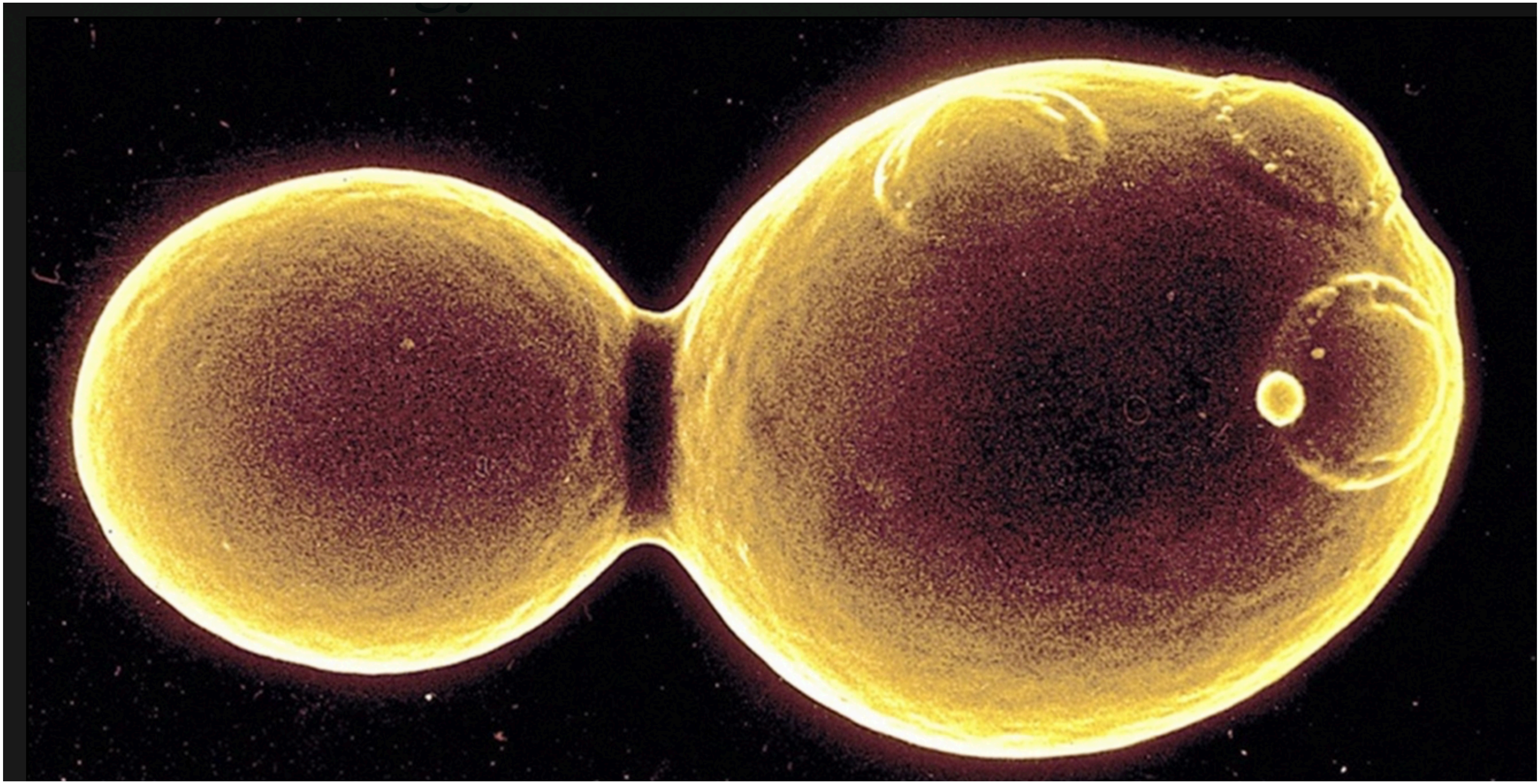
- Identifying units of conservation
  - Management unit (MU)
    - Shallower time scale
    - Usually defined as barriers to gene flow
    - Often isolated populations
    - Aligns well with population genetic ideas
  - Evolutionary Significant Unit (ESU)
    - Deeper time scale
    - Recognized through reciprocal monophyly
    - Can be tested using molecular phylogenetics
    - Often intraspecific phylogroups or subspecies
    - Aligns more closely with phylogenetic ideas, but within a single species

Phylogeography: Applications

- Species: Why important?
  - New species are ‘units’ of biodiversity acquisition
  - Need to know what they are to conserve them
  - Operational unit of conservation biology
- Species: Concepts
  - Biological Species Concept: groups of actually or potentially interbreeding individuals
  - Phylogenetic Species Concept: smallest diagnosable cluster of individual organisms with parental ancestry and descent (monophyletic)
- Species: Operationally, define as reproductively isolated lineages that are sufficiently evolutionarily distinct to have built up phylogenetic differences

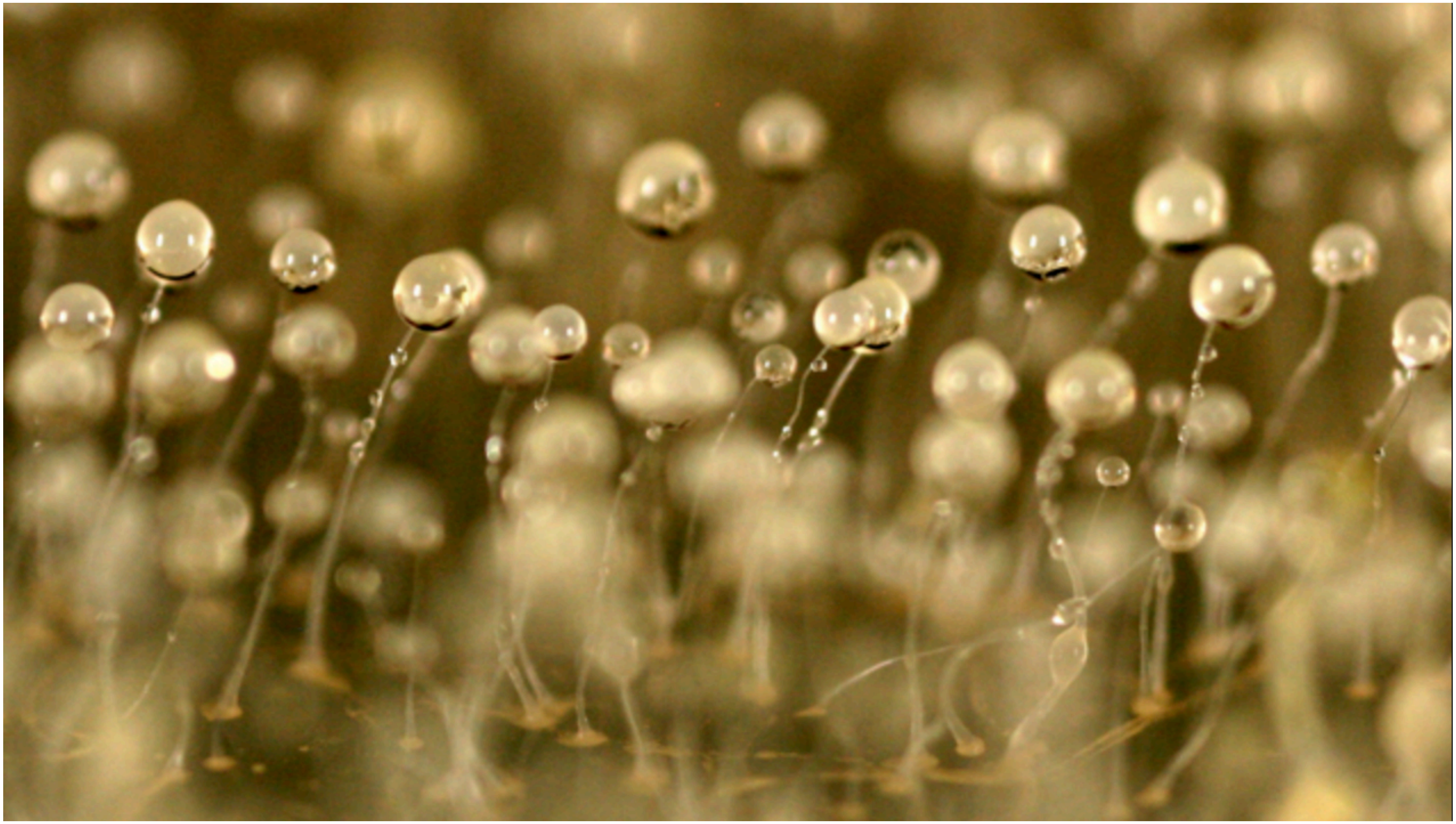
Phylogeography: Applications

- Individual: Why important?
  - Fundamental genetic unit in populations
  - How many individuals are there?
  - How many individuals contribute, and how much?
  - What are the source populations for individuals?
  - What do we conserve?
- Individual: Nuances
  - Ramet (physical) versus Genet (genetic)



Phylogeography: Applications

- Individual: Why important?
  - Fundamental genetic unit in populations
  - How many individuals are there?
  - How many individuals contribute, and how much?
  - What are the source populations for individuals?
  - What do we conserve?
- Individual: Nuances
  - Ramet (physical) versus Genet (genetic)
  - Chimeras



Phylogeography: Conservation

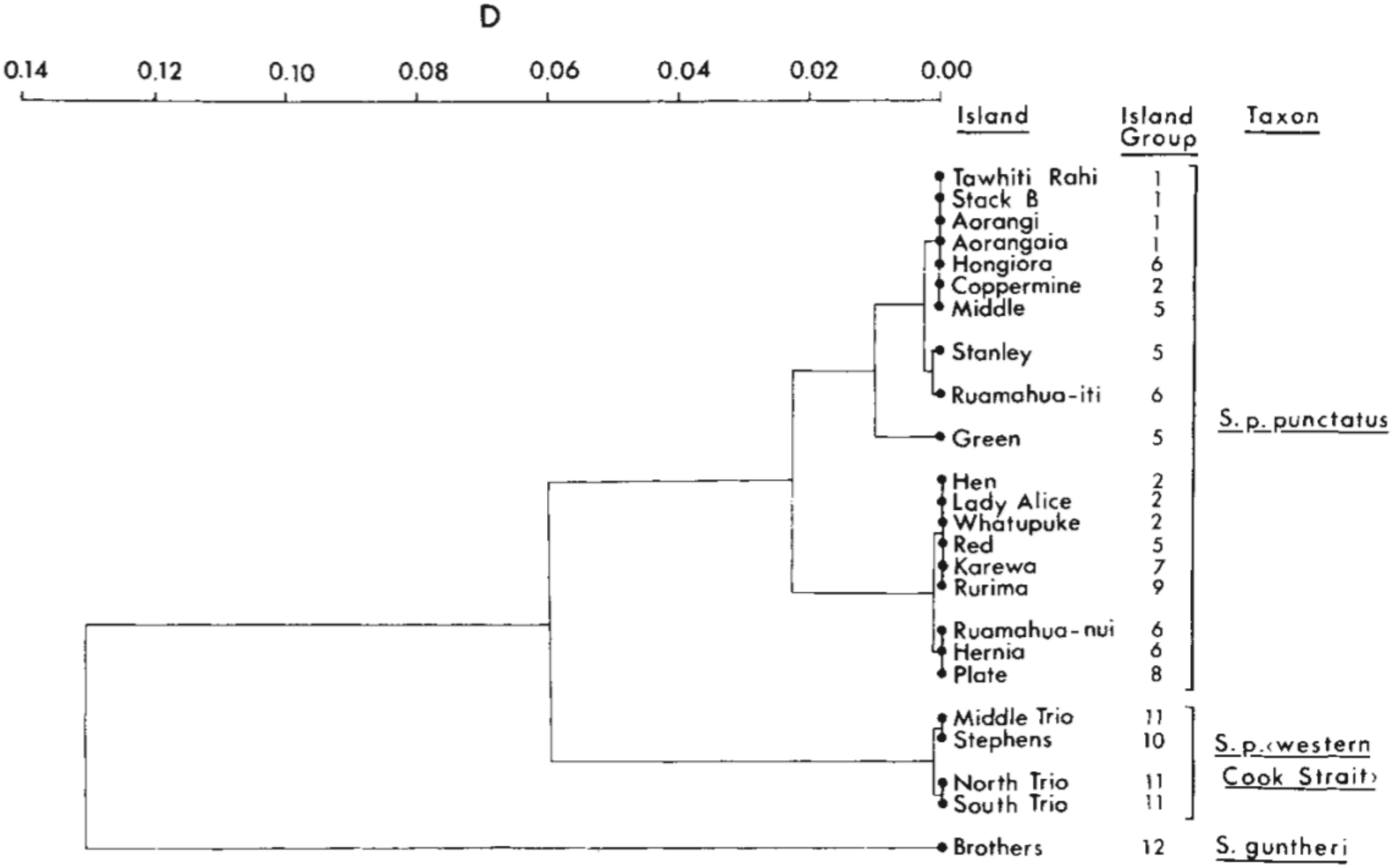
- Identify units of conservation

Tuatara

- Endemic to New Zealand
- Only surviving genus of one reptilian order
- Once widespread, now < 100,000
- Decline prior to European settlement

Tuatara

- Fully protected in 1895
- Protection only refers to one of the two extant species



Daugherty et al. 1990

Red Wolf

- Extensive range in SE US pre-settlement
- Extinct in the wild by 1980
- Captive breeding program started in 1973
  - 400 animals captured
  - 43 sent for breeding
  - 17 became the breeding stock
- In 2007, 300 red wolves alive
  - 207 in captivity
- 100-130 individuals in NC red wolf recovery area

Red Wolf

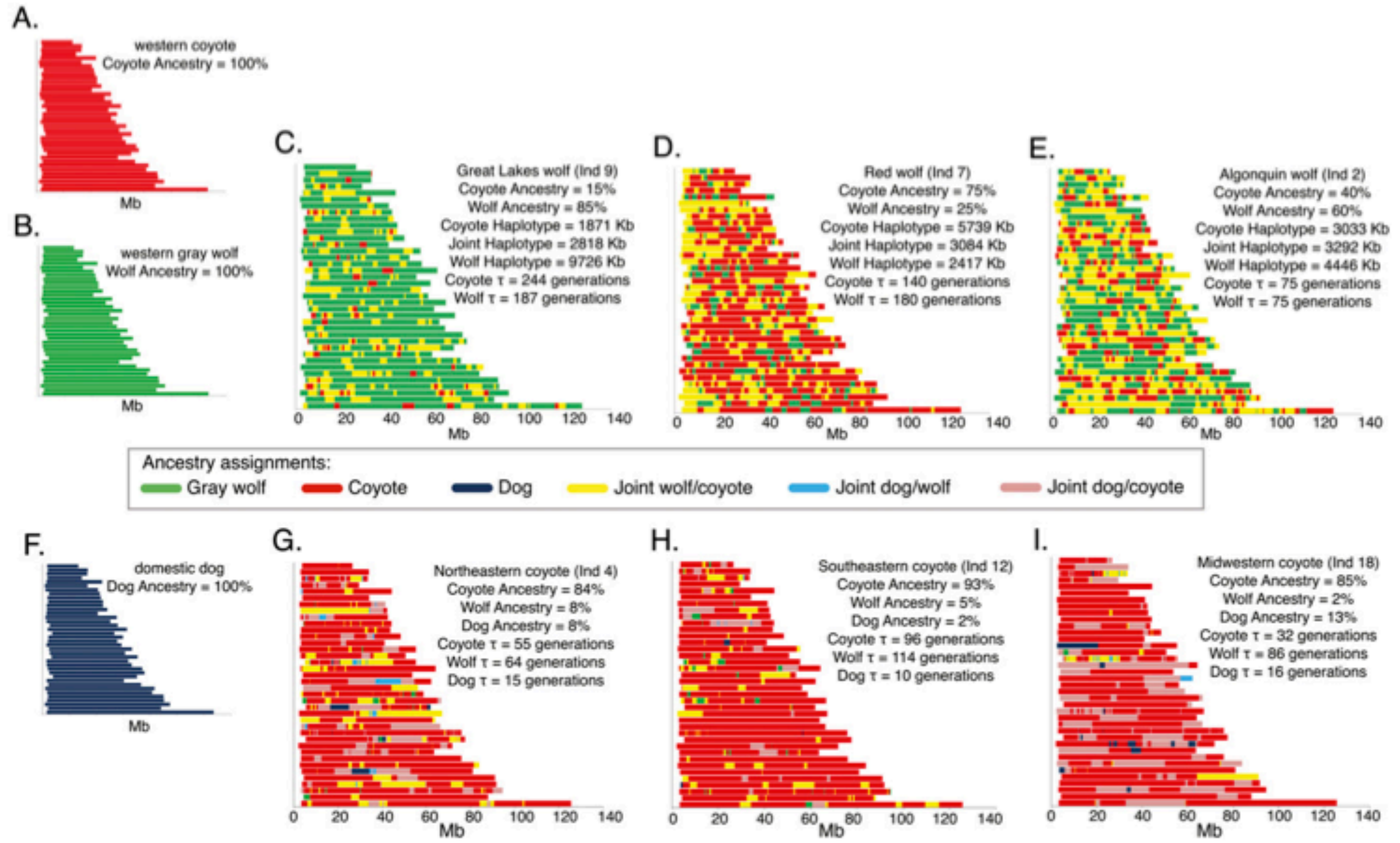


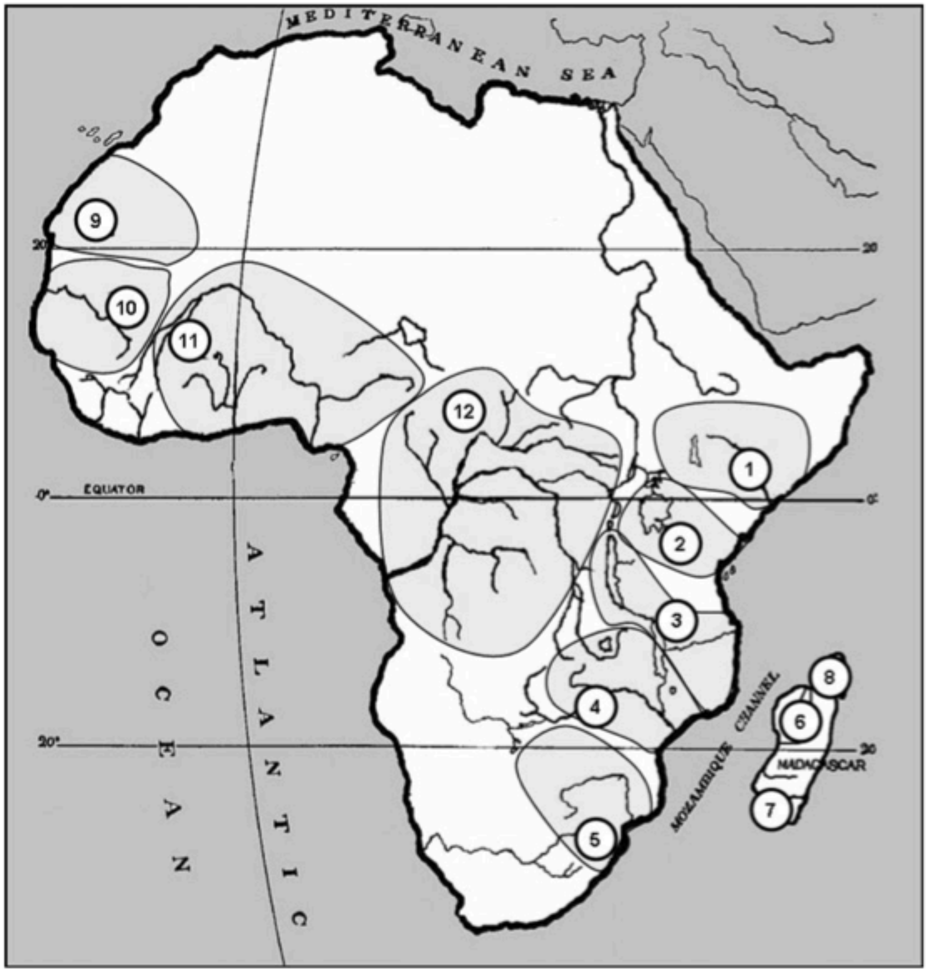
Figure 6. Ancestry analysis of 38 canine autosomes. Plots show ancestry blocks and their assignments for representative individuals of canid populations with average size of blocks, percent ancestry, and number of generations since most recent admixture (τ) indicated. Two-ancestor (coyote, *A*; gray wolf, *B*) analyses are presented for a Great Lakes wolf from Minnesota (*C*), a captive red wolf (*D*), and an Algonquin wolf (*E*). Three-ancestor analyses (coyote, *A*; gray wolf, *B*; dog, *F*) are presented for a Northeastern coyote from Vermont (*G*), a Southern coyote from Louisiana (*H*), and a Midwestern coyote from Ohio (*I*). Individual results are given in Supplemental Tables S5 and S6.

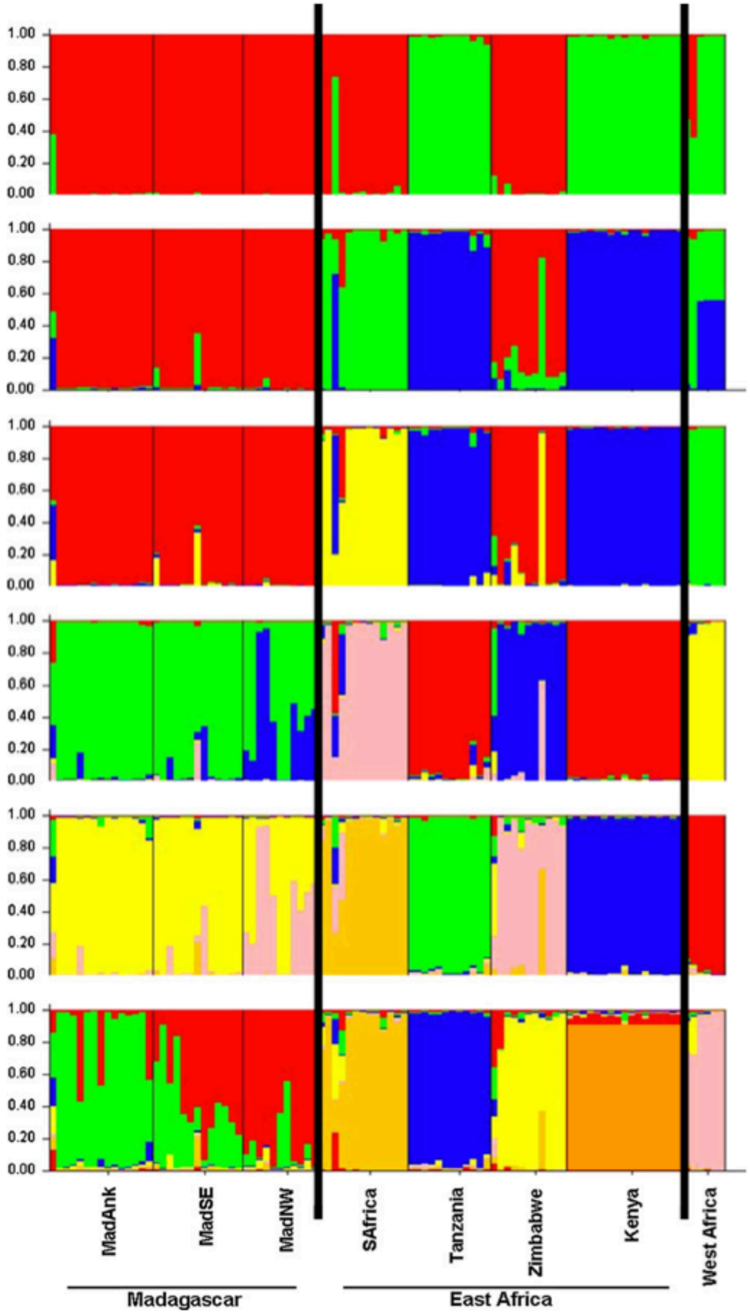
vonHoldt et al. 2011

Nile crocodile

- Largest crocodilian in Africa
- Top predator
- Hunting-induced decline late 1800s-mid 1900s
- 250,000-500,000 wild individuals remain
    - Central and West Africa populations sparse
- Currently threatened by pollution and habitat loss

Nile crocodile





Hekkala et al. 2010

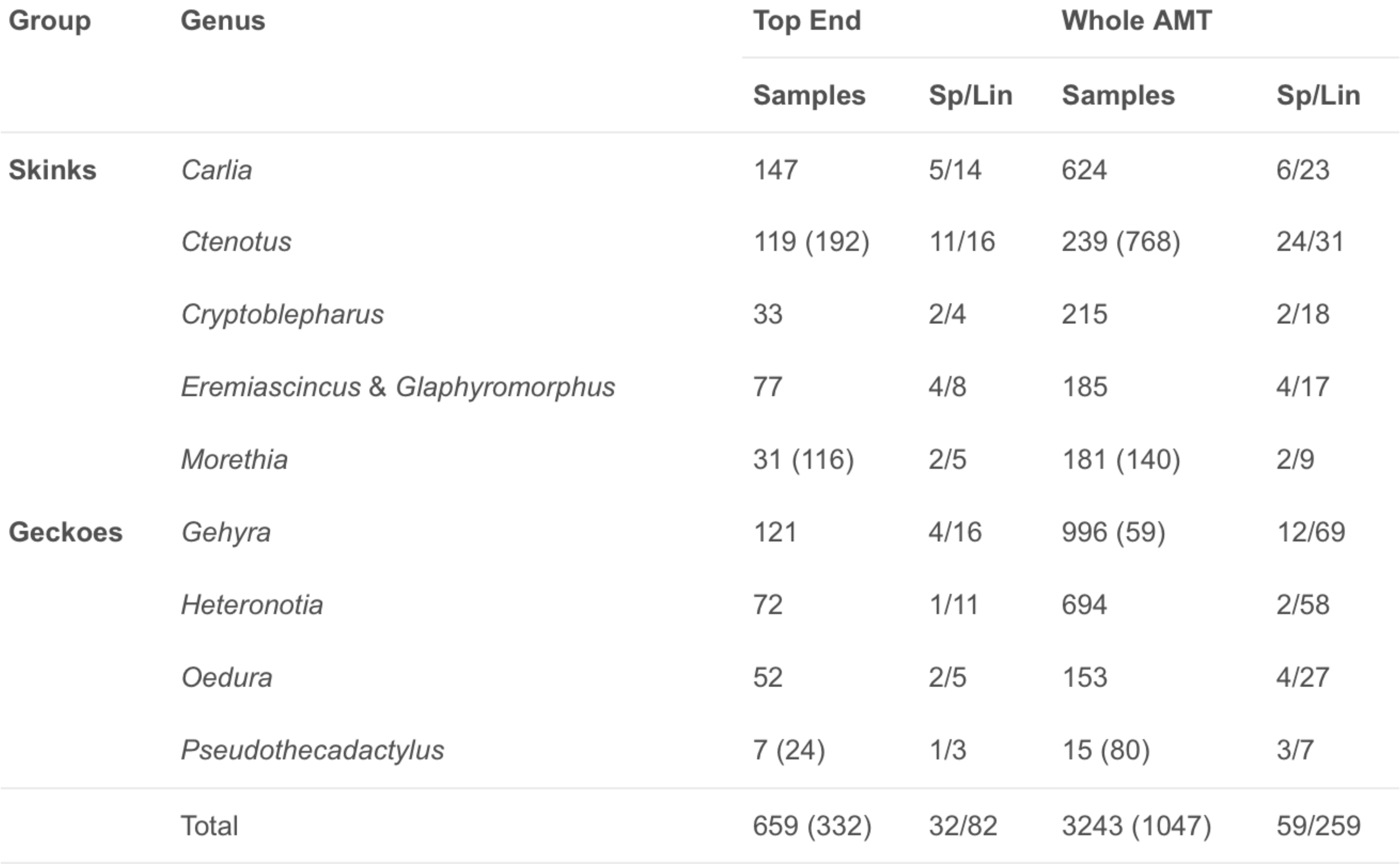
Phylogeography: Conservation

- Identify units of conservation
- Map conservation priorities

Australian Lizards

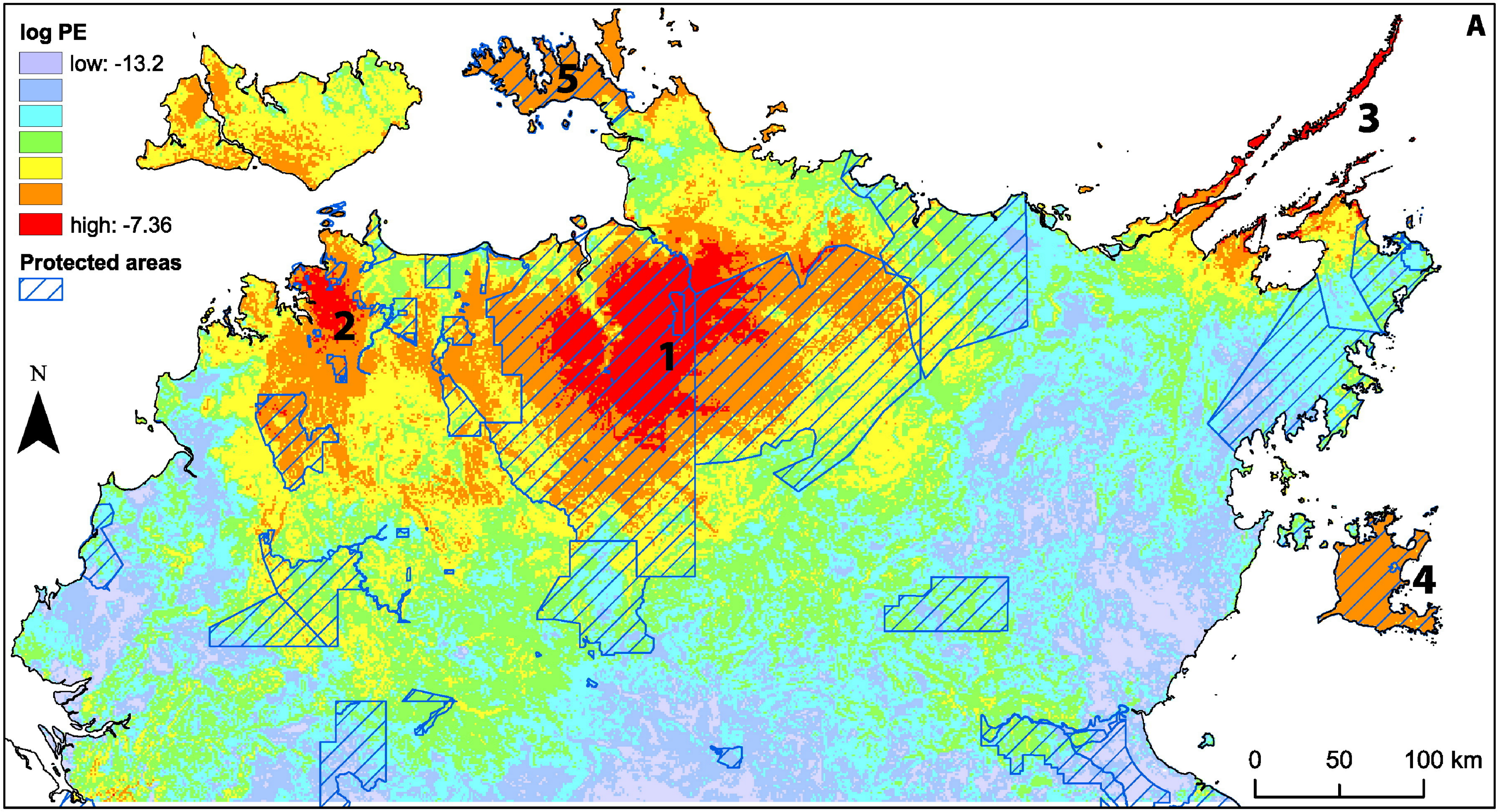
- mtDNA phylogeographic data from ten genera of lizards
- Goal: identify hotspots of phylogenetic endemism across the region

Australian Lizards

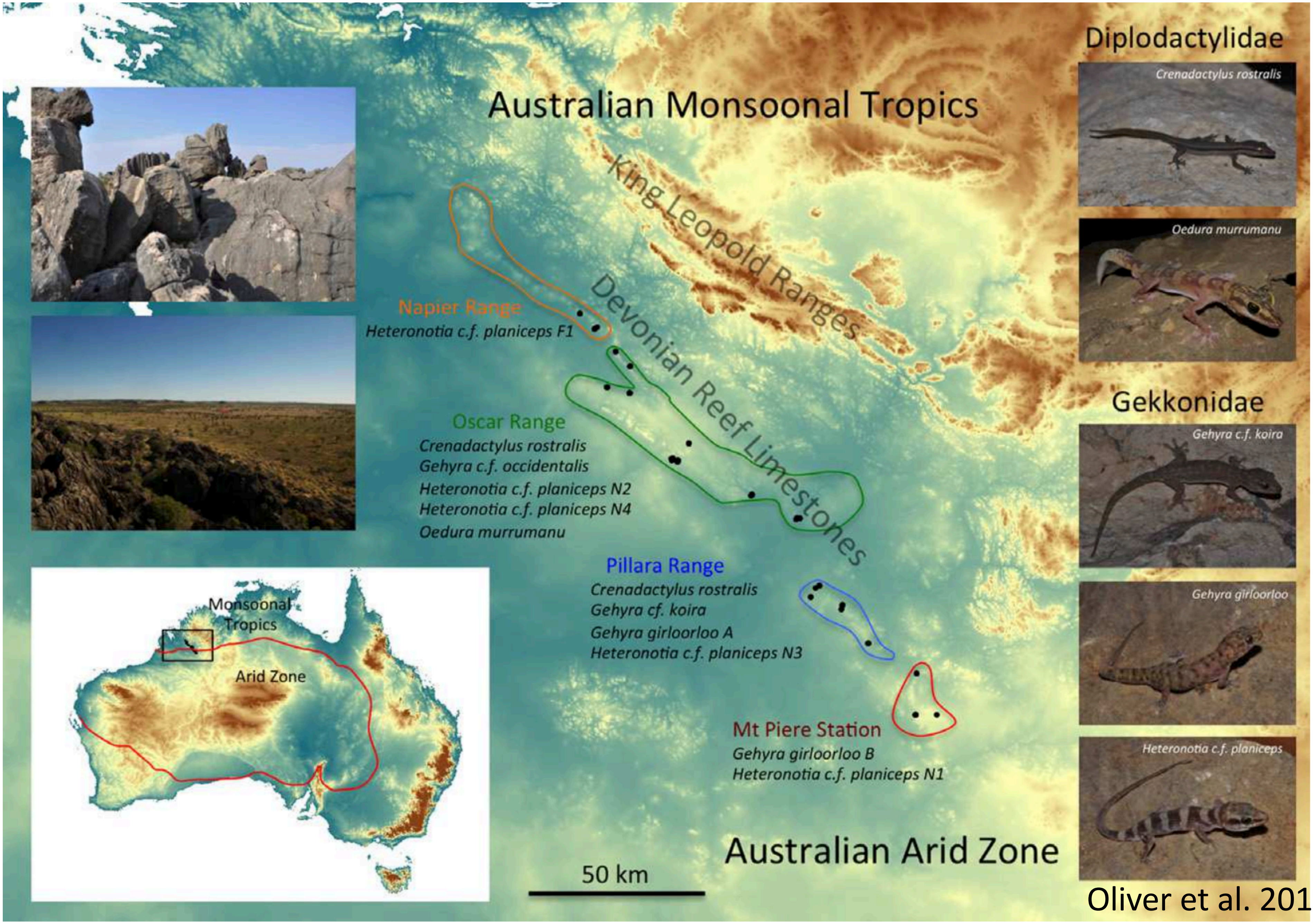


Group	**Genus**	**Top End**		**Whole AMT**	
		Samples	**Sp/Lin**	**Samples**	**Sp/Lin**
Skinks	*Carlia*	147	5/14	624	6/23
	Ctenotus	119 (192)	11/16	239 (768)	24/31
	Cryptoblepharus	33	2/4	215	2/18
	Eremiascincus & *Glaphyromorphus*	77	4/8	185	4/17
	Morethia	31 (116)	2/5	181 (140)	2/9
Geckoes	*Gehyra*	121	4/16	996 (59)	12/69
	Heteronotia	72	1/11	694	2/58
	Oedura	52	2/5	153	4/27
	Pseudothecadactylus	7 (24)	1/3	15 (80)	3/7
	Total	659 (332)	32/82	3243 (1047)	59/259

Australian Lizards

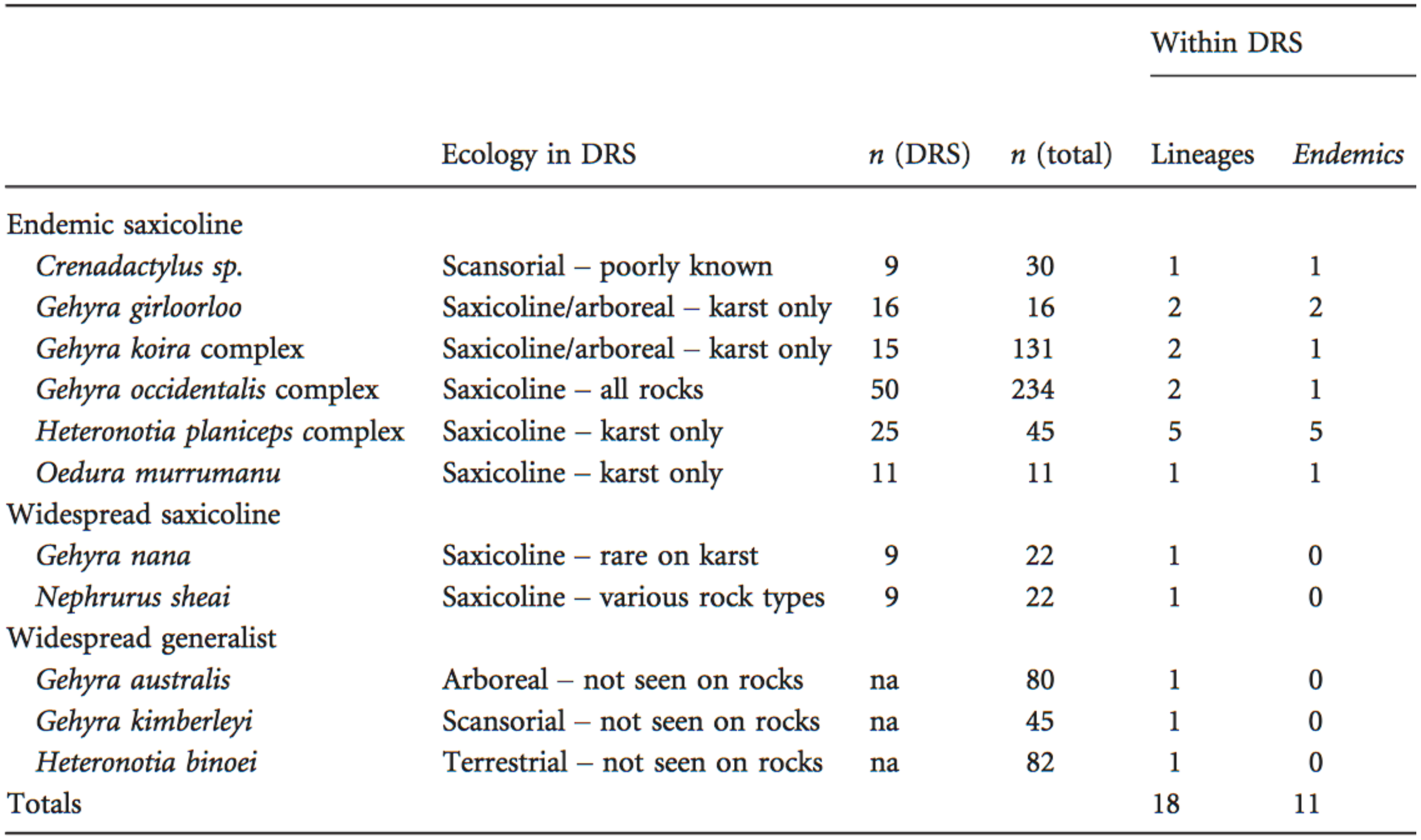


Australian geckos



Oliver et al. 2017

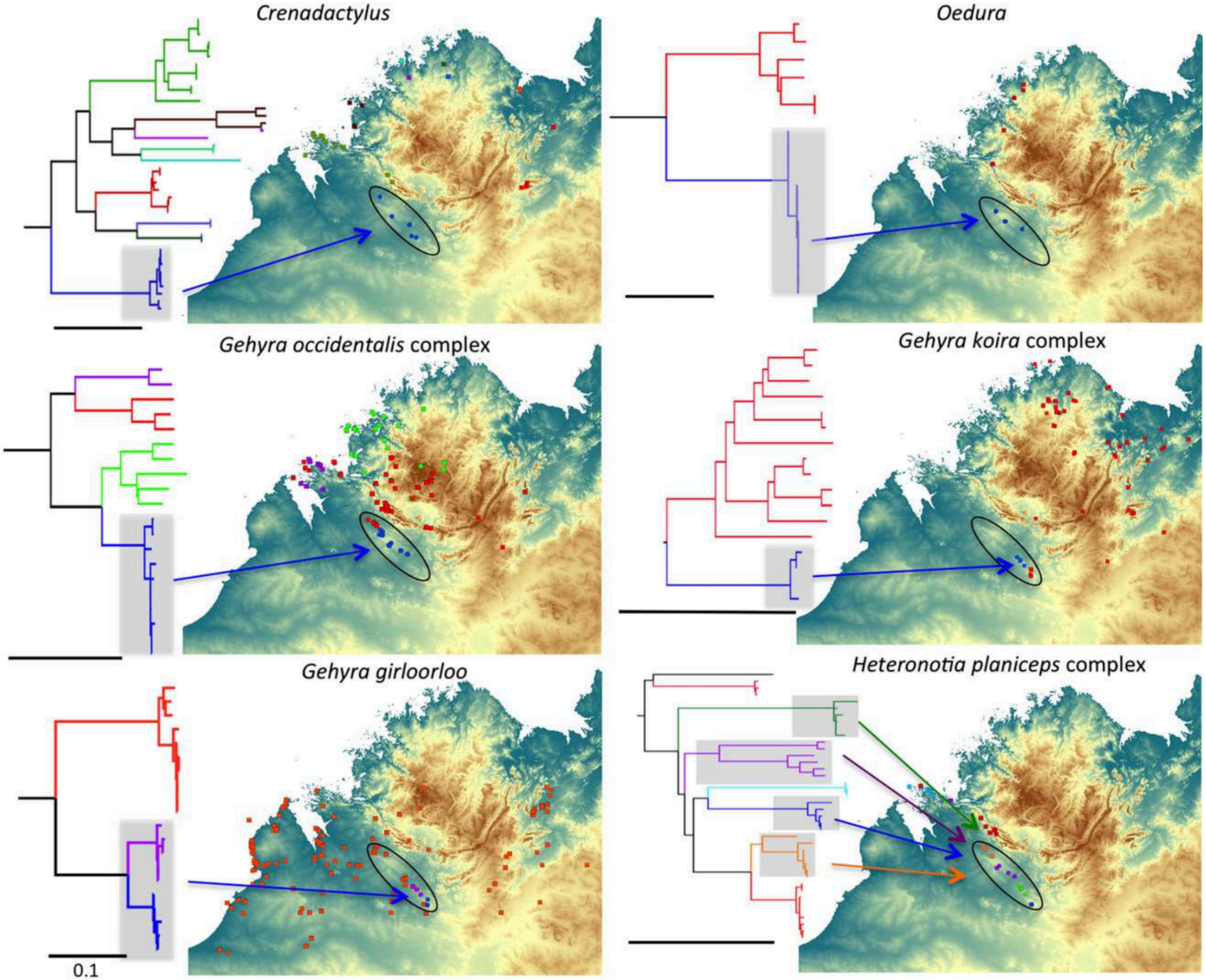
Australian geckos



				Within DRS	
	Ecology in DRS	*n* (DRS)	*n* (total)	Lineages	*Endemics*
Endemic saxicoline					
Crenadactylus sp.	Scansorial – poorly known	9	30	1	1
Gehyra girloorloo	Saxicoline/arboreal – karst only	16	16	2	2
Gehyra koira complex	Saxicoline/arboreal – karst only	15	131	2	1
Gehyra occidentalis complex	Saxicoline – all rocks	50	234	2	1
Heteronotia planiceps complex	Saxicoline – karst only	25	45	5	5
Oedura murrumanu	Saxicoline – karst only	11	11	1	1
Widespread saxicoline					
Gehyra nana	Saxicoline – rare on karst	9	22	1	0
Nephrurus sheai	Saxicoline – various rock types	9	22	1	0
Widespread generalist					
Gehyra australis	Arboreal – not seen on rocks	na	80	1	0
Gehyra kimberleyi	Scansorial – not seen on rocks	na	45	1	0
Heteronotia binoei	Terrestrial – not seen on rocks	na	82	1	0
Totals				18	11

Oliver et al. 2017

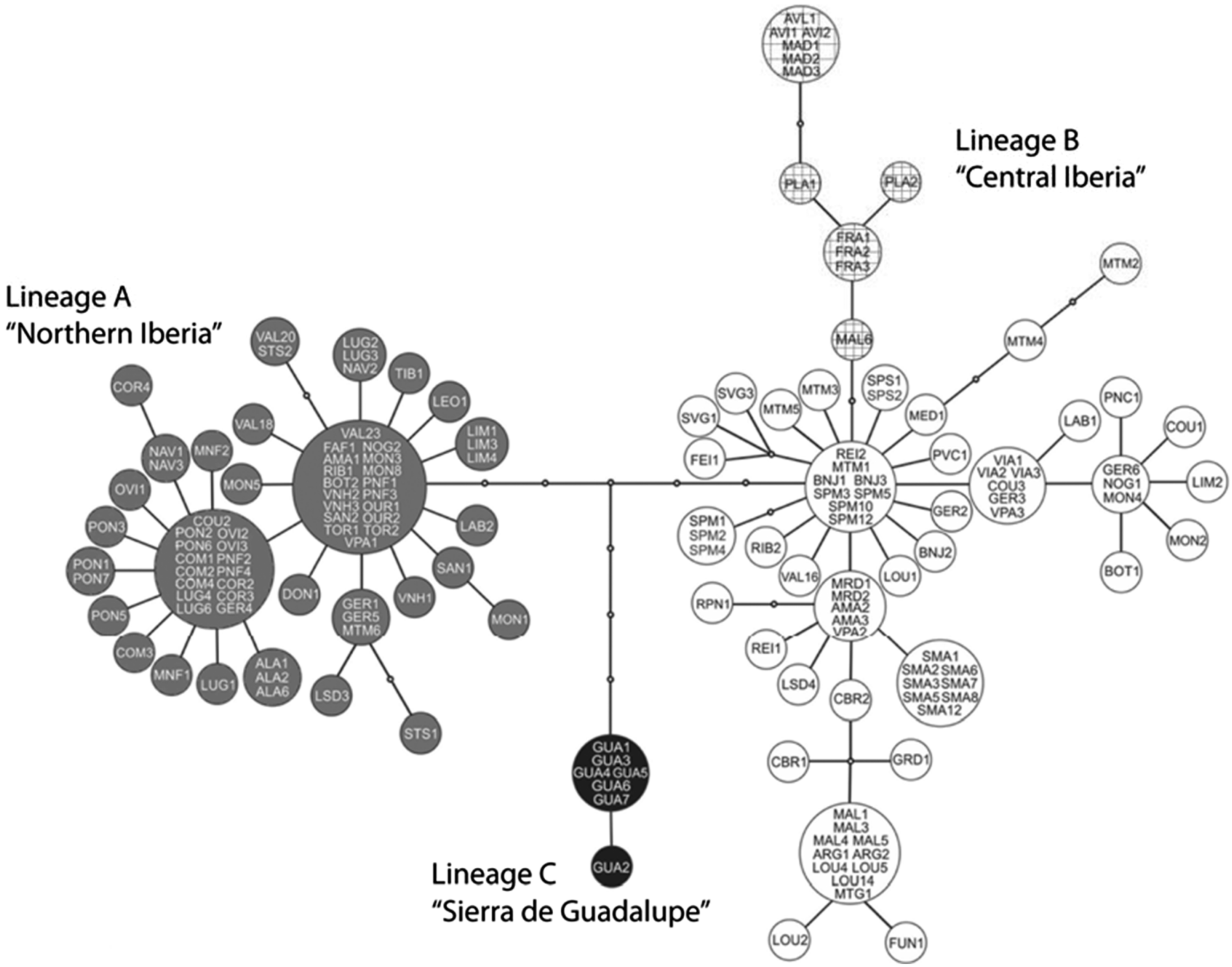
Australian geckos



l. 2017

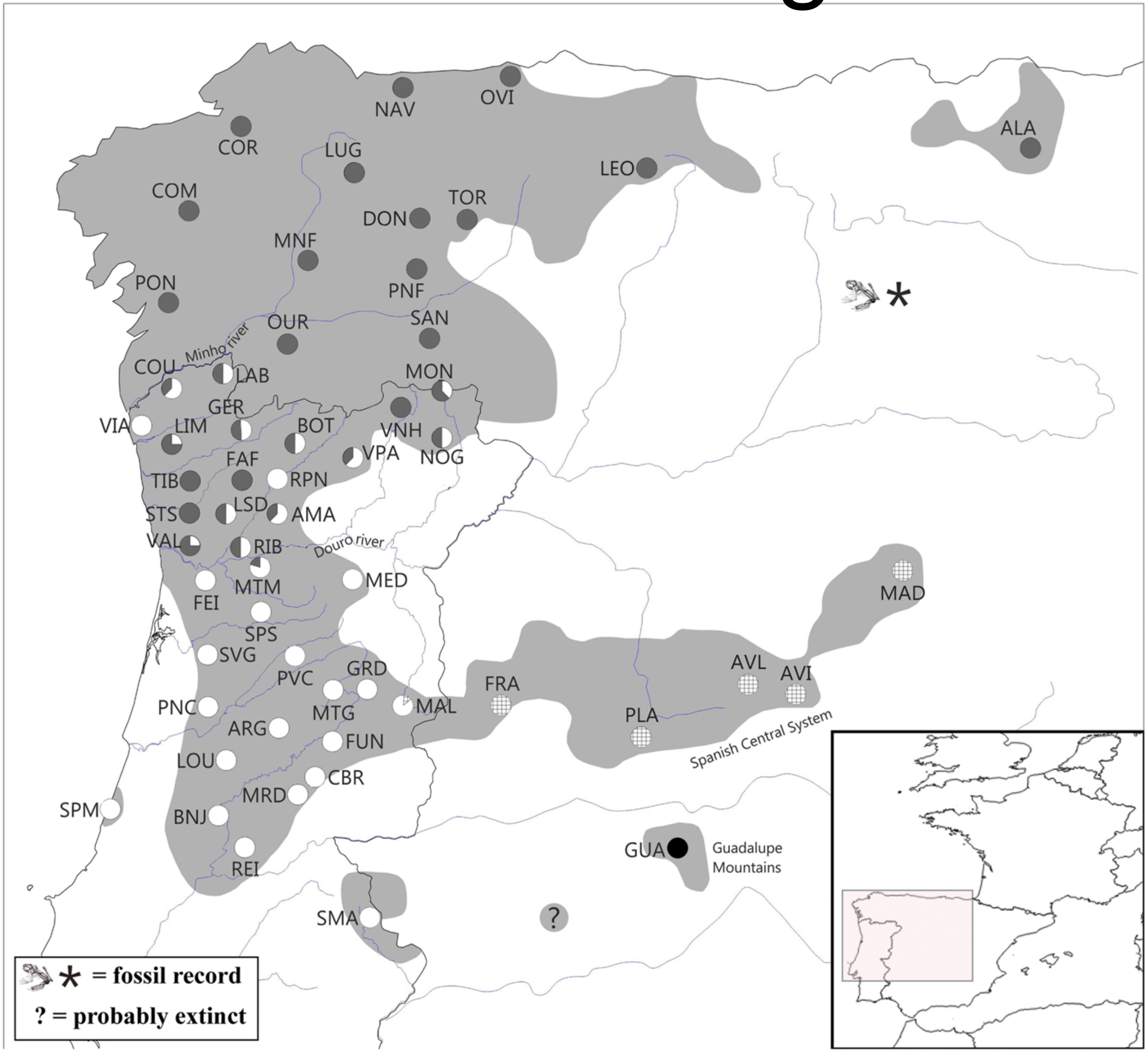
Iberian Frog

Iberian Frog



Teixeira et al. 2018

Iberian Frog



Teixeira et al. 2018