Some comments about detecting adaptive evolution using the OU model

Clay Cressler

Queens University,
Department of Biology

Aaron King

University of Michigan,
Ecology & Evolutionary Biology

Marguerite Butler

University of Hawaii, Department of Zoology

(*Ornstein-Uhlenbeck for Comparative Hypotheses)

Rapid Development of OU models in comparative analysis

* Hansen 1997 - First use of OU models for comparative hypotheses - fit an adaptive model to the evolution of hypsodonty in horses

* Butler & King 2004; Scales, King & Butler 2009; King & Butler 2009 - Multiple optimum OU model with discrete & punctate shifts among them to represent multiple 'biologically-inspired' hypotheses

* Hansen et al. 2008; Labra et al. 2009 - Introduces SLOUCH, a hybrid BM-OU model that allows a form of evolutionary regression between phenotype and covariates

* King and Butler 2009; Bartoszek et al. 2012 - Multivariate extensions to OUCH and SLOUCH

* Beaulieu et al. 2012 - OUwie introduces the ability to have both multiple optimum OU models as well as varying alpha and sigma across the tree

OU machinery used for different goals

Evaluation of Biologically-Informed Hypotheses

* Model selection here is an aid to reasoning about the importance of evolutionary mechanisms and drivers (not an end in itself)

* Relies heavily on the construction of a small number of well-reasoned, distinct hypotheses

* The larger the number of models considered, the less reliable the inferences obtained

(Burnham and Anderson 2002, Butler & King 2004, O'Meara et al. 2006, Revell et al. 2011, Bartoszek et al, 2012, Beaulieu et al. 2012)

Reconstruction of Ancestral States

* Infers the ancestral states an/or historical shifts in selective regimes based on the phylogeny and traits of extant organisms

* Avoids the a priori limitation of the range of possible histories

* Makes use of an exploration of a large number of alternative histories to find those most consistent with the data

(Hipp & Escudero 2010, Eastman et al. 2011, Ingram & Mahler 2013)

Some pesky lingering problems

Assumptions about the root state (X_0) in the OU model impact whether BM and OU are nested.

Fit a value for the ancestral state $X_0 = X(0)$.

- Early implementations introduced a parameter for the ancestral state $X_0 = X(0)$ in the OU model, as in the BM model.

- Results in BM being a special case of OU as $\alpha \to 0$.

- Problems: The data often have little information about the root state, estimates of X_0 are often inconsistent, and X_0 and α are often correlated (Ane 2008). (and we can get erratic results).

Assume that $X_0 =$ is distributed according to the stationary distribution of the OU process.

- $X_0 \sim$ normal ($\theta(0)$, $\sigma^2/2\alpha$) (Ho and Ane 2013, Hansen et al 2008)

- The model fit is better behaved

- BM is no longer nested within OU

Some pesky lingering problems

Scaling of Parameters and Time dependence makes it difficult to compare results among studies

Dimensionless Parameter Combinations:

* dimensionless time
* dimensionless Trait value
* dimensionless optimum values
* *η = Selection opportunity — aT depth of tree relative to characteristic timescale of adaptation*
* *ϕ = Discriminability Ratio — sqrt(2α)Δθ/σ : the smallest difference between selective optima relative to the standard deviation of the stationary trait distribution*
* *sqrt(η) ϕ = Signal-to-Noise Ratio (SNR)*

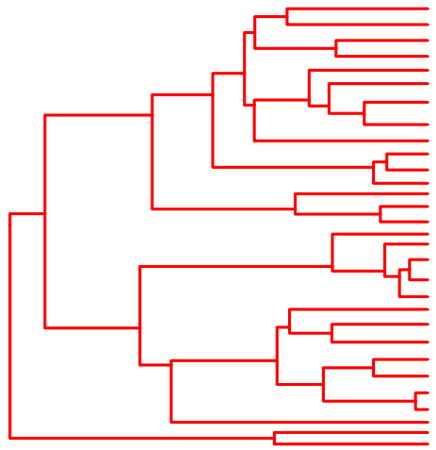
How well do OU models work for testing comparative hypotheses?

Cressler, King, and Butler, 2015

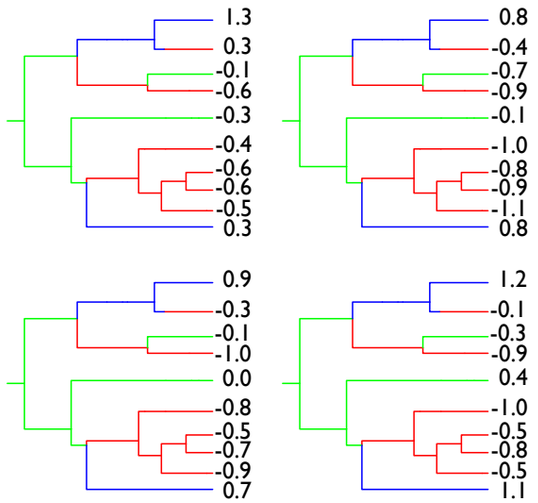
Massive Simulation Study

- 220 combinations of η, ϕ, tree size
- 40 sets of regime paintings per combo
- 40 trees of each size per regime painting
- 40 sets of simulated phenotypic data per tree
- 14.08 x 10^6 simulated datasets fit to each of 6 models:
- BM, OU1, OU2a,b, OU3 (true model), NP

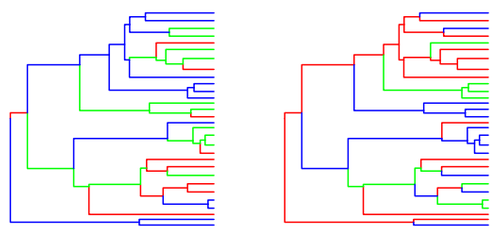
1. Choose values for tree size, selection strength, and noise intensity.



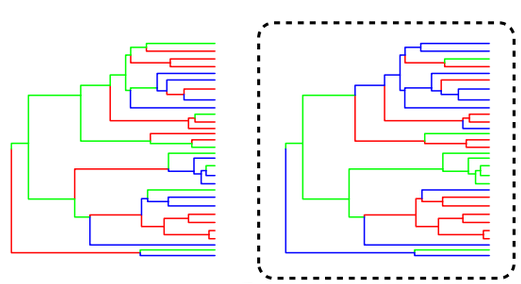
2. For each parameter set, generate 40 regime paintings of the full tree.



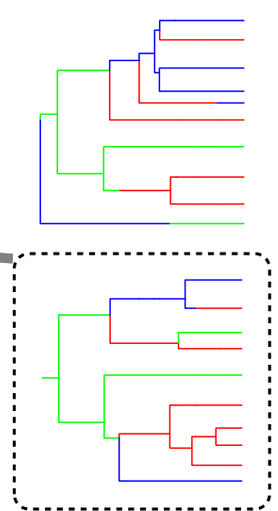
5. Fit the models to each phenotypic dataset and record the AICc and parameter estimates.



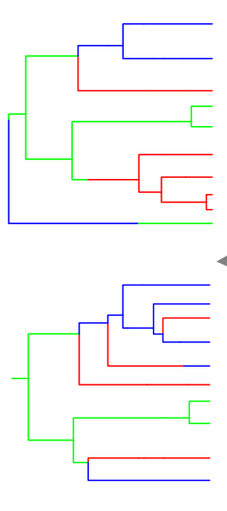
6. Calculate power, precision, and accuracy across the 40 phenotypic datasets.



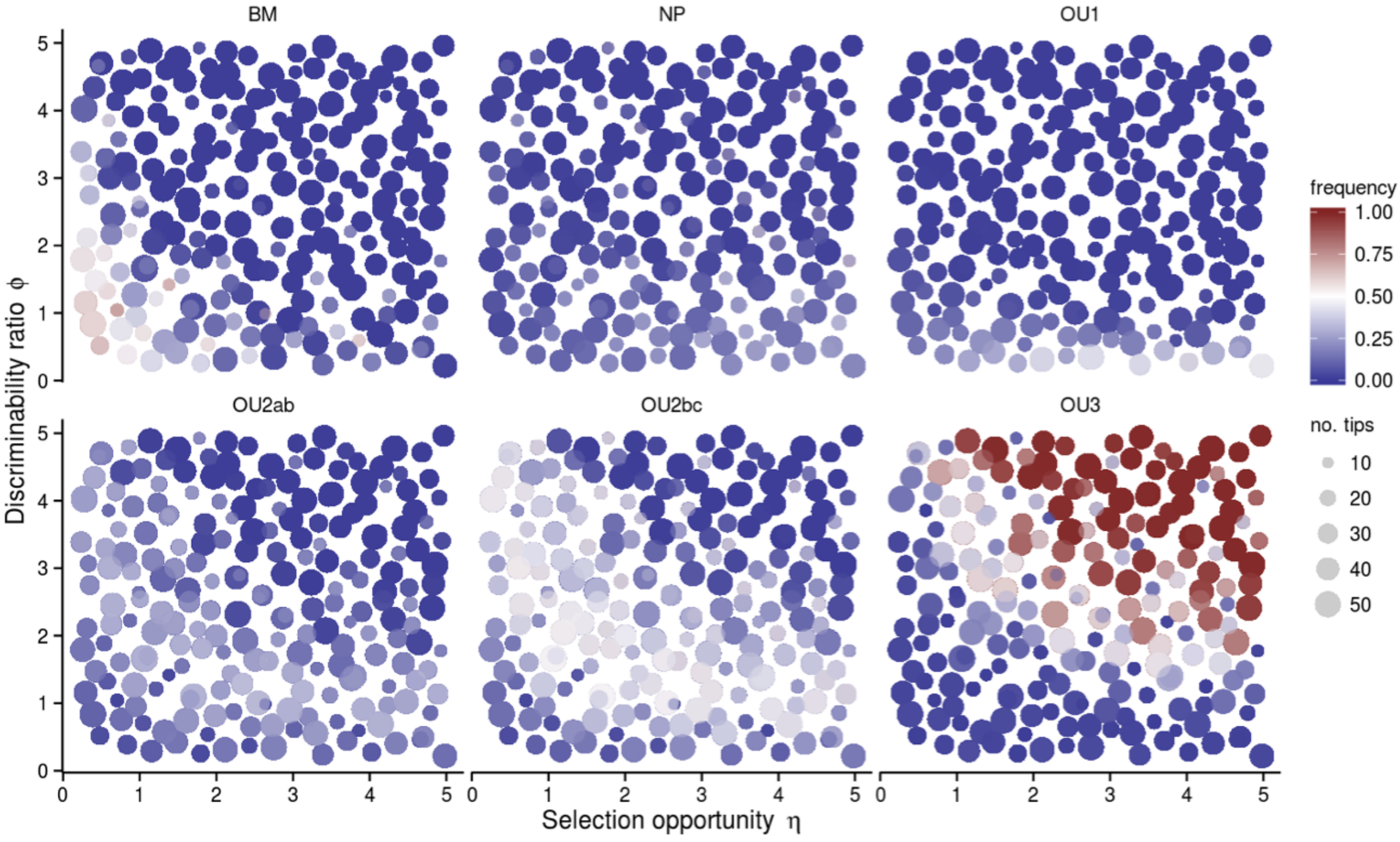
4. For each subtree, generate 40 phenotypic datasets with the specified selection strength and noise intensity.



3. For each regime painting, choose 40 subtrees of the specified size and calculate tree and regime metrics.

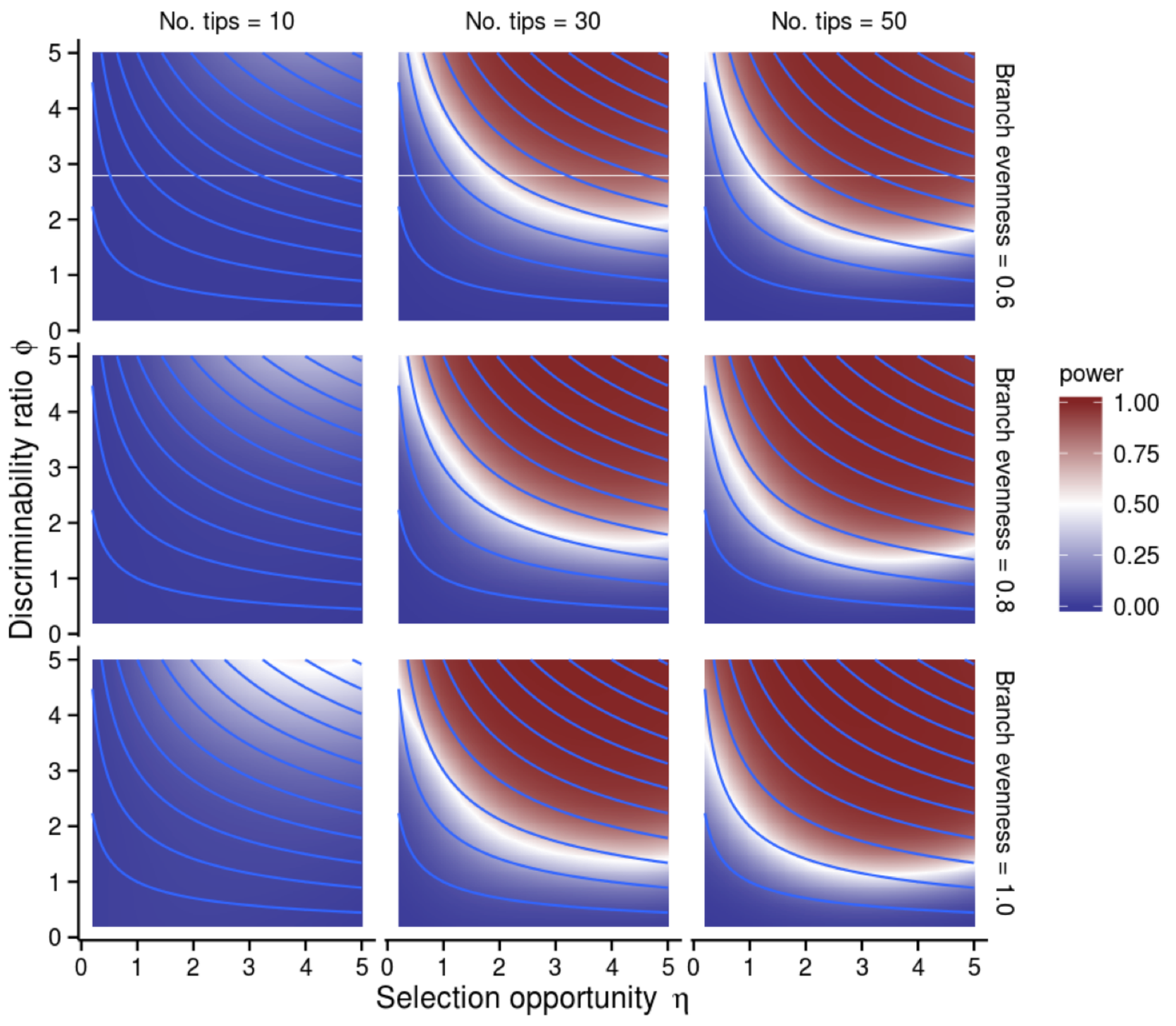


Model support varies with selection opportunity and discriminability ratio



Overall, model selection works pretty well!

Power as a function of Discriminability Ratio, Selection Opportunity, tree size, branch evenness



Parameters are difficult to estimate

theta

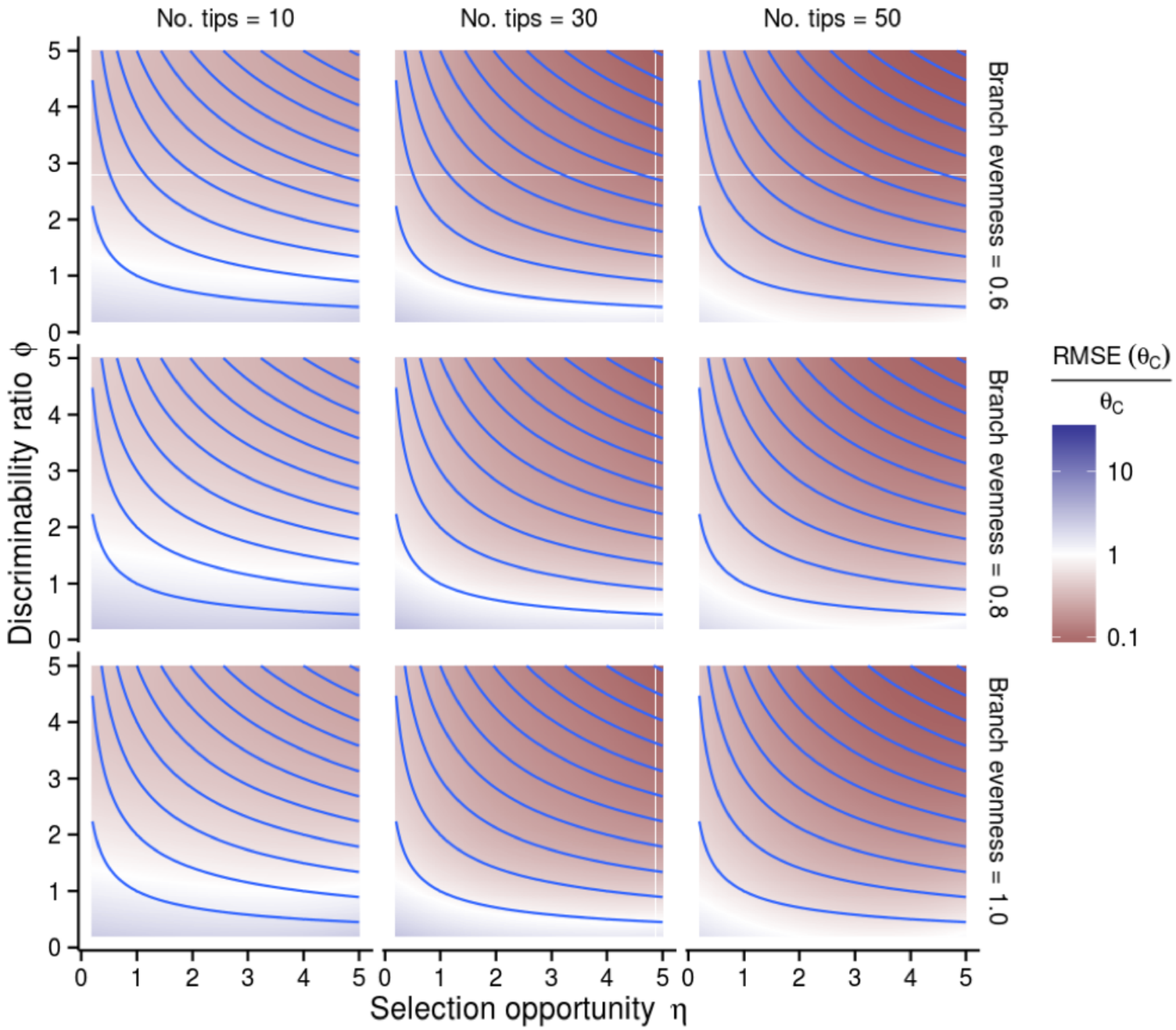


FIGURE 4. The relative root mean square error (RMSE) in the estimate of the selective optimum θ_C as η and ϕ are varied for a range of tree size and branch evenness values. Red hues correspond to less than 100% relative error; blue, to greater than 100%. Contours show the levels of SNR. Predictions are based on the best-fitting GAM regression.

A cautionary tale

The Good:

- When selection opportunity and discriminability ratio are large enough, adaptation can be reliably detected
- When discriminability ratio is large enough, power can be high even when selection opportunity is small. In this case, tree size and SNR are good predictors of model power.

The Bad:

- Some parameters are really hard to estimate. Discriminability ratio, in general, is better estimated that selection opportunity.
- In some regions of parameter space, model selection can be quite reliable, yet parameter estimates are far off.

The Ugly:

- There is no way of knowing for sure whether you're in the "low performance" region of parameter space
- However, we explored an extremely wide range of parameter space. Most comparative biologists studying adaptation would not be likely to pose hypotheses for datasets in which the optima are difficult to distinguish. (how would they even know there were multiple optima there?)
- Data mining using OU tools may not be advisable. (at least be very conservative!)

A
B
C
Discriminability ratio ϕ
Selection opportunity η
6.52
6.82
6.04
5.89
BM vs. OU3
NP vs. OU3
OU1 vs. OU3
OU2ab vs. OU3
OU2bc vs. OU3
Density
AIC$_c$ difference

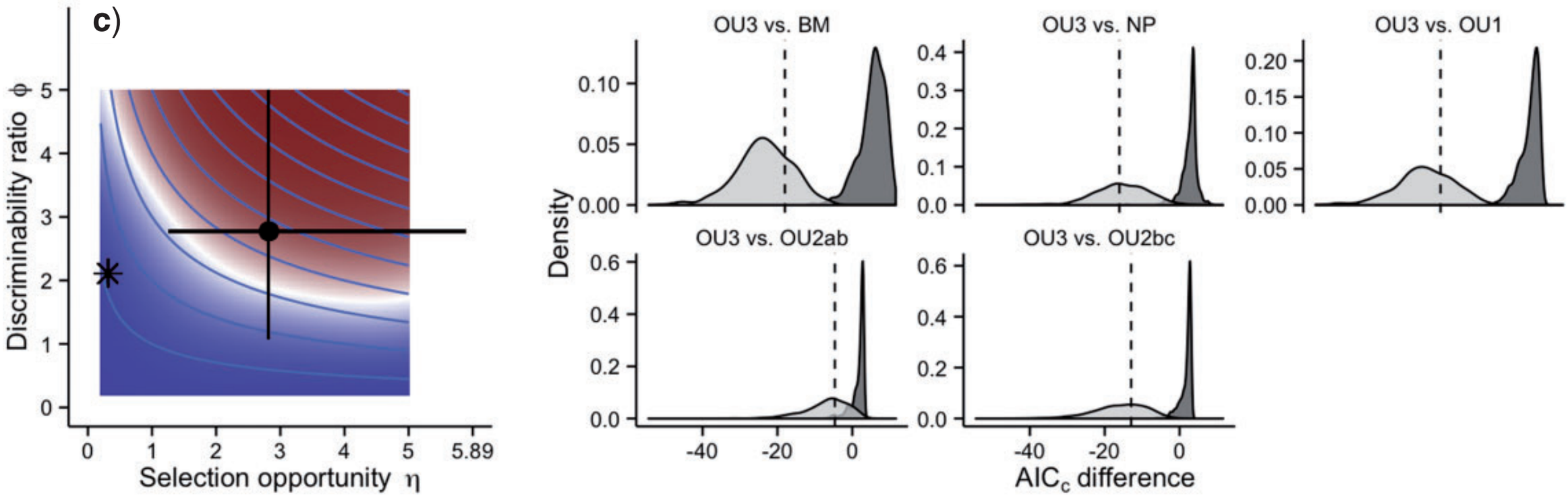


FIGURE 8. Three scenarios illustrating various parameter estimation and model selection outcomes, as described in the text. Scenario a is drawn from a region where power is high and parameter estimates reliable. Scenarios b and c are drawn from regions where power is low and parameter estimation unreliable. In Scenario b, both the true parameters and the estimates lie in regions of low power. In Scenario c, however, the overestimation of η and ϕ tempts one to the specious conclusion that the true parameters are in a region of high power and accuracy. The plots at left indicate true parameter values (asterisks) and ML estimates (filled circles). The crosshairs shows 95% bootstrap confidence intervals. The plots at right show the phylogenetic Monte Carlo results for pairwise comparison of alternative models based on 1000 bootstrap replicates. The density plots show the distribution of AIC_c differences between the indicated models (OU3-simpler model). In each case, the light gray density shows the distribution of AIC_c differences when the data-generating process is OU3 at its MLE; the dark gray differences result when the data-generating process is the simpler model at its MLE. The vertical dashed line shows the observed AIC_c difference on the original data. Comparing the observed AIC_c difference to the dark gray density gives an indication of the statistical significance of the observed AIC_c difference. Lack of overlap between the light gray and dark gray densities indicates high power of the hypothesis test.