The Biogeography of Speciation: new genomic insights about reinforcement using biological collections

Stephen A. Sefick

2016-12-02

Outline

Topic

Motivation

1. Speciation

Motivation

1. Speciation
   - Speciation results in biodiversity

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced

2. Biogeography important to speciation process

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced
2. Biogeography important to speciation process
3. Genomics

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced
2. Biogeography important to speciation process
3. Genomics
   - since 2006 near exponential increase in genomic data

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced
2. Biogeography important to speciation process
3. Genomics
   - since 2006 near exponential increase in genomic data
   - Recently computationaly feasable to identify genetic variants

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced
2. Biogeography important to speciation process
3. Genomics
   - since 2006 near exponential increase in genomic data
   - Recently computationaly feasable to identify genetic variants
     - simple

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced
2. Biogeography important to speciation process
3. Genomics
   - since 2006 near exponential increase in genomic data
   - Recently computationaly feasable to identify genetic variants
     - simple
     - complex

Motivation

1. Speciation
   - Speciation results in biodiversity
   - Biodiversity declining
   - Important to study how biodiversity is produced
2. Biogeography important to speciation process
3. Genomics
   - since 2006 near exponential increase in genomic data
   - Recently computationaly feasable to identify genetic variants
     - simple
     - complex
4. Transform our understanding of the relationship of biogeography with speciation

Types of speciation

Types of speciation

1. Allopatric Speciation

Types of speciation

1. Allopatric Speciation
   - interupted species range (i.e., stream)

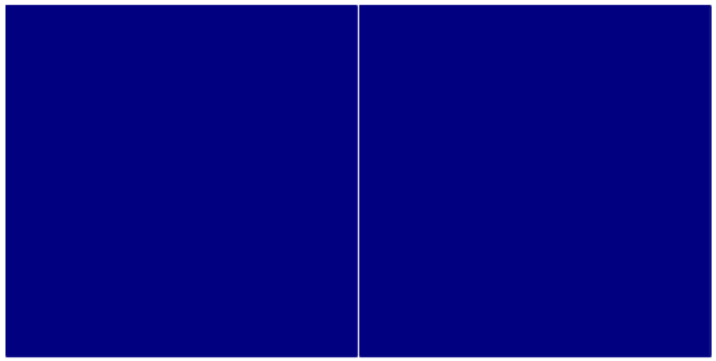
Types of speciation

1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow

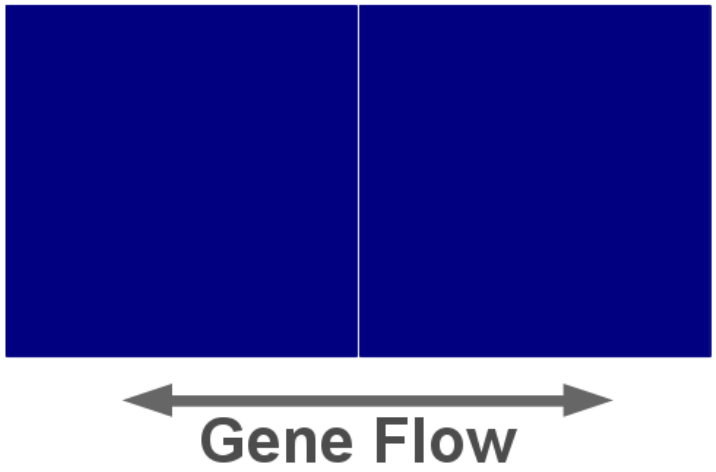
Types of speciation

1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow
   - 2 incipient species

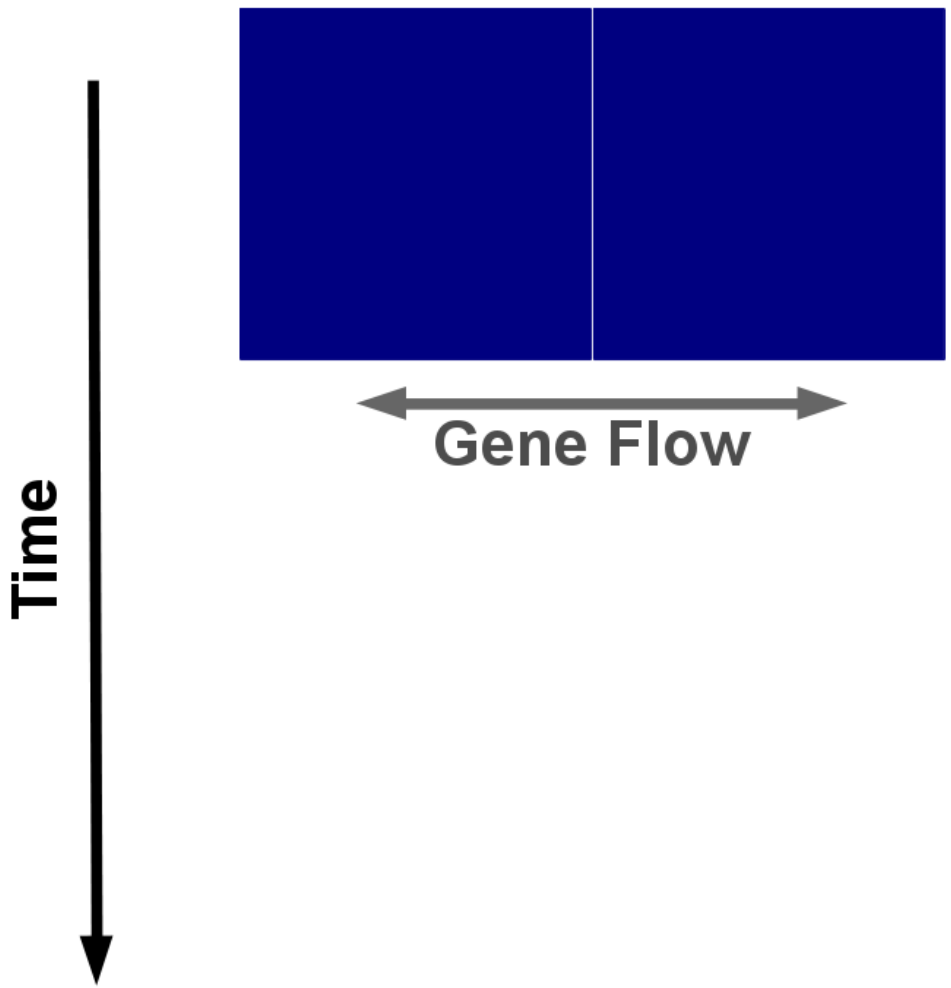
Allopatric Speciation



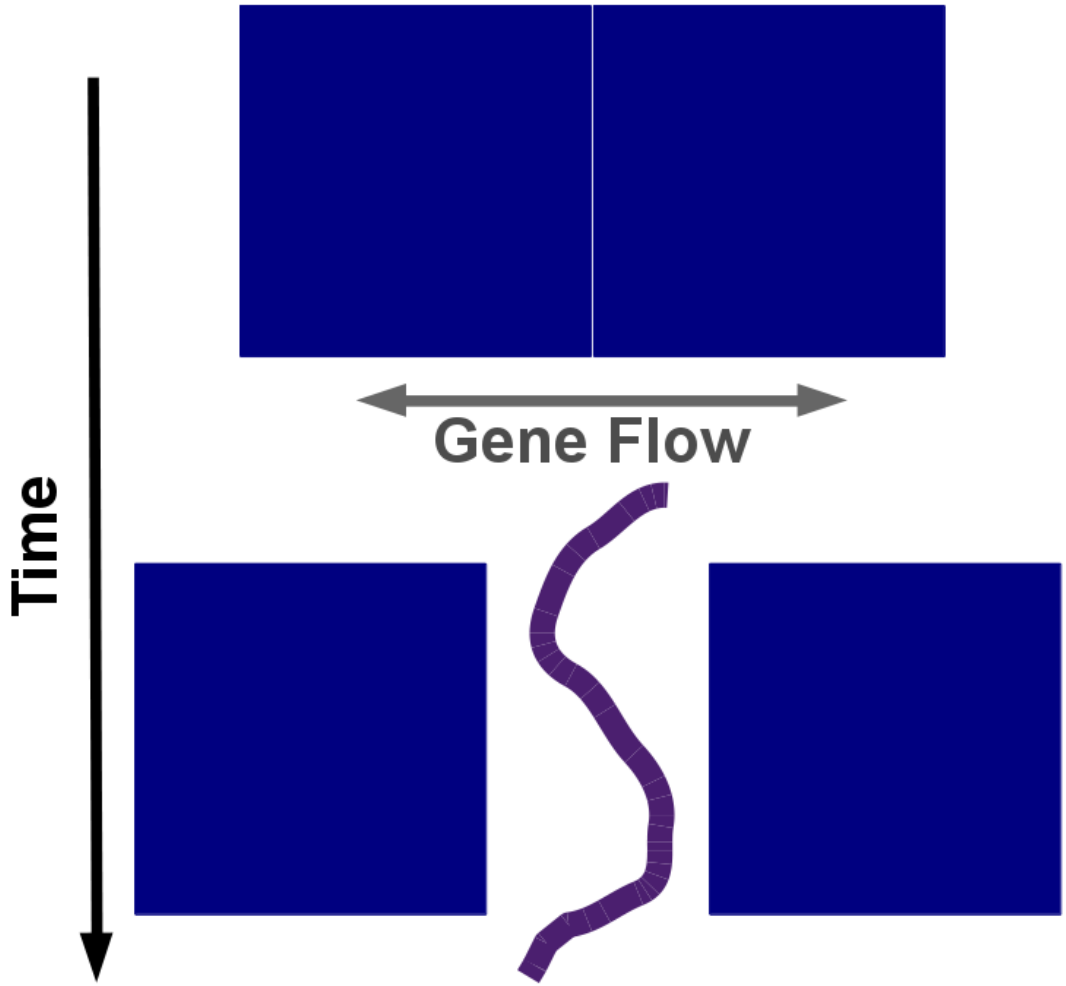
Allopatric Speciation



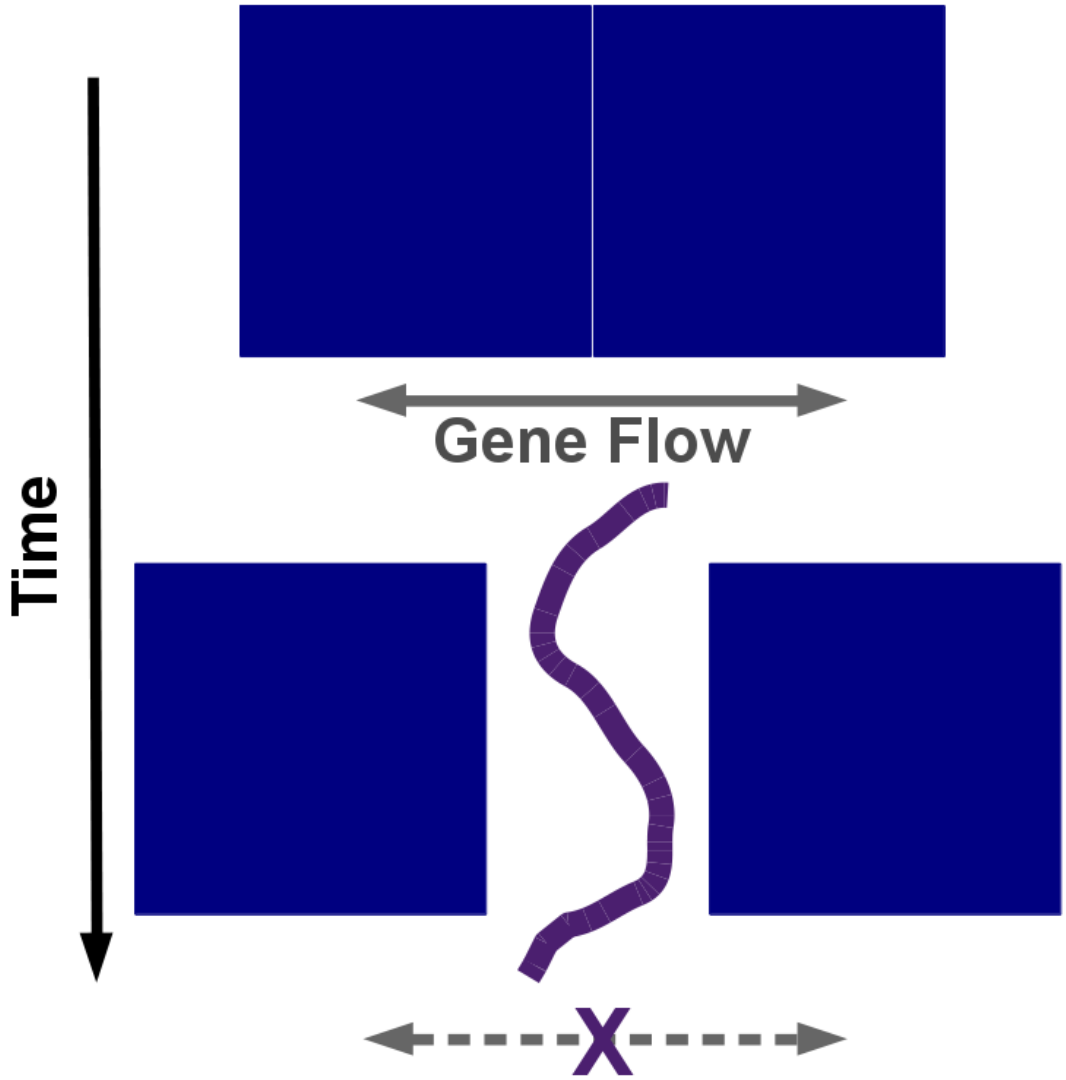
Allopatric Speciation



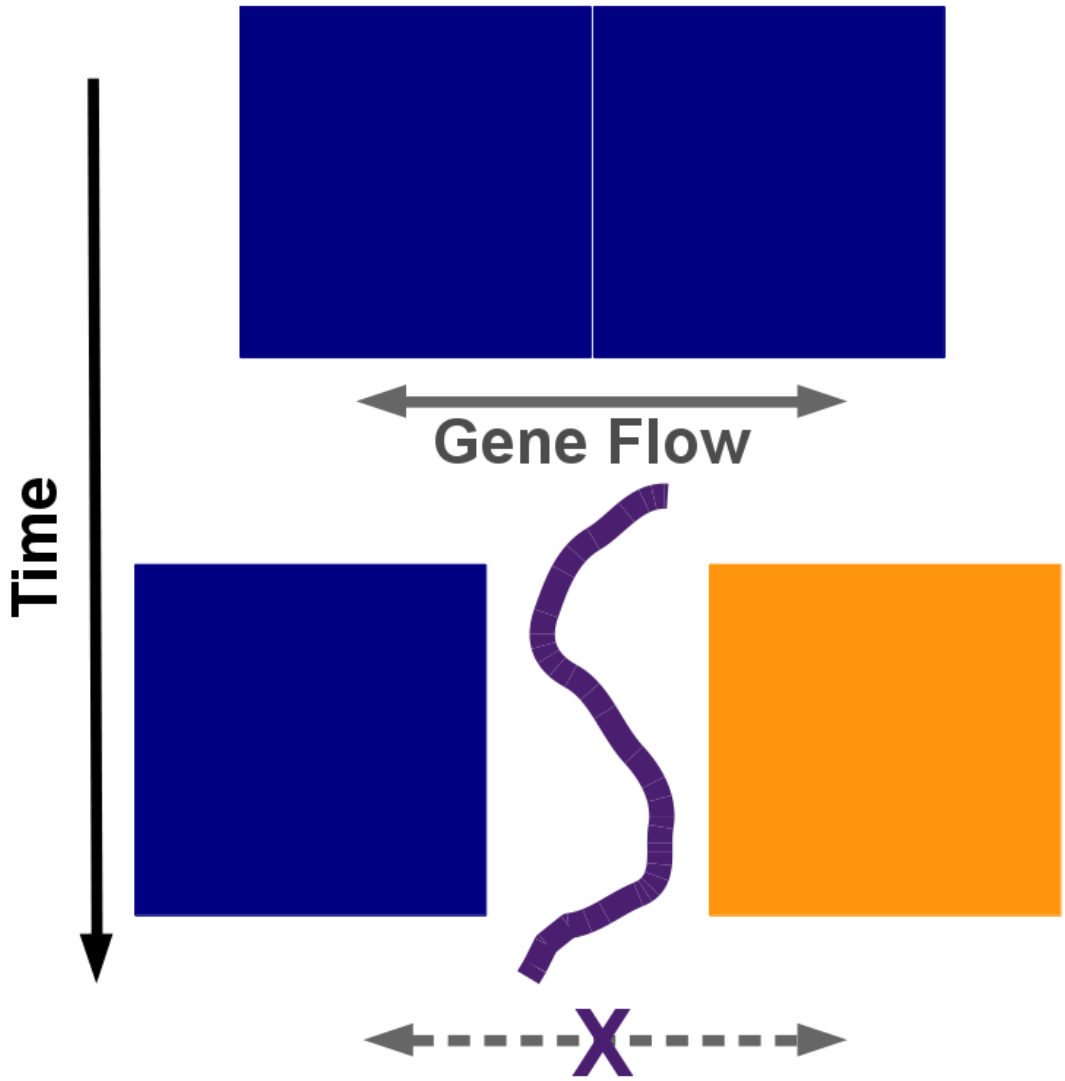
Allopatric Speciation



Allopatric Speciation



Allopatric Speciation



Types of speciation

2 main types of speciation in a biogeographic context

1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow
   - 2 incipient species

1. Sympatric Speciation

Types of speciation

2 main types of speciation in a biogeographic context

1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow
   - 2 incipient species

1. Sympatric Speciation
   - no species range interuption

Types of speciation

2 main types of speciation in a biogeographic context

1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow
   - 2 incipient species

1. Sympatric Speciation
   - no species range interuption
   - proceeds with gene flow

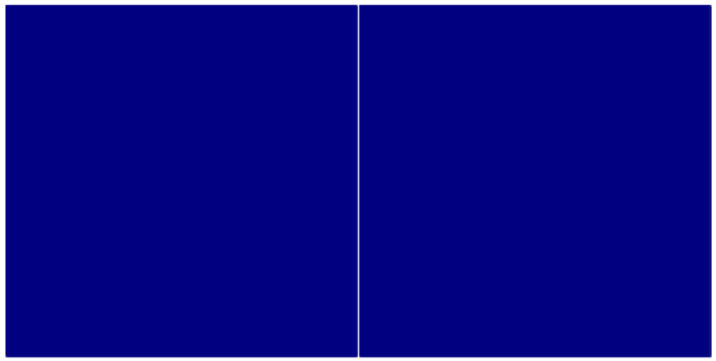
Types of speciation

2 main types of speciation in a biogeographic context

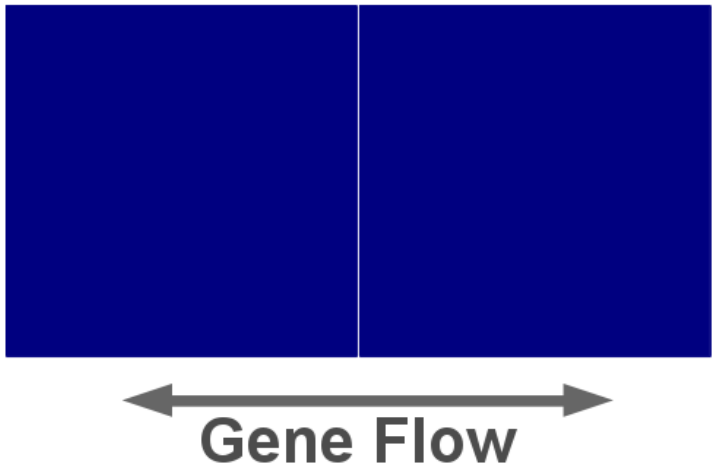
1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow
   - 2 incipient species

1. Sympatric Speciation
   - no species range interuption
   - proceeds with gene flow
   - 2 incipient species

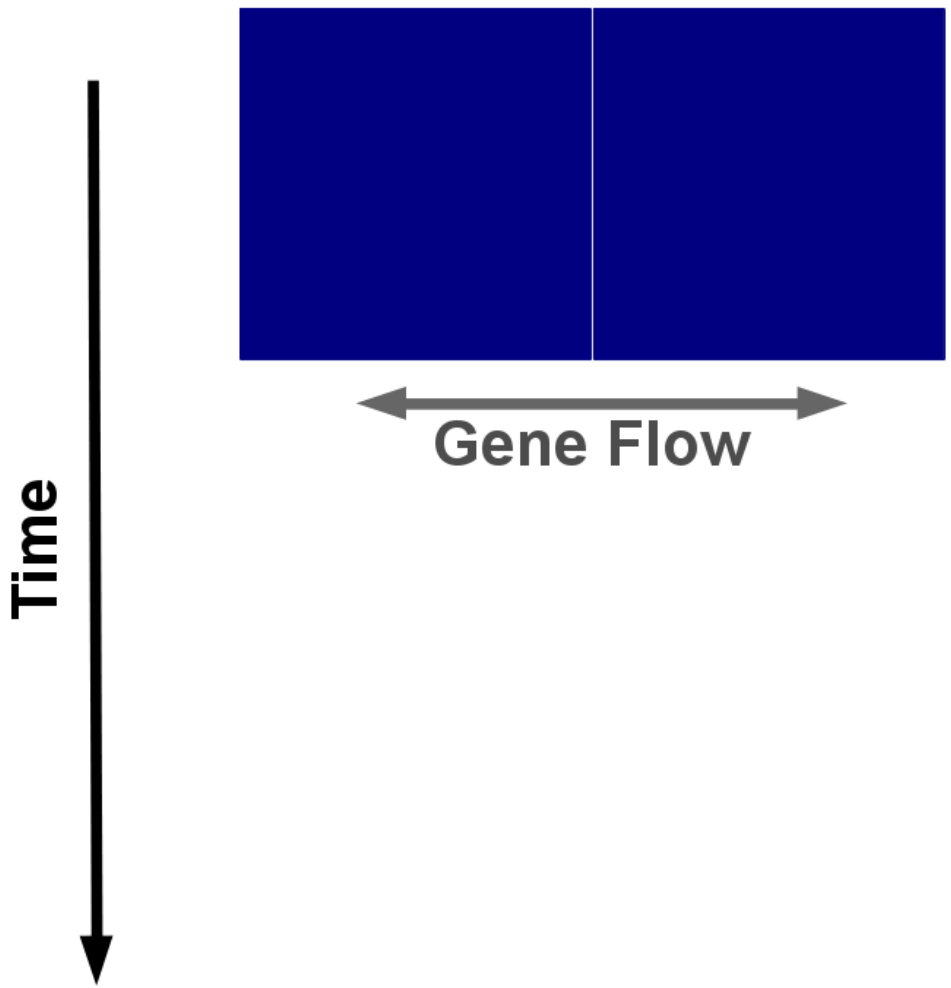
Sympatric Speciation



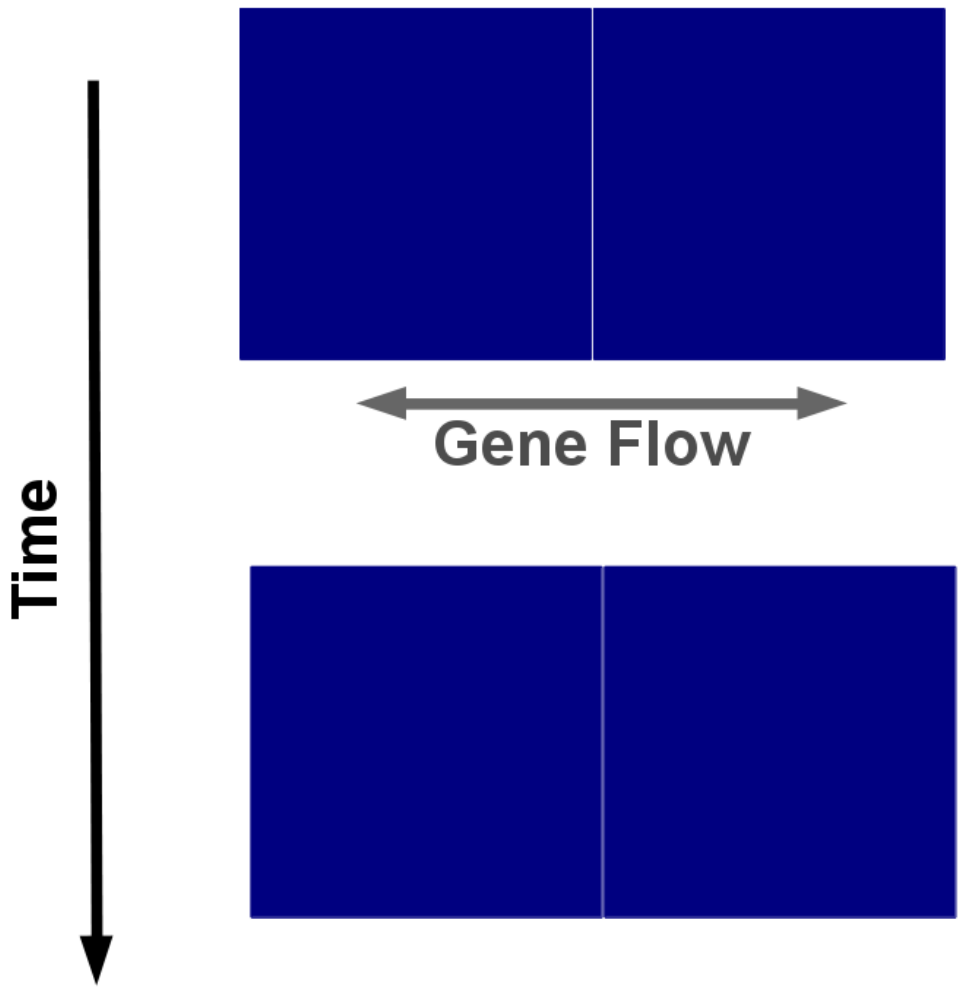
Sympatric Speciation



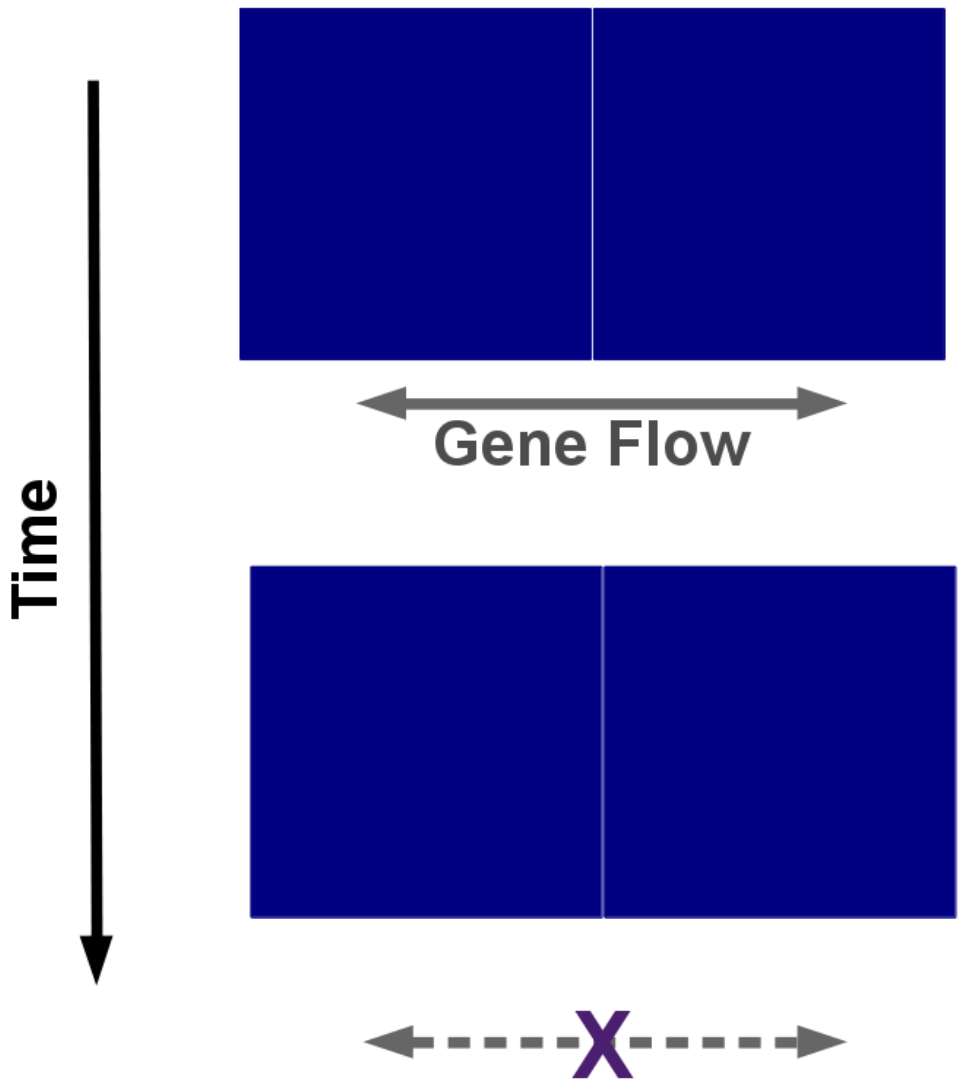
Sympatric Speciation



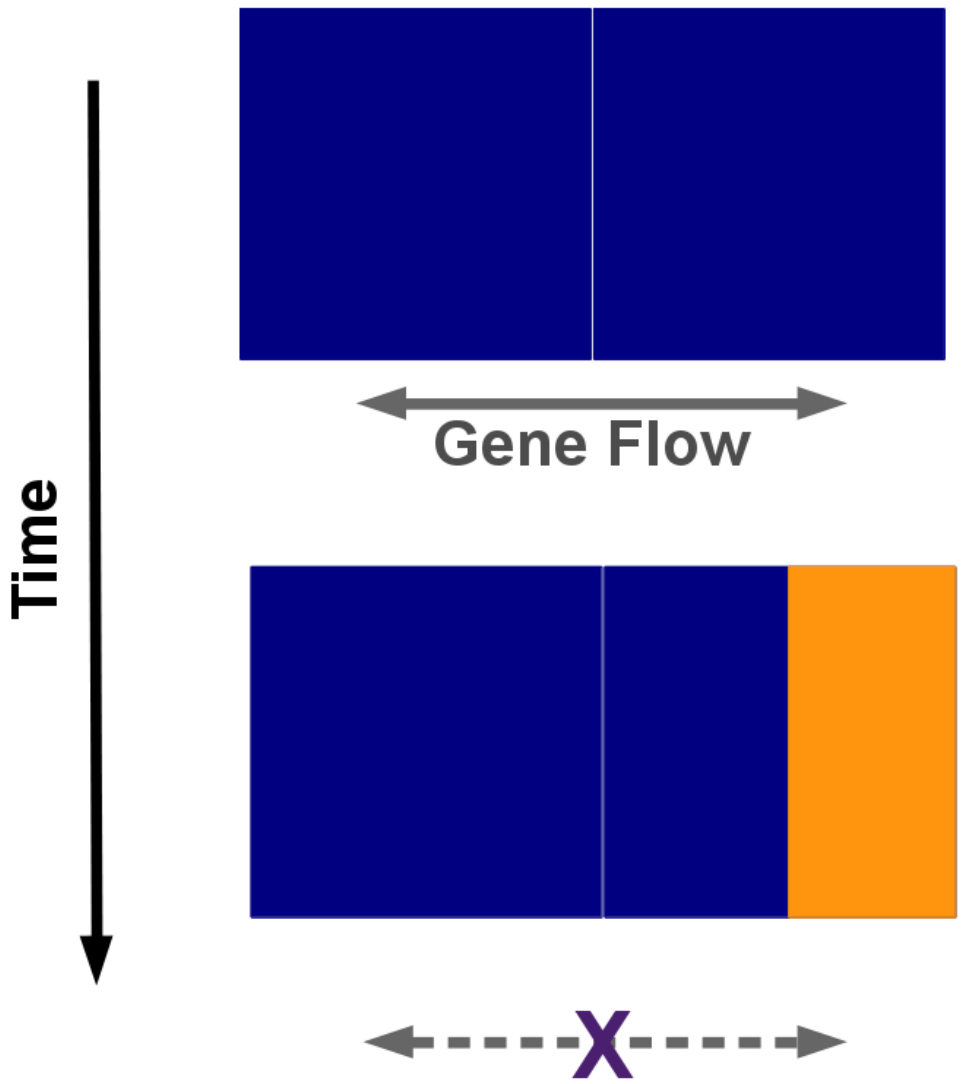
Sympatric Speciation



Sympatric Speciation



Sympatric Speciation



Types of speciation

1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow
   - 2 incipient species

2. Sympatric Speciation
   - no geographic interuption
   - proceeds with geneflow
   - 2 incipient species

Types of speciation

1. Allopatric Speciation
   - interupted species range (i.e., stream)
   - decreased migration and gene flow
   - 2 incipient species

2. Sympatric Speciation
   - no geographic interuption
   - proceeds with geneflow
   - 2 incipient species
   - Reinforcement recently shown to be important

What is reinforcement?

What is reinforcement?

1. 2 previously allopatric populations come into contact

What is reinforcement?

1. 2 previously allopatric populations come into contact
2. Speciation process not complete

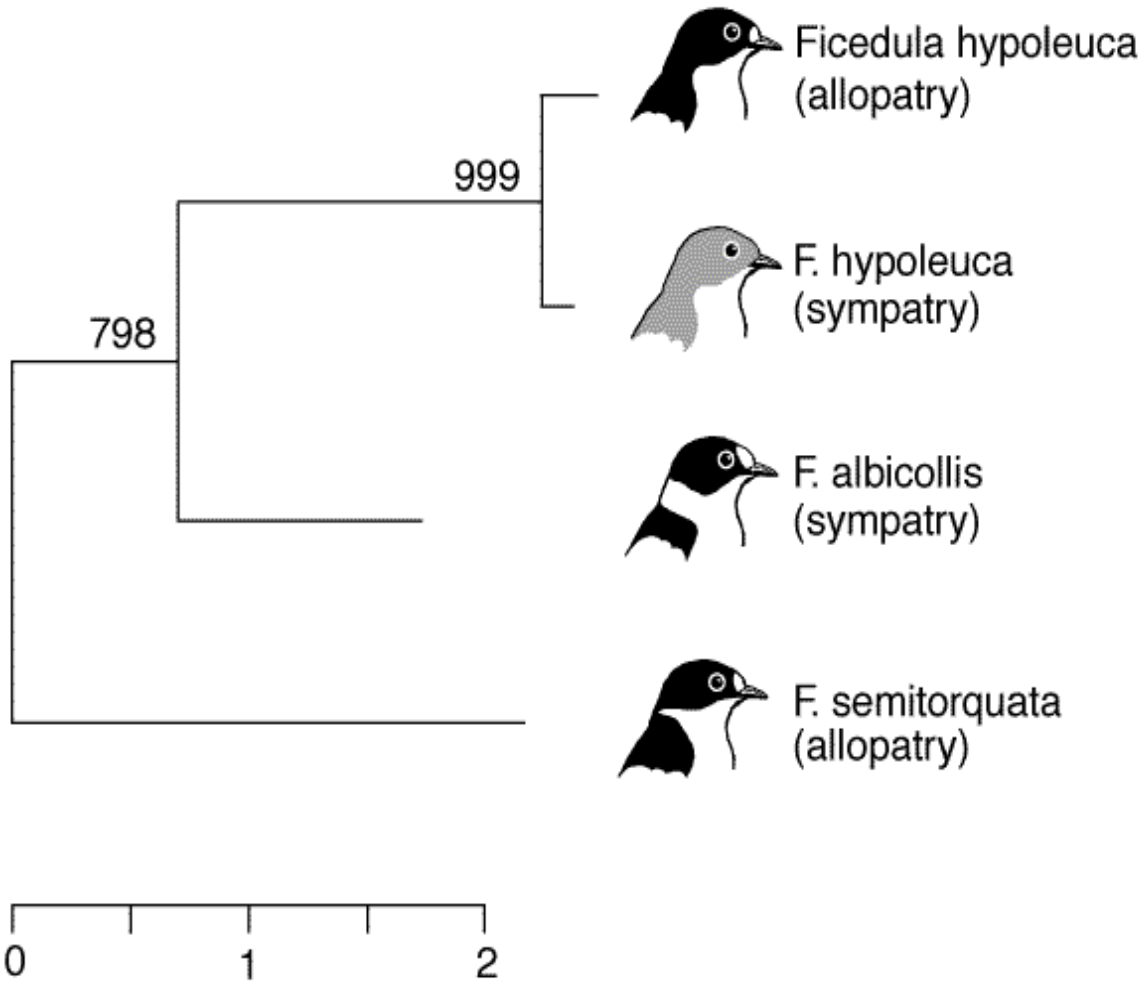
What is reinforcement?

1. 2 previously allopatric populations come into contact
2. Speciation process not complete
3. There is selection against hybrids

What is reinforcement?

1. 2 previously allopatric populations come into contact
2. Speciation process not complete
3. There is selection against hybrids
    - Reproductive character displacement

Flycatchers



Stre et al. 1997

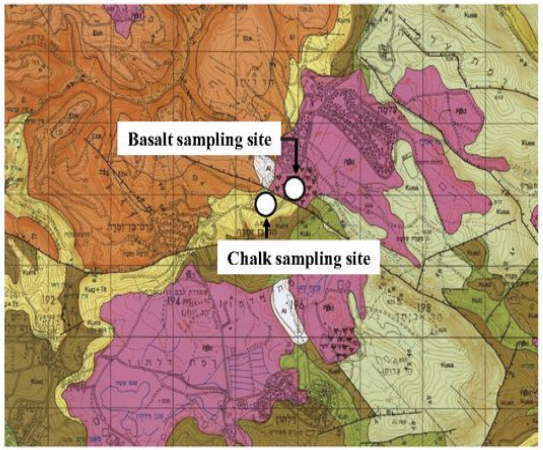
What is reinforcement?

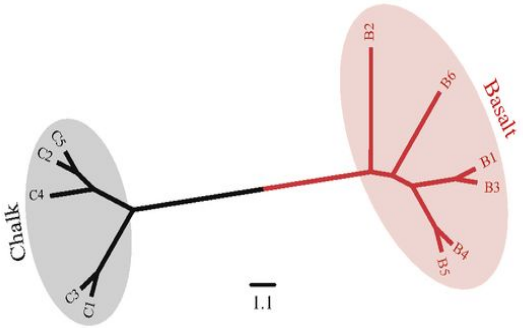
1. 2 previously allopatric populations come into contact
2. Speciation process not complete
3. There is selection against hybrids
    - Reproductive character displacement

What is reinforcement?

1. 2 previously allopatric populations come into contact
2. Speciation process not complete
3. There is selection against hybrids
    - Reproductive character displacement
    - Ecological character displacement

Naked mole rat





Li et al. 2015

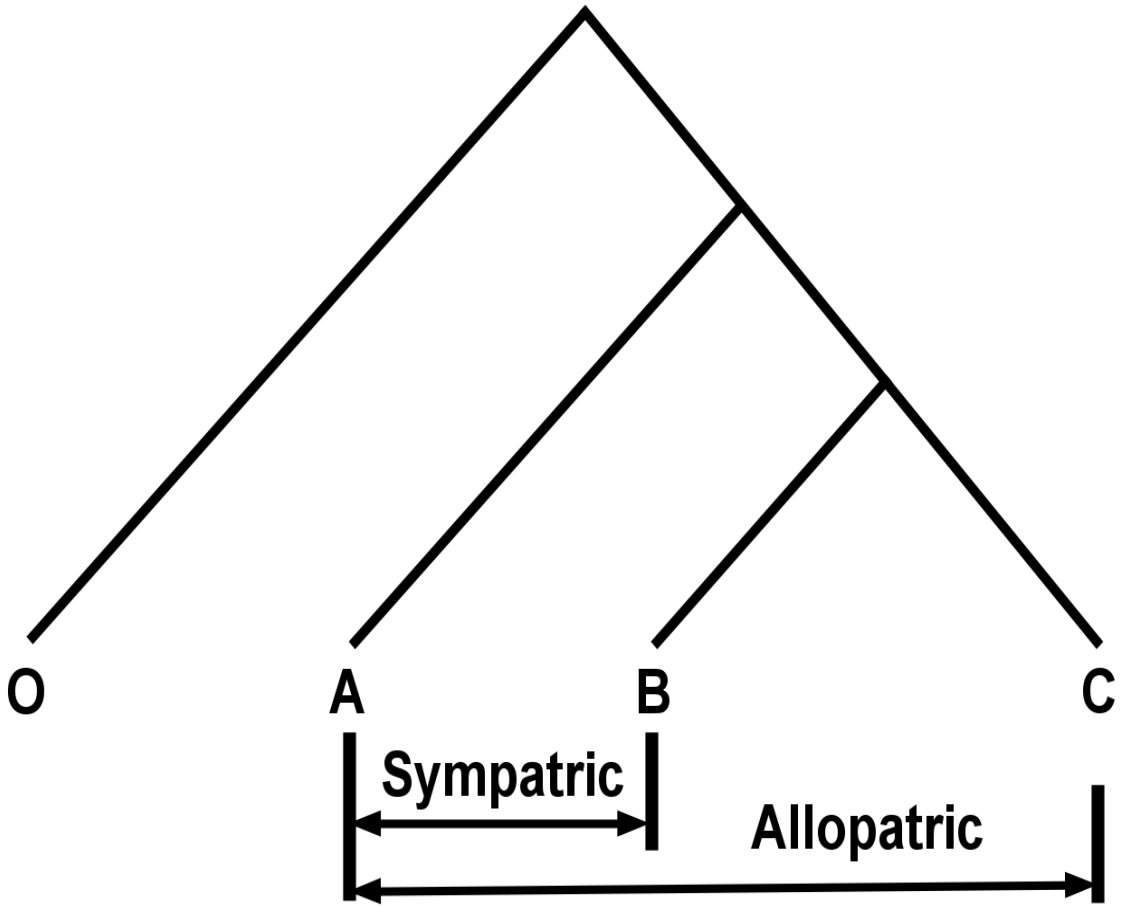
What is reinforcement?

1. 2 previously allopatric populations come into contact
2. Speciation process not complete
3. There is selection against hybrids
   - Reproductive character displacement
   - Ecological character displacement

What is reinforcement?

1. 2 previously allopatric populations come into contact
2. Speciation process not complete
3. There is selection against hybrids
   - Reproductive character displacement
   - Ecological character displacement
4. 2 incipient species result

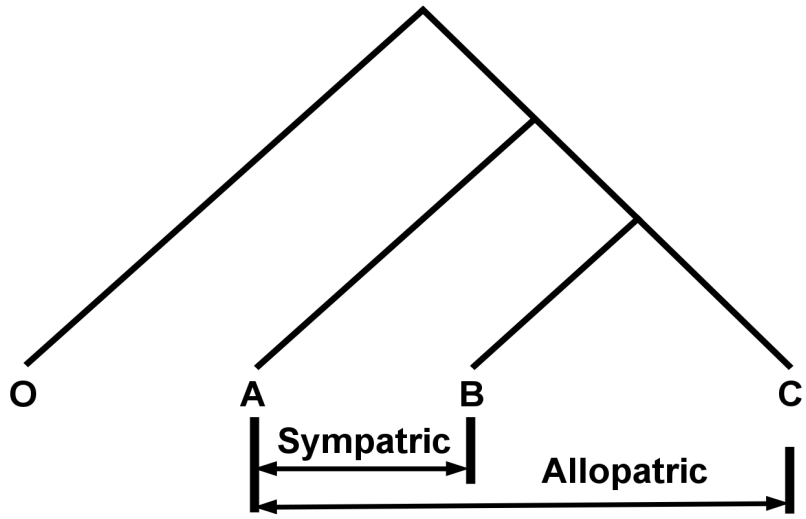
Biogeographic and phylogenetic test for reinforcement



Noor 1997

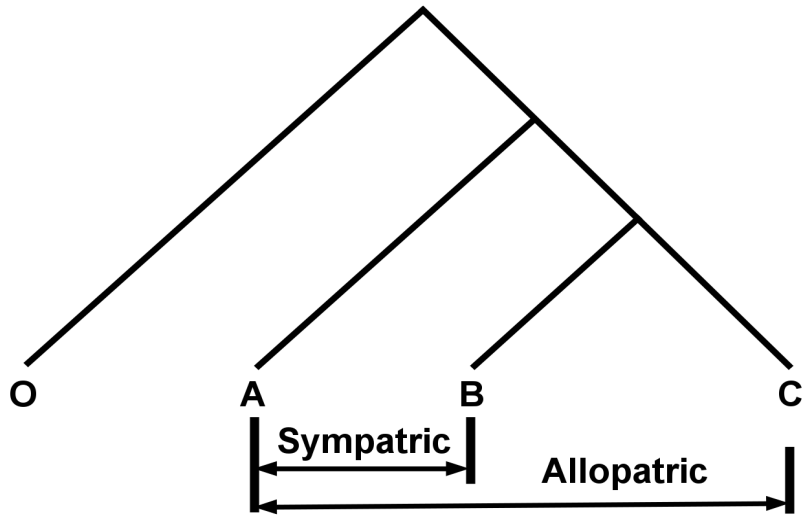
Biogeographic and phylogenetic test for reinforcement

- Species B more often different from A than C



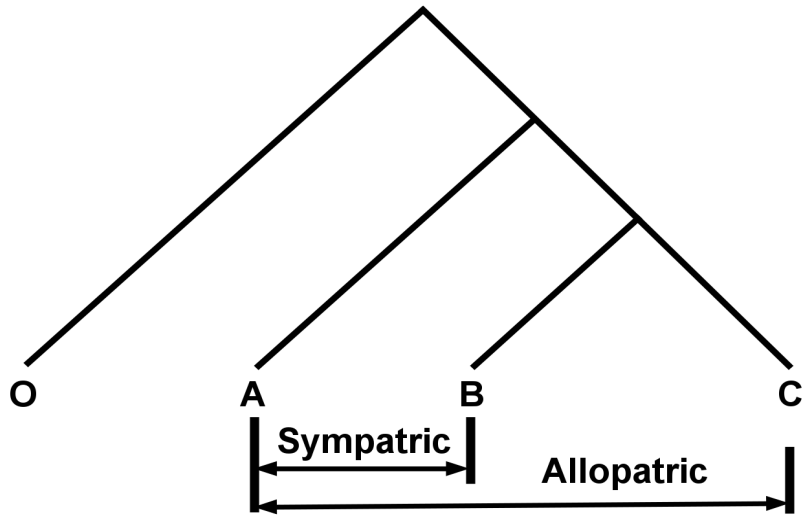
Biogeographic and phylogenetic test for reinforcement

- Species B more often different from A than C
- Originally for reproductive isolation



Biogeographic and phylogenetic test for reinforcement

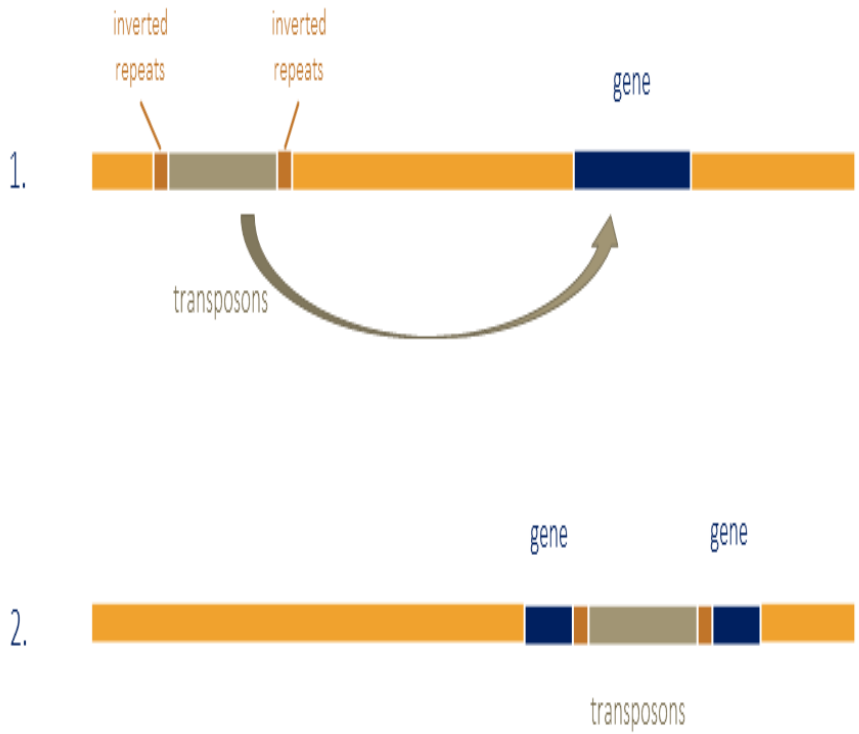
- Species B more often different from A than C
- Originally for reproductive isolation
- Logic of test can be applied to genetic variants



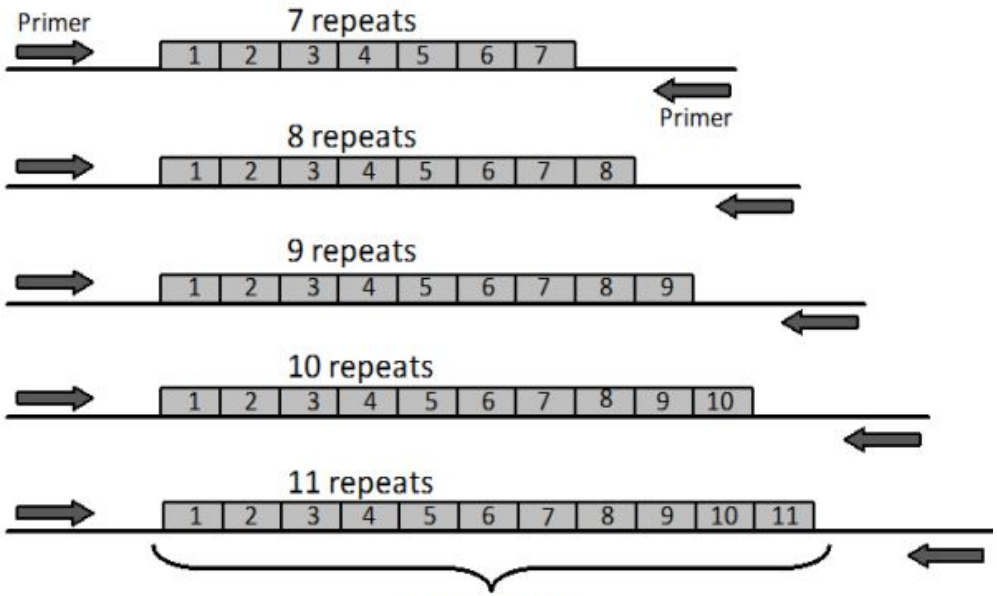
Genetic variants

- What exactly are genetic variants?

Genetic variants: Transposable elements



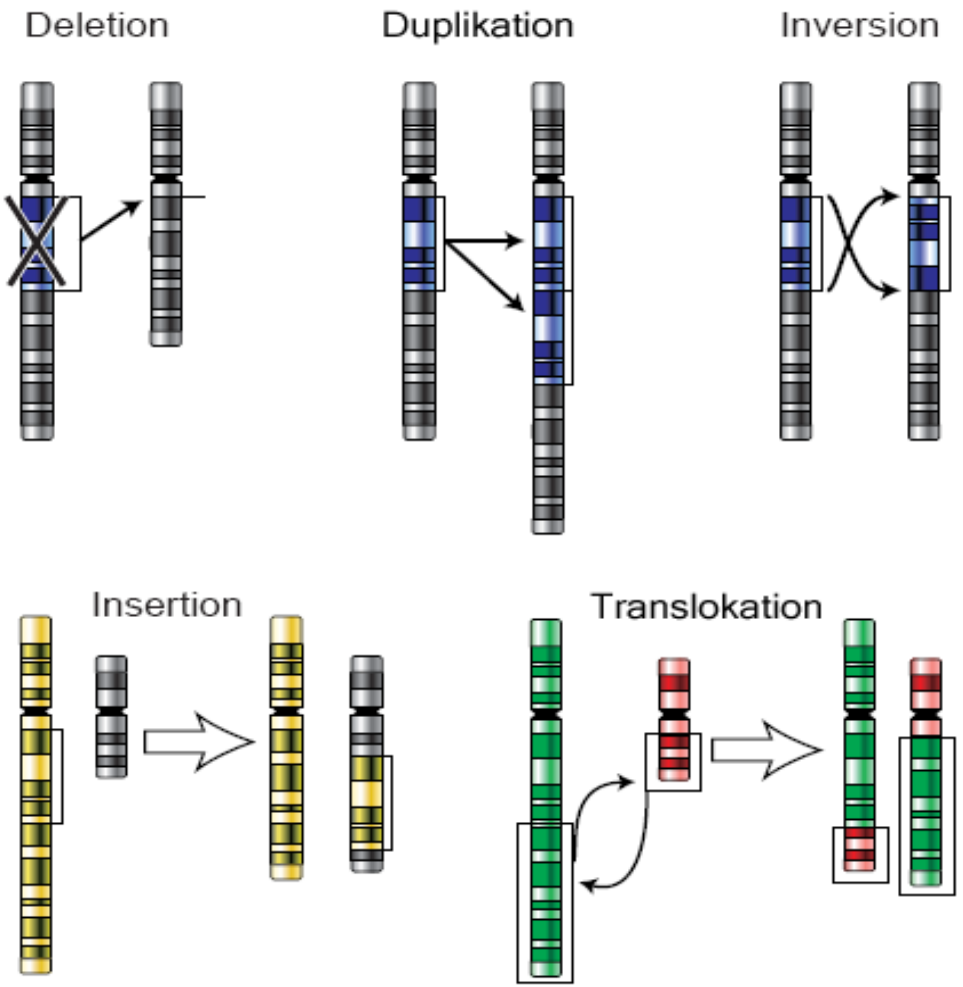
Genetic variants: Short tandem repeats



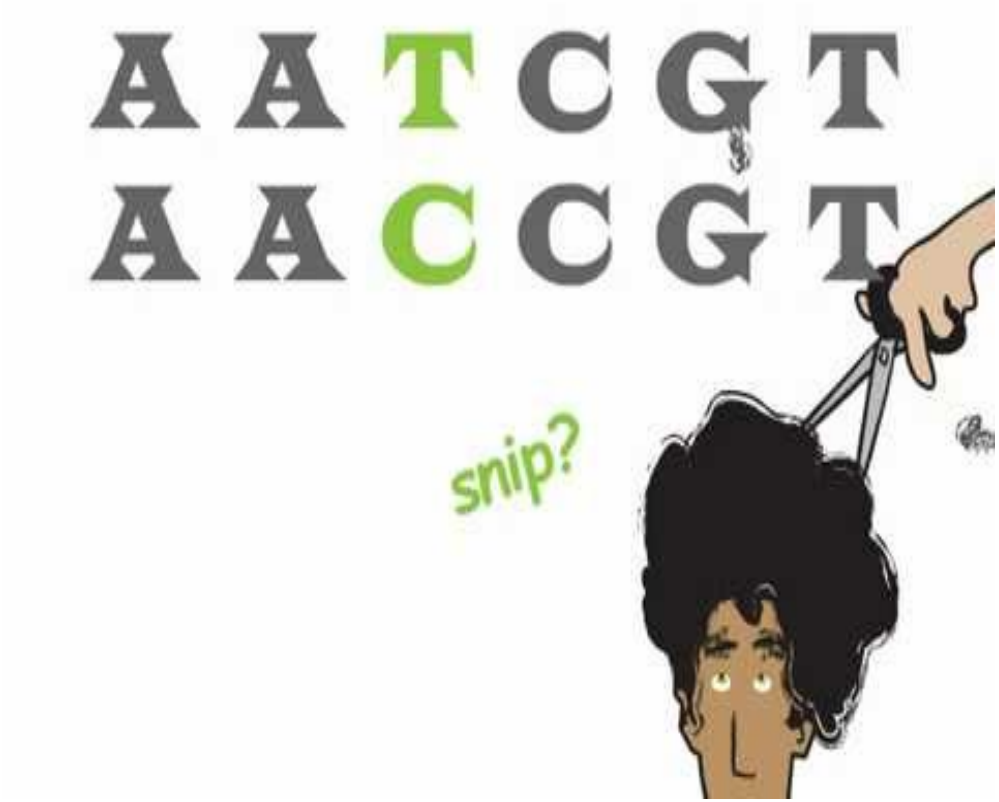
1 2 3 ...

- 2-nucletotide repeat unit : (CA)(CA)(CA)····
- 3 -nucletotide repeat unit : (GCC)(GCC)(GCC) ····
- 4 -nucletotide repeat unit : (AATG)(AATG)(AATG) ····
- 5 -nucletotide repeat unit : (AGAAA)(AGAAA) ····

Genetic variants: Structural Variants



Genetic variants: Single nucleotide polymorphisms



Objectives and Hypotheses

- Objective: Use a diverse set of taxa with genomic data in online biological collections to investigate the relationship of genetic variants with biogeography.

Objectives and Hypotheses

- Objective: Use a diverse set of taxa with genomic data in online biological collections to investigate the relationship of genetic variants with biogeography.

- H1: Complex genomic variants will show the pattern of reinforcement.

Objectives and Hypotheses

- Objective: Use a diverse set of taxa with genomic data in online biological collections to investigate the relationship of genetic variants with biogeography.

- H1: Complex genomic variants will show the pattern of reinforcement.

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

Topic

Identify species groups

Identify species groups

- Conducted literature review

Identify species groups

- Conducted literature review
  - phylogeny

Identify species groups

- Conducted literature review
  - phylogeny
  - biogeographic context

Identify species groups

- Conducted literature review
  - phylogeny
  - biogeographic context
  - whole genome sequencing

Identify species groups

- Conducted literature review
  - phylogeny
  - biogeographic context
  - whole genome sequencing
- Publically avaliable data (e.g., NCBI)

Identify species groups

- Conducted literature review
  - phylogeny
  - biogeographic context
  - whole genome sequencing
- Publically avaliable data (e.g., NCBI)
- Appropriate phylogeny, biogeographic context, and whole genome sequencing

Organism	Acession Numbers
Mosquitoes (Anopheles)	NCBI: PRJNA6751 and PRJNA254046
Horses (Equus)	ENA: PRJEB7446
Butterflies (Heliconius)	ENA: ERP002440
Flycatchers (Ficula)	ENA: PRJEB7359
Dogs (Canis)	Authors Contacted
Cichlids	NCBI: PRJNA78915, PRJNA60369,PRJNA60363, and PRJNA78185

Genetic Variants

- Download data (NCBI)

Genetic Variants

- Download data (NCBI)
- Use HPCs

Genetic Variants

- Download data (NCBI)
- Use HPCs
- GATK pipline (variant/variant filtration)

Genetic Variants

- Download data (NCBI)
- Use HPCs
- GATK pipline (variant/variant filtration)
  - SNPs

Genetic Variants

- Download data (NCBI)
- Use HPCs
- GATK pipline (variant/variant filtration)
  - SNPs
- Computationionally identify complex genetic variants

Genetic Variants

- Download data (NCBI)
- Use HPCs
- GATK pipline (variant/variant filtration)
  - SNPs
- Computationionally identify complex genetic variants
  - TEs

Genetic Variants

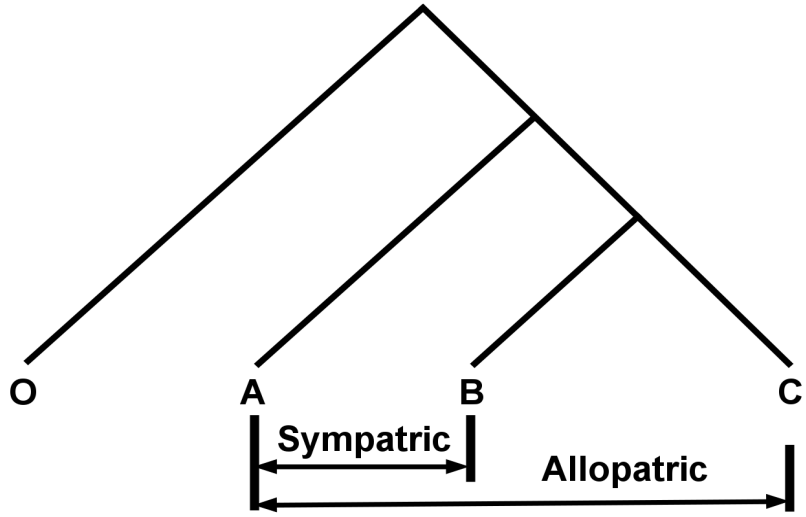
- Download data (NCBI)
- Use HPCs
- GATK pipline (variant/variant filtration)
  - SNPs
- Computationionally identify complex genetic variants
  - TEs
  - STRs

Genetic Variants

- Download data (NCBI)
- Use HPCs
- GATK pipline (variant/variant filtration)
  - SNPs
- Computationionally identify complex genetic variants
  - TEs
  - STRs
  - SVs

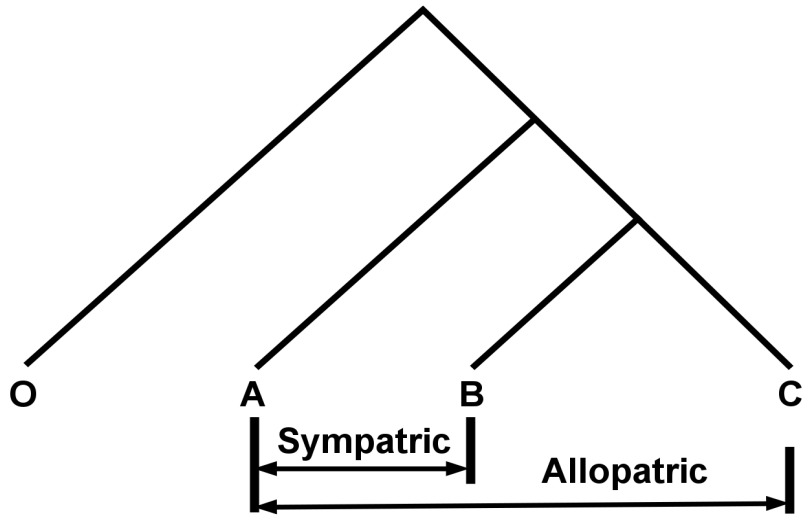
Phylogenetic test for reinforcement

- Restrictions:



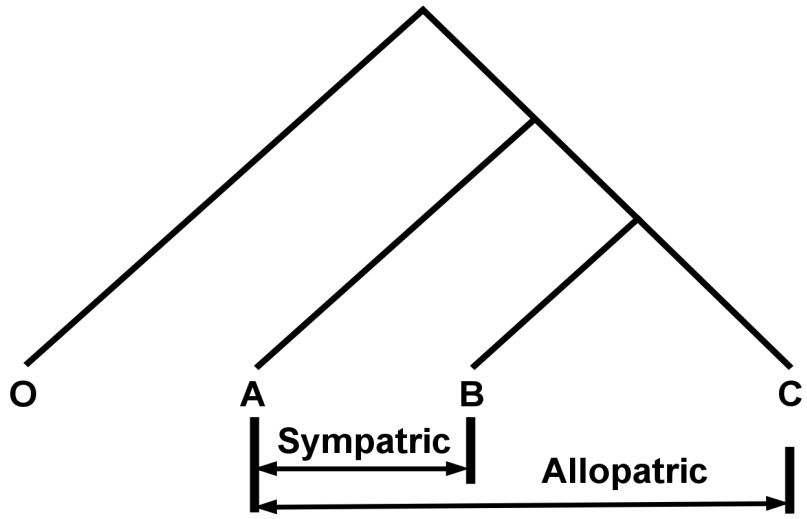
Phylogenetic test for reinforcement

- Restrictions:
  - C Allopatric to all other closely related species



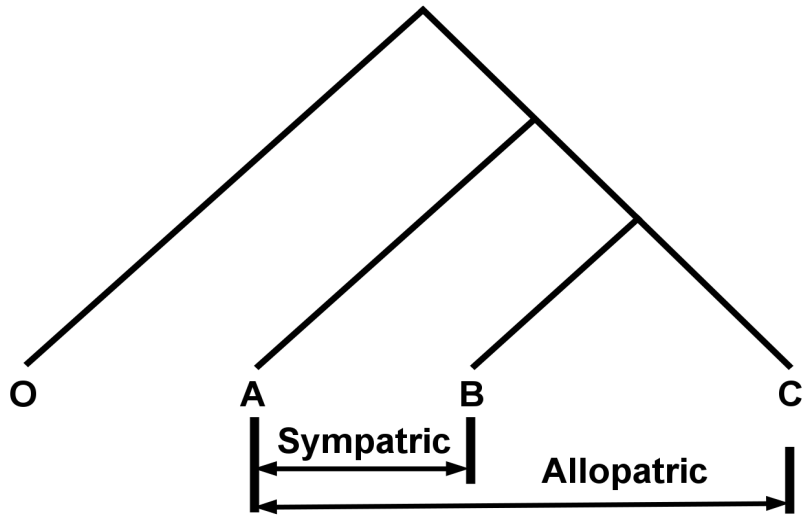
Phylogenetic test for reinforcement

- Restrictions:
  - C Allopatric to all other closely related species
  - but B and C some overlap



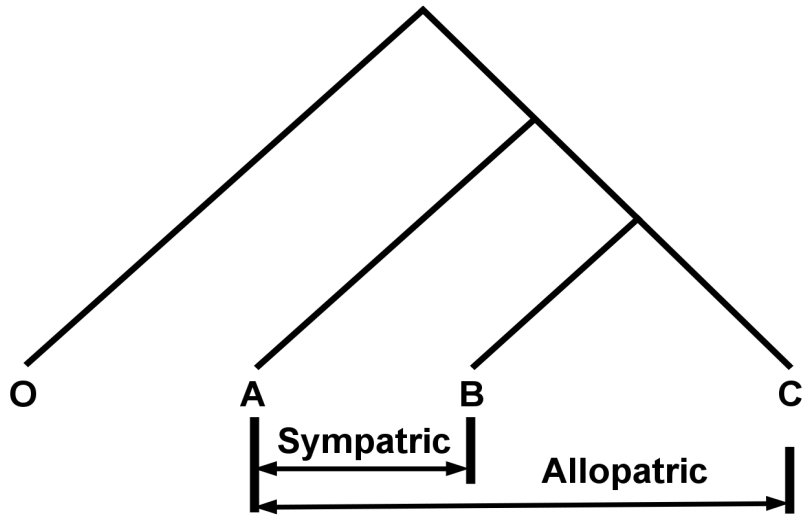
Phylogenetic test for reinforcement

- Restrictions:
  - C Allopatric to all other closely related species
  - but B and C some overlap
    - effects of sympatry shared



Phylogenetic test for reinforcement

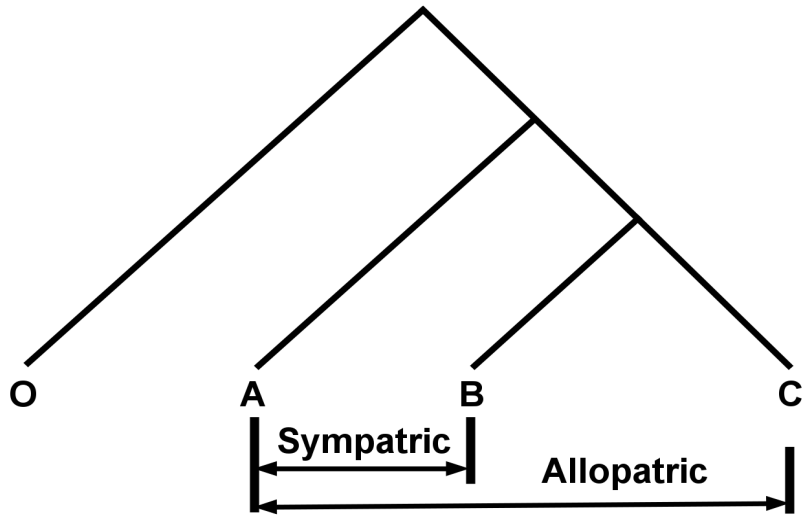
- Restrictions:
  - C Allopatric to all other closely related species
  - but B and C some overlap
    - effects of sympatry shared
- outgroup (O) to restrict search for genetic variants to those derived after most recent common ancestor



Phylogenetic test for reinforcement

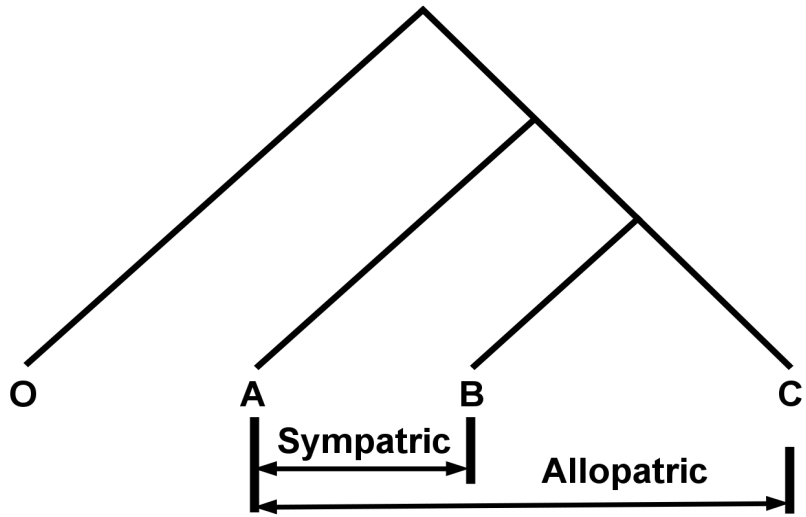
- Restrictions:
  - C Allopatric to all other closely related species
  - but B and C some overlap
    - effects of sympatry shared
- outgroup (O) to restrict search for genetic variants to those derived after most recent common ancestor
-

$$D = \frac{\sum_{n=1}^{i} Sym_i - Allo_i}{\sum_{n=1}^{i} Sym_i + Allo_i}$$



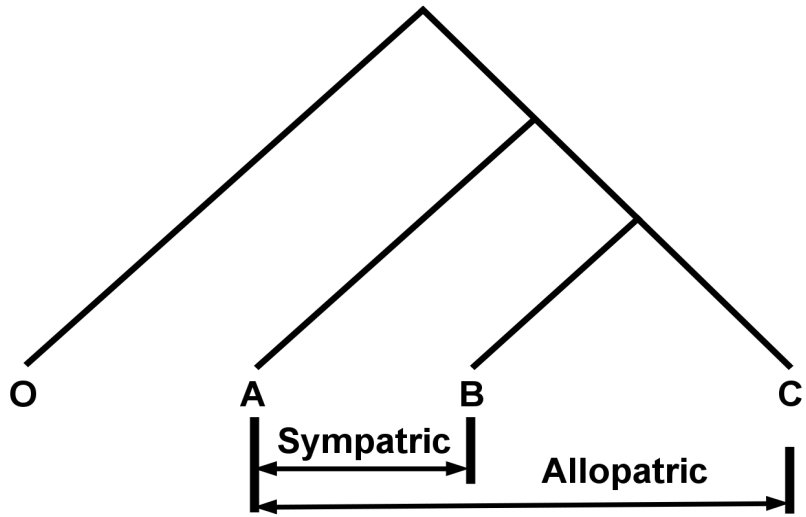
Phylogenetic test for reinforcement

- Restrictions:
  - C Allopatric to all other closely related species
  - but B and C some overlap
    - effects of sympatry shared
- outgroup (O) to restrict search for genetic variants to those derived after most recent common ancestor
- $$D = \frac{\sum_{n=1}^{i} Sym_i - Allo_i}{\sum_{n=1}^{i} Sym_i + Allo_i}$$
- $1 > D > -1$



Phylogenetic test for reinforcement

- Restrictions:
  - C Allopatric to all other closely related species
  - but B and C some overlap
    - effects of sympatry shared
- outgroup (O) to restrict search for genetic variants to those derived after most recent common ancestor
- $$D = \frac{\sum_{n=1}^{i} Sym_i - Allo_i}{\sum_{n=1}^{i} Sym_i + Allo_i}$$
- $1 > D >$ -1
  - Reinforcement $D > 0$



Reinforcement: Complex Genetic Variants

- H1: Complex genomic variants will show the pattern of reinforcement.

Reinforcement: Complex Genetic Variants

- H1: Complex genomic variants will show the pattern of reinforcement.

- Genome wide D

Reinforcement: Complex Genetic Variants

- H1: Complex genomic variants will show the pattern of reinforcement.

- Genome wide D
- Novel permutation proceedure to assess significance

Reinforcement: Complex Genetic Variants

- H1: Complex genomic variants will show the pattern of reinforcement.

- Genome wide D
- Novel permutation proceedure to assess significance
- Predictions

Reinforcement: Complex Genetic Variants

- H1: Complex genomic variants will show the pattern of reinforcement.

- Genome wide D
- Novel permutation proceedure to assess significance
- Predictions
  - Reinforcement $D > 0$

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$
  - Use GO enrichment analysis

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$
  - Use GO enrichment analysis
- Predictions

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$
  - Use GO enrichment analysis
- Predictions
  - Significantly enriched GO terms ecological character displacement

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$
  - Use GO enrichment analysis
- Predictions
  - Significantly enriched GO terms ecological character displacement
    - e.g., Habitat preferences

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$
  - Use GO enrichment analysis
- Predictions
  - Significantly enriched GO terms ecological character displacement
    - e.g., Habitat preferences
- Alternatives

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$
  - Use GO enrichment analysis
- Predictions
  - Significantly enriched GO terms ecological character displacement
    - e.g., Habitat preferences
- Alternatives
  - Significantly enriched GO terms reproductive character displacement

Reinforcement: Ecological Character Displacement

- H2: SNPs representative of reinforcement will be associated with functions indicative of ecological character displacement.

- D in sliding windows
- Novel permutation proceedure to assess significance
- Use Gene Ontology to assign functions where $D > 0$
  - Use GO enrichment analysis
- Predictions
  - Significantly enriched GO terms ecological character displacement
    - e.g., Habitat preferences
- Alternatives
  - Significantly enriched GO terms reproductive character displacement
    - e.g., sperm development

Topic

Significance

- taxonomic breadth

Significance

- taxonomic breadth
- genetic variants (complex variants and SNPs)

Significance

- taxonomic breadth
- genetic variants (complex variants and SNPs)
- Potential to transform our understanding of speciation's relationship with biogeography

Significance

- taxonomic breadth
- genetic variants (complex variants and SNPs)
- Potential to transform our understanding of speciation's relationship with biogeography
- Reveal impotant basic insights into how biodiversity is produced through speciation

Topic

7th/8th Graders

- Speciation module AUMNH Junior Curator Camp

7th/8th Graders

- Speciation module AUMNH Junior Curator Camp
- Computer Based Bioinformatics lesson

7th/8th Graders

- Speciation module AUMNH Junior Curator Camp
- Computer Based Bioinformatics lesson
- Program to connect online and traditional biological collections

Topic

Questions