Trait evolution on phylogenetic networks

Tony Jhwueng		**Brian O'Meara**
tonyjhwueng.info		brianomeara.info
Feng-Chia University		@omearabrian
Taichung, Taiwan		U. of Tennessee, Knoxville

Sharing encouraged!

Acknowledgments

- National Institute for Mathematical and Biological Synthesis (NIMBioS)
- National Science Foundation
- bioRxiv
- Paul Bastide
- Cécile Ané
- Luke Harmon
- Anonymous reviewers of *Systematic Biology*

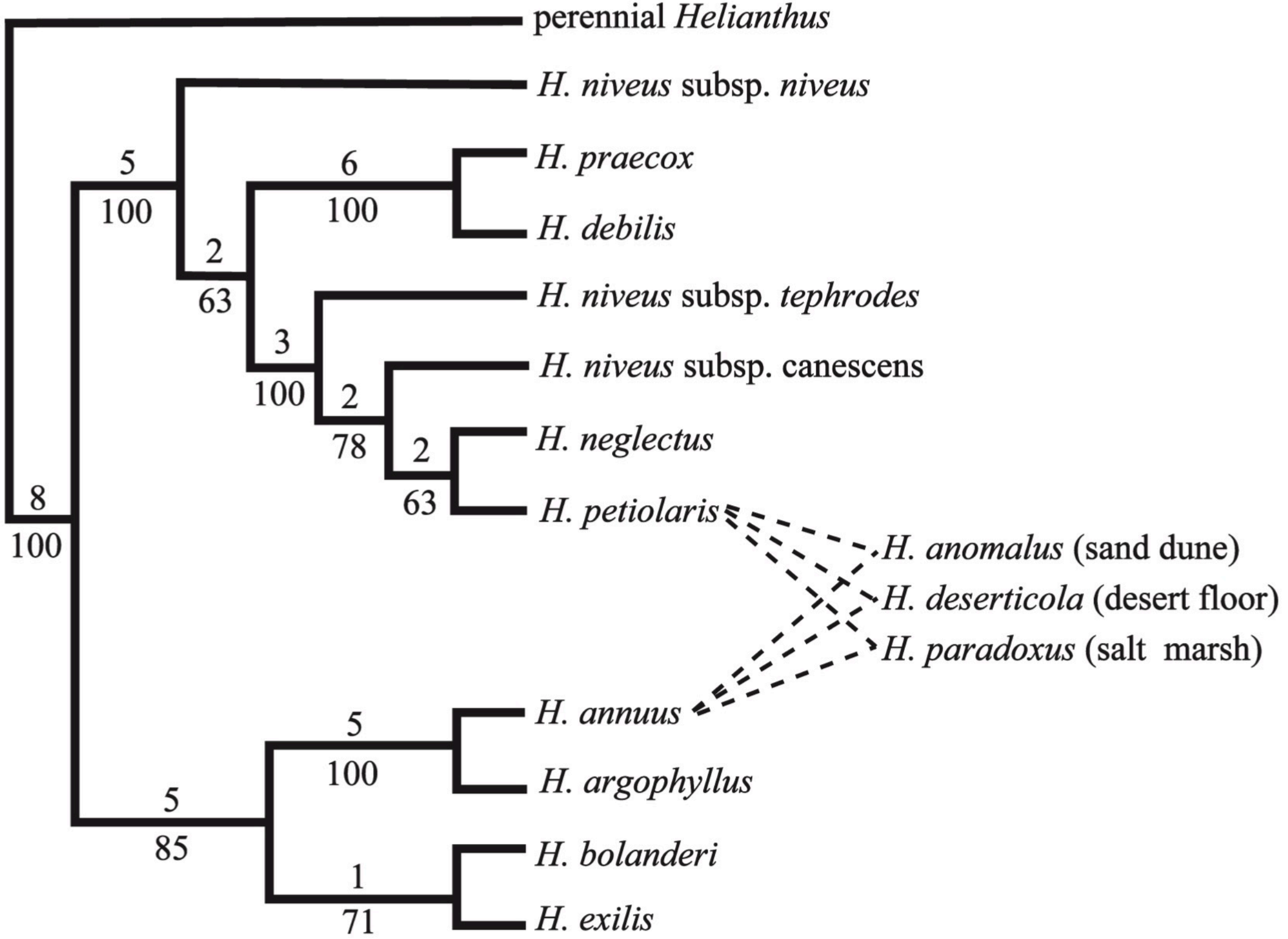


Figure 1. Phylogenetic tree for *Helianthus* sect. *Helianthus* based on combined chloroplast DNA and nuclear ribosomal DNA data (redrawn from Rieseberg, 1991, fig. 9). The number of mutations are given above and bootstrap percentages below each branch. Dashed lines indicate parentage of homoploid hybrid species.

Rieseberg 2006

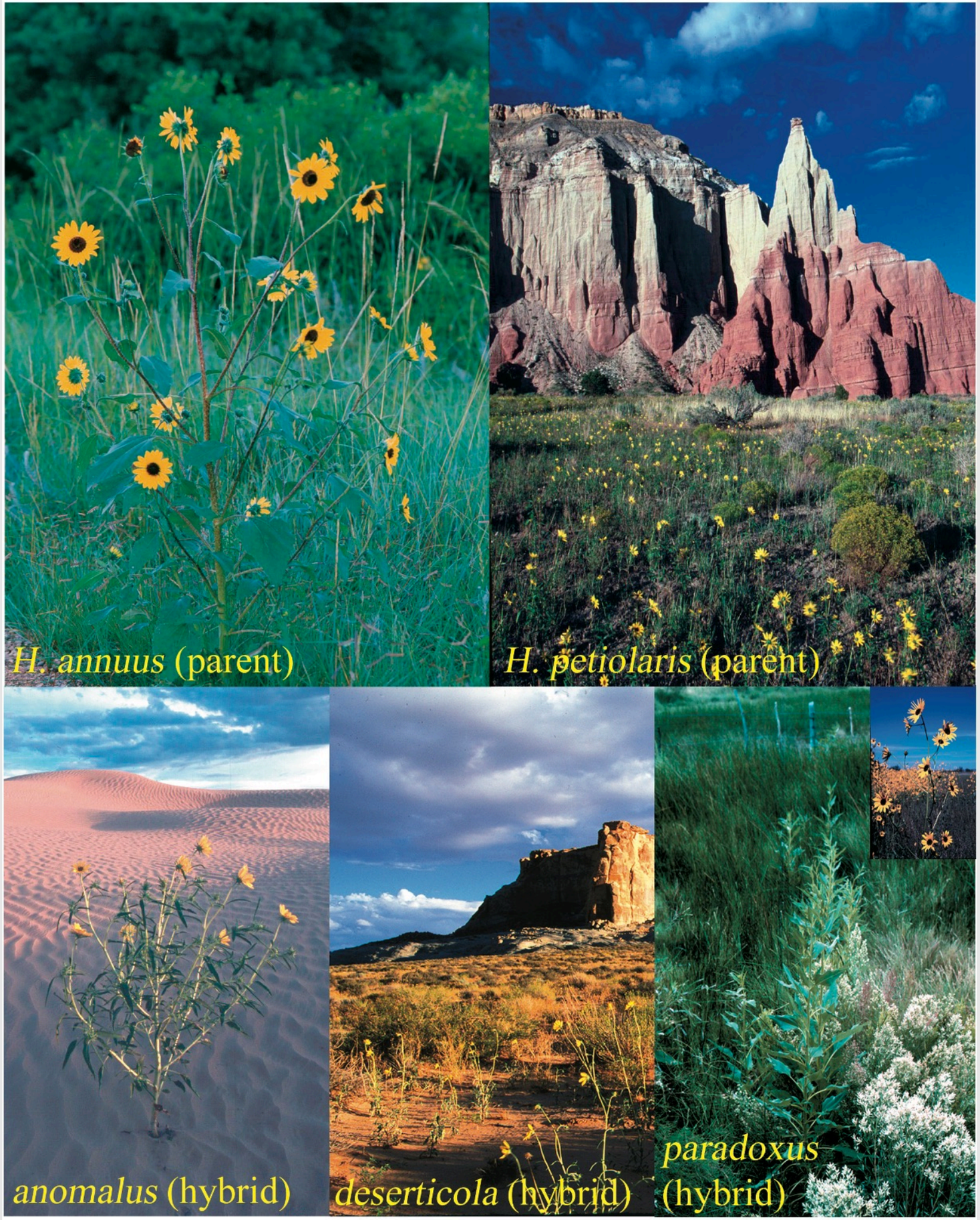
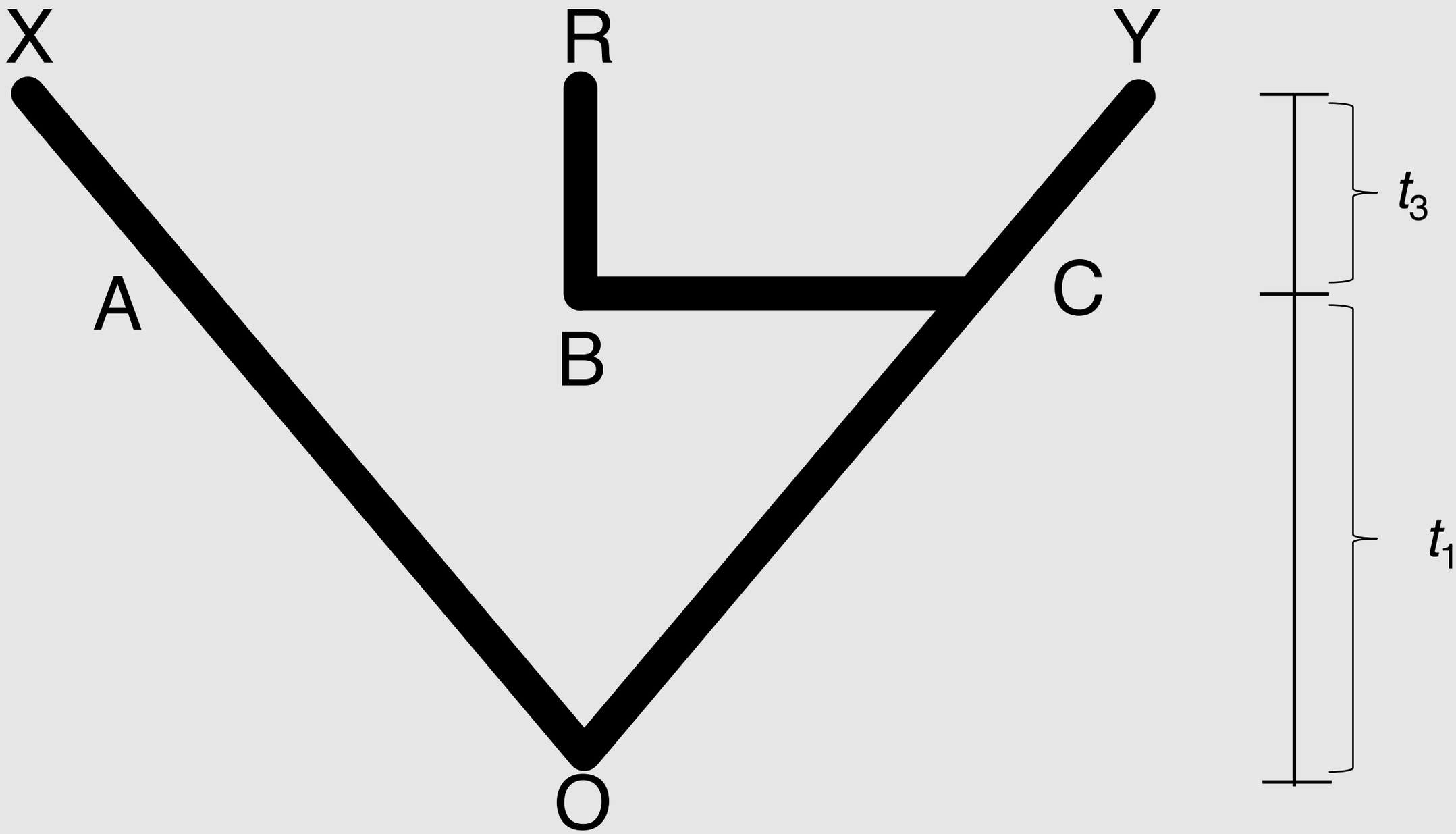


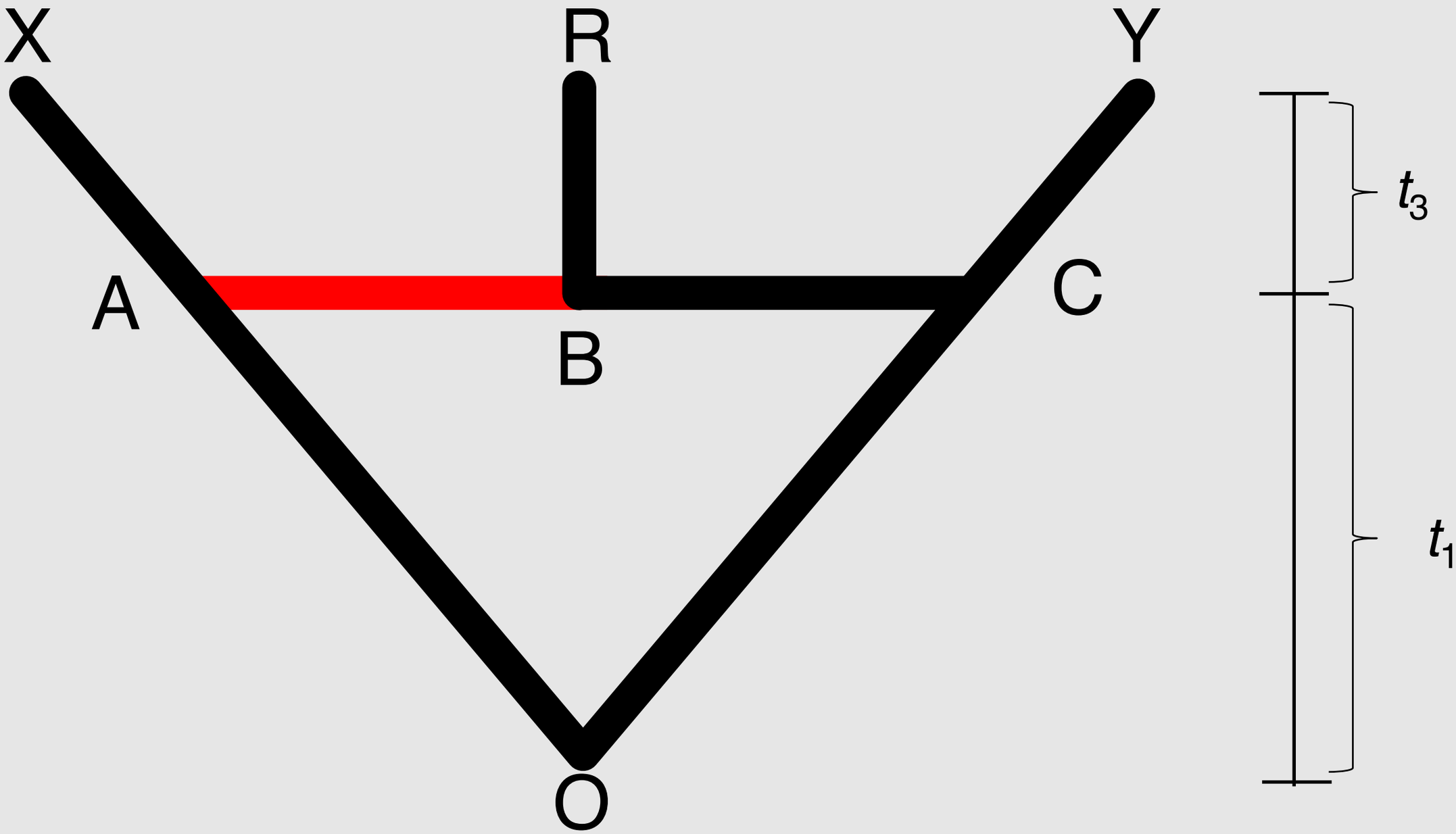
Figure 3. Photographs of the two parental species and their three hybrid derivative species in typical habitats. Photographs by Jason Rick.

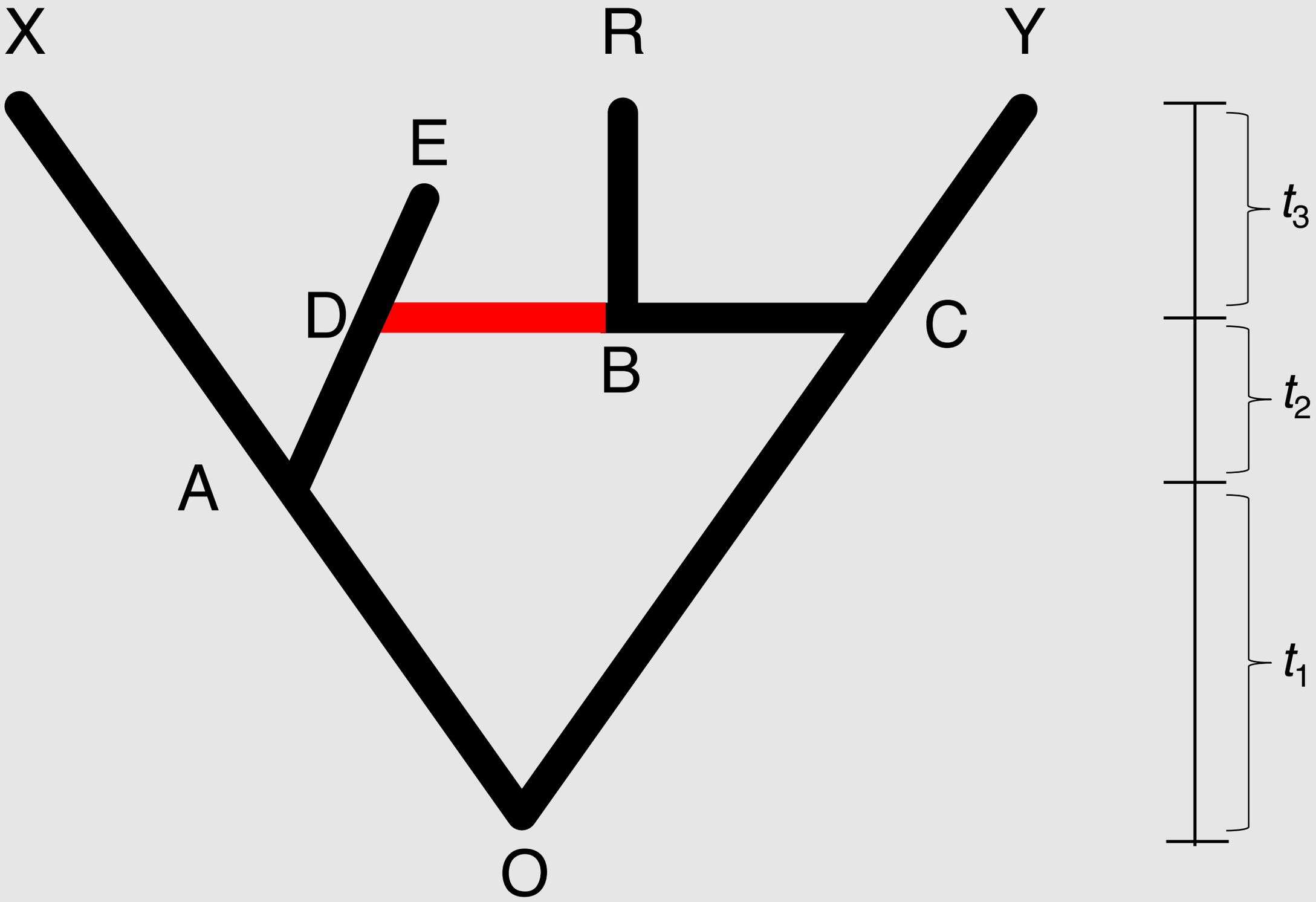
Rieseberg 2006
Photos by James Rick

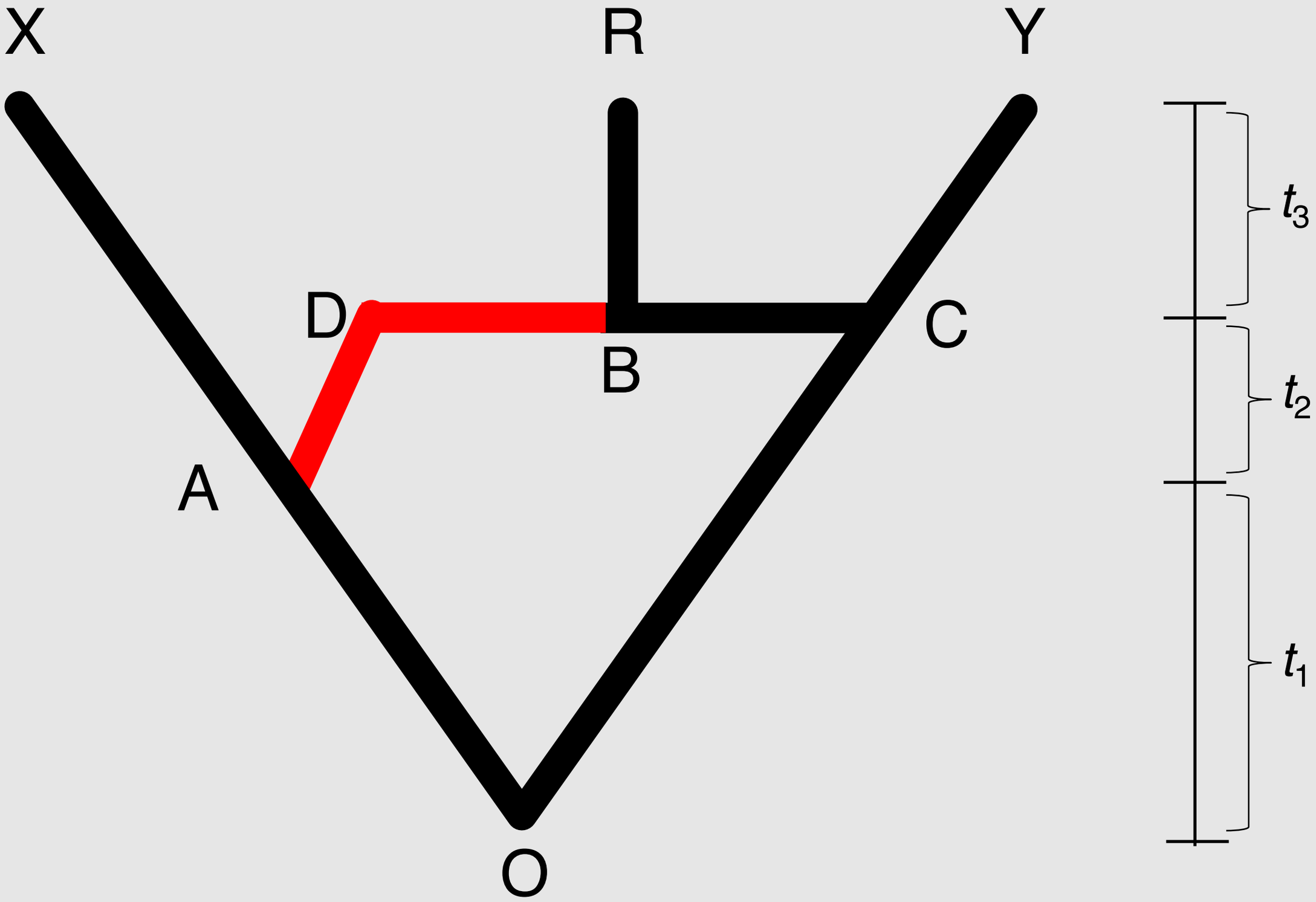
Model requirements

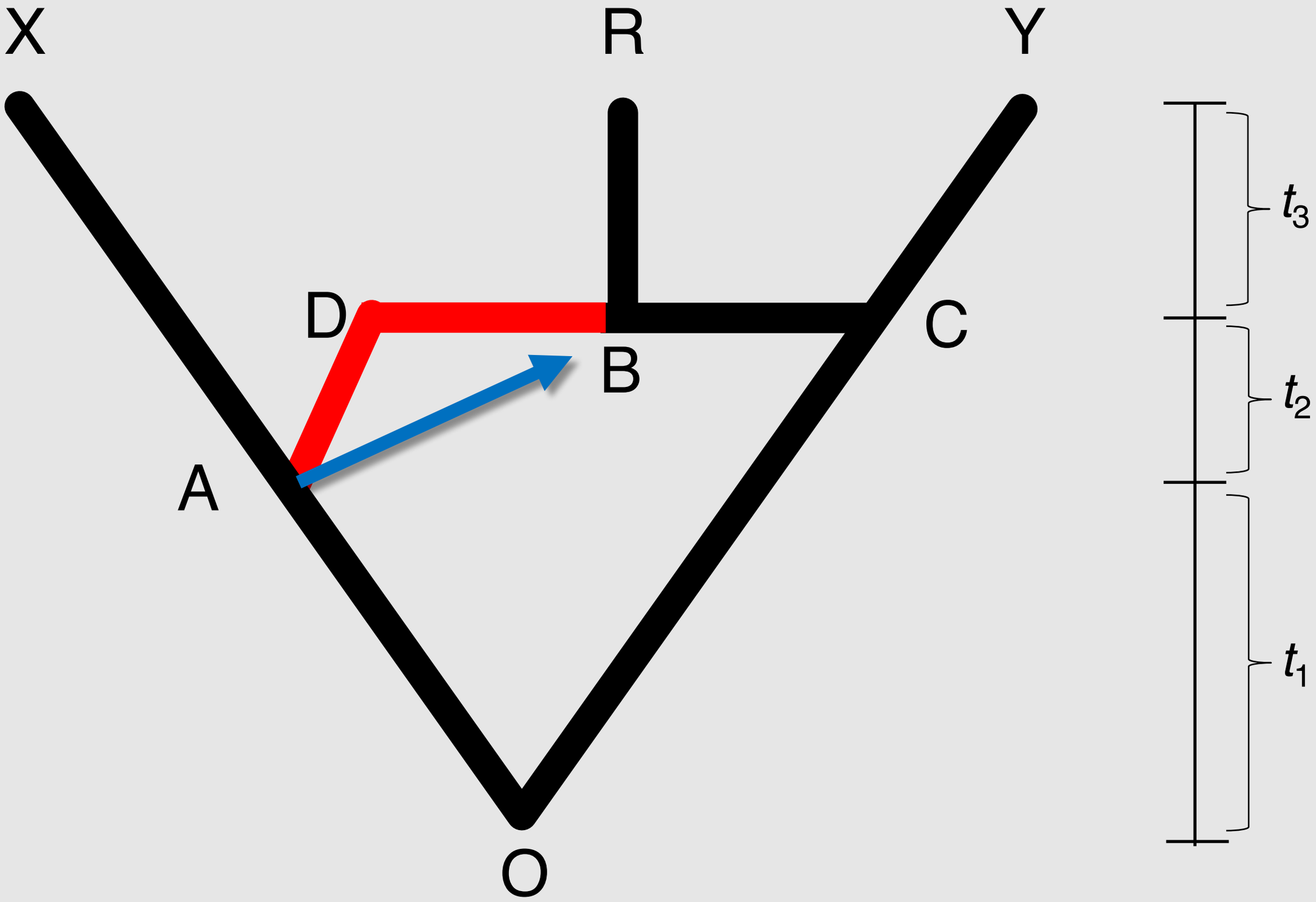
- Work on networks, not just trees
- Get parameter estimates
- Estimate mean shifts (i.e., hybrids bigger than parents)
- Estimate increased variation (hybrids weirder than parents)
- Incorporate measurement error (life happens)

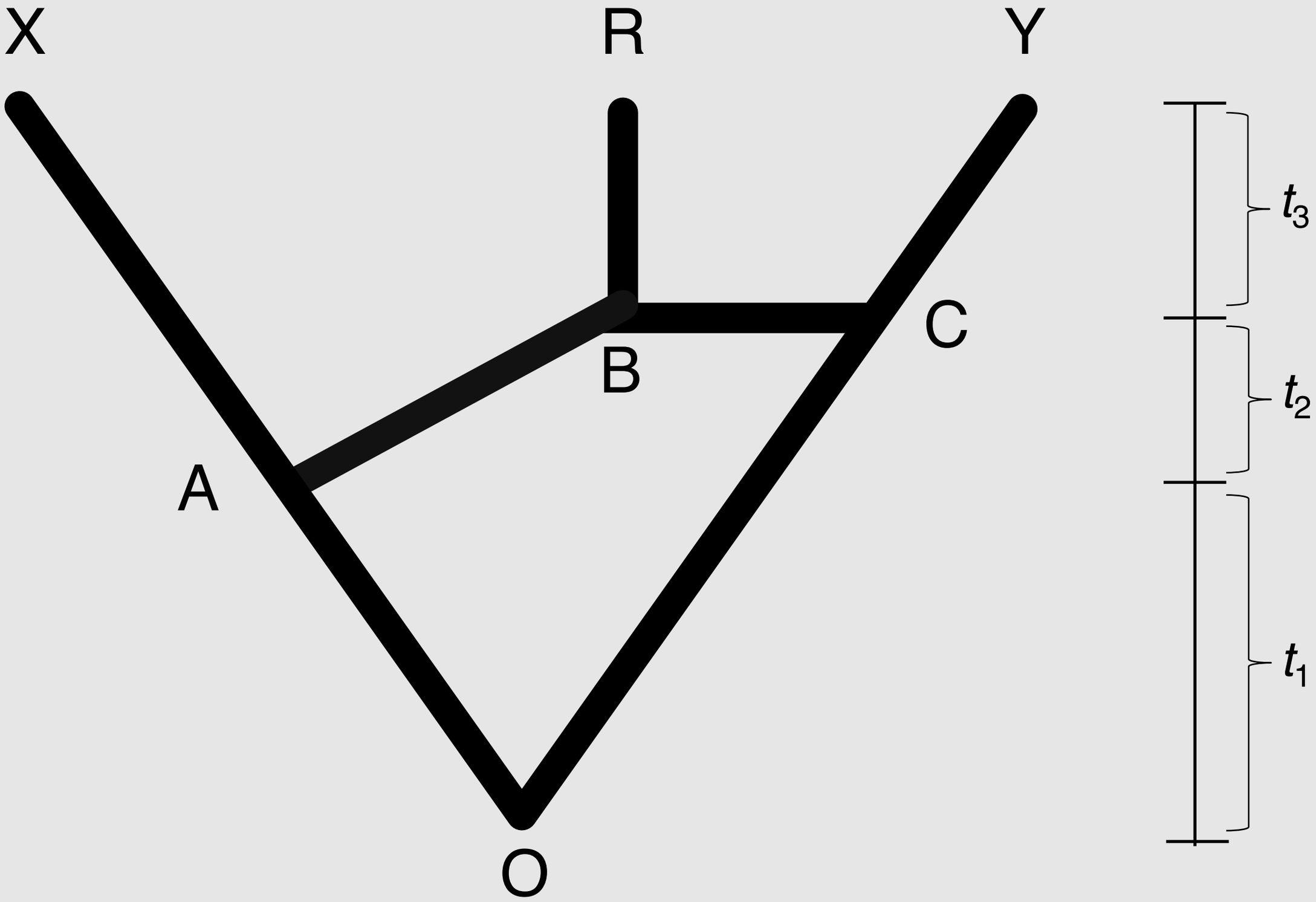

X
A
R
Y
B
C
O
t_3
t_1+t_2


X
R
Y
A
B
C
O
t_3
t_1+t_2

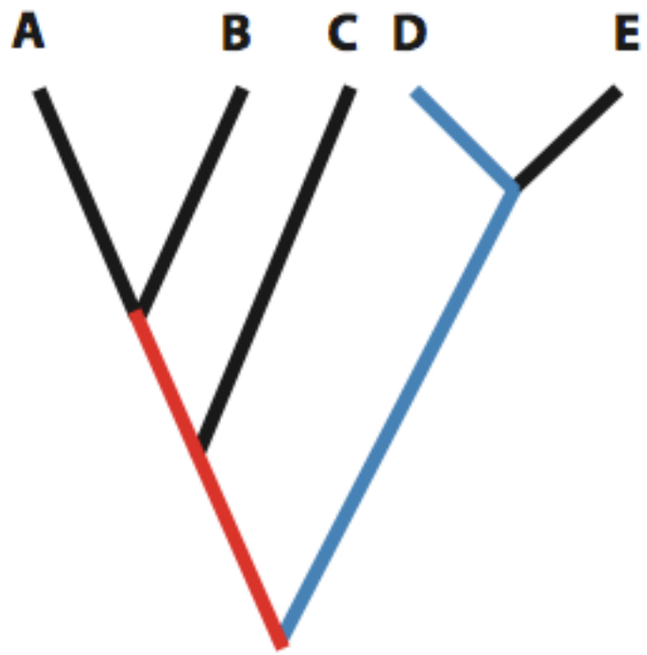

X
R
Y
E
D
C
B
A
O
t_3
t_2
t_1


X
R
Y
D
C
B
A
O
t_3
t_2
t_1


X
R
Y
D
C
B
A
O
t_3
t_2
t_1


X
R
Y
A
B
C
O
t_3
t_2
t_1

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$



Variance-covariance matrix

	A	B	C	D	E
A	var(A)	cov(A,B)	cov(A,C)	cov(A,D)	cov(A,E)
B	cov(B,A)	var(B)	cov(B,C)	cov(B,D)	cov(B,E)
C	cov(C,A)	cov(C,B)	var(C)	cov(C,D)	cov(C,E)
D	cov(D,A)	cov(D,B)	cov(D,C)	var(D)	cov(D,E)
E	cov(E,A)	cov(E,B)	cov(E,C)	cov(E,D)	var(E)

Mean vector

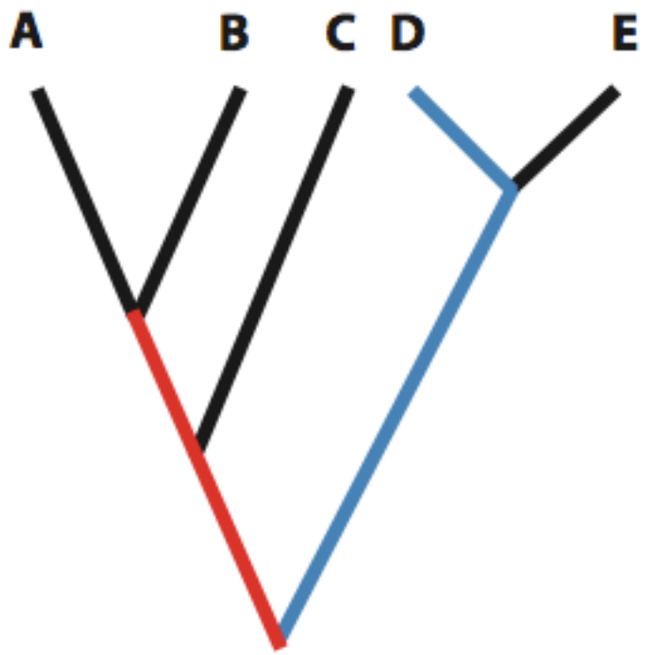
A	μ_A
B	μ_B
C	μ_C
D	μ_D
E	μ_E

Figure 4

Multivariate normal distribution. The figure shows a tree, the tree's variance-covariance matrix, and the vector of means (which, under Brownian motion, would equal the root state). Highlighted are the branches leading to covariance between taxa A and B (*red*) and the branches leading to variance in D (*blue*).

O'Meara 2012

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$



Variance-covariance matrix

	A	B	C	D	E
A	var(A)	cov(A,B)	cov(A,C)	cov(A,D)	cov(A,E)
B	cov(B,A)	var(B)	cov(B,C)	cov(B,D)	cov(B,E)
C	cov(C,A)	cov(C,B)	var(C)	cov(C,D)	cov(C,E)
D	cov(D,A)	cov(D,B)	cov(D,C)	var(D)	cov(D,E)
E	cov(E,A)	cov(E,B)	cov(E,C)	cov(E,D)	var(E)

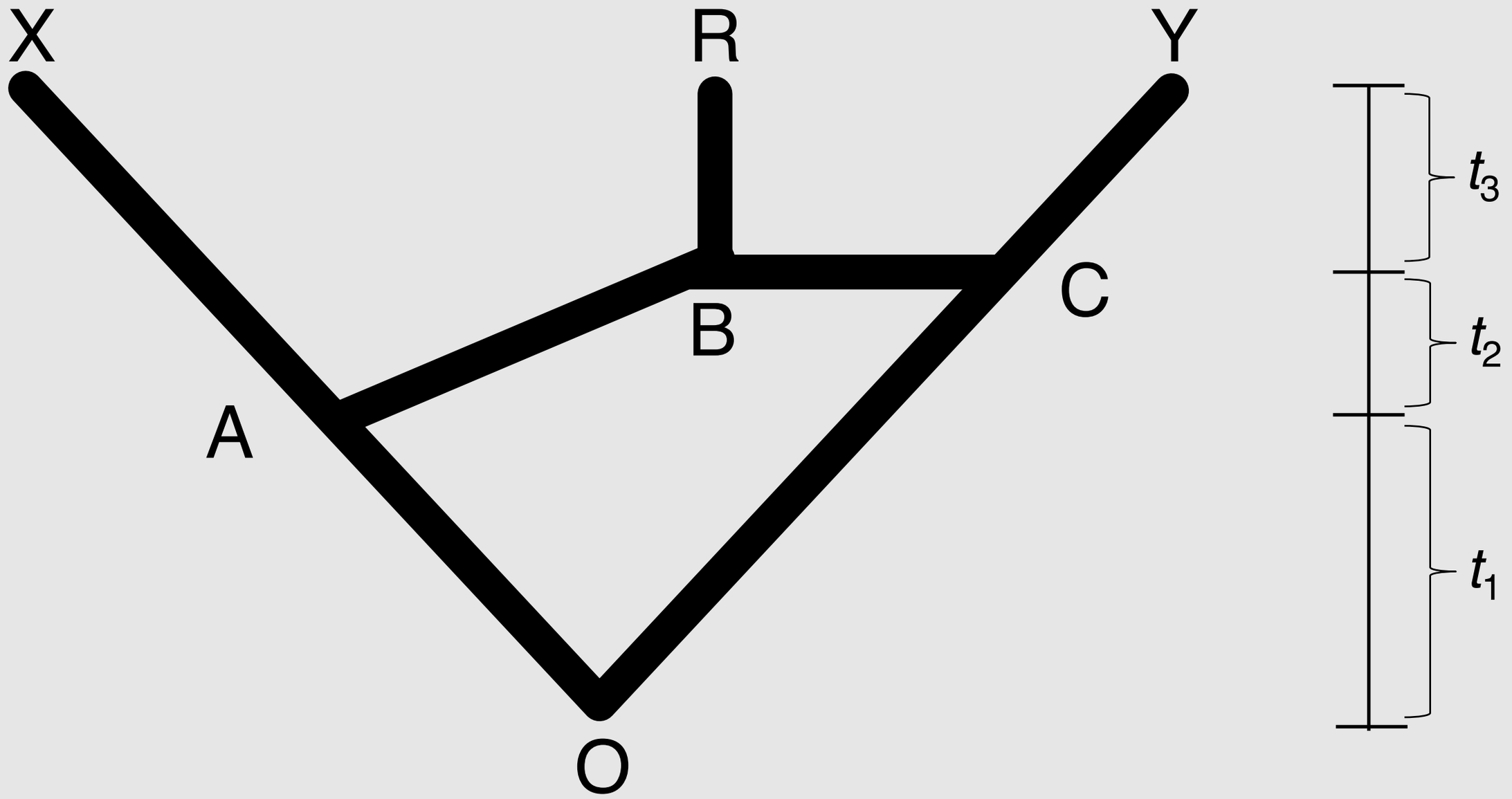
Mean vector

A	μ_A
B	μ_B
C	μ_C
D	μ_D
E	μ_E

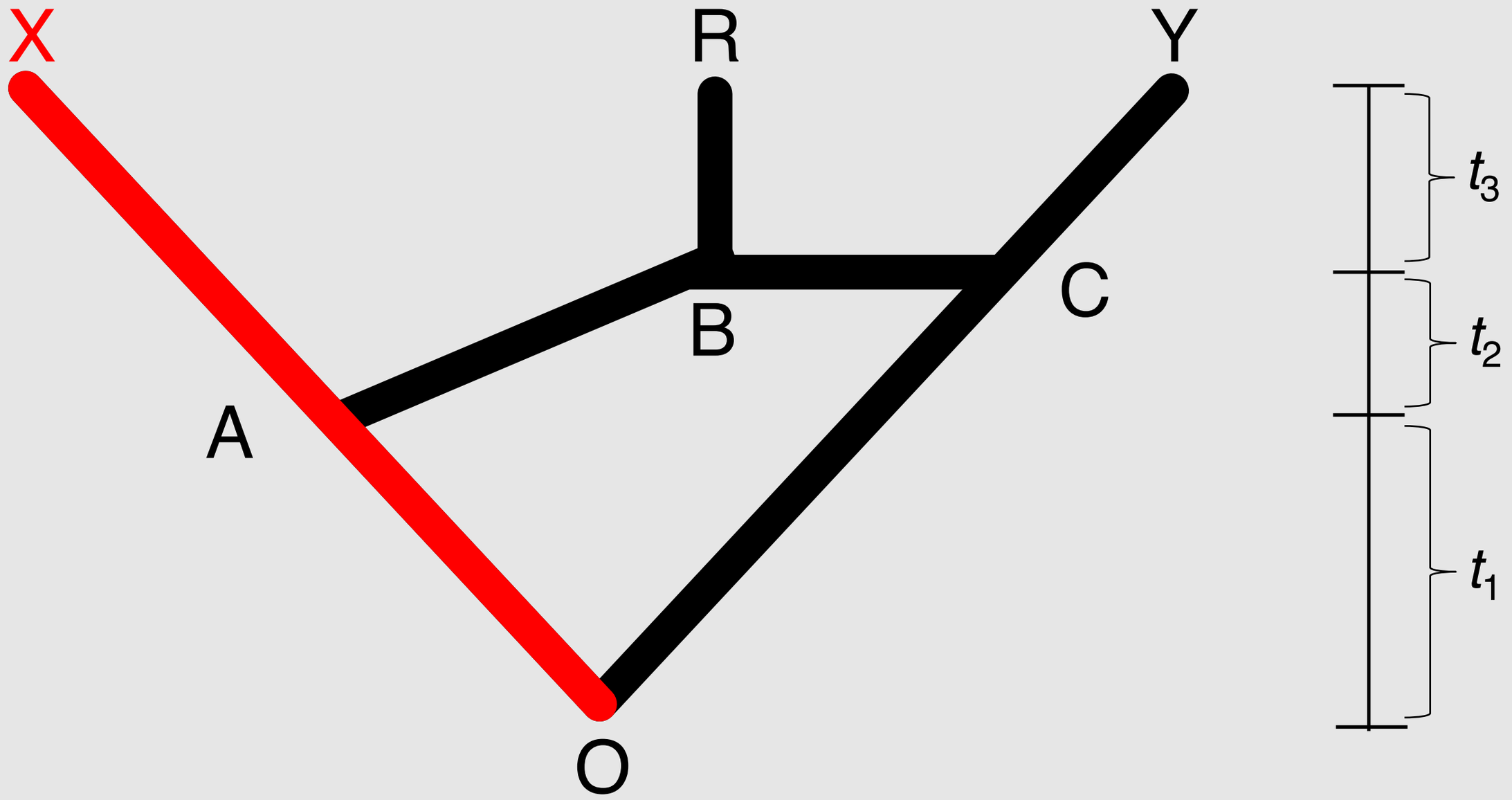
Figure 4

Multivariate normal distribution. The figure shows a tree, the tree's variance-covariance matrix, and the vector of means (which, under Brownian motion, would equal the root state). Highlighted are the branches leading to covariance between taxa A and B (*red*) and the branches leading to variance in D (*blue*).

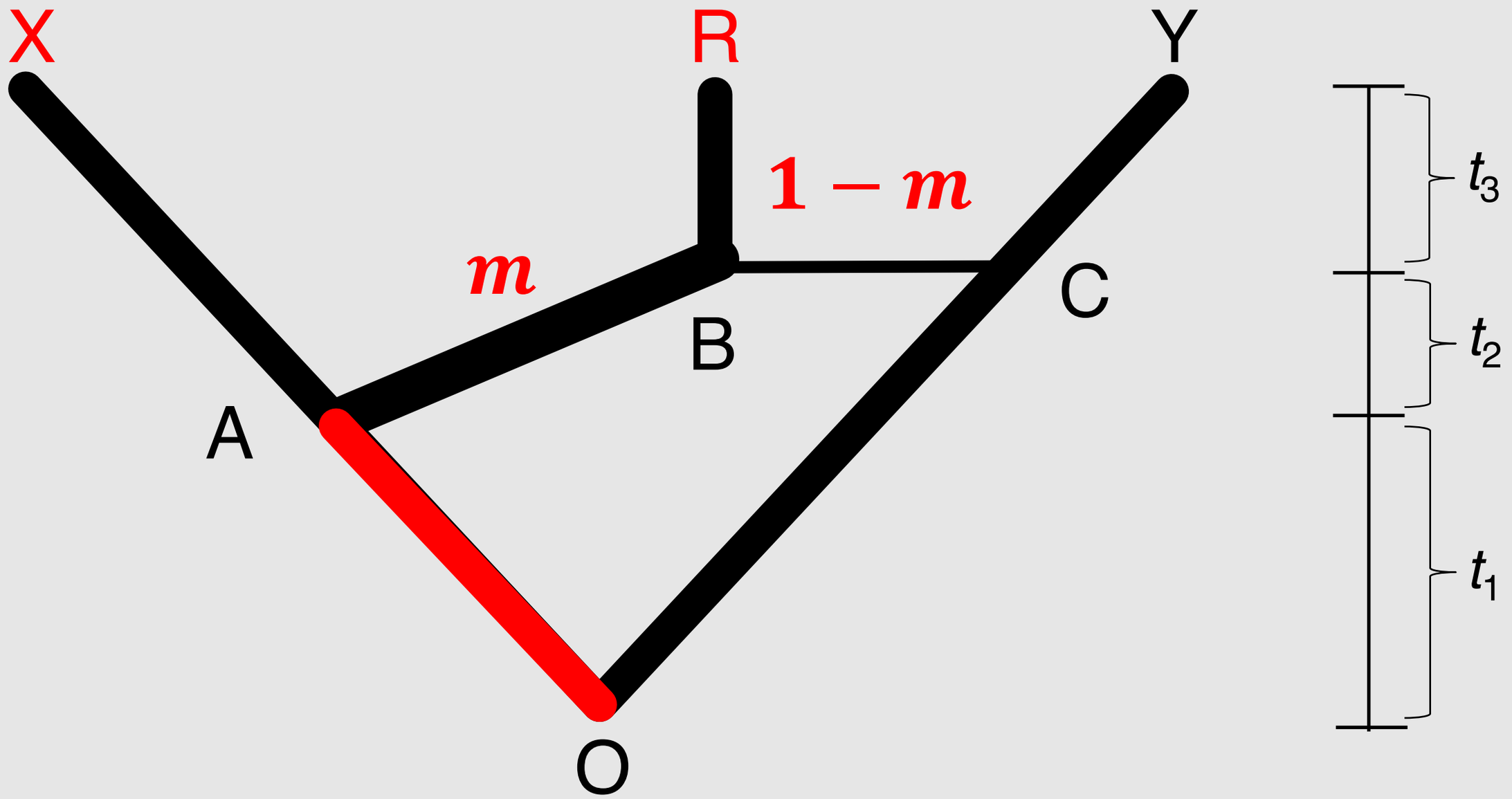
O'Meara 2012



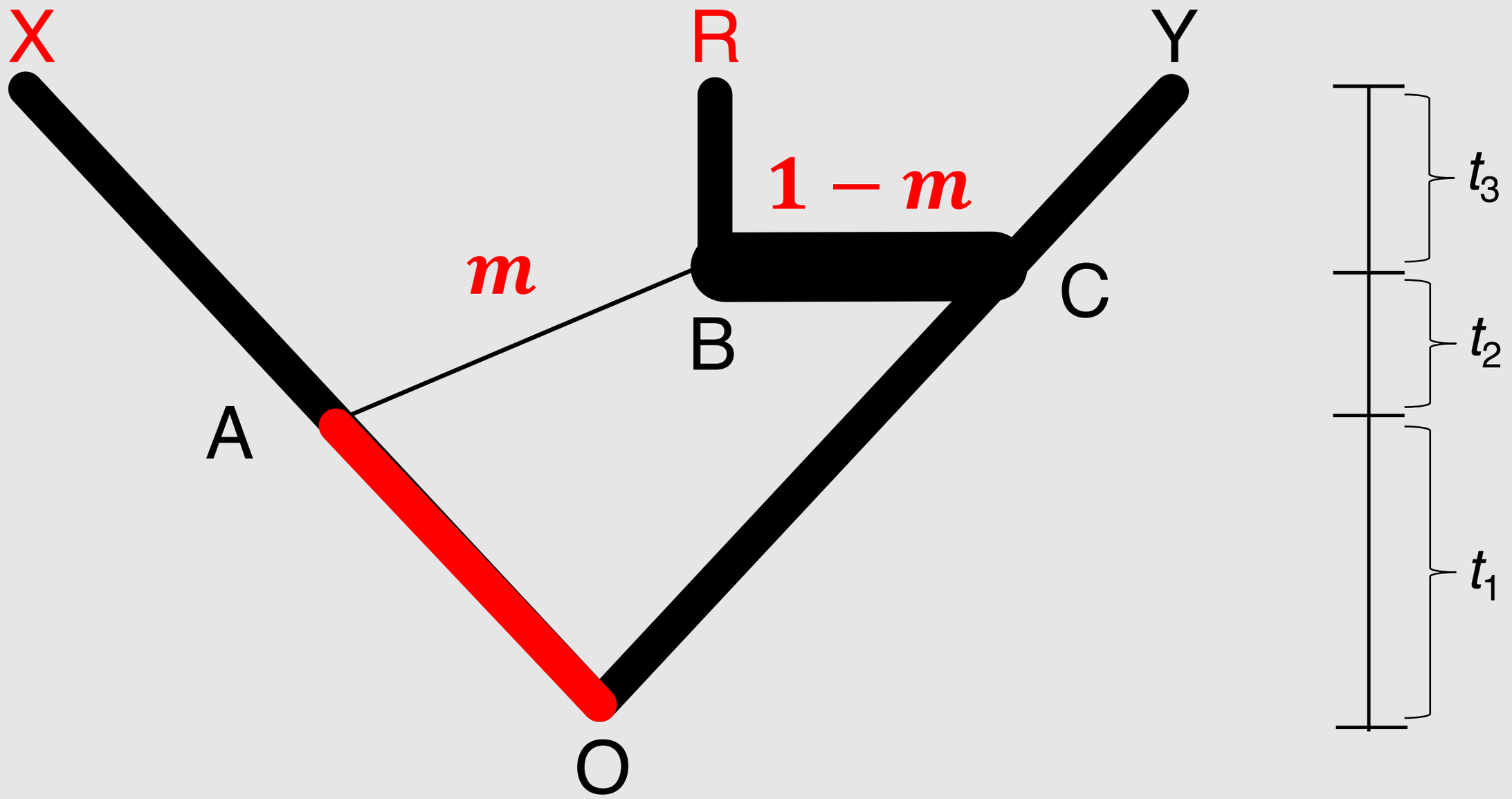
$$
\begin{array}{c}
 \\
X \\
R \\
Y
\end{array}
\begin{array}{c}
\begin{array}{ccc} X & R & Y \end{array} \\
\begin{bmatrix}
\sigma^2(t_1+t_2+t_3) & \sigma^2 m t_1 & 0 \\
\sigma^2 m t_1 & \sigma^2(m^2+(1-m)^2)(t_1+t_2)+\sigma^2 t_3+v_H & \sigma^2(1-m)(t_1+t_2) \\
0 & \sigma^2(1-m)(t_1+t_2) & \sigma^2(t_1+t_2+t_3)
\end{bmatrix}
\end{array}
$$



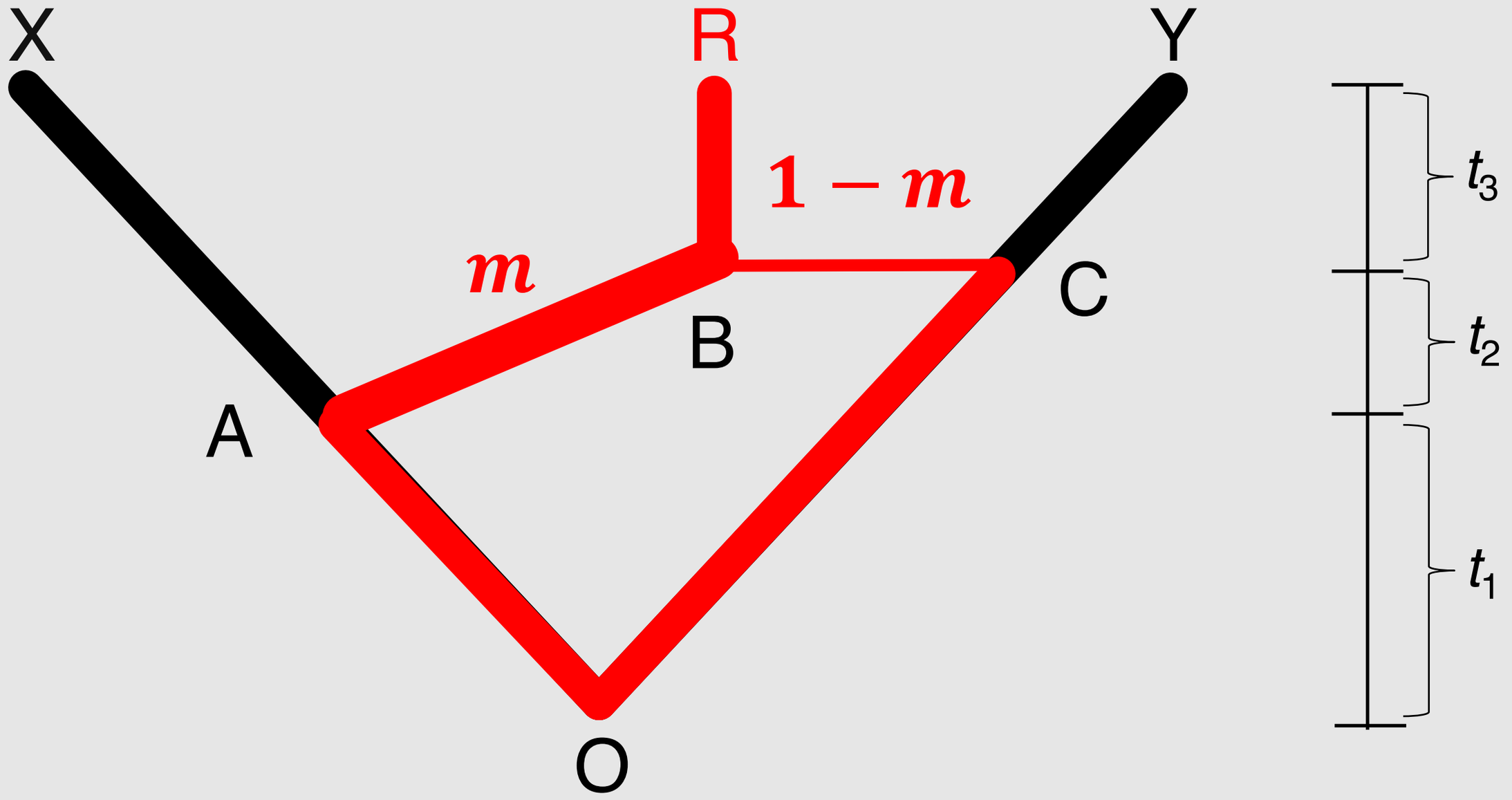
	X	R	Y
X	$\boldsymbol{\sigma^2(t_1+t_2+t_3)}$	$\sigma^2 m t_1$	0
R	$\sigma^2 m t_1$	$\sigma^2(m^2+(1-m)^2)(t_1+t_2)+\sigma^2 t_3+v_H$	$\sigma^2(1-m)(t_1+t_2)$
Y	0	$\sigma^2(1-m)(t_1+t_2)$	$\sigma^2(t_1+t_2+t_3)$



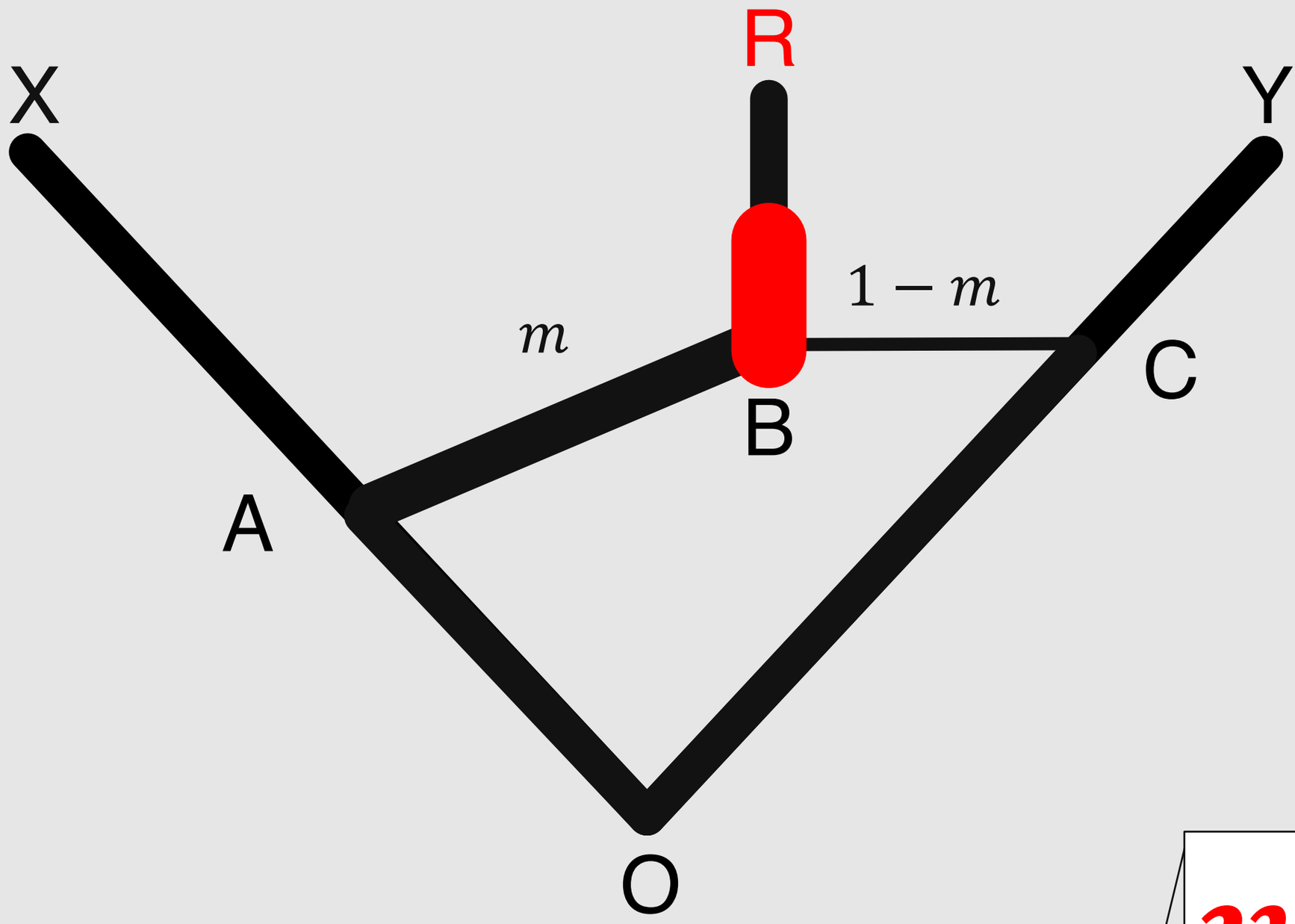
	X	R	Y
X	$\sigma^2(t_1+t_2+t_3)$	$\sigma^2 m t_1$	0
R	$\boldsymbol{\sigma^2 m t_1}$	$\sigma^2(m^2+(1-m)^2)(t_1+t_2)+\sigma^2 t_3+v_H$	$\sigma^2(1-m)(t_1+t_2)$
Y	0	$\sigma^2(1-m)(t_1+t_2)$	$\sigma^2(t_1+t_2+t_3)$

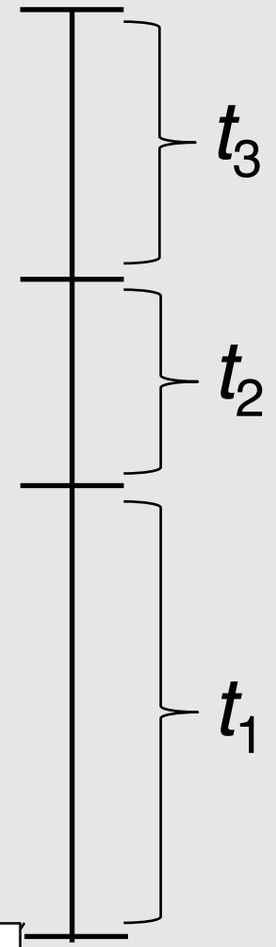


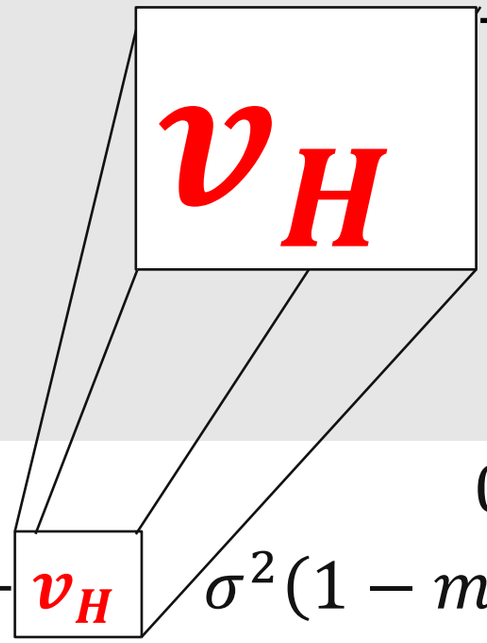
	X	R	Y
X	$\sigma^2(t_1+t_2+t_3)$	$\sigma^2 m t_1$	0
R	$\boldsymbol{\sigma^2 m t_1}$	$\sigma^2(m^2+(1-m)^2)(t_1+t_2)+\sigma^2 t_3+v_H$	$\sigma^2(1-m)(t_1+t_2)$
Y	0	$\sigma^2(1-m)(t_1+t_2)$	$\sigma^2(t_1+t_2+t_3)$



	X	R	Y
X	$\sigma^2(t_1+t_2+t_3)$	$\sigma^2 m t_1$	0
R	$\sigma^2 m t_1$	$\boldsymbol{\sigma^2(m^2+(1-m)^2)(t_1+t_2)+\sigma^2 t_3+v_H}$	$\sigma^2(1-m)(t_1+t_2)$
Y	0	$\sigma^2(1-m)(t_1+t_2)$	$\sigma^2(t_1+t_2+t_3)$

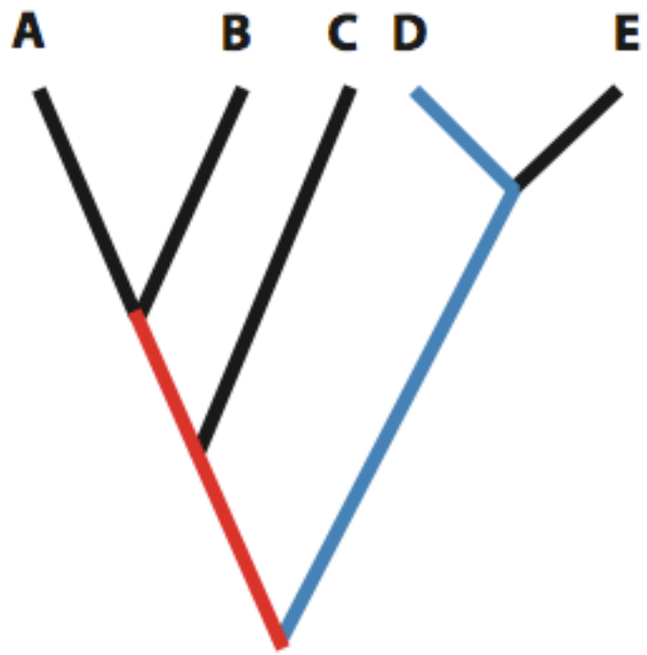






$$
\begin{array}{c} \\ X \\ R \\ Y \end{array}
\begin{array}{c}
\begin{array}{ccc} X & R & Y \end{array} \\
\begin{bmatrix}
\sigma^2(t_1+t_2+t_3) & \sigma^2 m t_1 & 0 \\
\sigma^2 m t_1 & \sigma^2(m^2+(1-m)^2)(t_1+t_2)+\sigma^2 t_3+v_H & \sigma^2(1-m)(t_1+t_2) \\
0 & \sigma^2(1-m)(t_1+t_2) & \sigma^2(t_1+t_2+t_3)
\end{bmatrix}
\end{array}
$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$



Variance-covariance matrix

	A	B	C	D	E
A	var(A)	cov(A,B)	cov(A,C)	cov(A,D)	cov(A,E)
B	cov(B,A)	var(B)	cov(B,C)	cov(B,D)	cov(B,E)
C	cov(C,A)	cov(C,B)	var(C)	cov(C,D)	cov(C,E)
D	cov(D,A)	cov(D,B)	cov(D,C)	var(D)	cov(D,E)
E	cov(E,A)	cov(E,B)	cov(E,C)	cov(E,D)	var(E)

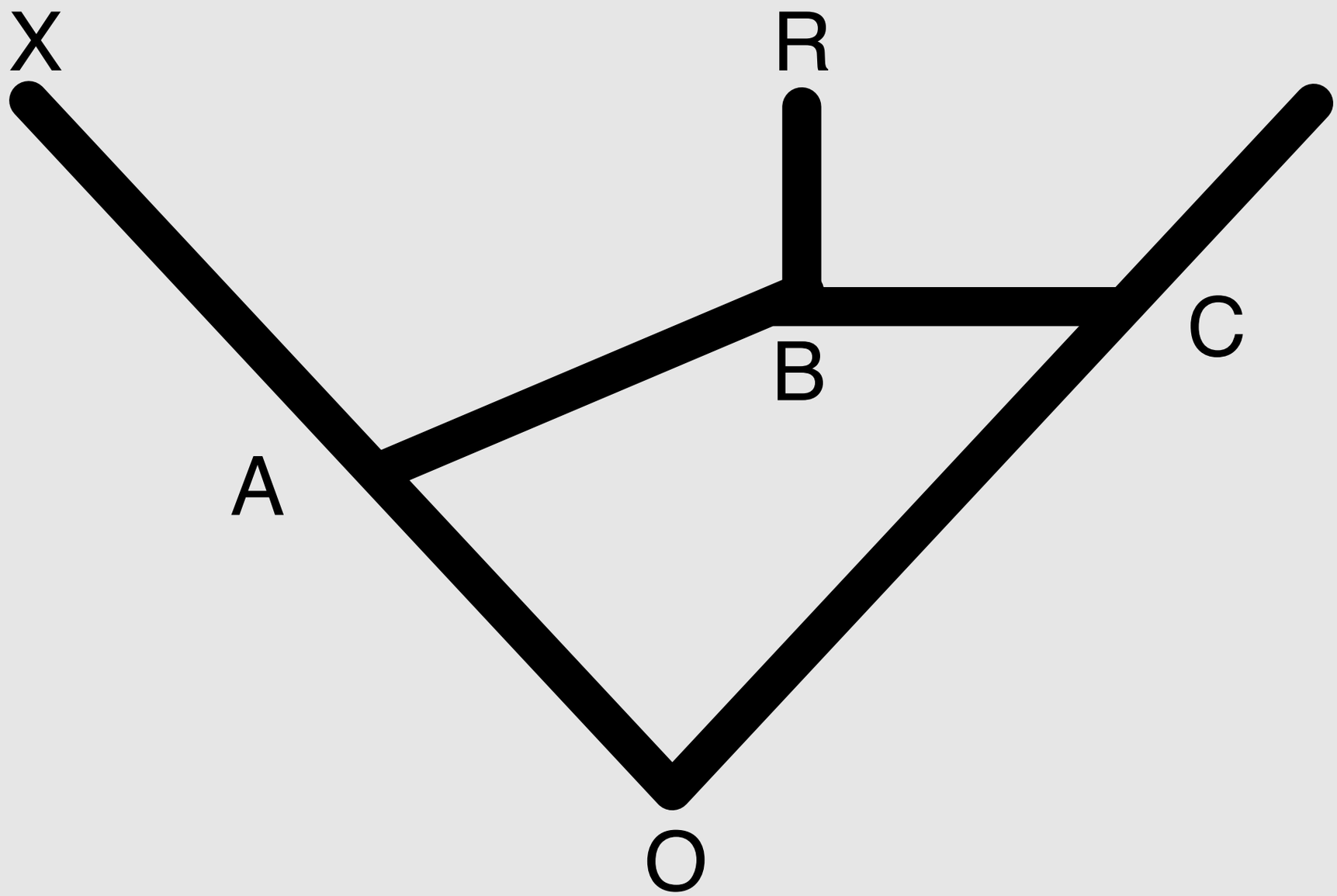
Mean vector

A	μ_A
B	μ_B
C	μ_C
D	μ_D
E	μ_E

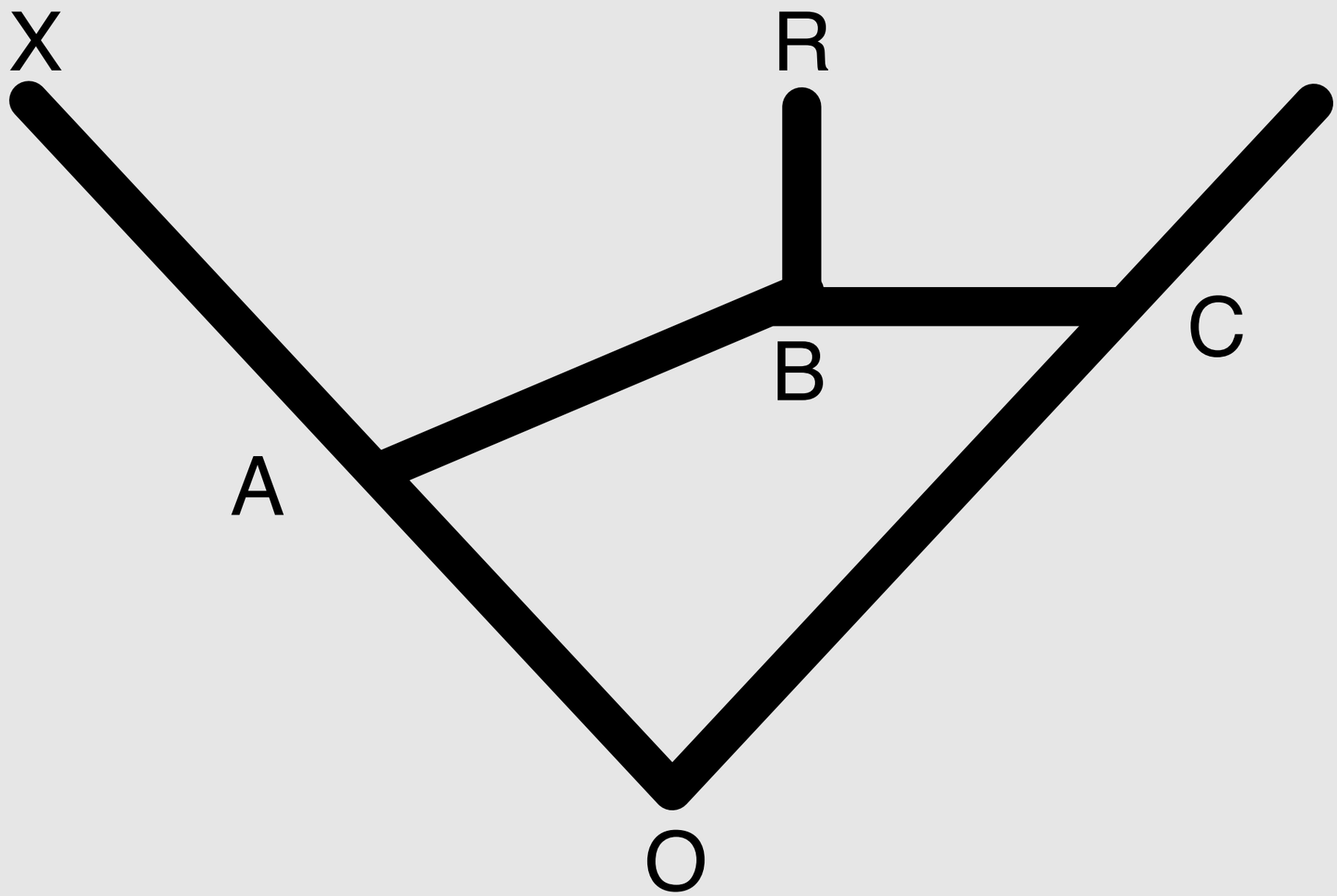
Figure 4

Multivariate normal distribution. The figure shows a tree, the tree's variance-covariance matrix, and the vector of means (which, under Brownian motion, would equal the root state). Highlighted are the branches leading to covariance between taxa A and B (*red*) and the branches leading to variance in D (*blue*).

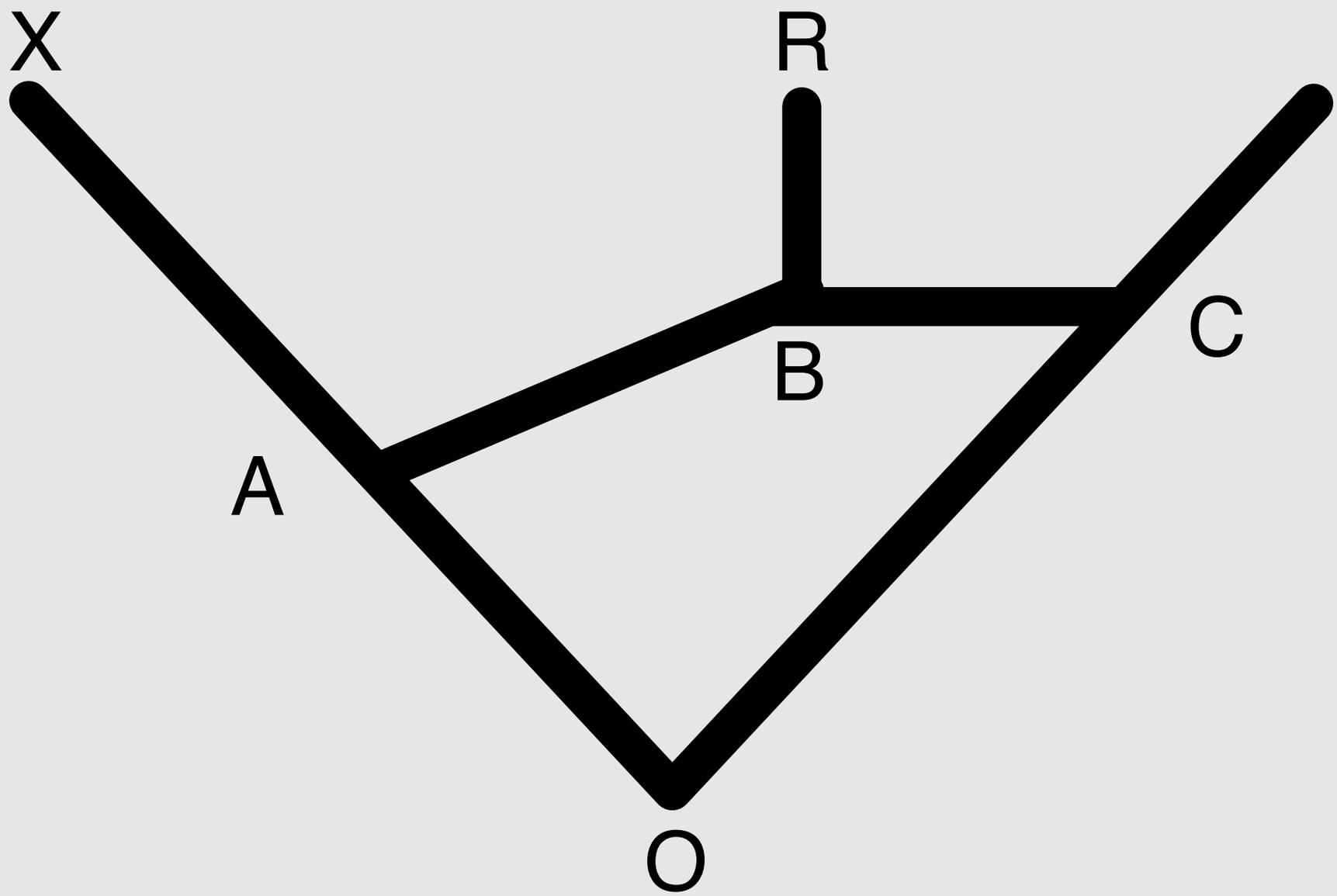
O'Meara 2012



$$\mu_B = [\, m\mu_D + (1-m)\mu_C] + \log\beta$$

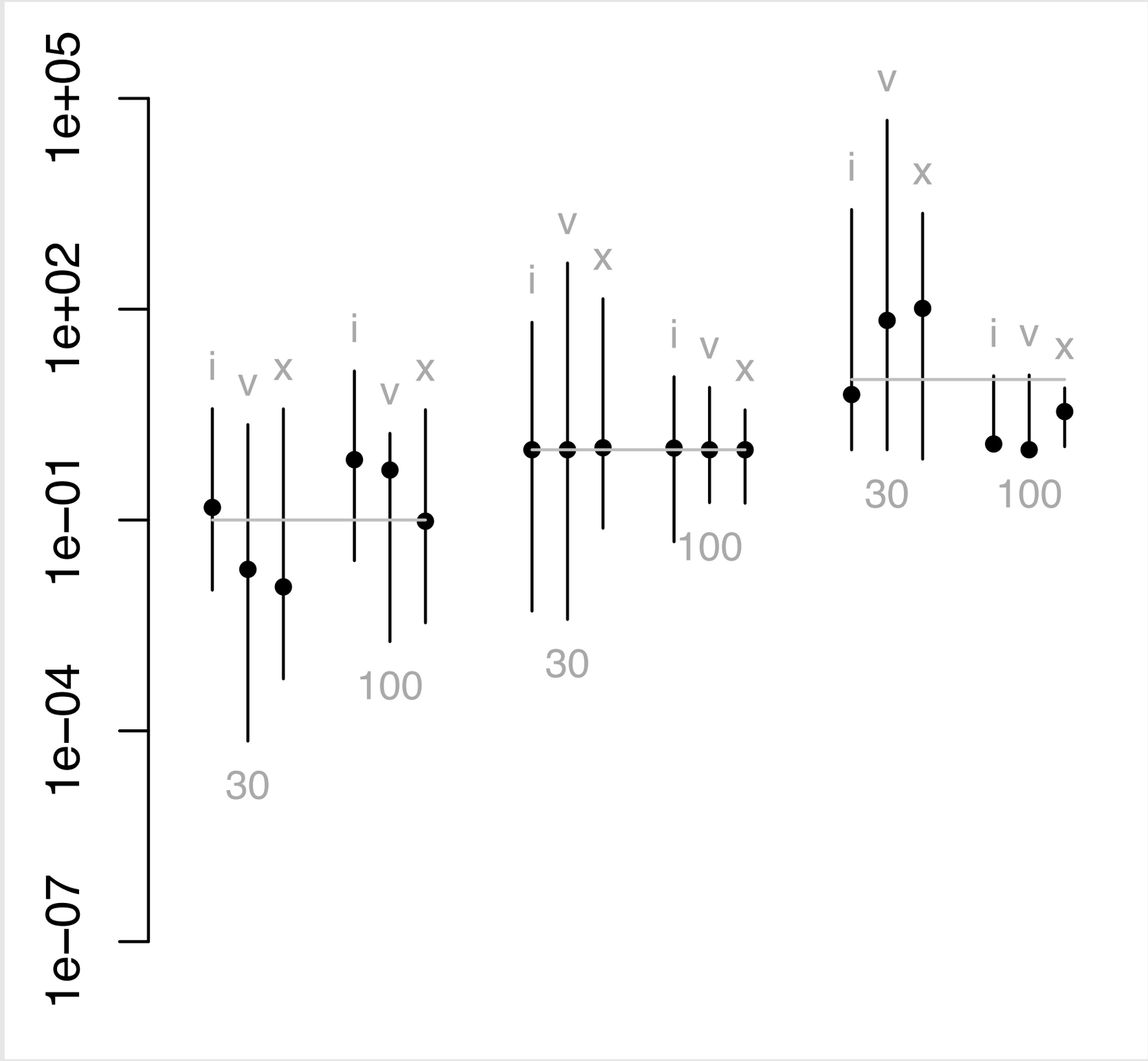


$$\mu_B = [\, m\mu_D + (1-m)\mu_C] + \log\beta$$

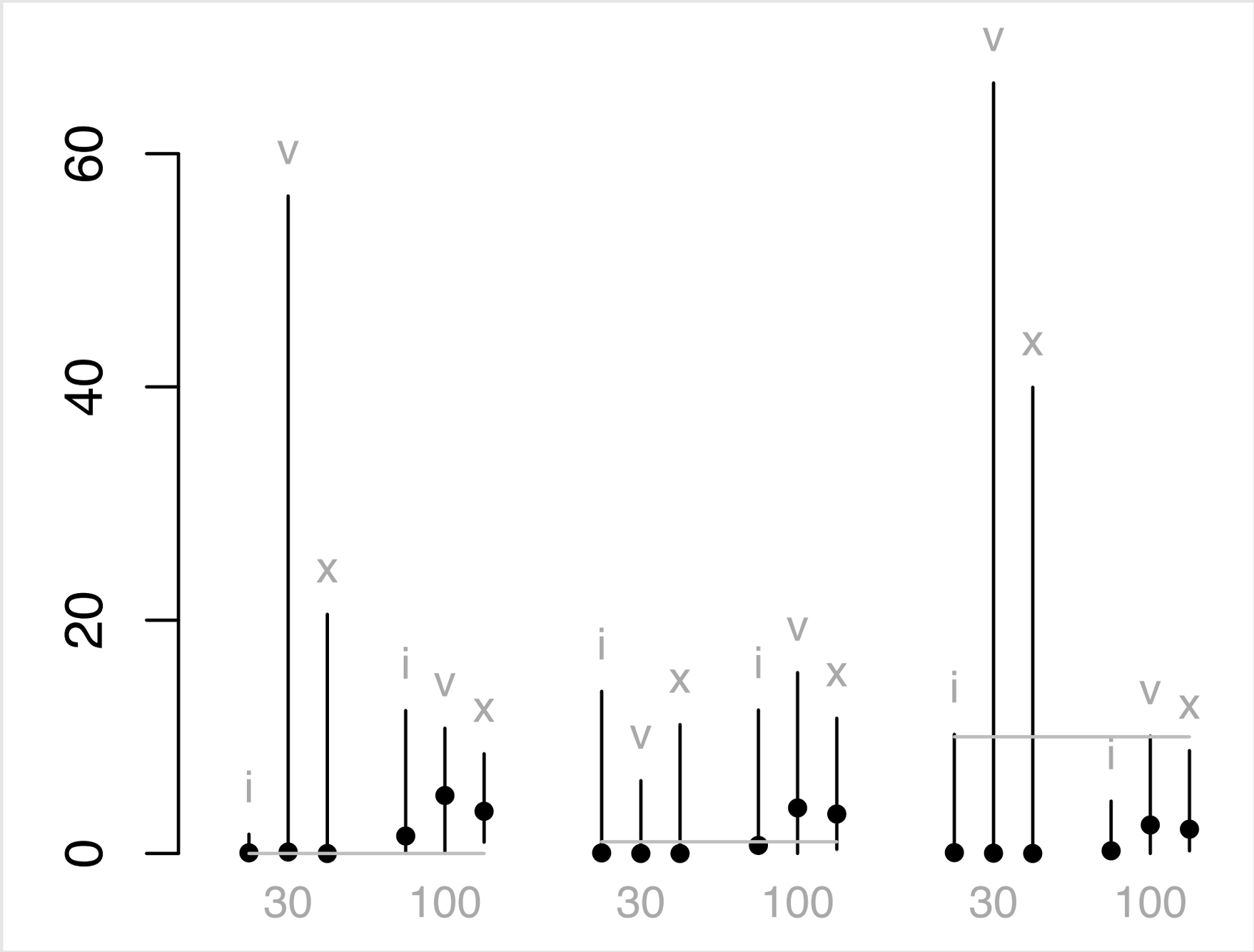


$$\mu_B = [\, m\mu_D + (1-m)\mu_C] + \log\beta$$

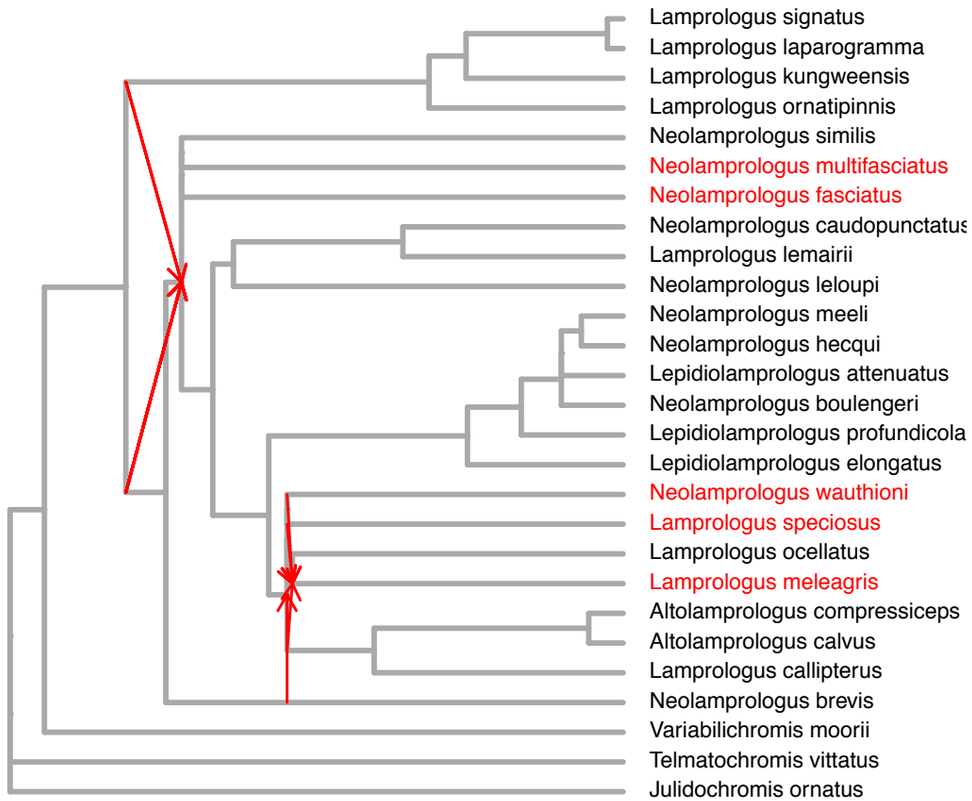
$$\mu_B = [\, m\mu_D + (1-m)\mu_C] + \log \beta$$



v_H

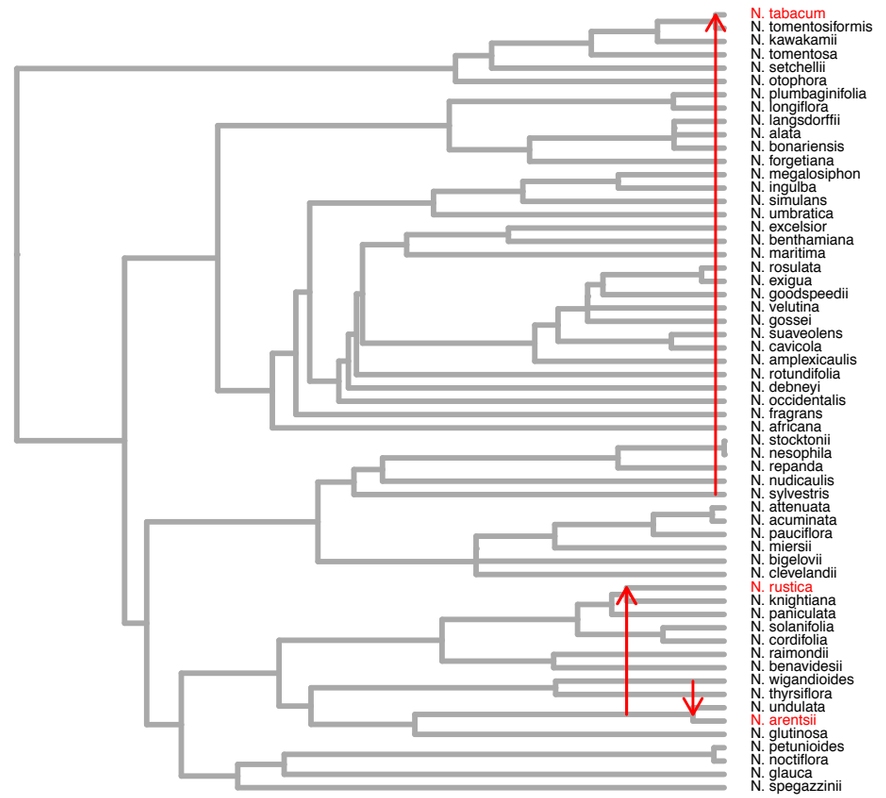


Cichlid



Network: Kobmüller et al. (2007)
Traits: standard length from fishbase (Froese and Pauly, 2010) using rfishbase (Boettiger et al. 2012)

Nicotiana



Network: Chase et al. (2003)
Traits: seedling growth under mannitol treatment (Komori et al. 2000)

Cichlids

Akaike weight	σ^2	μ	β	vH	SE	variance over tree
	$(\log(cm))^2/MY$	$\log(cm)^2$	unitless scalar	$\log(cm)^2$	log(cm)	$\log(cm)^2$
0.525	0 (0, 0.093)	2.152 (1.737, 2.454)	1 (fixed)	0 (fixed)	0.324 (0.091, 0.558)	0 (0, 0.524)
0.129	0 (0, 0.08)	2.155 (1.751, 2.44)	1 (fixed)	0 (0, 1.052)	0.31 (0.104, 0.536)	0 (0, 0.448)
0.284	0 (0, 0.079)	2.216 (1.82, 2.509)	0.708 (0.419, 1.196)	0 (fixed)	0.306 (0.117, 0.527)	0 (0, 0.446)
0.062	0 (0, 0.068)	2.216 (1.848, 2.497)	0.707 (0.434, 1.192)	0 (0, 0.524)	0.305 (0.128, 0.501)	0 (0, 0.384)
Model average	0 (0, 0.086)	2.175 (1.77, 2.47)	**0.899 (0.8, 1.068)**	0 (0, 0.169)	0.316 (0.102, 0.543)	0 (0, 0.483)

Nicotiana

Akaike weight	σ^2 log(rel. growth)^2/MY	μ log(rel.growth)^2	β unitless scalar	vH log(rel.growth)^2	SE log(rel.growth)	variance over tree log(rel.growth)^2
0.084	0.04 (0.008, 0.083)	2.601 (2.02, 3.191)	1 (fixed)	0 (fixed)	0.472 (0.318, 0.801)	0.347 (0.071, 0.73)
0.036	0.053 (0.019, 0.132)	2.582 (1.945, 3.234)	1 (fixed)	0.412 (0.105, 5.568)	0.393 (0.2, 0.723)	0.462 (0.171, 1.156)
0.716	0.076 (0.037, 0.091)	2.601 (2.119, 3.188)	2.358 (1.522, 2.748)	0 (fixed)	0.329 (0.279, 0.479)	0.664 (0.324, 0.803)
0.165	0.08 (0.036, 0.119)	2.511 (2.045, 2.91)	2.42 (1.448, 2.982)	0.025 (0.007, 0.495)	0.332 (0.243, 0.502)	0.701 (0.314, 1.044)
Model average	0.072 (0.034, 0.097)	2.586 (2.093, 3.144)	**2.206** **(1.447, 2.578)**	0.019 (0.005, 0.279)	0.344 (0.274, 0.519)	0.636 (0.296, 0.849)

```
Package: BMhyb
Version: 1.3.5
Date: 2016-08-22
Type: Package
Title: PCM for Hybridization
Author: Dwueng-Chwuan Jhwueng <djhwueng@umail.iu.edu> and Brian O'Meara
    <omeara.brian@gmail.com>
Maintainer: Dwueng-Chwuan Jhwueng <djhwueng@umail.iu.edu>
Imports: corpcor, numDeriv, geiger, ape, TreeSim, phytools, phylobase,
        mvtnorm, grDevices, graphics, methods, stats
Description: The BMhyb package analyzes the phenotypic evolution of species
    of hybrid origin on a phylogenetic network. This package can detect the hybrid
    vigor effect, a burst of variation at formation, and the relative portion of
    heritability from its parents. Parameters are estimated by maximum likelihood.
    Users need to enter a comparative data set, a phylogeny, and information on gene
    flow leading to hybrids.
License: GPL (>=2)
RoxygenNote: 5.0.1
Built: R 3.4.0; ; 2017-06-26 07:16:09 UTC; unix
```

```
Package: BMhyb
Version: 1.3.5
Date: 2016-08-22
Type: Package
Title: PCM for Hybridization
Author: Dwueng-Chwuan Jhwueng <djhv
    <omeara.brian@gmail.com>
Maintainer: Dwueng-Chwuan Jhwueng <
Imports: corpcor, numDeriv, geiger,
        mvtnorm, grDevices, graphic
Description: The BMhyb package anal
    of hybrid origin on a phylogene
    vigor effect, a burst of variat
    heritability from its parents.
    Users need to enter a comparati
    flow leading to hybrids.
License: GPL (>=2)
RoxygenNote: 5.0.1
Built: R 3.4.0; ; 2017-06-26 07:16:
```

Cold Spring Harbor Laboratory

bioRχiv beta

THE PREPRINT SERVER FOR BIOLOGY

HOME | Search

New Results

Trait Evolution on Phylogenetic Networks

Dwueng-Chwuan Jhwueng, Brian O'Meara

doi: https://doi.org/10.1101/023986

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract

Species may evolve on a reticulate network due to hybridization or other gene flow rather than on a strictly bifurcating tree, but comparative methods to deal with trait evolution on a network are lacking. We create such a method, which uses a Brownian motion model. Our method seeks to separately or jointly detect a bias in trait value coming from hybridization (β) and a burst of variation at the time of hybridization (v_H) associated with the hybridization event, as well as traditional Brownian motion parameters of ancestral state (μ) and rate of evolution (σ^2) of Brownian motion, as well as measurement error of the tips (SE). We test the method with extensive simulations. We also apply the model to two empirical examples, cichlid body size and Nicotiana drought tolerance, and find substantial measurement error and a hint that hybrids have greater drought tolerance in the latter case. The new methods are available in CRAN R package BMhyd.

Caveats

- Some numerical issues (matrix condition makes inversion imprecise sometimes)
- Large uncertainty in some param estimates
- Peer-reviewed, but **not yet accepted**. So don’t use yet for real work