

OU: a top-down look

Stochastic motion rate

Ornstein-Uhlenbeck mean

Continuous state

Brian O'Meara

<http://www.brianomeara.info>

 @omearabrian

In what ways can a continuous trait change in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards some value

In what ways can a continuous trait **change** in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards some value

$$dX_T =$$

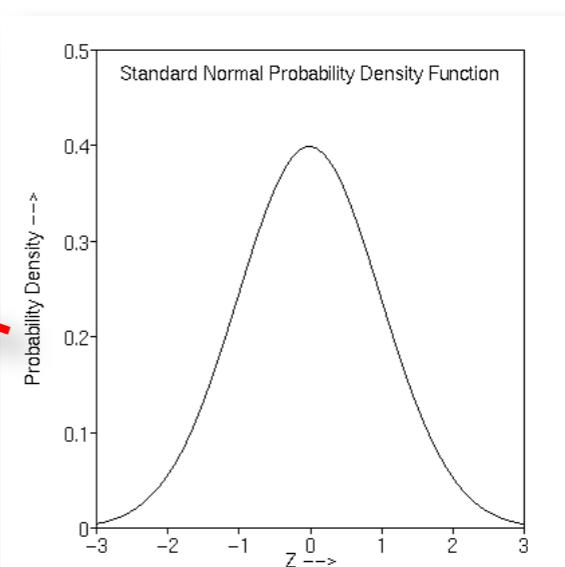
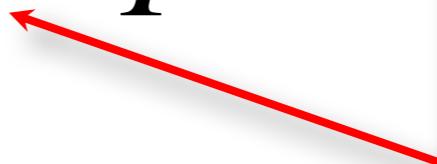
In what ways can a continuous trait change in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards some value

$$dX_T =$$

$$dW_T$$



In what ways can a continuous trait change in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards some value

$$dX_T = \sigma \times dW_T$$

Rate of wiggle

In what ways can a continuous trait change in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards **some value**

$$dX_T = \sigma \times dW_T \quad \theta$$

In what ways can a continuous trait change in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards some value

$$dX_T = \sigma \times dW_T + (\theta - X_T)$$

Adds the entire difference

In what ways can a continuous trait change in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards some value

$$dX_T = \sigma \times dW_T + \alpha(\theta - X_T)$$

Allows directional change less than 100% (even zero)

In what ways can a continuous trait change in an instant of time?

Randomly: increase or decrease slightly by chance
and/or

Directionally: be pulled towards some value

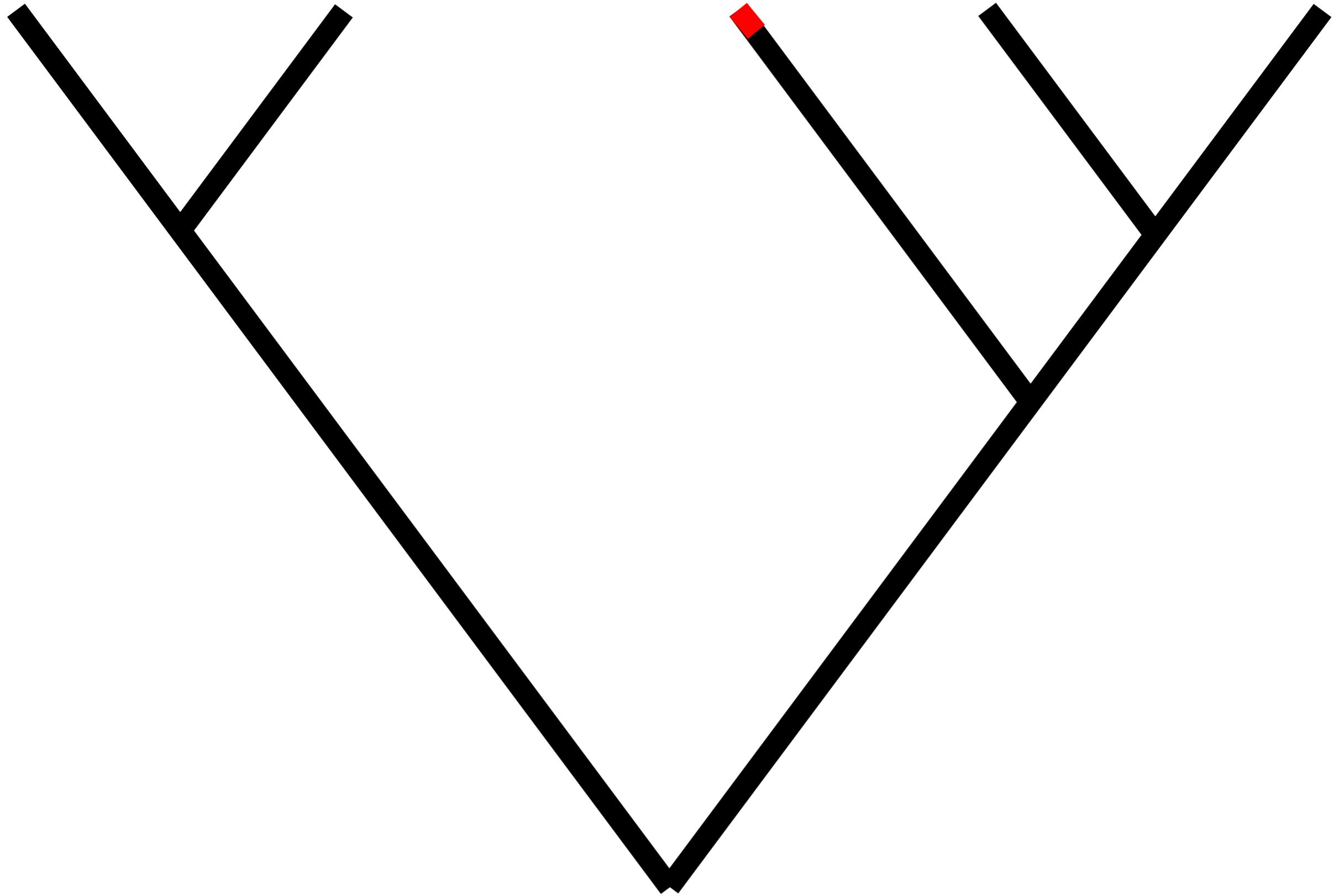
$$dX_T = \sigma \times dW_T + \alpha(\theta - X_T)$$

Ornstein-Uhlenbeck process

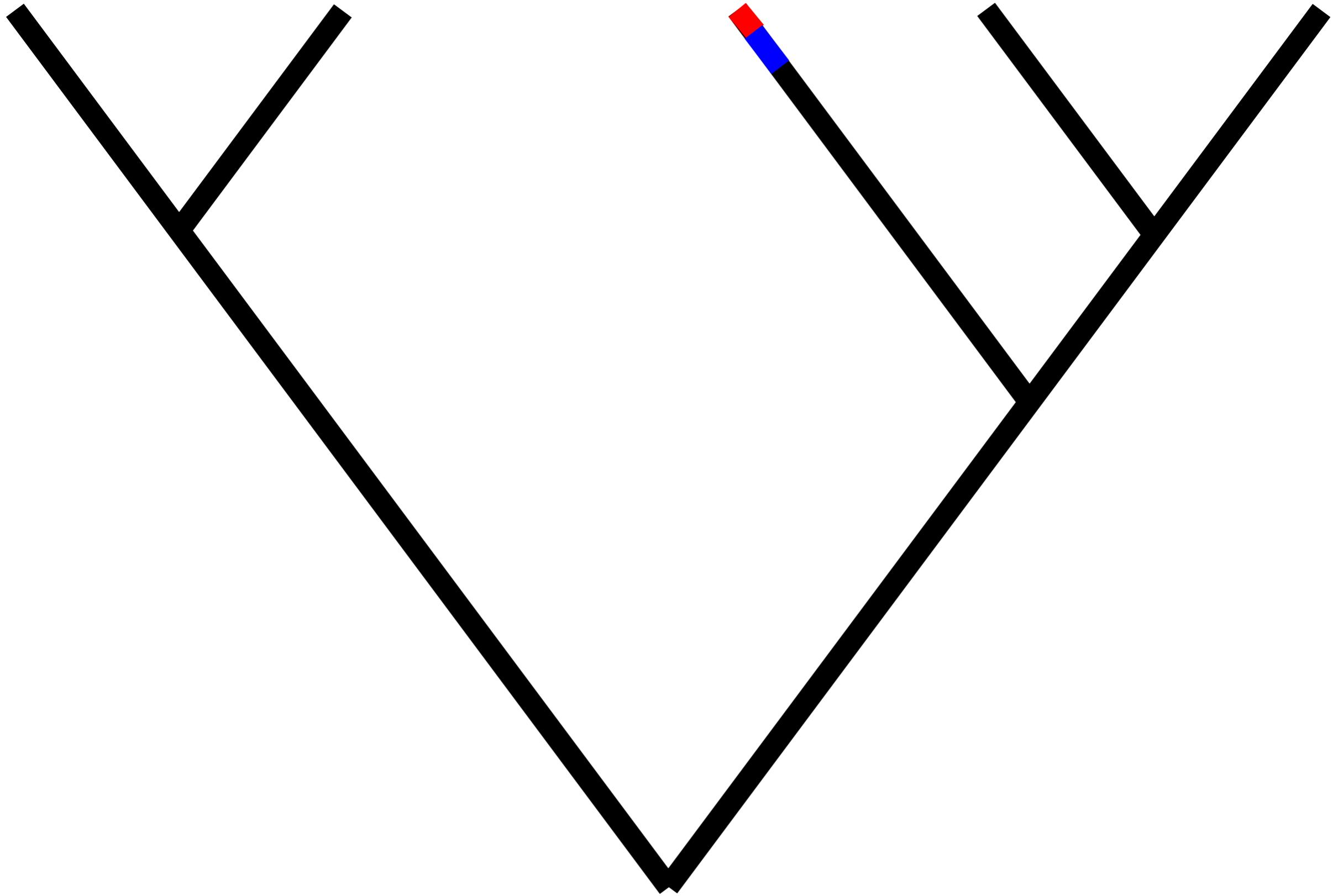


$$dX_T = \sigma \times dW_T + \alpha(\theta - X_T)$$

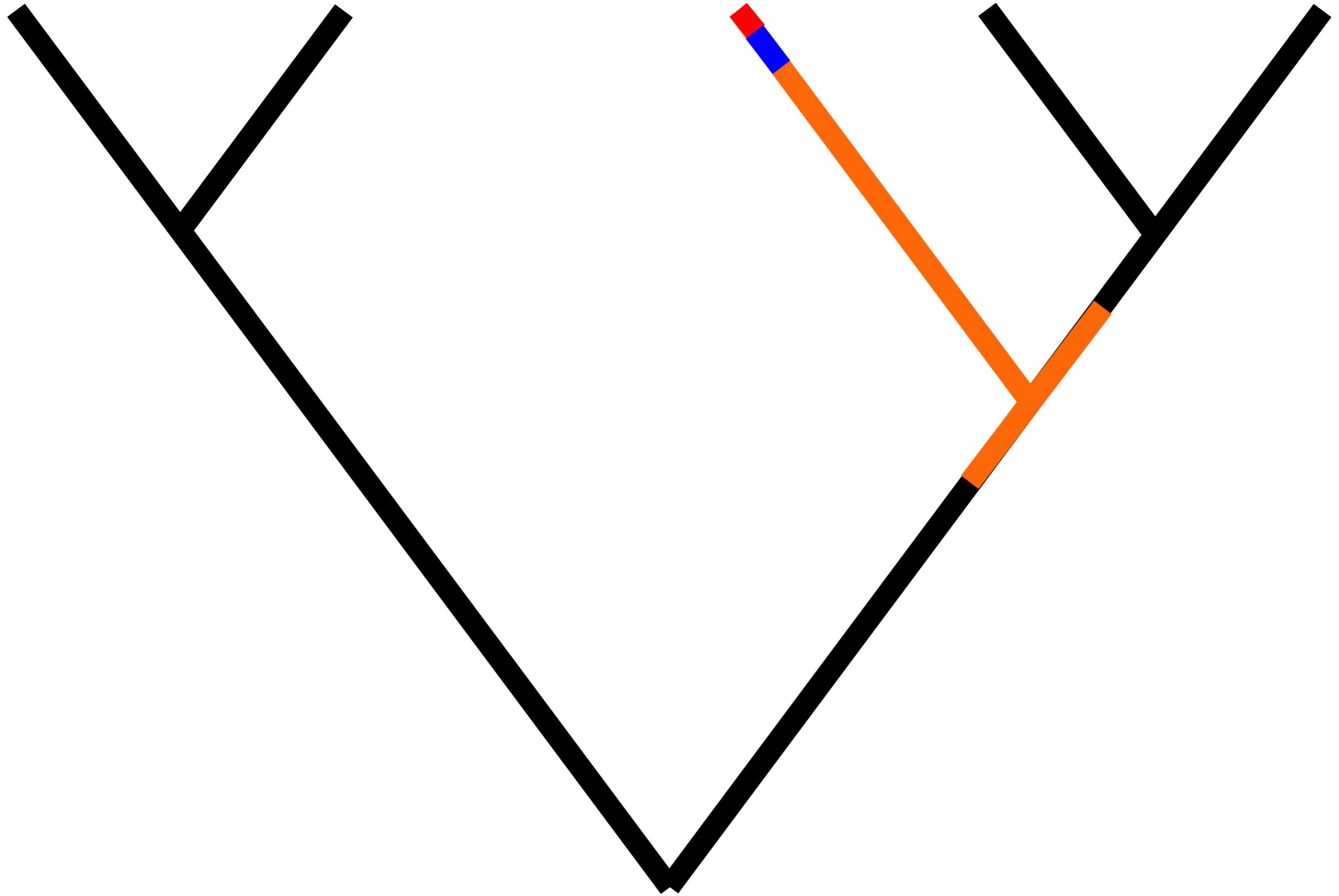
Ornstein-Uhlenbeck process



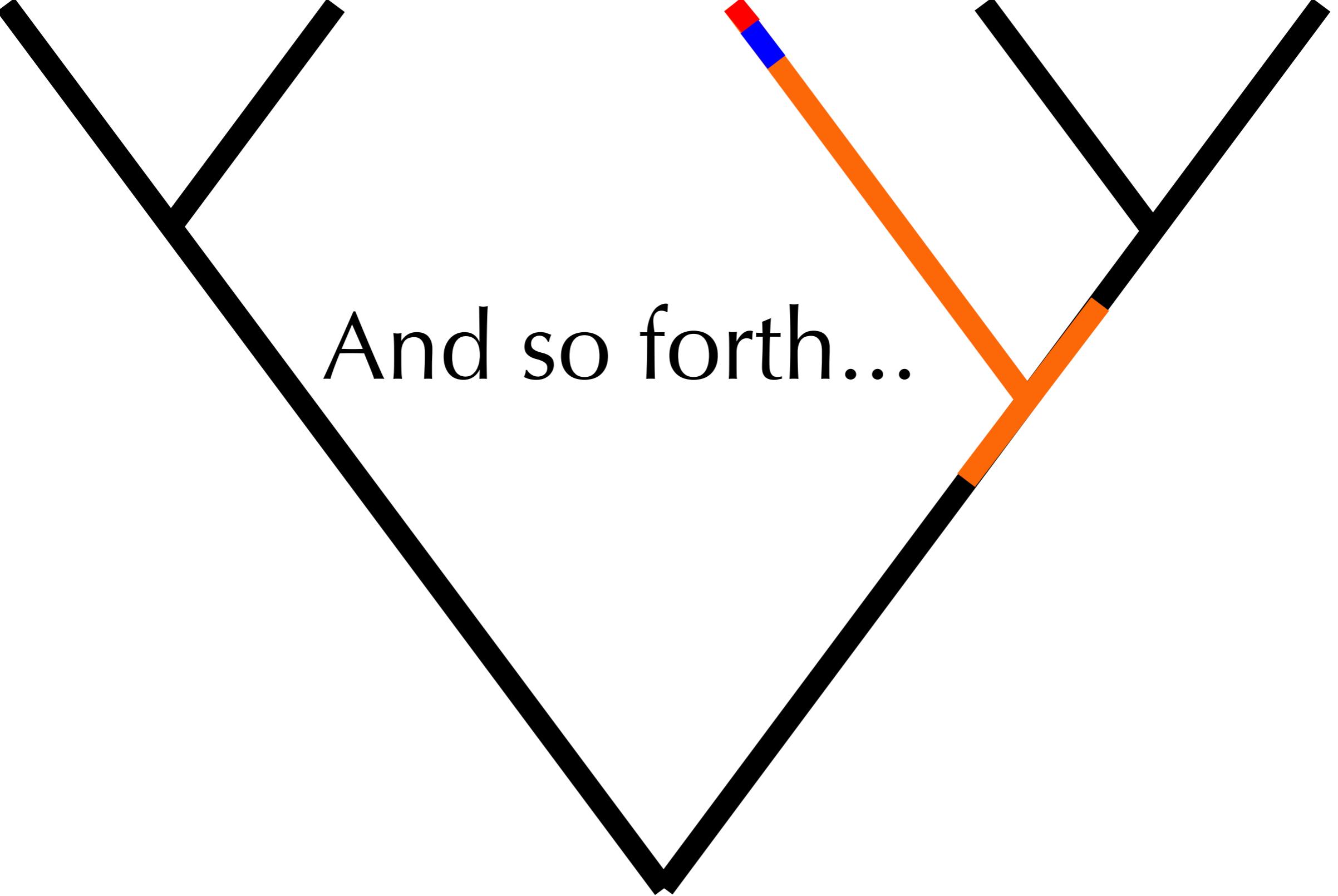
$$dX_{T_1} = \sigma_1 \times dW_T + \alpha_1 (\theta_1 - X_{T_1})$$



$$dX_{T_2} = \sigma_2 \times dW_T + \alpha_2 (\theta_2 - X_{T_2})$$



$$dX_{T_3} = \sigma_3 \times dW_T + \alpha_3 (\theta_3 - X_{T_3})$$



And so forth...

$$dX_{T_3} = \sigma_3 \times dW_T + \alpha_3 (\theta_3 - X_{T_3})$$

$$\frac{\sigma_i}{\alpha_i} \theta_i$$

$$dX_{T_i} = \sigma_i \times dW_T + \alpha_i (\theta_i - X_{T_i})$$

Brownian rate OU attraction OU mean

	Single rate Brownian motion	Multiple mean Ornstein-Uhlenbeck	Multiple rate Brownian motion	Multiple everything
σ_i	all equal	all equal	some vary	some vary
α_i	0	all equal	0	some vary
θ_i	NA	some vary	NA	some vary
Independent contrasts (Felsenstein, 1985), ANCML (Schluter et al, 1998)	Hansen, 1997; OUCH (Butler & King, 2004), SURFACE (Ingram & Mahler, 2012)	Brownie (O'Meara et al., 2006, Thomas et al., 2006), AUTEUR (Eastman et al. 2011)	OUwie (Beaulieu et al. 2012), bayou (Uyeda et al. 2014)	

$$dX_{T_i} = \sigma \times dW_T + \alpha (\theta_i - X_{T_i})$$

Brownian rate OU attraction OU mean

Model	Optima (mean = M)	Sigma (variance = V)	Attraction (A)
BM ₁		1	
BM _S		≥2	
OU ₁	1	1	1
OU _M	≥2	1	1
OU _{MA}	≥2	1	≥2
OU _{MV}	≥2	≥2	1
OU _{MVA}	≥2	≥2	≥2

Modeling stabilizing selection: expanding
the Ornstein-Uhlenbeck model of adaptive
evolution

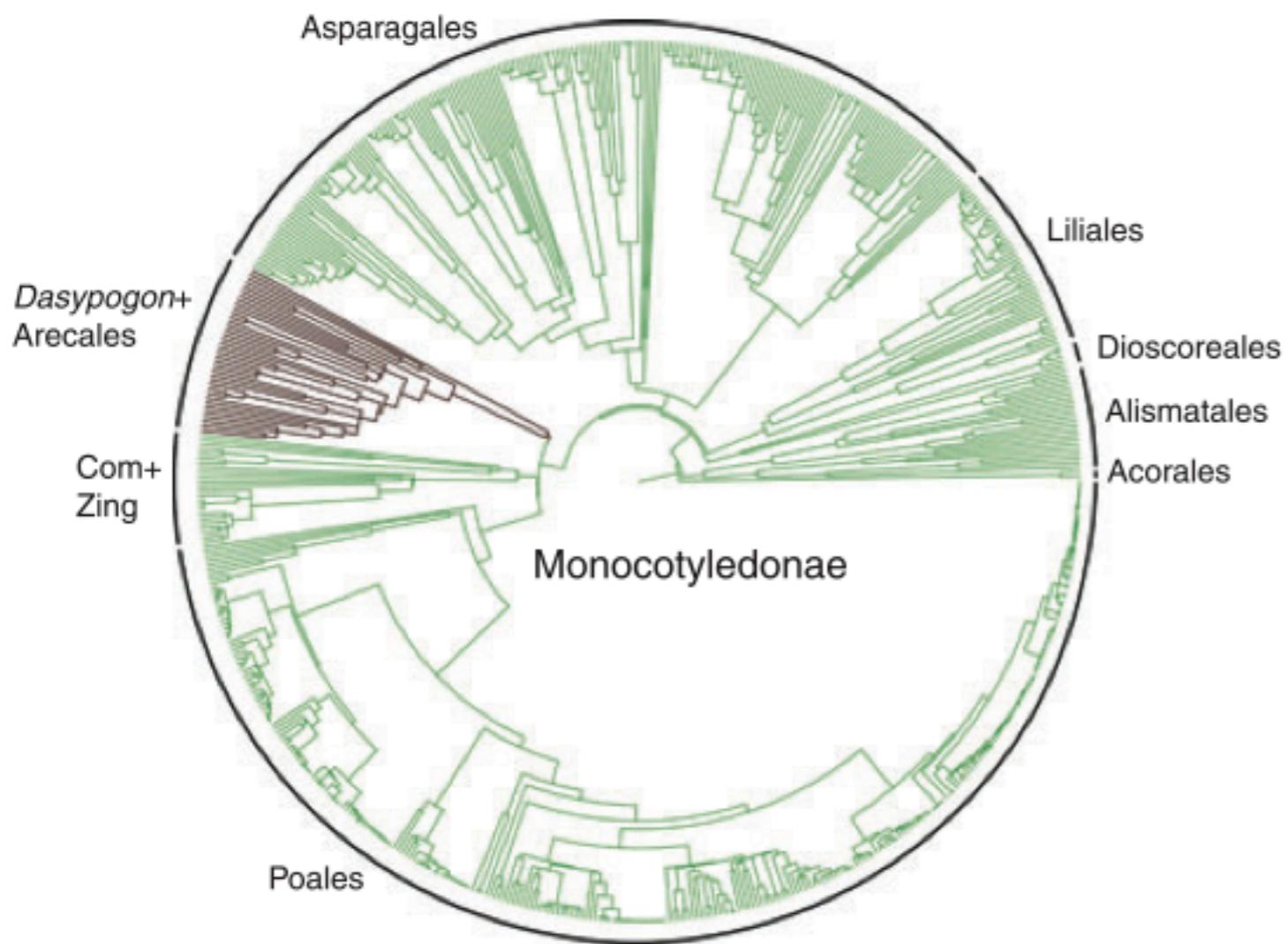
Modeling “stabilizing selection”: expanding
the Ornstein-Uhlenbeck model of adaptive
evolution

TABLE 1. A summary of some phenotypic covariance structures expected under microevolutionary models discussed in the text. Variables are defined in the main text, and some of the formulas are simplified for clarity.

Model	$\text{Cov}[\mathbf{X}_i, \mathbf{X}_j] \approx$	Correlation between species phenotypes
Random genetic drift:		
alone	$(\mathbf{G}/N_e)t_z$	Decreases linearly with increasing phylogenetic distance (but may take other forms)
drift-mutation balance	$2\mathbf{G}_m t_z$	
Directional selection:		
with random genetic drift	$(\mathbf{G}/N_e)t_z$	Decreases linearly with increasing phylogenetic distance.
fluctuating environment	$\mathbf{G}\mathbf{V}_s \mathbf{G}' t_z$	Trend cannot be detected from comparative data alone.
Stabilizing selection and drift:		
univariate	$\hat{\mathbf{V}} \text{Exp}[-w\mathbf{V}_A t_{ij}]$	Decreases exponentially with increasing phylogenetic distance.
multivariate	$\mathbf{Q}(t_{iz}) \hat{\mathbf{V}} \mathbf{Q}'(t_{jz})$	Decreases with increasing phylogenetic distance as a sum of exponentials.
Stabilizing selection + environmental change:		
Brownian motion of optimum	$\mathbf{E}_\theta t_z (+ \text{exponentials})$	Decreases linearly with increasing phylogenetic distance.
Lynch-Lande model	$\mathbf{Q}(t_{iz}) \mathbf{E}_\theta \mathbf{Q}'(t_{jz})$	Decreases with increasing phylogenetic distance as a sum of exponentials.
Punctuated phenotypic change	$\left(\mathbf{H} + \frac{\sigma^2}{\mu} \mathbf{h} \mathbf{h}' \right) \frac{t_z}{\mu}$	Decreases linearly with increasing phylogenetic distance.
Change correlated with speciation events	$\mathbf{H} \mathbf{N}_z$	Decreases with ratio of shared to total speciation events.

Model	$\text{Cov}[\mathbf{X}_1, \mathbf{X}_j] \approx$
Random genetic drift: alone	❤️ $(\mathbf{G}/N_e)t_z$
drift-mutation balance	❤️ $2\mathbf{G}_m t_z$
Directional selection: with random genetic drift	❤️ $(\mathbf{G}/N_e)t_z$
fluctuating environment	❤️ $\mathbf{G}\mathbf{V}_s\mathbf{G}'t_z$
Stabilizing selection and drift: univariate	👎 $\hat{\mathbf{V}}\text{Exp}[-w\mathbf{V}_A t_{ij}]$
multivariate	👎 $\mathbf{Q}(t_{iz})\hat{\mathbf{V}}\mathbf{Q}'(t_{jz})$
Stabilizing selection + environmental change: Brownian motion of optimum Lynch-Lande model	❤️ $E_\theta t_z (+ \text{exponentials})$ 👎 $\mathbf{Q}(t_{iz})E_\theta\mathbf{Q}'(t_{jz})$
Punctuated phenotypic change	❤️ $\left(\mathbf{H} + \frac{\sigma^2}{\mu} \mathbf{h}\mathbf{h}'\right) \frac{t_z}{\mu}$
Change correlated with speciation events	🤷‍♀️ $\mathbf{H}\mathbf{N}_z$

A



B

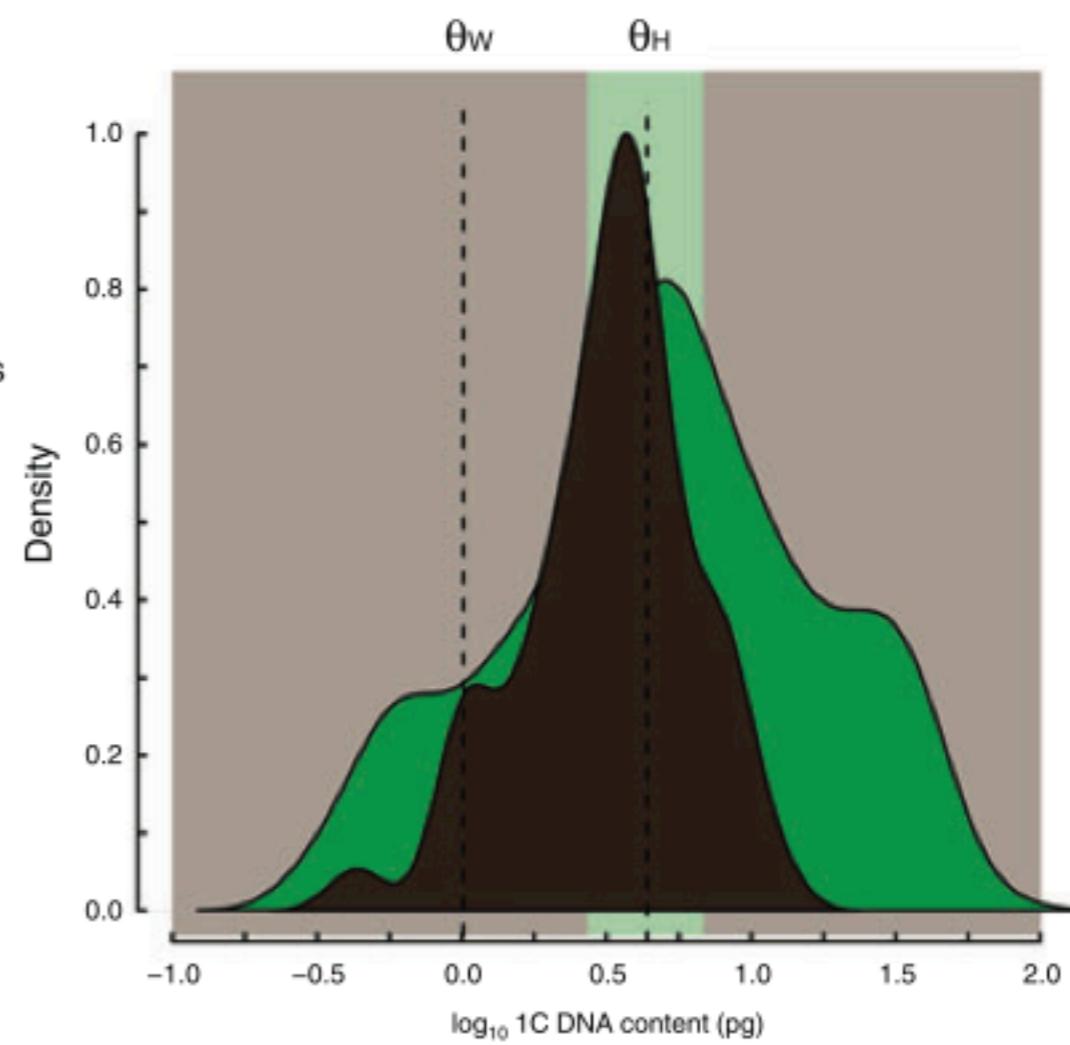


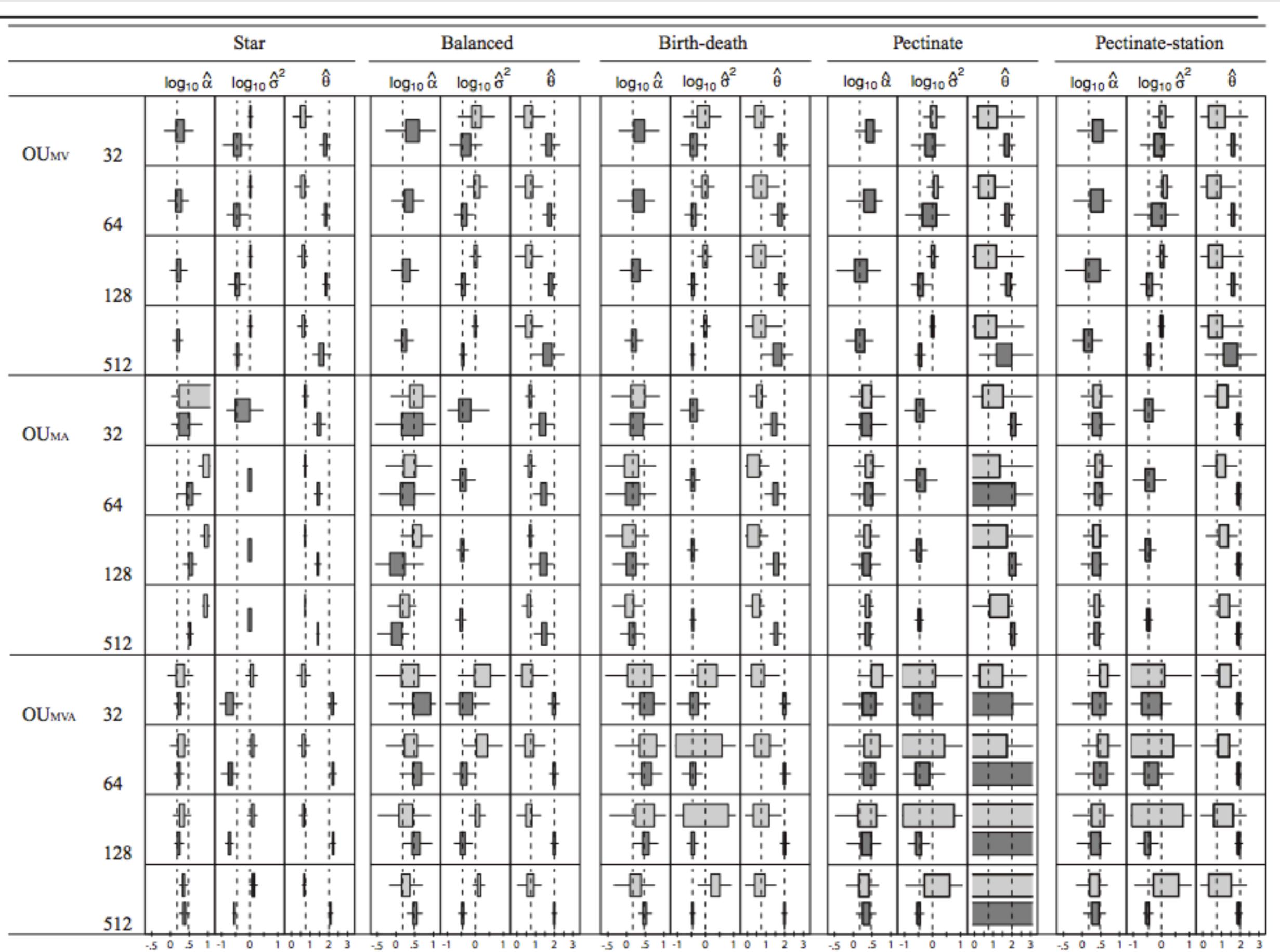
Figure 2. (A) Time-calibrated phylogeny of Monocotyledonae (monocots). The phylogeny is taken from a maximum likelihood analysis of 590 species based on combined analysis *atpB*, *matK*, *ndhF*, *rbcL*, and *trnL-F*. The major clades of monocots are labeled, and estimates of the likeliest growth form state (woody = brown; herbaceous = green) across all branches in the tree. Com + Zing represents the combined clade of Commelinaceae and Zingiberaceae. (B) The distributions of 1C DNA content among growth form, with the optimum value for woody plants (θ_W) estimated to be larger than the optimum inferred for herbaceous lineages (θ_H). However, the optimum value for woody lineages was not identifiable, which may be explained by very weak selection ($\alpha < 0.001$) operating within this regime.

Table 2. The fit of alternative models of genome size evolution in monocots. The best model, based on ΔAIC and Akaike weights, was the OU_{MVA} , which estimated a separate θ , α , and σ^2 for woody and herbaceous monocot lineages.

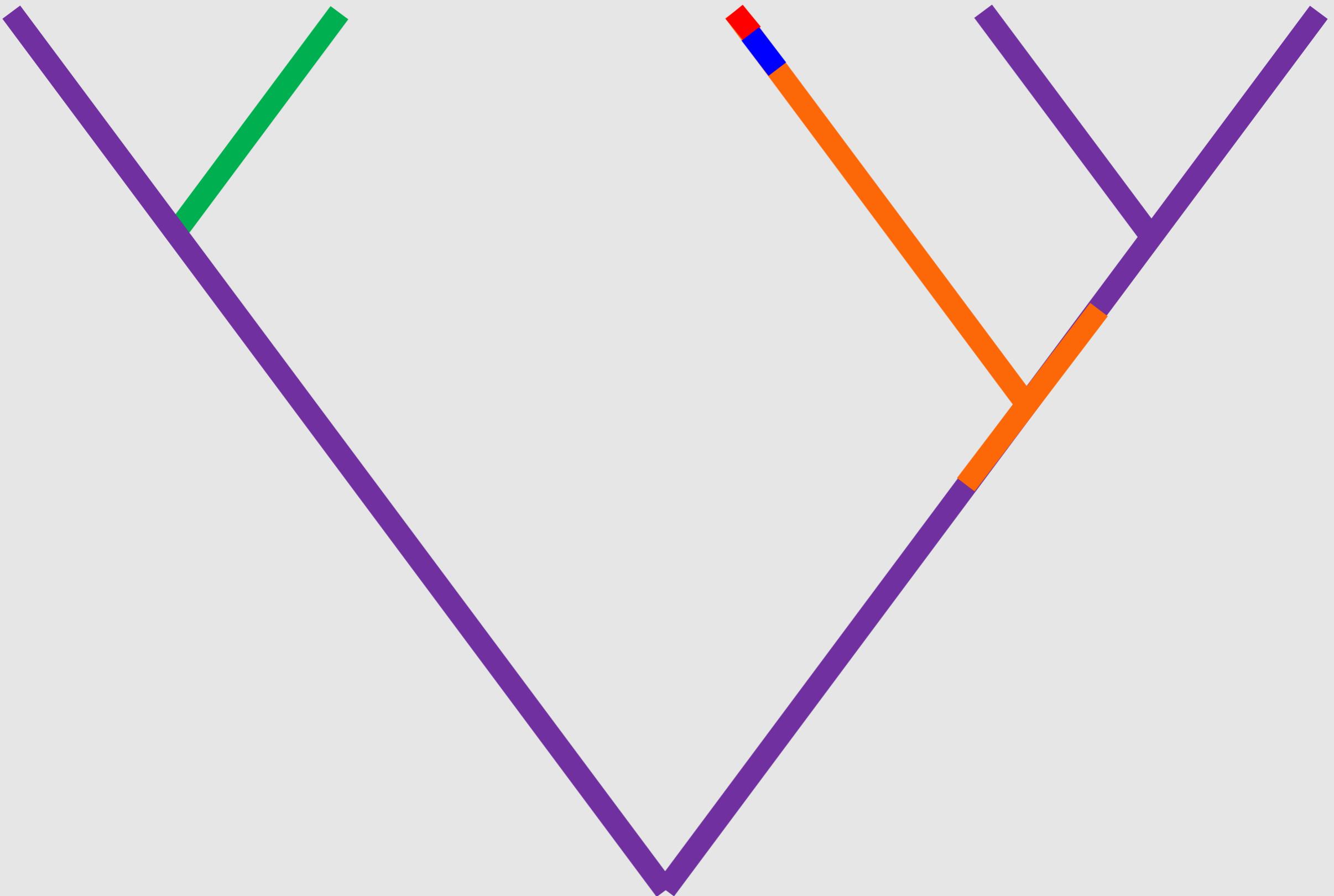
Model	-lnL	AIC	ΔAIC	w_i
BM1	-227.6	459.1	156.2	<0.01
BMS	-203.0	412.0	109.1	<0.01
OU1	-160.0	326.1	23.2	<0.01
OU _M	-159.2	326.5	23.6	<0.01
OU _{MV}	-147.3	304.6	1.7	0.290
OU _{MA}	-159.2	328.5	25.6	<0.01
OU _{MVA}	-145.3	302.9	0.0	0.678

Table 3. Parameter estimates and their associated 95% confidence interval (CI) for the OU_{MVA} model, the model that best fit the genome size data. Each CI was obtained by multiplying each approximate standard error by the critical value in the *t*-distribution where the cumulative probability is equal to 0.975 (i.e., $t(0.975, \infty) = 1.96$).

	Herb Estimate	95% CI	Woody Estimate	95% CI
α	3.85	± 0.955	<0.001	$\pm <0.01$
σ^2	2.51	± 0.376	0.531	± 0.281
θ	0.618	± 0.143	<0.001	$\pm \infty$



How do we figure out painting?



How do we figure out painting?

- a priori biological hypotheses
- Reconstruct a discrete trait, use that to paint continuous regimes
- Slice by taxonomy (monocots vs other plants)
- Slice by time (pre and post KT)
- Try a bunch (i.e. SURFACE)

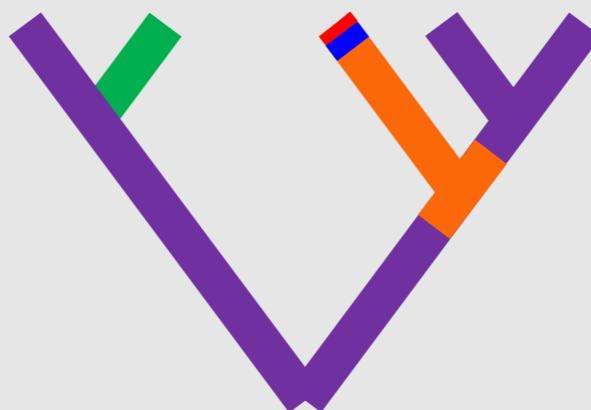


Table 2: Parameters estimated for the five models comparing character displacement with alternative hypotheses

	BM	OU(1)	OU(3)	OU(4)	OU(LP)
α		0	.32	14.67	2.49
σ	.21	.21	.20	.47	.22
θ_0	2.95	2.95	3.99	... ^a	.86
θ_{small}		... ^a	-1.40	2.58	2.75
θ_{medium}			.18	3.11	3.24
θ_{large}			2.71	3.30	3.56
$\theta_{\text{ancestral}}$				2.83	

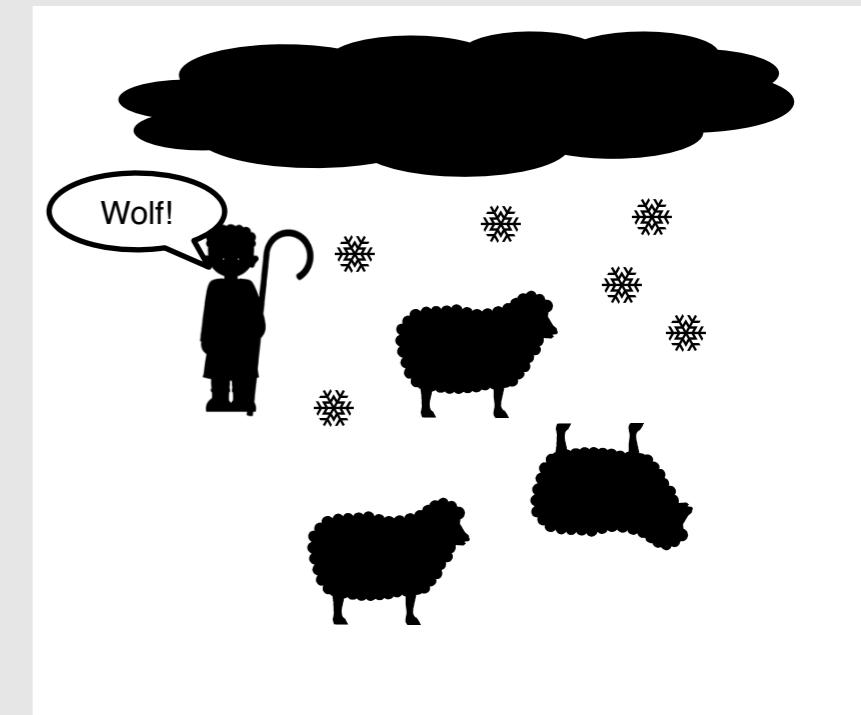
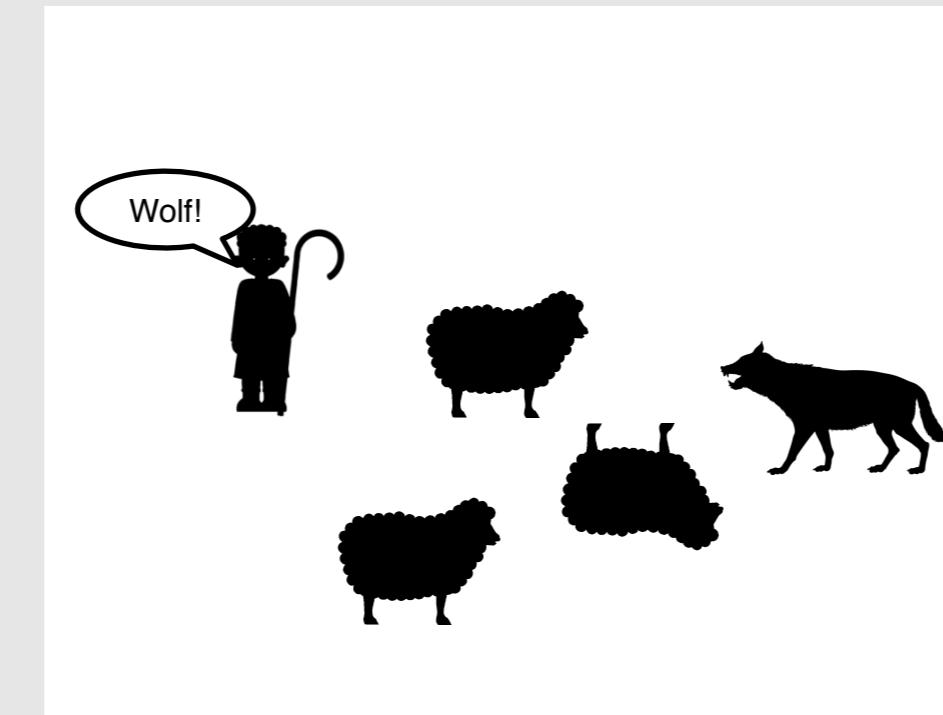
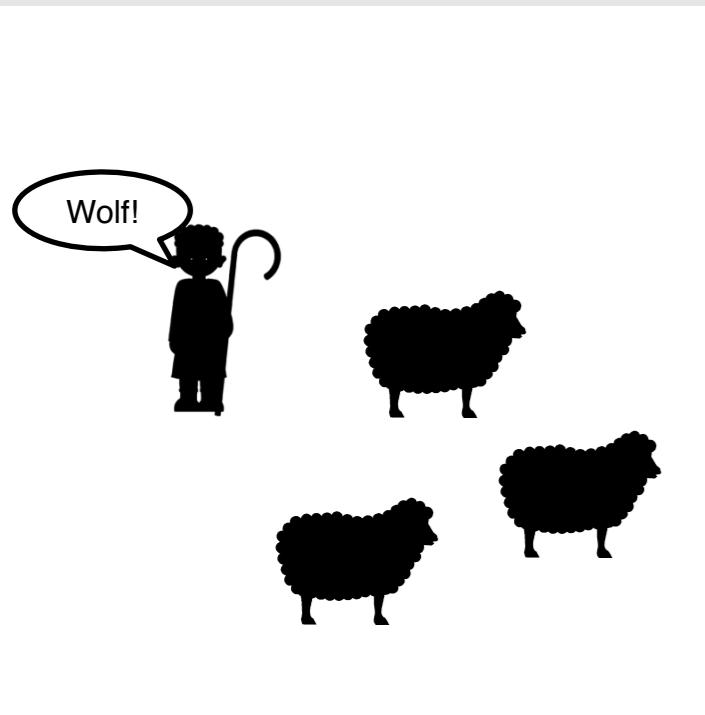
Phylogenetic half life
 = Tree height scaling $\times \ln(2) / \alpha$
 = 30 MY $\times \ln(2) / 2.49$
 = 8.4 MY

to go halfway from log(25 mm) to log(35 mm) head lengths: 0.6 mm/MY, or 0.24% of the starting value/MY

Model rejection / selection

- Is Model B better than Model A
- Can I reject a null?

Model rejection



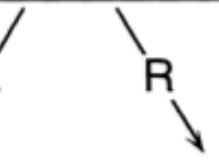
Type I error:
Townspeople
incorrectly
accept a wolf
is there (reject
a true null).

Type II error:
Townspeople incorrectly
deny a wolf is there
(incorrectly accept a
false null).

Type ? error

Equal Base Frequencies

JC69 vs. F81



Transition Rate Equals Transversion Rate

JC69 vs. K80

F81 vs. HKY85



Rates Equal Among Sites

JC69
vs.
JC69+Γ

K80
vs.
K80+Γ

F81
vs.
F81+Γ

HKY85
vs.
HKY85+Γ



Molecular Clock

JC69	JC69+Γ	K80	K80+Γ	F81	F81+Γ	HKY85	HKY85+Γ
vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.
JC69c	JC69+Γc	K80c	K80+Γc	F81c	F81+Γc	HKY85c	HKY85+Γc

Table 2 The results of likelihood ratio tests performed on the albumin DNA data from five vertebrates

Null hypothesis	Models compared	$\log L_0$	$\log L_1$	$-2 \log \Lambda$	d.f.	P
Equal base frequencies	H_0 : JC69 H_1 : F81	-7675.86	-7667.08	17.56	3	2.78×10^{-5}
Transition rate equals transversion rate	H_0 : F81 H_1 : HKY85	-7667.08	-7628.03	78.10	1	9.75×10^{-19}
Equal rates among sites	H_0 : HKY85 H_1 : HKY85+Γ	-7628.03	-7568.56	118.94	1	0
Molecular clock	H_0 : HKY85+Γc H_1 : HKY85+Γ	-7573.81	-7568.56	10.5	3	1.47×10^{-2}

L_0 and L_1 denote the likelihoods under the null (H_0) and alternative (H_1) hypotheses, respectively. P represents the probability of obtaining the observed value of the likelihood ratio test statistic ($-2 \log \Lambda$) if the null hypothesis were true. Because multiple tests are performed, the significance value for rejection of the null hypothesis should be adjusted using a Bonferroni correction (hence, the significance level for rejection of the null hypothesis is set to 1.25×10^{-2}).

Figure 4 The hierarchy of hypotheses examined for the albumin data from five vertebrates. The parameters of the models are explained in Table 1. At each level, the null hypothesis is either accepted, "A," or rejected, "R."

Reality



Reality \neq Model

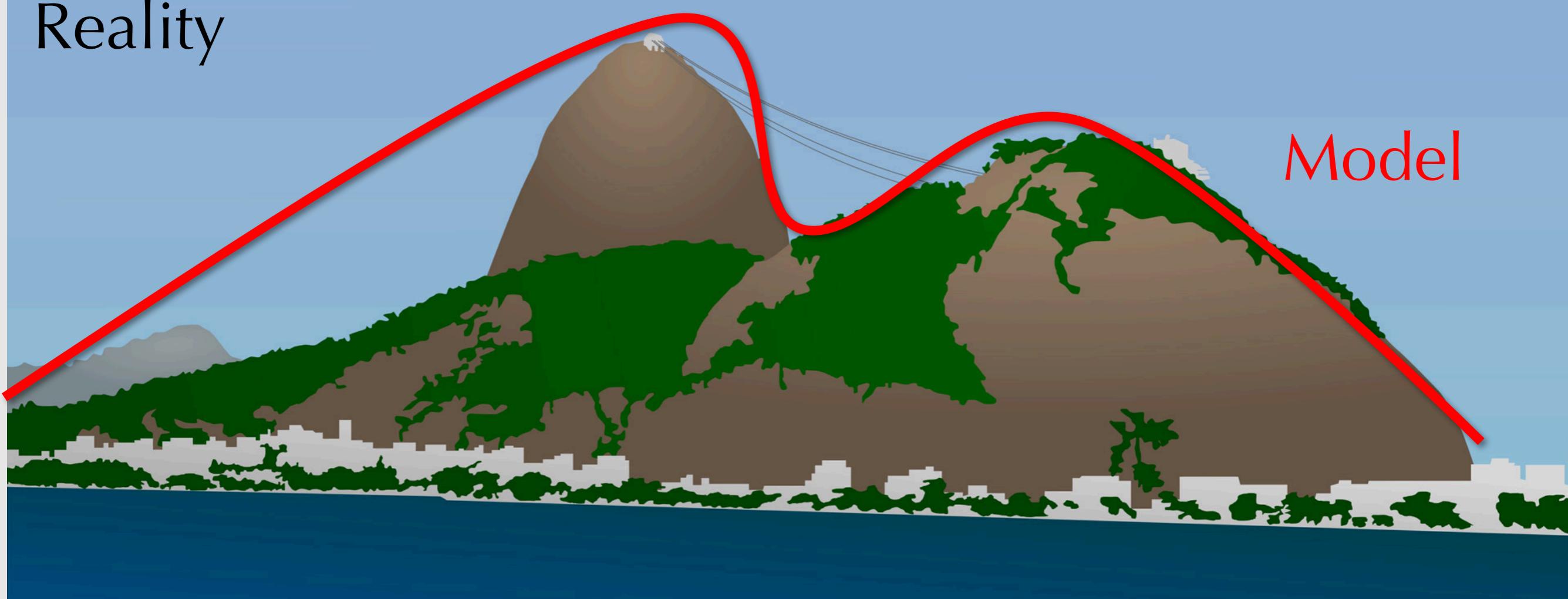
Q: How much information about reality does the model lose?

A1: Kullback-Leibler distance

A2: Akaike Information Criterion estimates KL distance

Reality

Model



A2: Akaike Information Criterion estimates KL distance

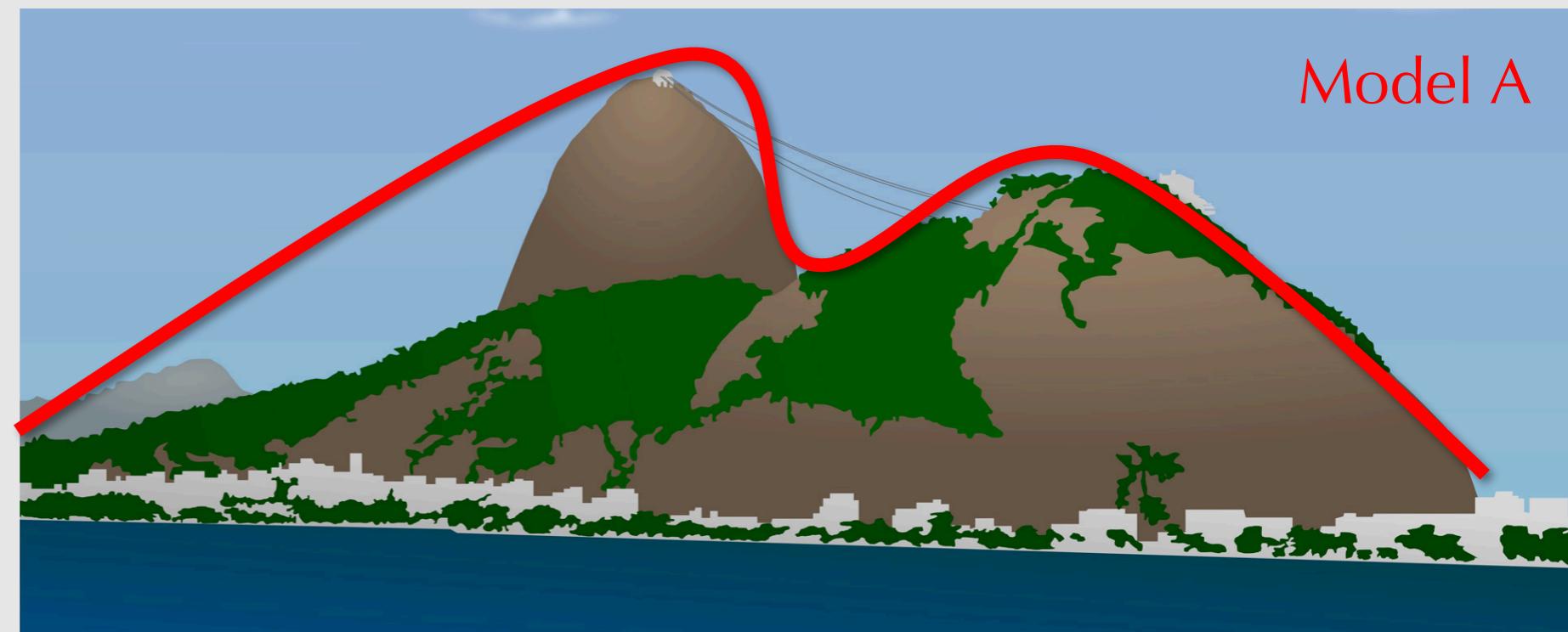
AIC: $-2 \log(\text{likelihood}) + 2 (\# \text{ free parameters})$

Two models

$AIC_A = 402$

$\Delta AIC_A = 5$

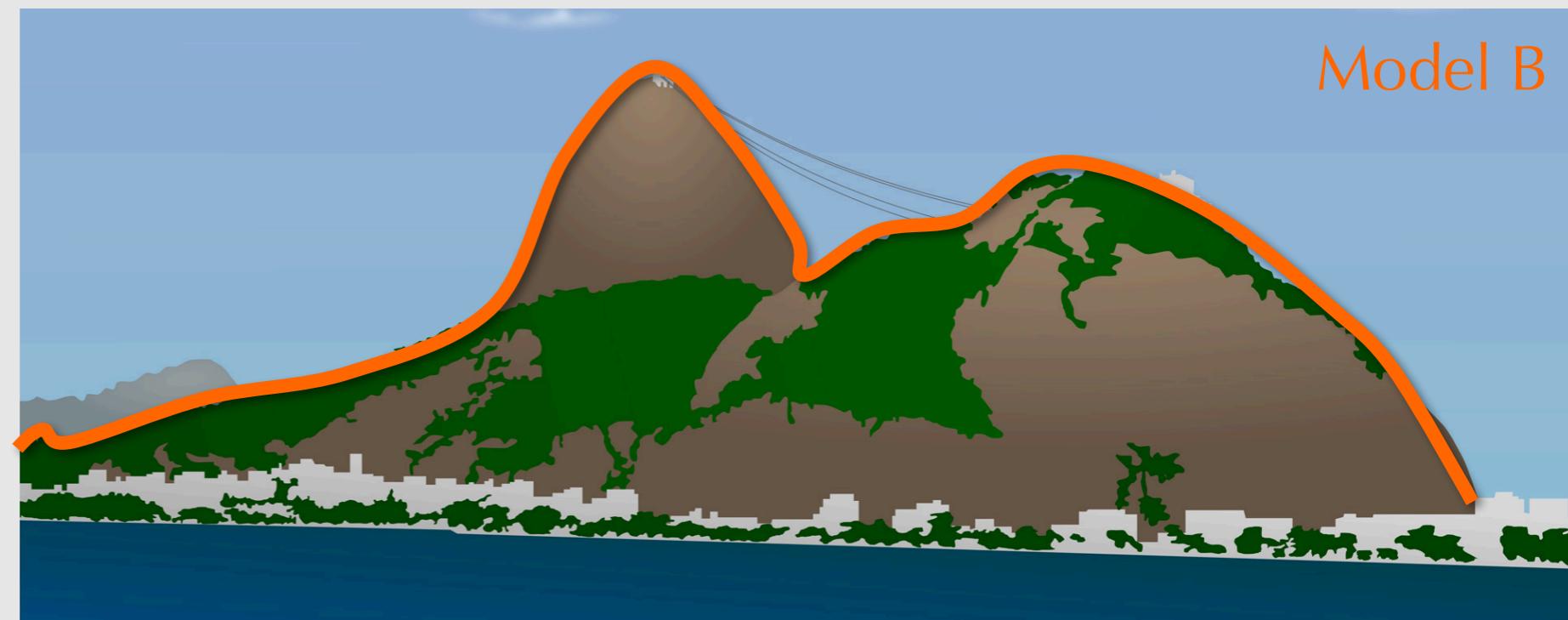
$weight_A = 0.08$



$AIC_B = 397$

$\Delta AIC_B = 0$

$weight_B = 0.92$



Akaike Information Criterion: Want the model that minimizes the information lost. Complex idea, simple to calculate ($2 \times \# \text{parameters} - 2 \ln(L)$). Generally, one subtracts the smallest AIC from a set of models from the values for all models to get ΔAIC (so best, lowest AIC, model has $\Delta \text{AIC} \equiv 0$).

Advantages: doesn't require nested models. Based on information theory. Allows model weighting.

AIC

ΔAIC	Level of empirical support for model
0 – 2	Substantial
4 – 7	Considerably less
10+	Essentially none

Things folk tend to misunderstand with AIC

- It is not a significance measure.
 - No, really. Not even $\Delta 2$. Just stop.
- The best model isn't always the true model.
- The true model doesn't have to be in the set of models.

k = # free parameters

n = number of data points

AIC: $-2 \log(\text{likelihood}) + 2 k$

AICc: $-2 \log(\text{likelihood}) + 2 k * (n / (n - k - 1))$

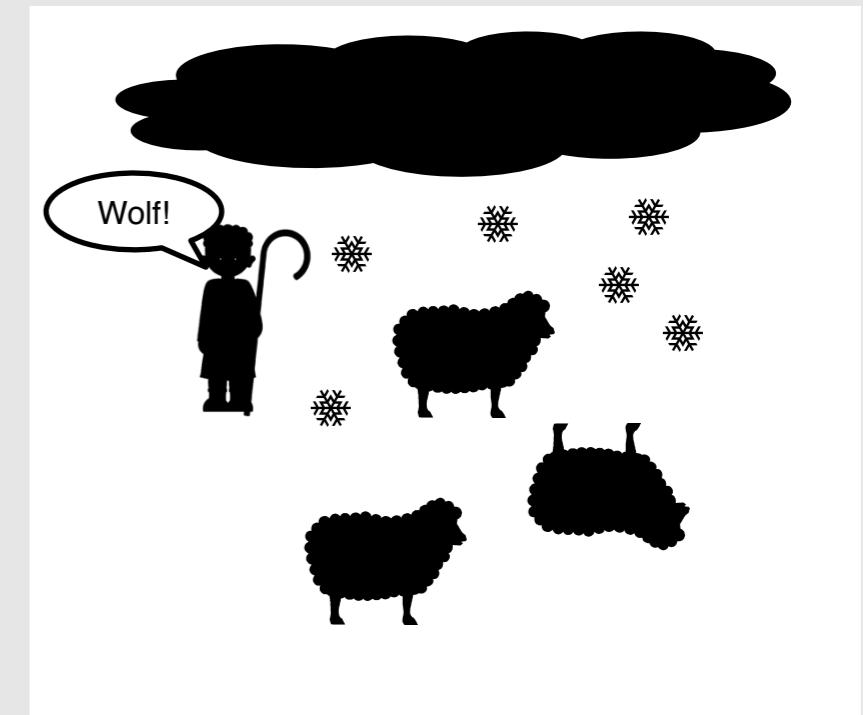
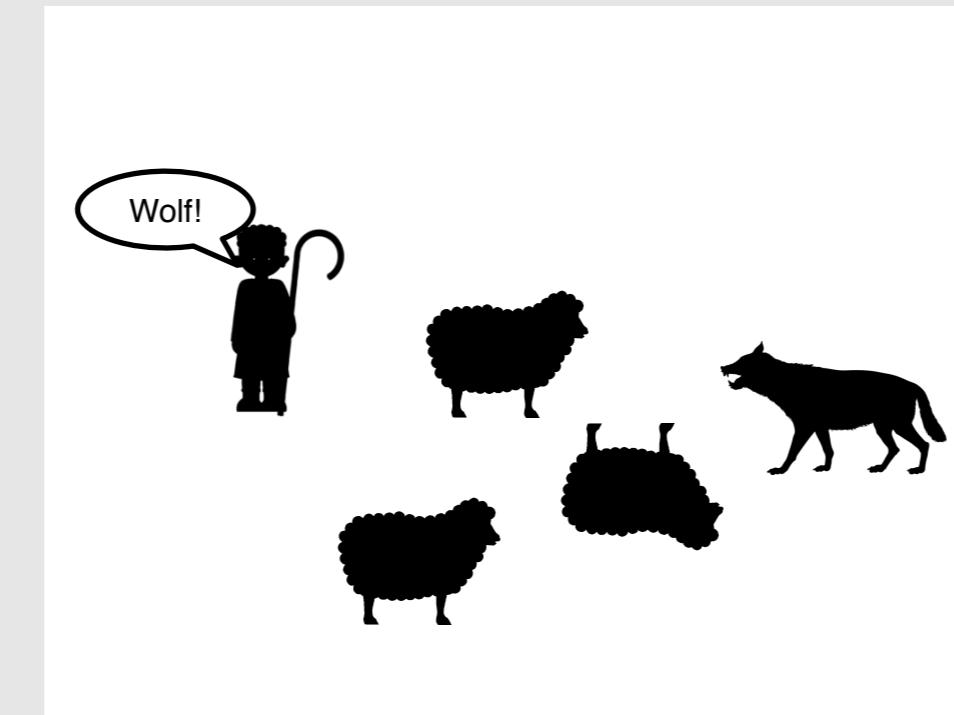
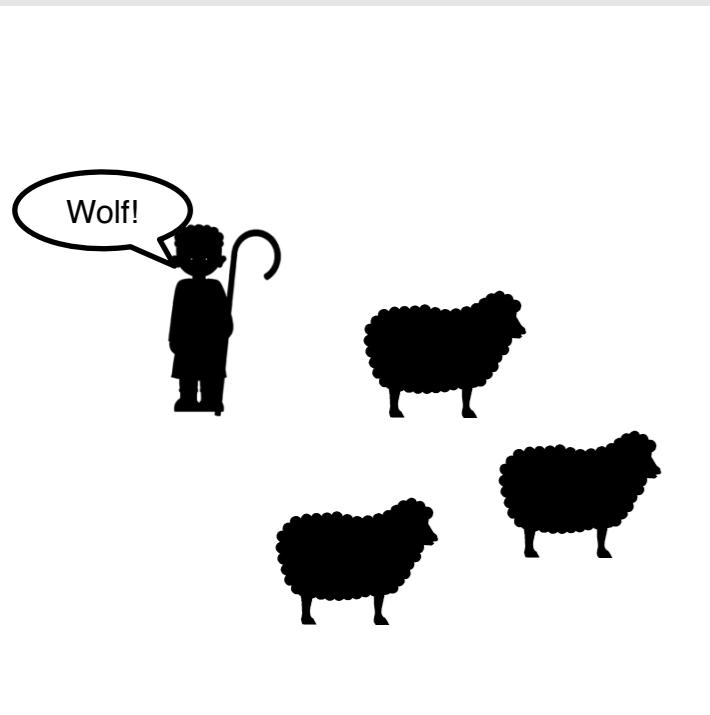
BIC: $-2 \log(\text{likelihood}) + \log(n) * k$

AICc has a correction for small sample size.

BIC is Bayesian Information Criterion (but easiest to think of as another information criterion). Regular AIC tends to overfit.

What is n ?

Model rejection



Type I error:
Townspeople
incorrectly
accept a wolf
is there (reject
a true null).

Type II error:
Townspeople incorrectly
deny a wolf is there
(incorrectly accept a
false null).

Type ? error

Parameter estimation

Hypothesis 1

When carotenoids are simply strictly limiting color signal elaboration, we expect a BM model **with similar rates of change (σ^2)** between lineages with and without carotenoid-dependent traits to be favored. Although commonly interpreted as a “pure drift” model, a BM model may also reflect the pattern of lineages evolving toward phylogenetically structured, lineage-specific adaptive optima, which are randomly structured in relation to the considered factors (Revell et al. 2008)....

Hypothesis 2

When carotenoids are a limiting resource, we expect a single optimum OU model to be favored. Selection to maintain or increase plasma carotenoid level is independent of the presence of carotenoid-dependent ornaments, because carotenoids provide physiological resource benefits irrespective of the presence of sexual signaling. Thus we expect strong attraction toward this optimum, **with possibly higher attraction (or lower evolutionary rates)** in lineages that exhibit carotenoid-dependent traits, because sexual selection within these linages is predicted to enhance selective pressures, thereby maintaining carotenoid levels closer to the shared optimal value.

Hypothesis 3

When carotenoids are not a limiting resource and are mostly used as a pigment, we expect an OU model with **two optima—one lower optimum with low attraction (weak stabilizing selection) in lineages lacking carotenoid-dependent traits and a higher optimum with stronger attraction in lineages that do exhibit carotenoid-dependent traits**—to be preferred. If carotenoids are not a physiological resource, natural selection should not drive the evolution of plasma carotenoid levels to be high in the absence of carotenoid-dependent sexual signals. In contrast, in lineages with carotenoid-dependent traits, sexual selection is expected to select for higher carotenoid levels (higher optima), which will mostly be used for signal pigmentation.

Parameter estimation

Model	Preferred, %	AICc	ΔAICc	wAICc
Plumage carotenoid regime models:				
BMS	0	211.89 (191.43–235.73)	84.7 (61.74–110.25)	0
OUM	73.20	127.58 (112.15–141.82)	.39 (0–3.60)	.38 (.09–.54)
OUMA	15.08	128.40 (113.18–142.32)	1.20 (0–2.52)	.25 (.12–.44)
OUMV	5.64	128.61 (113.23–142.35)	1.42 (0–4.01)	.22 (.08–.37)
OUMVA	5.64	129.72 (114.27–143.51)	2.52 (0–4.10)	.15 (.07–.52)

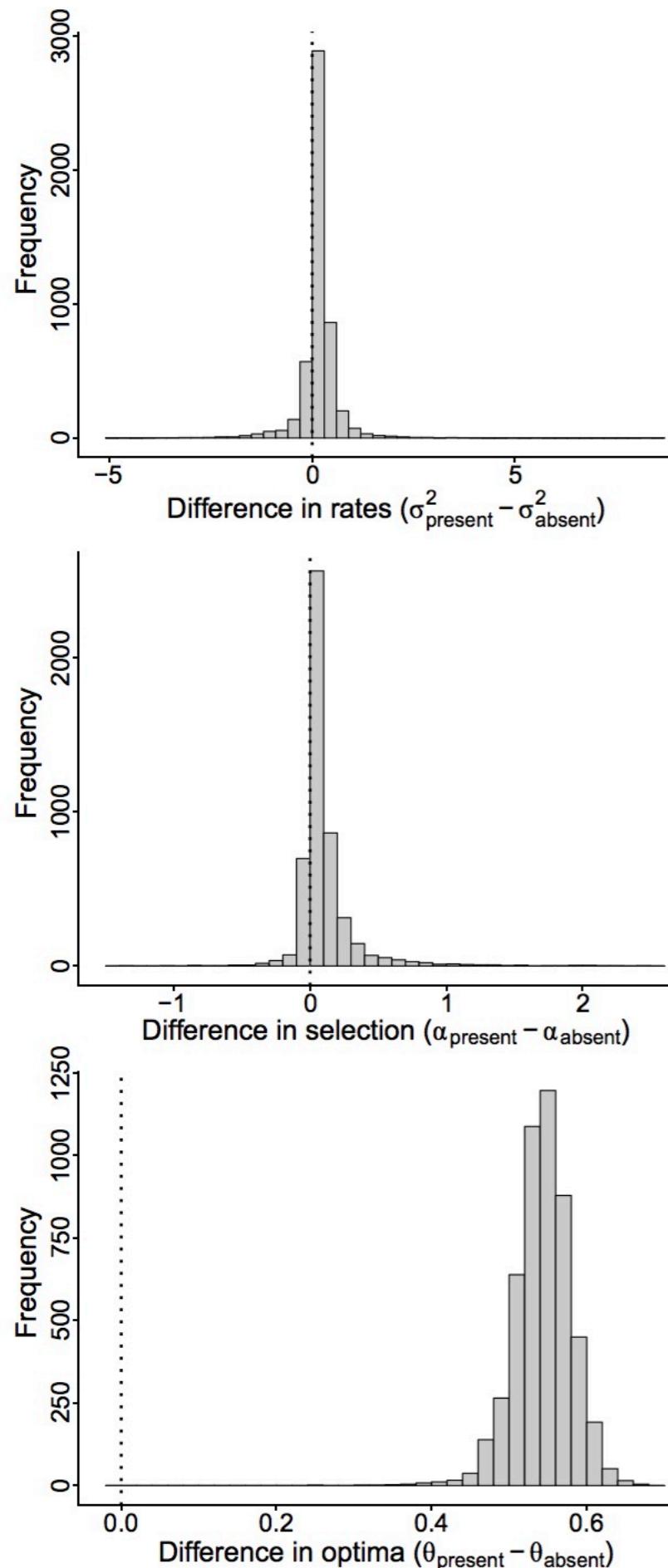


Figure 3: Histogram of the difference in the model-averaged parameters (σ^2 , α , and θ) for the evolution of circulating carotenoid levels in relation to selective regimes determined by the presence or absence of carotenoid-dependent plumage coloration across the posterior sample of the 5,000 trees used. The vertical dotted line indicates no difference in parameters (at zero). Optima of carotenoid levels are higher for lineages that exhibit carotenoid-dependent plumage (the tail-area probability of $\theta_{\text{present}} - \theta_{\text{absent}}$ relative to zero $<.001$), but no differences were observed between the attractions toward the optima (α) and rates (σ^2).

