

Species as datapoints

Comparative Methods

Biology 683

Heath Blackmon



Joe Felsenstein

The Problem

Does body size impact brain size?

Does flight effect speciation rate?

Does the evolution of some traits make species more likely to go extinct?

Do structures in predators and prey coevolve?

Are the expression level of some genes maintained over evolutionary time scales?

How many times have eyes evolved?

Some questions are hard to ask with experimental studies!

The Problem

An example from R

Solutions

Calculating independent contrasts

Only good for two continuous traits

Treating a phylogeny as a covariance matrix

PGLS

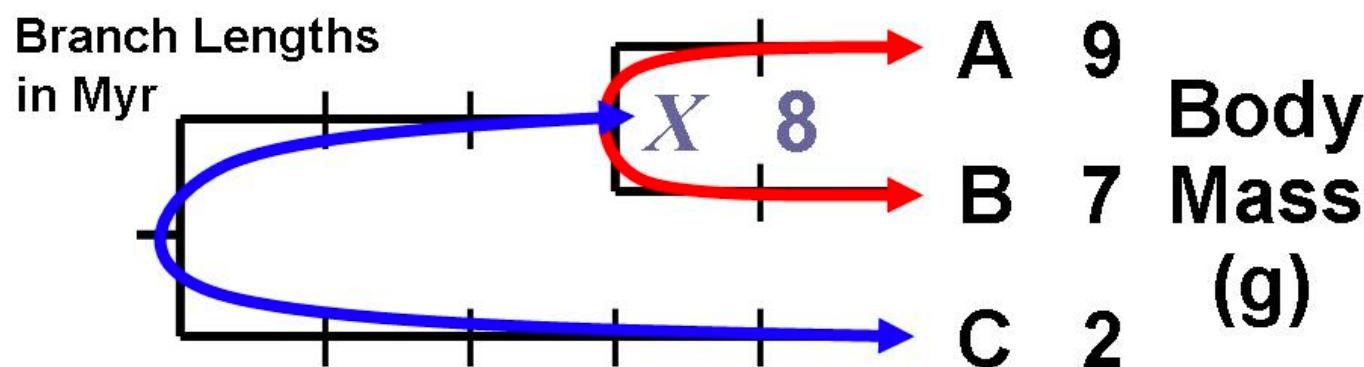
Creating a process based model that explicitly incorporates the phylogeny into the likelihood calculation

BAMM

Diversitree

Simulation using a monte carlo approach

Independent Contrasts



Identify and Compute Independent Contrasts

**Compute square roots of sums of
(corrected) branch lengths = S.D.**

| Contrast | Value | S.D. |
|----------|-------|------|
|----------|-------|------|

| | | |
|-----|---|---|
| A-B | 2 | 2 |
|-----|---|---|

| | | |
|-----|---|---|
| X-C | 6 | 3 |
|-----|---|---|

| Standardized Contrast | | |
|-----------------------|---|--|
| | 1 | |
| | 2 | |

```
library(ape)  
pic(tree, x)
```

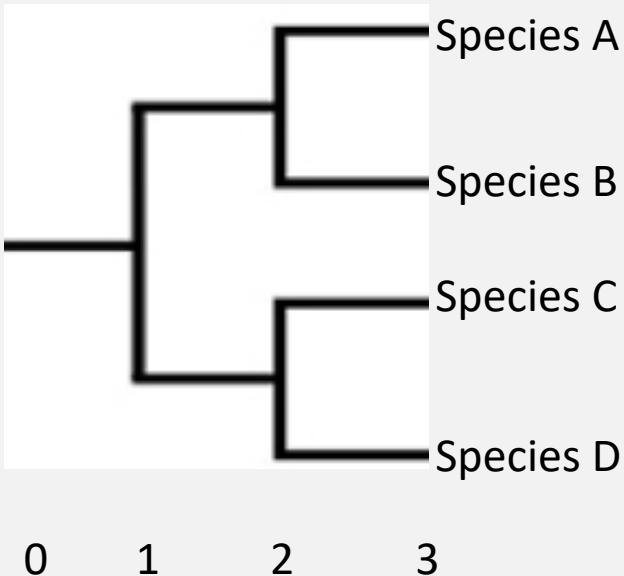
Assumptions

If our trait does not fit a Brownian motion model then we may have biased results.

If our branch lengths are inaccurate we will have error.

If we misplace taxa we will have error and reduced power.

Phylogeny as a covariance matrix



| | A | B | C | D |
|---|---|------|------|------|
| A | 1 | 0.66 | 0.33 | 0.33 |
| B | | 1 | 0.33 | 0.33 |
| C | | | 1 | 0.66 |
| D | | | | 1 |

Mathematically it can be shown that phylogenetic independent contrasts are a special case of this approach

Phylogeny as a covariance matrix

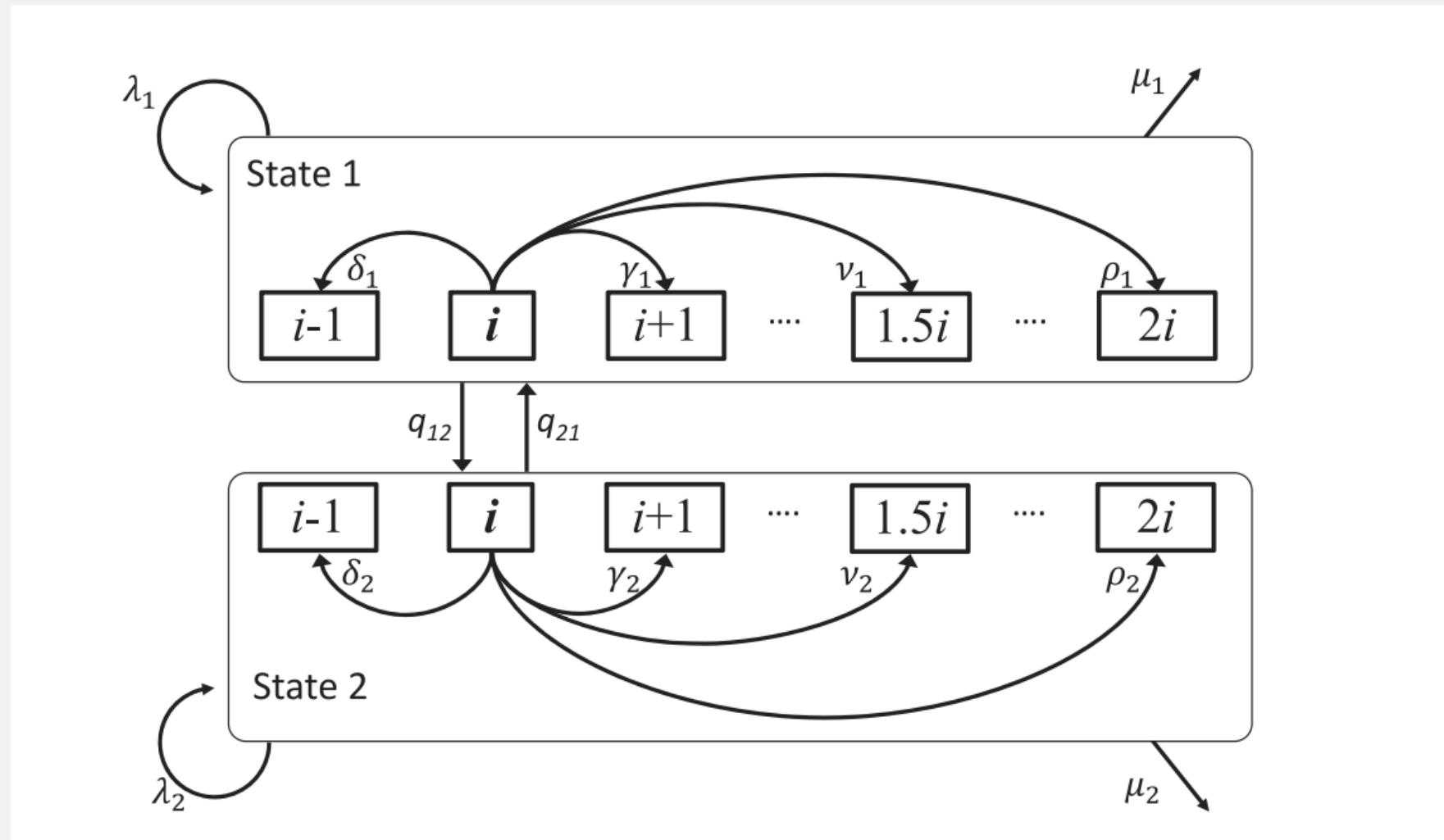
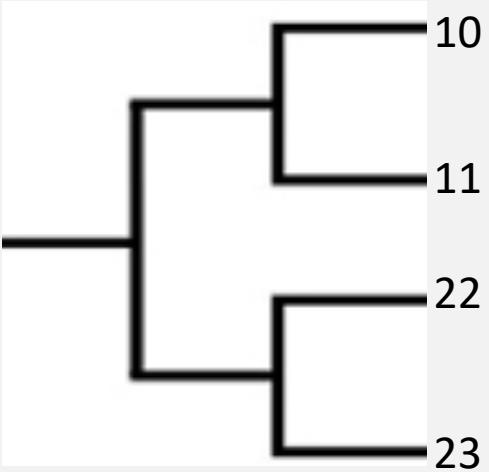
```
library(ape)
Library(nlme)
pglsModel <- gls(y ~ x,
                  correlation = corBrownian(phy),
                  data = data)
```

This is a less restricted approach because it allows us to fit different correlation structures. Still can be difficult/impossible to deal with discrete variables.

Process based model

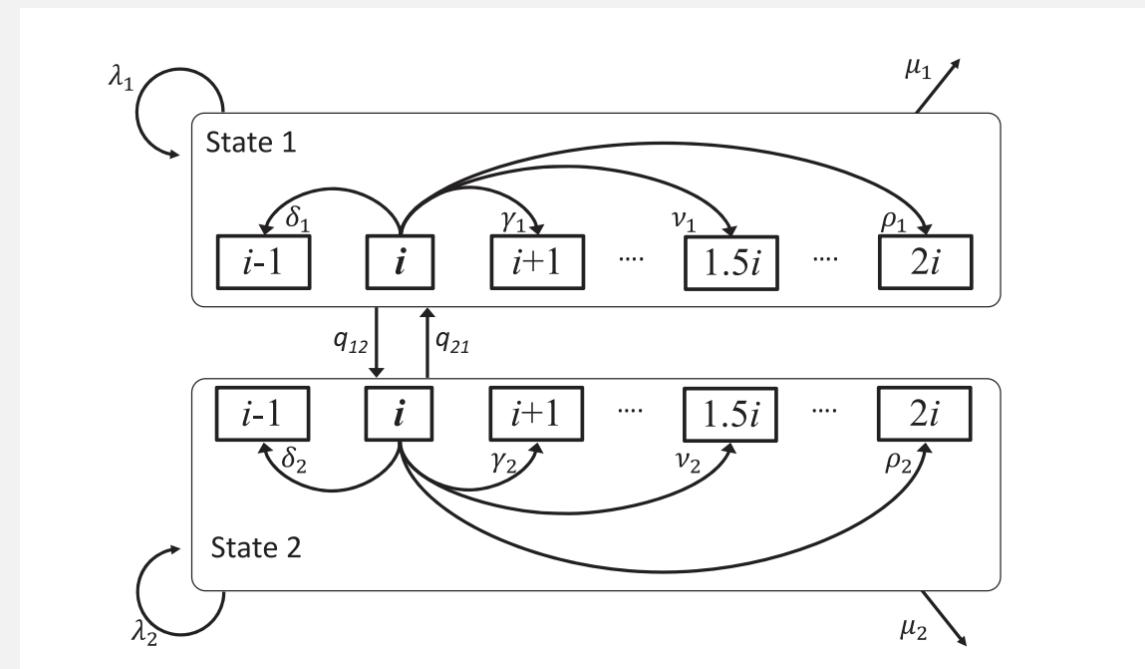
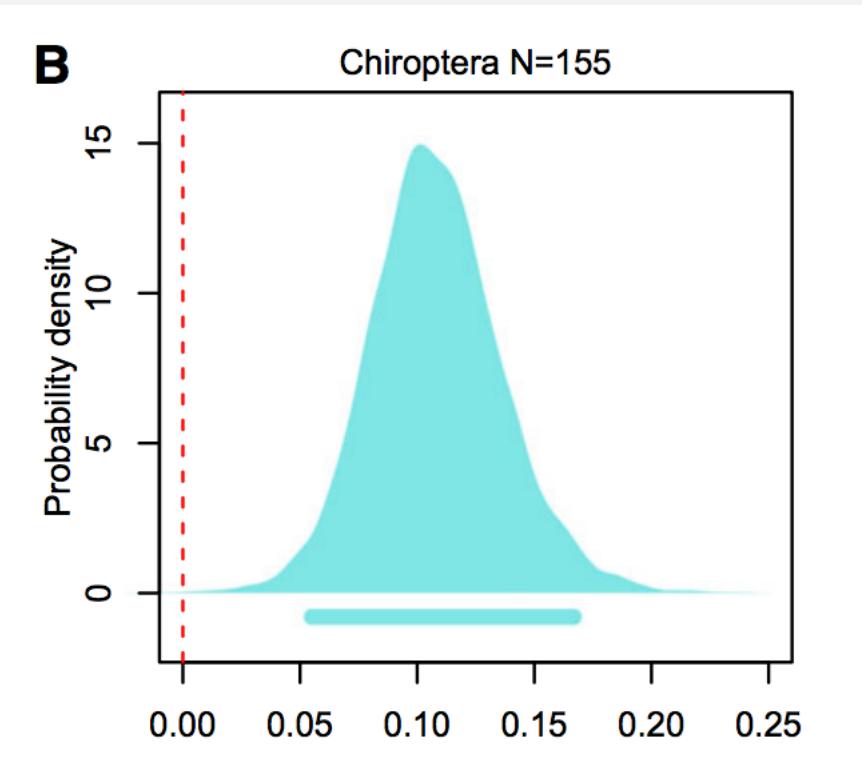
Process based models can be developed where we calculate the likelihood of our data given a model and the tip states. These models can include things like speciation, extinction, sampling bias, discrete and continuous characters.

Example with chromosomes

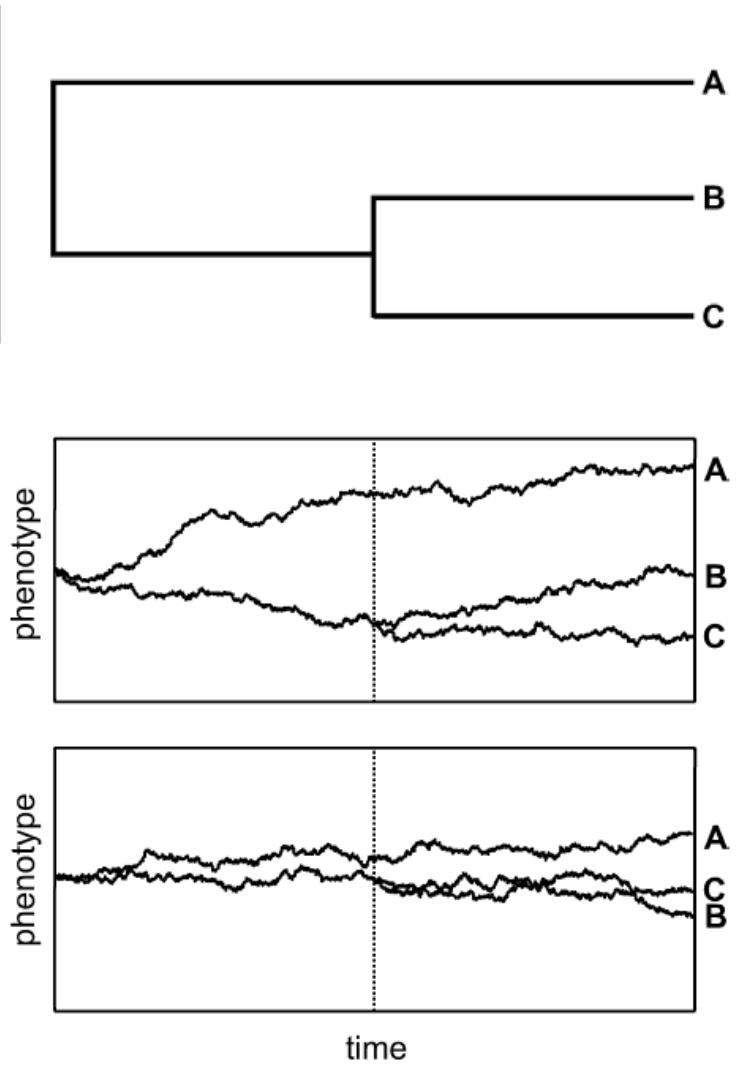


Example with chromosomes

With process based models we are usually interested in the parameters of the model and do model comparison to find the best model or Bayesian analyses where we compare parameters in the model



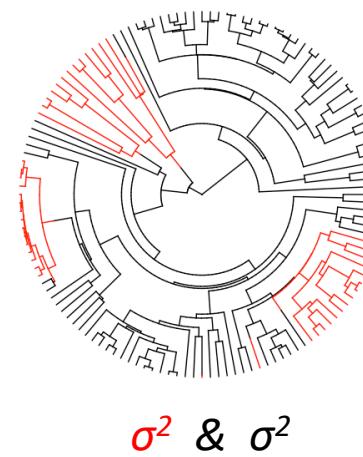
Example from beetles



*Single parameter:
 σ^2 the “drift” parameter*



vs.

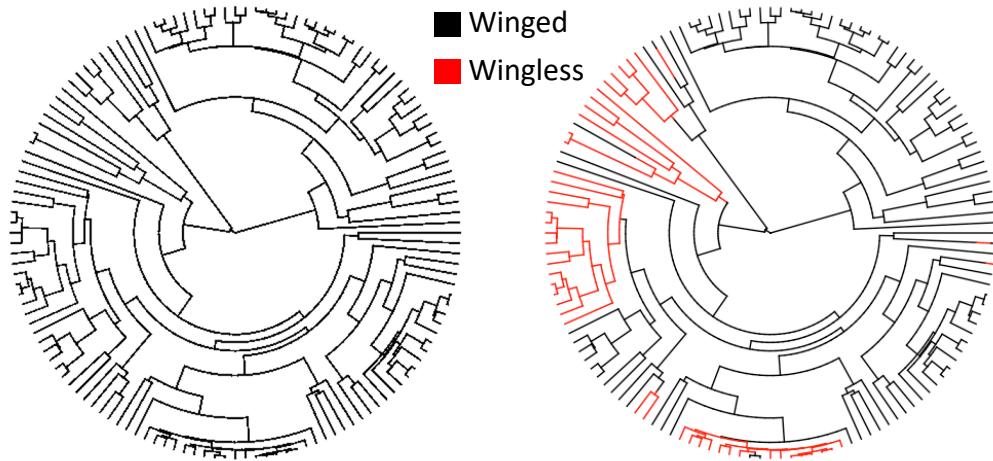


σ^2 & σ^2

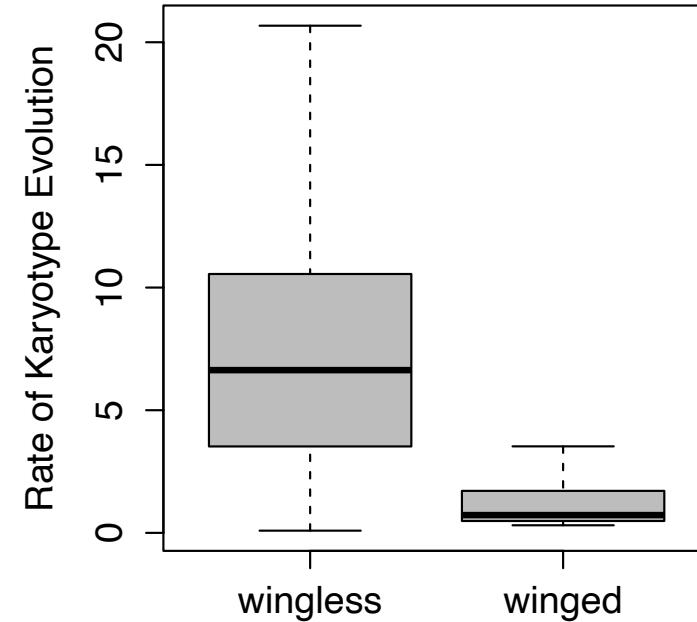
Example from beetles

Effect of wing loss on rate of chromosome evolution

2 rate model preferred on all
500 trees max p-val < 0.01



Winged $\sigma^2 = 1.5$
Wingless $\sigma^2 = 9.1$



Estimation via Monte Carlo

- Using regular stats test then simulating a lot of data and getting a null distribution of stats (like the f-statistic if you are doing anova). These are implemented for most things you might want.

APE, Geiger, and Phytools

- Simulating data and getting a null distribution for any value you are interested in.

Estimation via Monte Carlo

Using existing stats:

```
library(geiger)
fit <- aov.phylo(y~x, phy, nsim=100)
```

Creating your own tests and Monte Carlos isn't hard

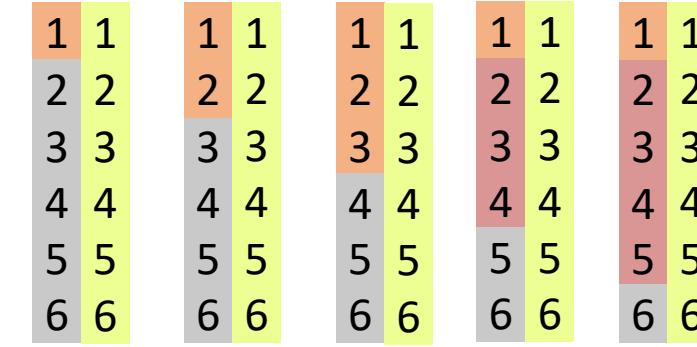
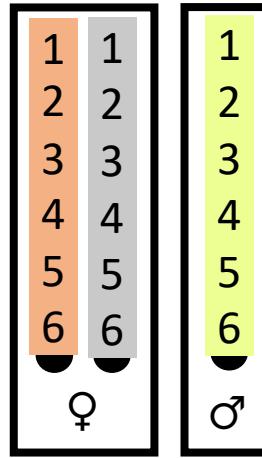
An example from Hymenoptera

Hypothesis: Eusociality should lead to selection for higher recombination and by extension higher chromosome number.

An example from Hymenoptera

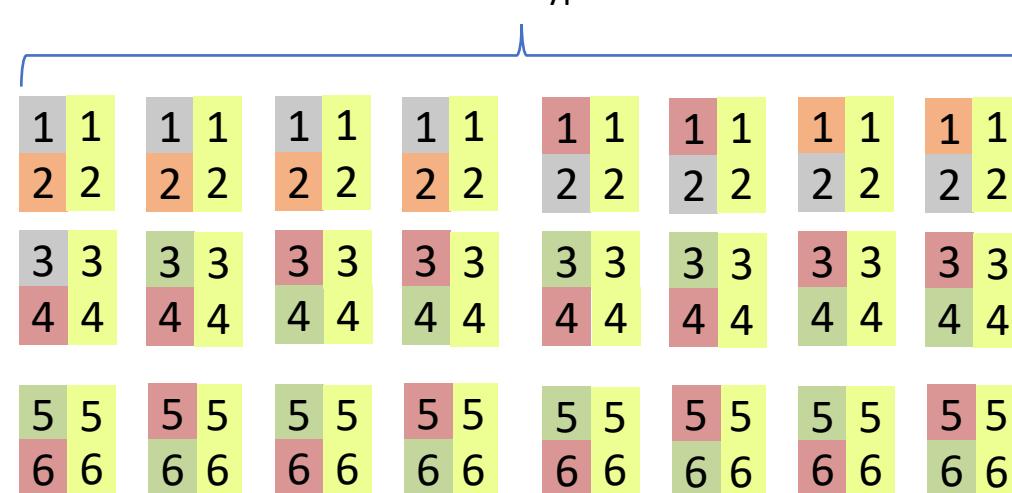
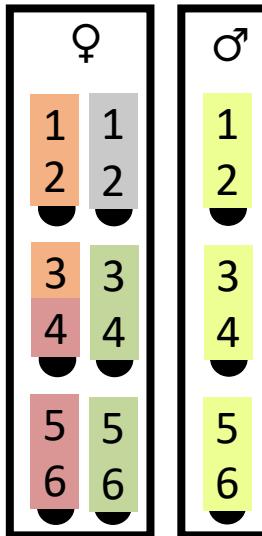
On average one crossover per arm per meiosis:

With 1 chromosome:



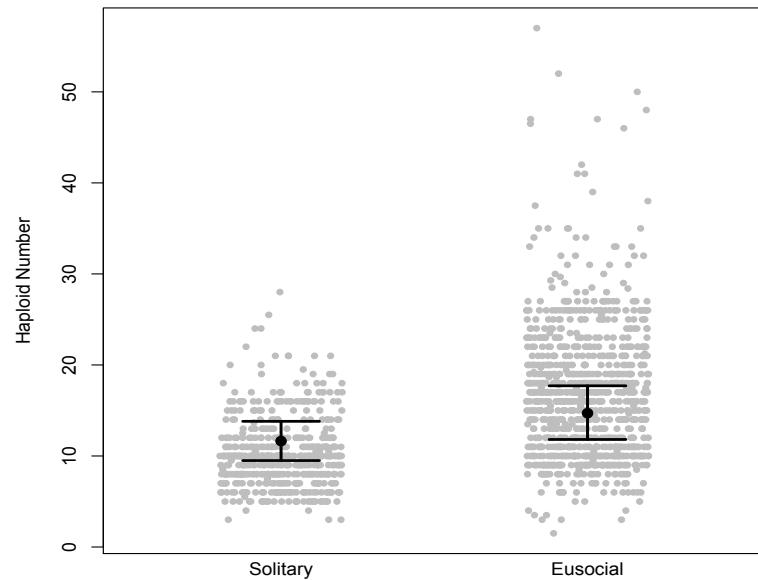
5 Genotypes

With 3 chromosomes:



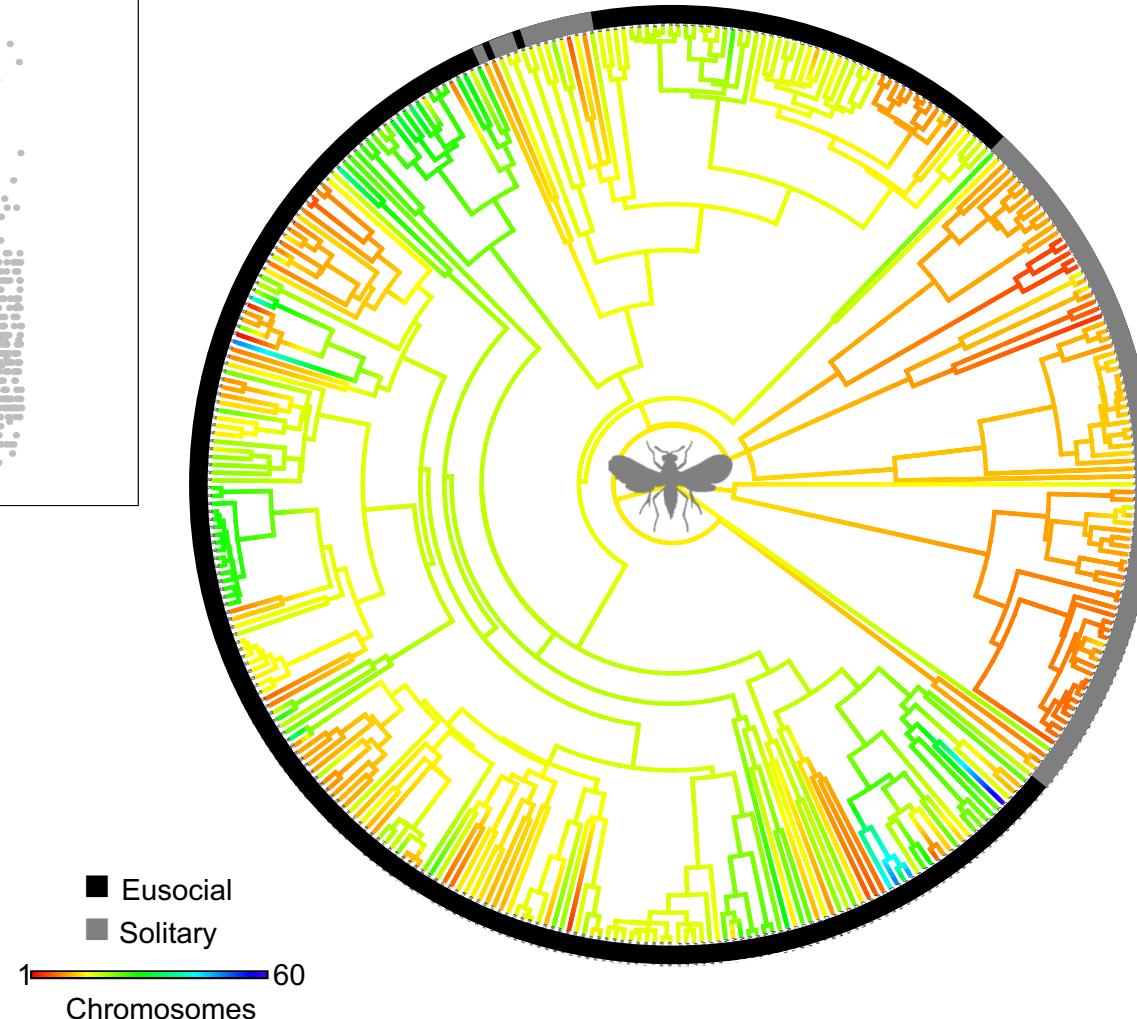
8 Genotypes

An example from Hymenoptera



Standard ANOVA
p-value < 0.001

Phylo - ANOVA
F-statistic = 70.5,
p-value = 0.23

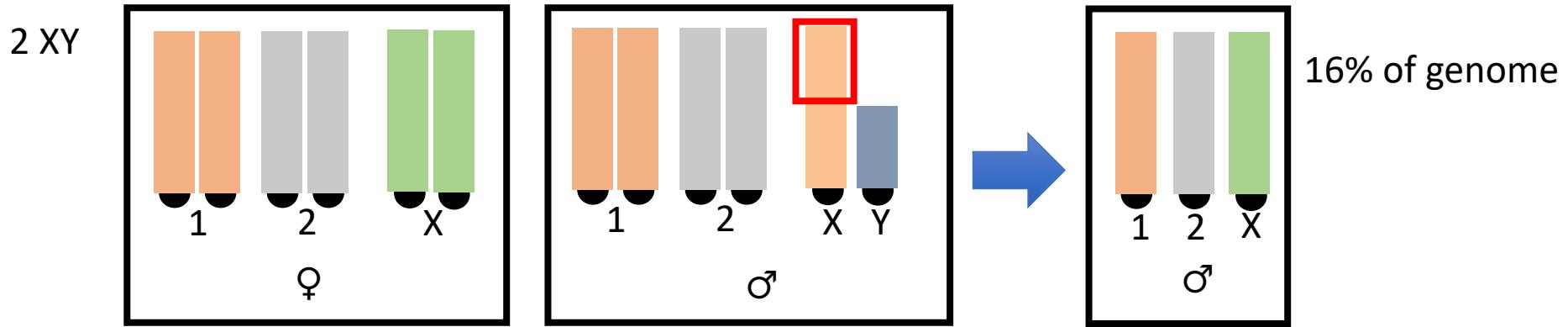


Origins of haplodiploidy

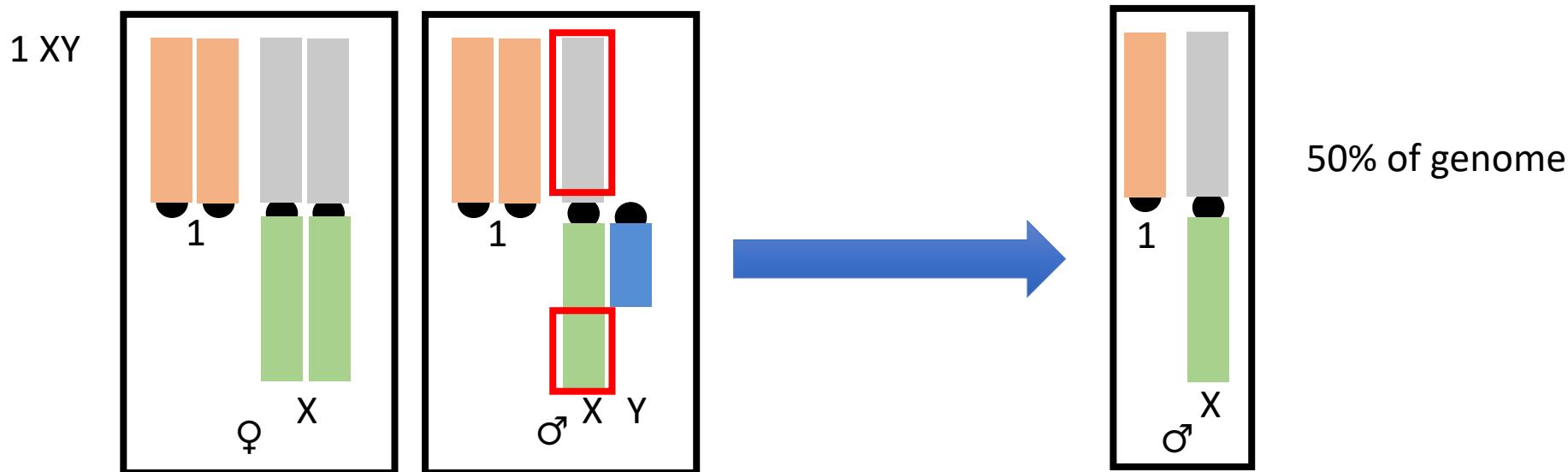
Bull's hypothesis

Species with relatively few chromosomes should experience transitions to haplodiploidy more frequently than species with many chromosomes.

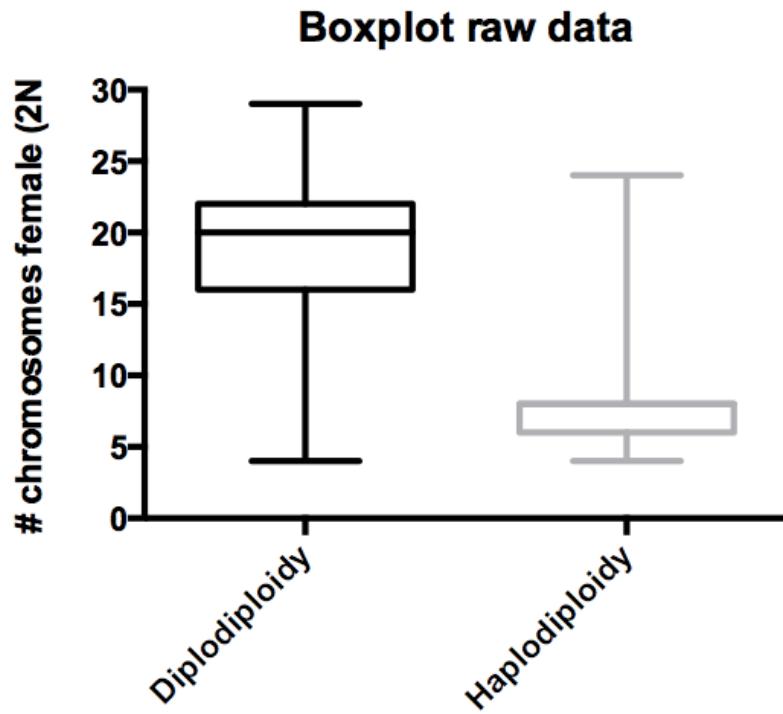
Origins of haplodiploidy



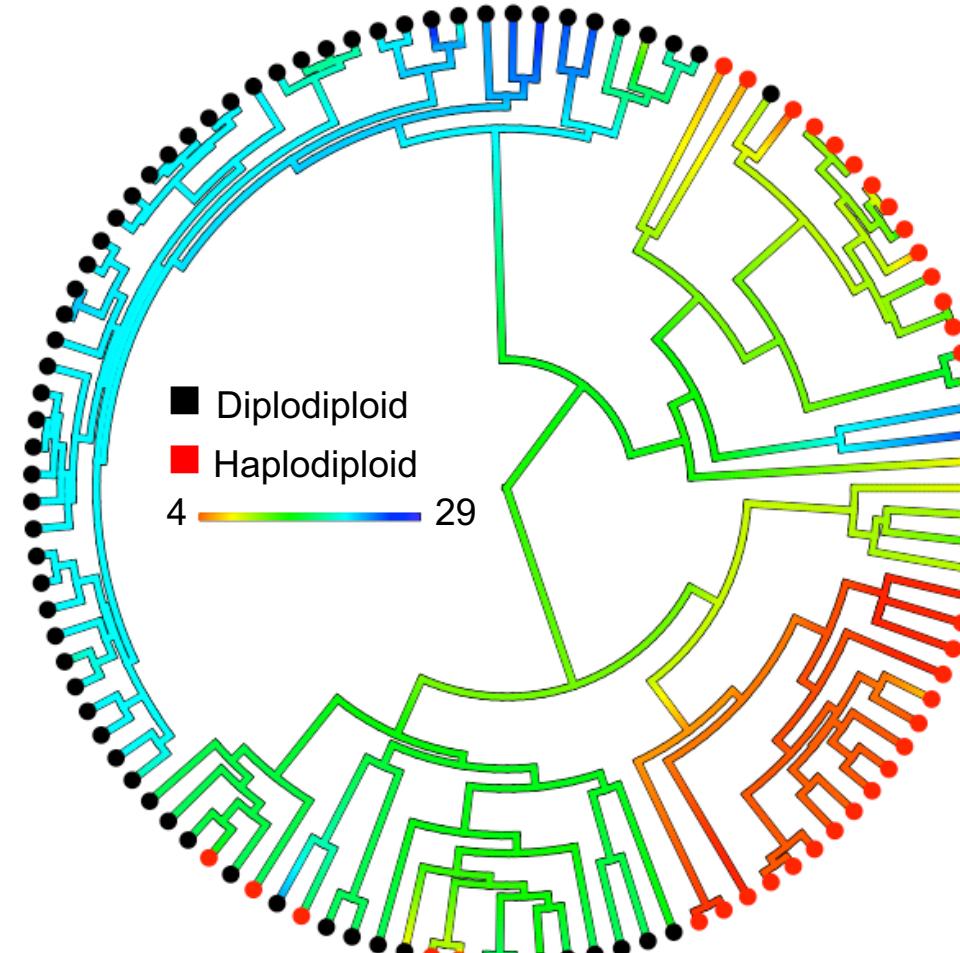
Fusion between chromosome 2 and X; followed by degeneration of neoY



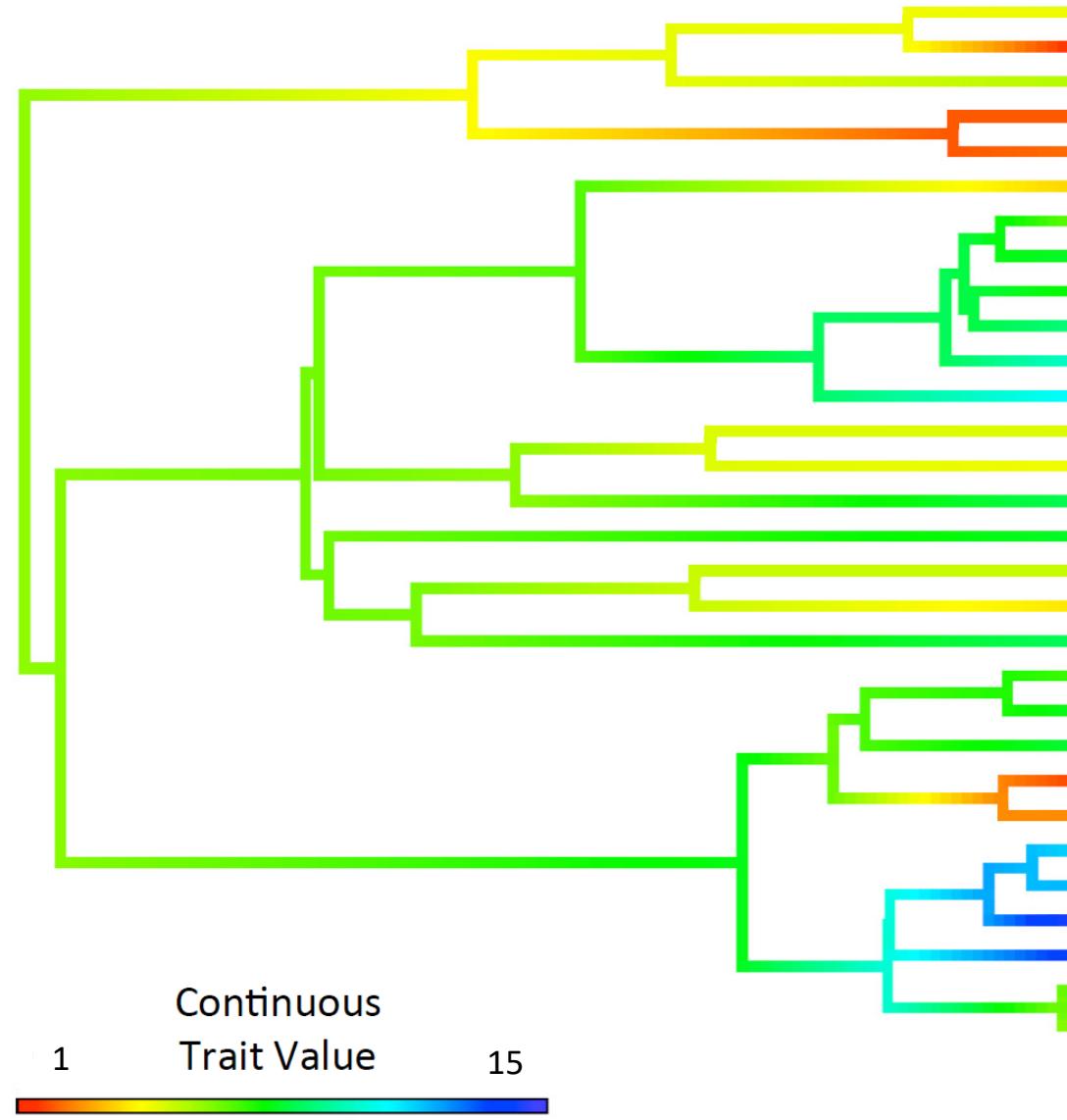
Origins of haplodiploidy



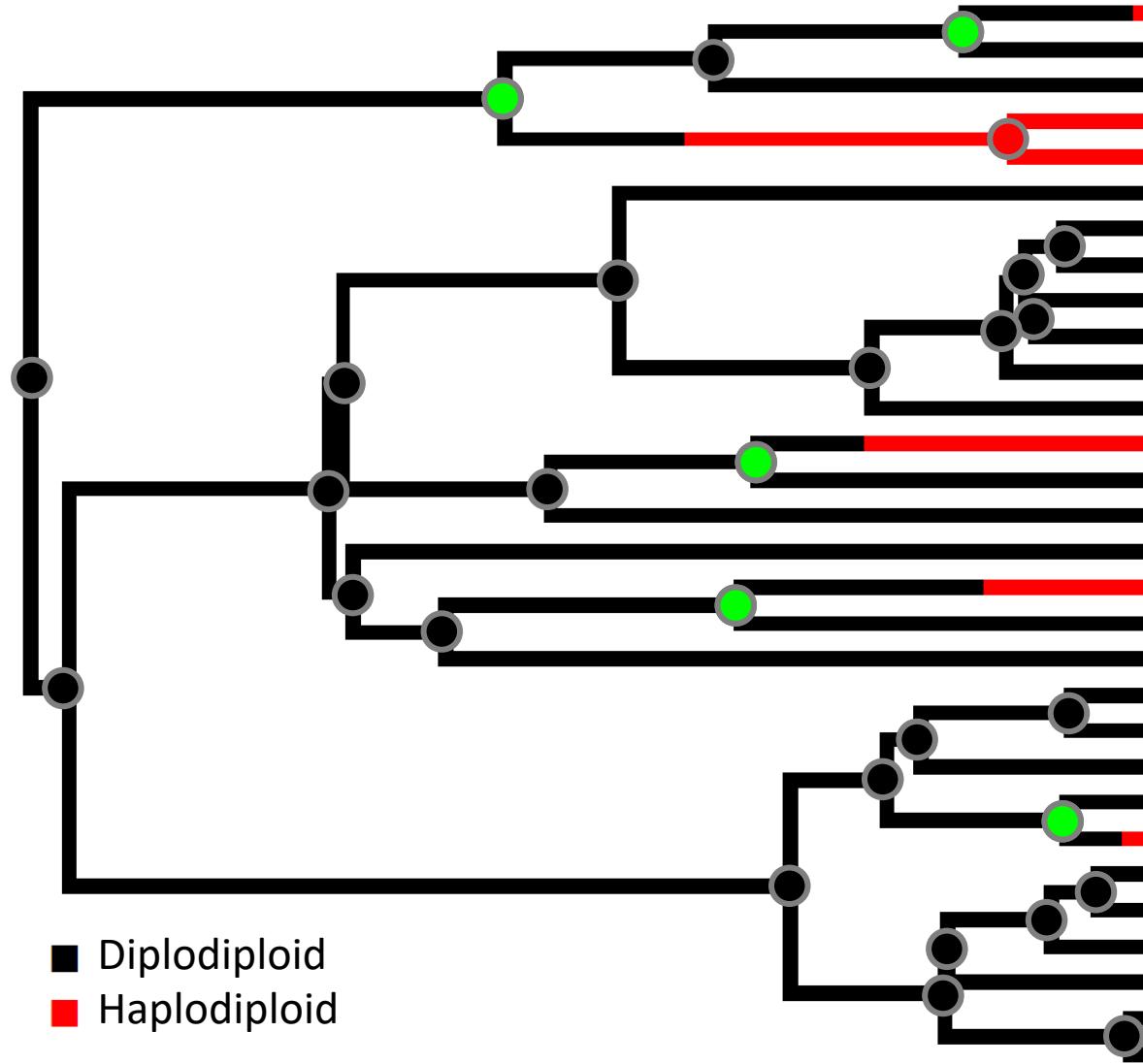
Phylo - ANOVA
F-statistic = 97.8,
p-value < 0.001



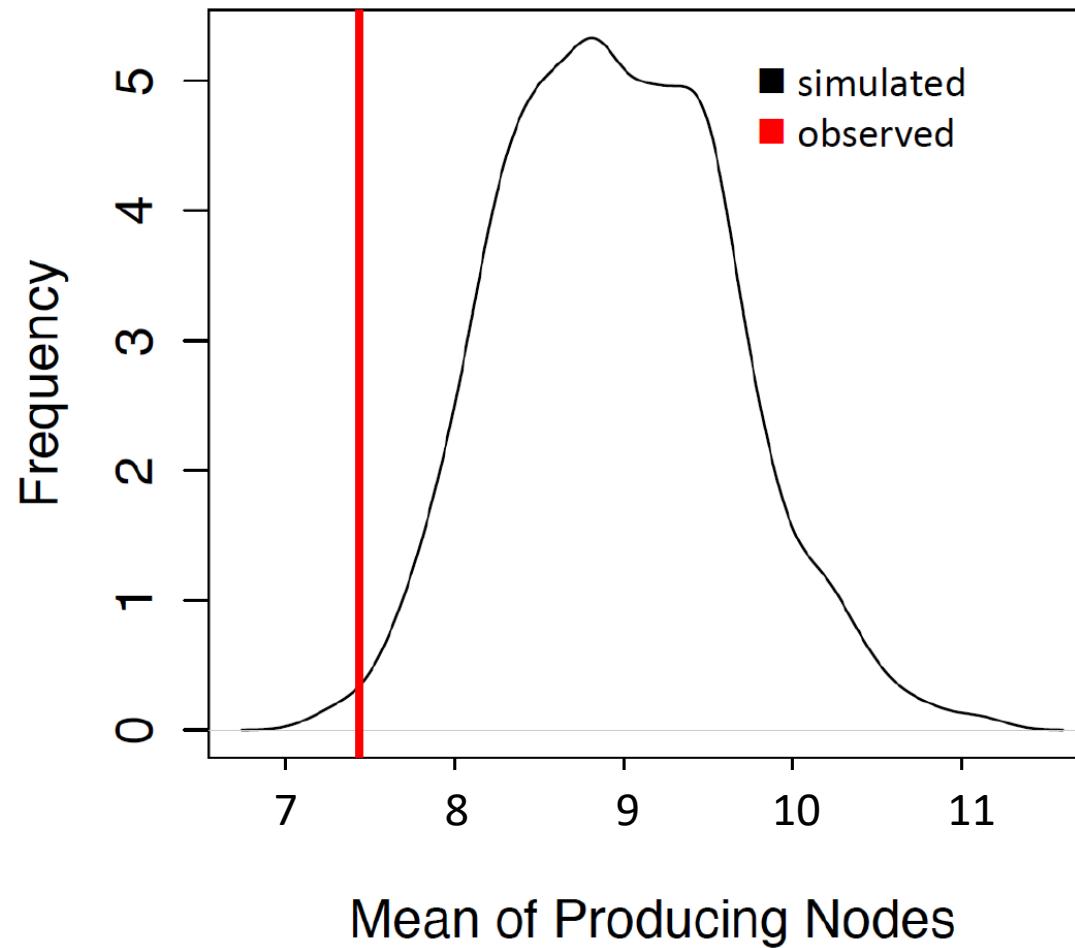
Ancestral condition test



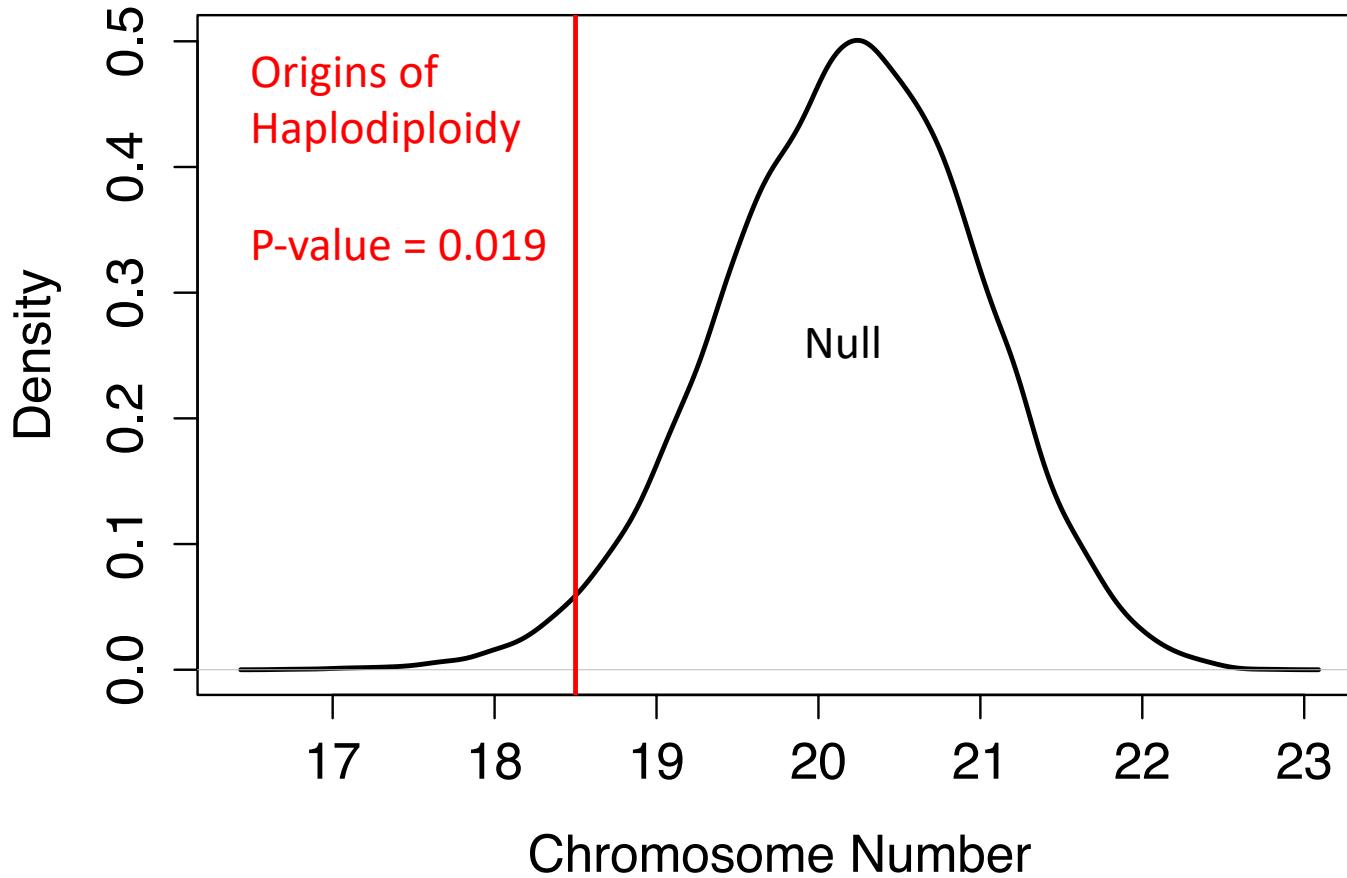
Ancestral condition test



Ancestral condition test



Section 4: Origins of haplodiploidy



Resources

[R CRAN phylogenetics task view](#)

[Inferring phylogenies](#)

Bodega Bay Phylogenetics Workshop

Me if you've already googled a little