Species as data points

Outline for today

- The problem with species data
- Phylogenetic signal in ecological traits
- Why phylogeny matters in comparative study
- Phylogenetically independent contrasts (PICs)
- A linear model approach
- A method for discrete data (and issues)

An example of species data

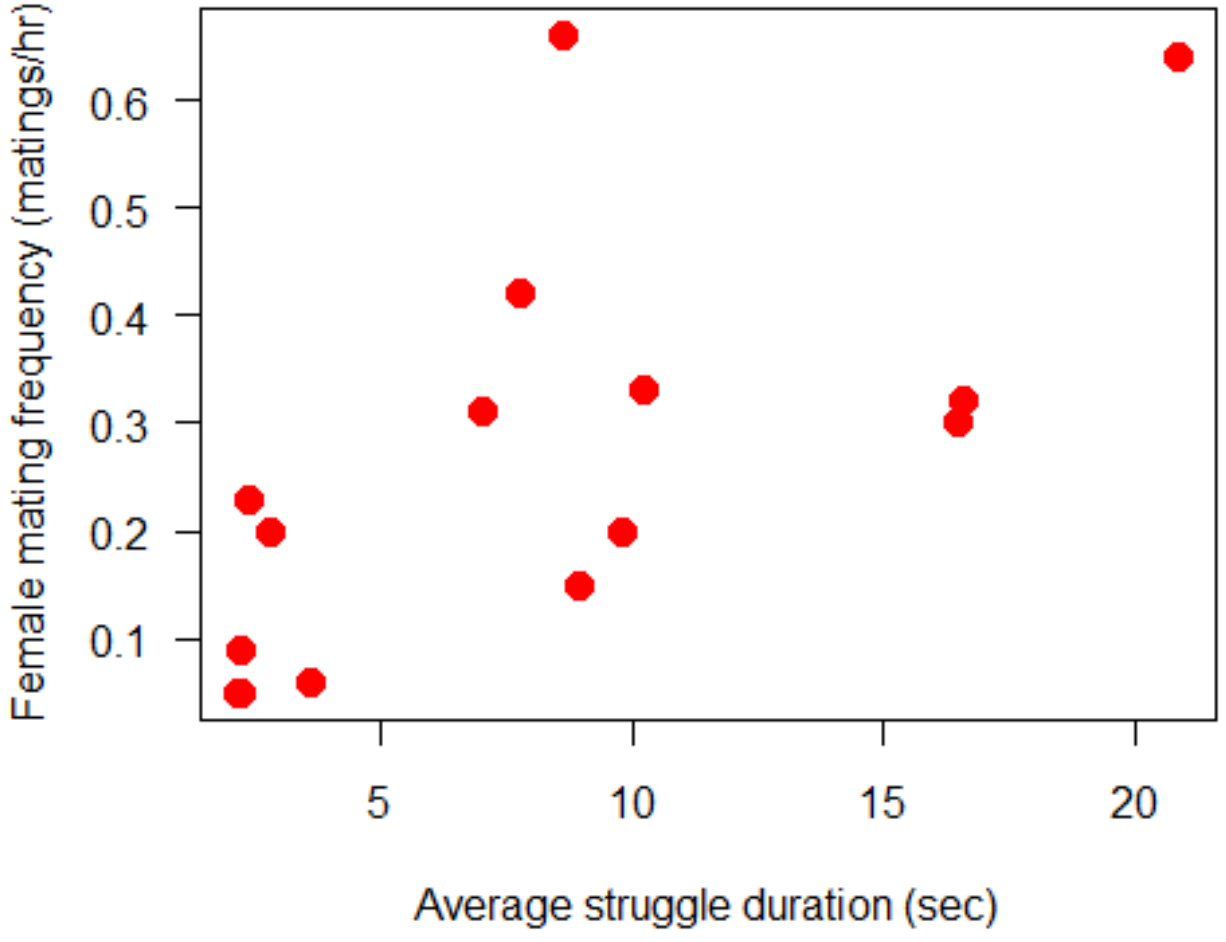
Mating behaviors in 15 species of water striders (*Gerris*) (Rowe and Arnqvist 2002).

Biology: Males chase females, who flee by skating away. If a male grasps a female, she initiates a series of leaps, rolls, and summersaults that usually toss him off. Males of some species have clasping genitalia that allow them to stay on longer, but females of these species often have spines or other devices that make it difficult for males to grasp her. Mating takes place after a female stops struggling.

Variables to correlate are average duration of female struggles for each species, which are the periods of evasive action by females in response to lunges or grasps by males; and average mating frequency of females, measured under controlled lab conditions.

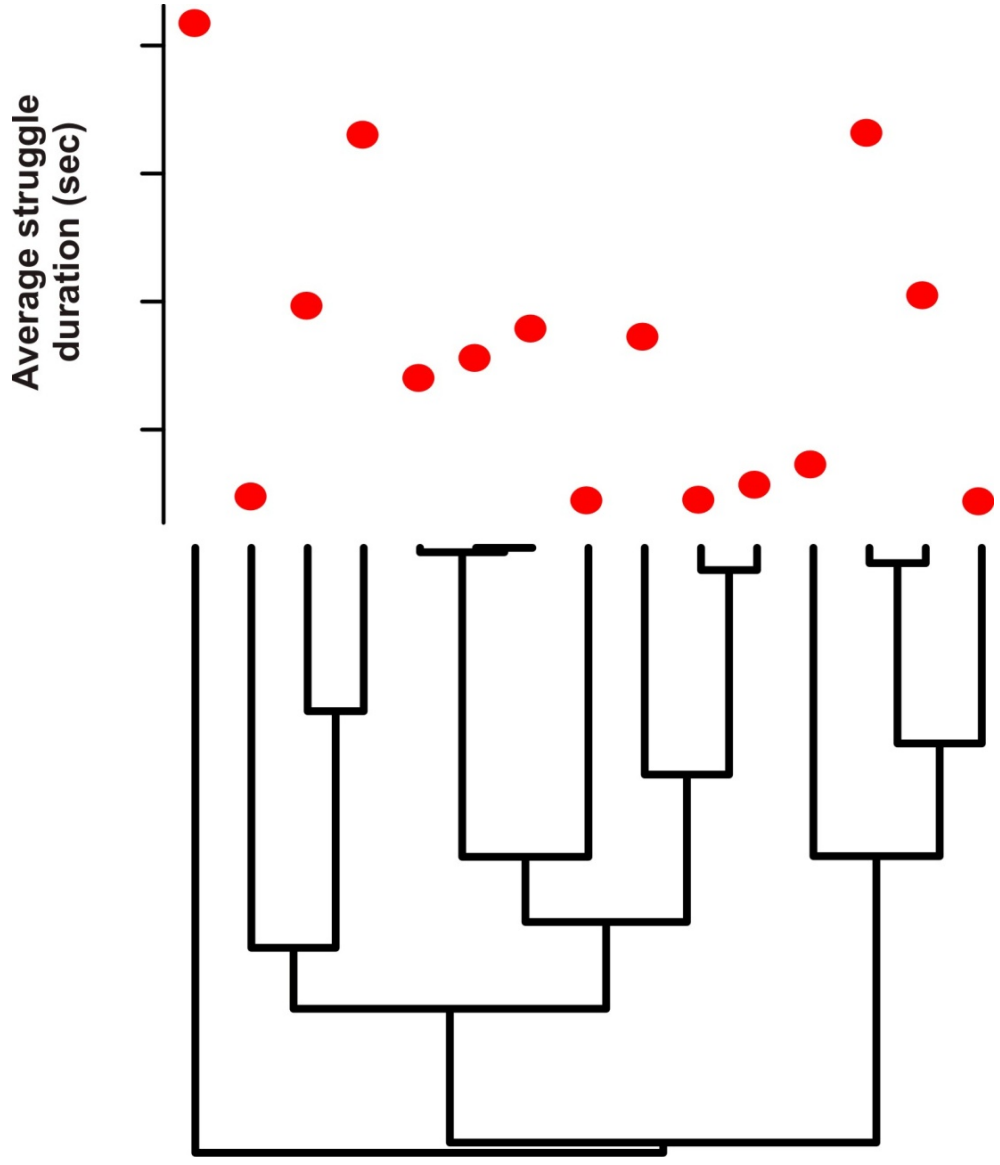
An example of species data

Data reveal a positive association between the two variables.
We would like to estimate the strength of the correlation.



The problem with species data

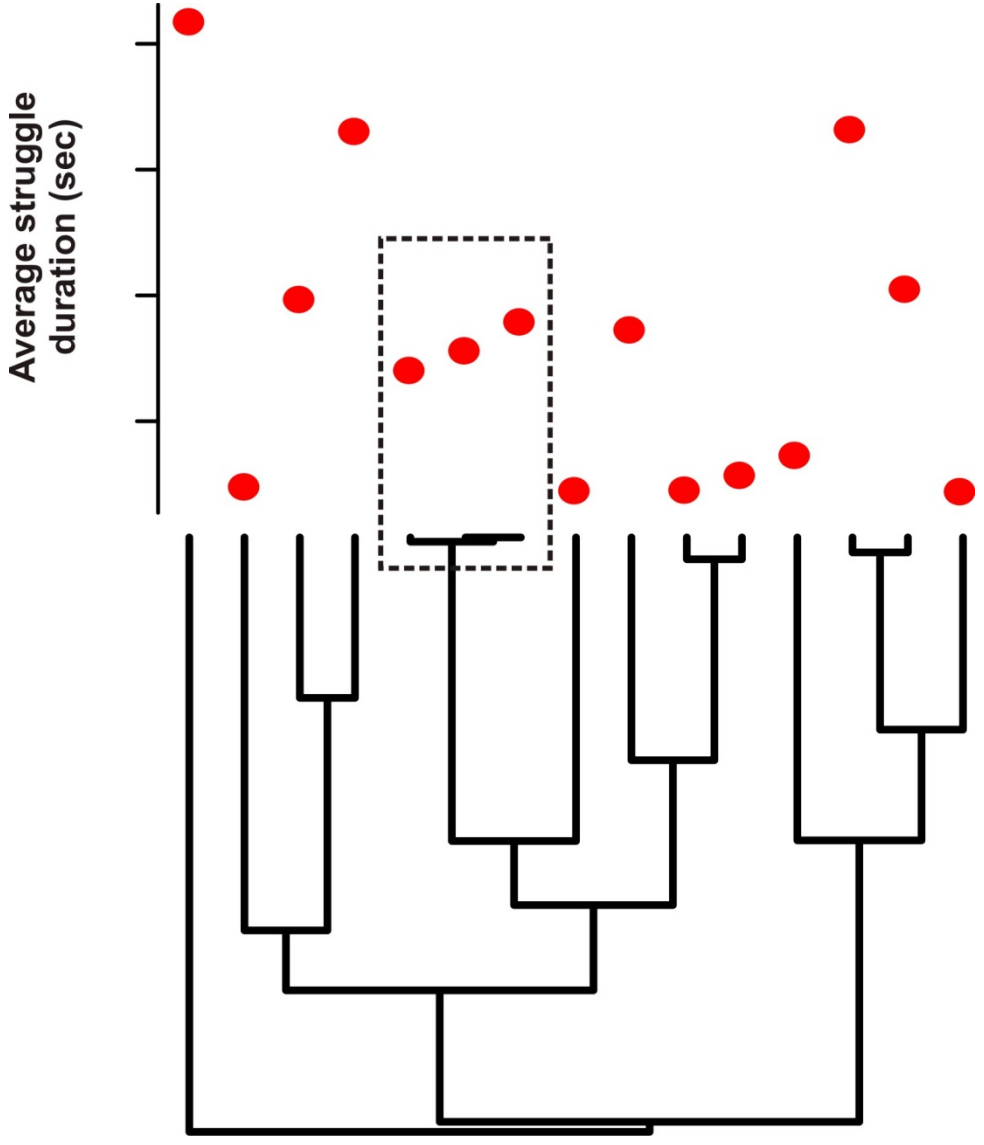
The data points (species) are not independent.



Phylogeny of *Gerris*

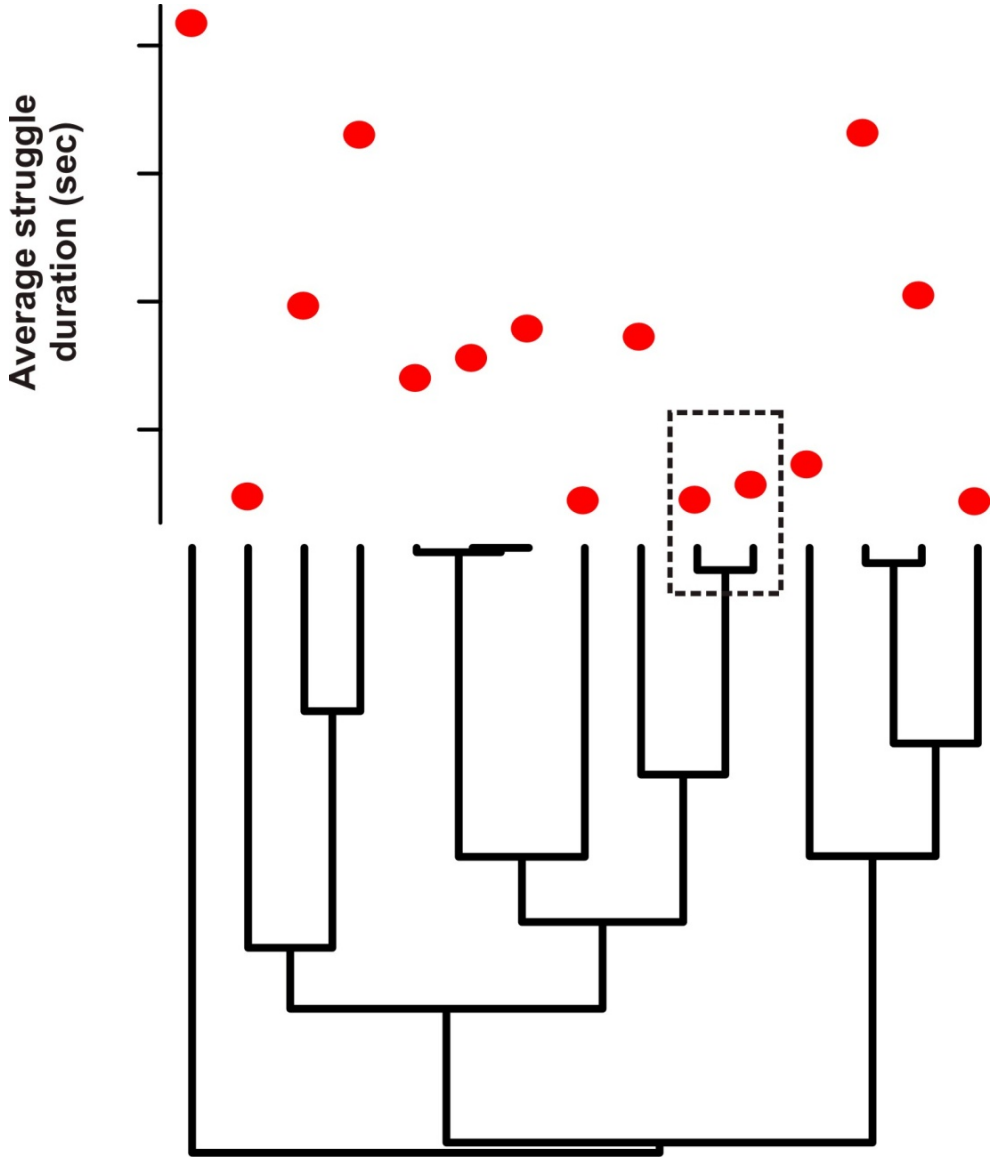
The problem with species data

Closely related species tend to have similar trait values.



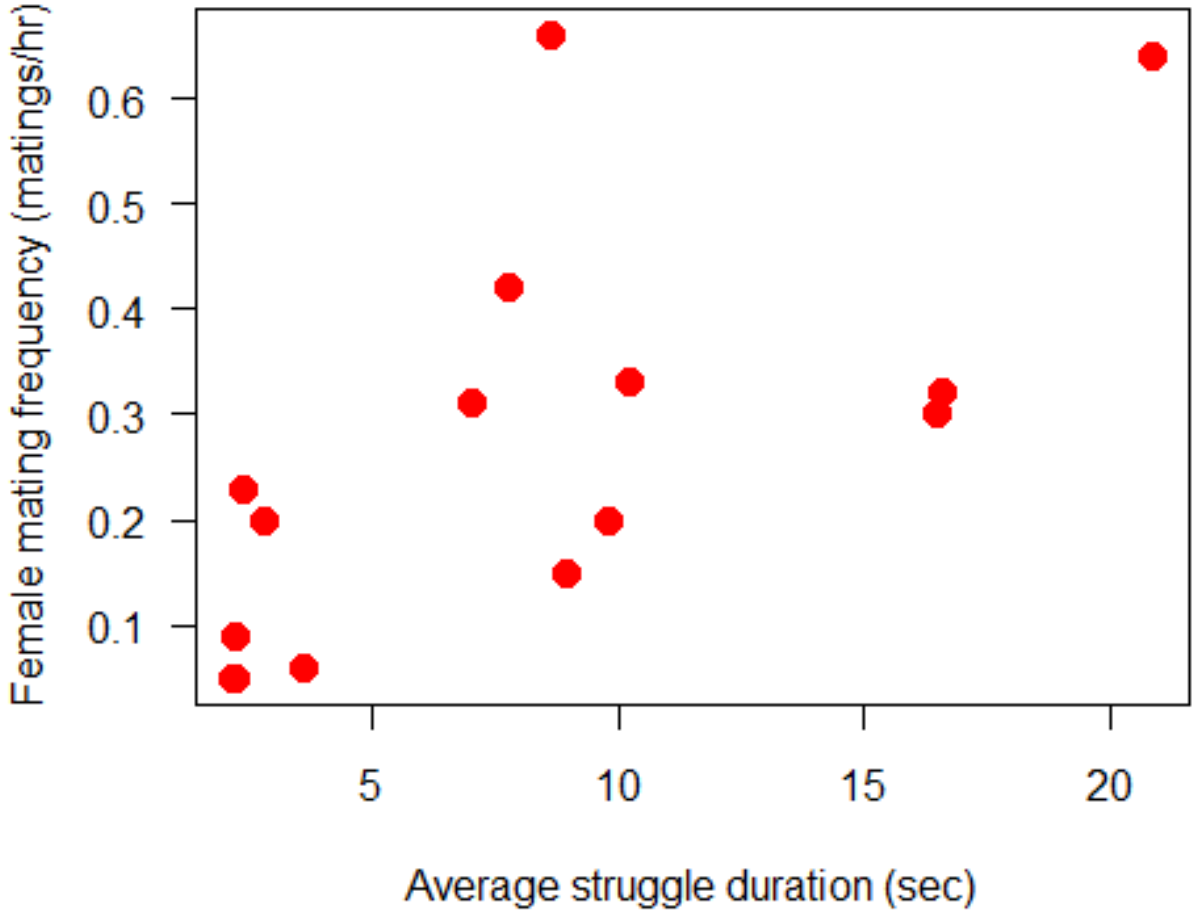
The problem with species data

This tendency is called “phylogenetic signal”.



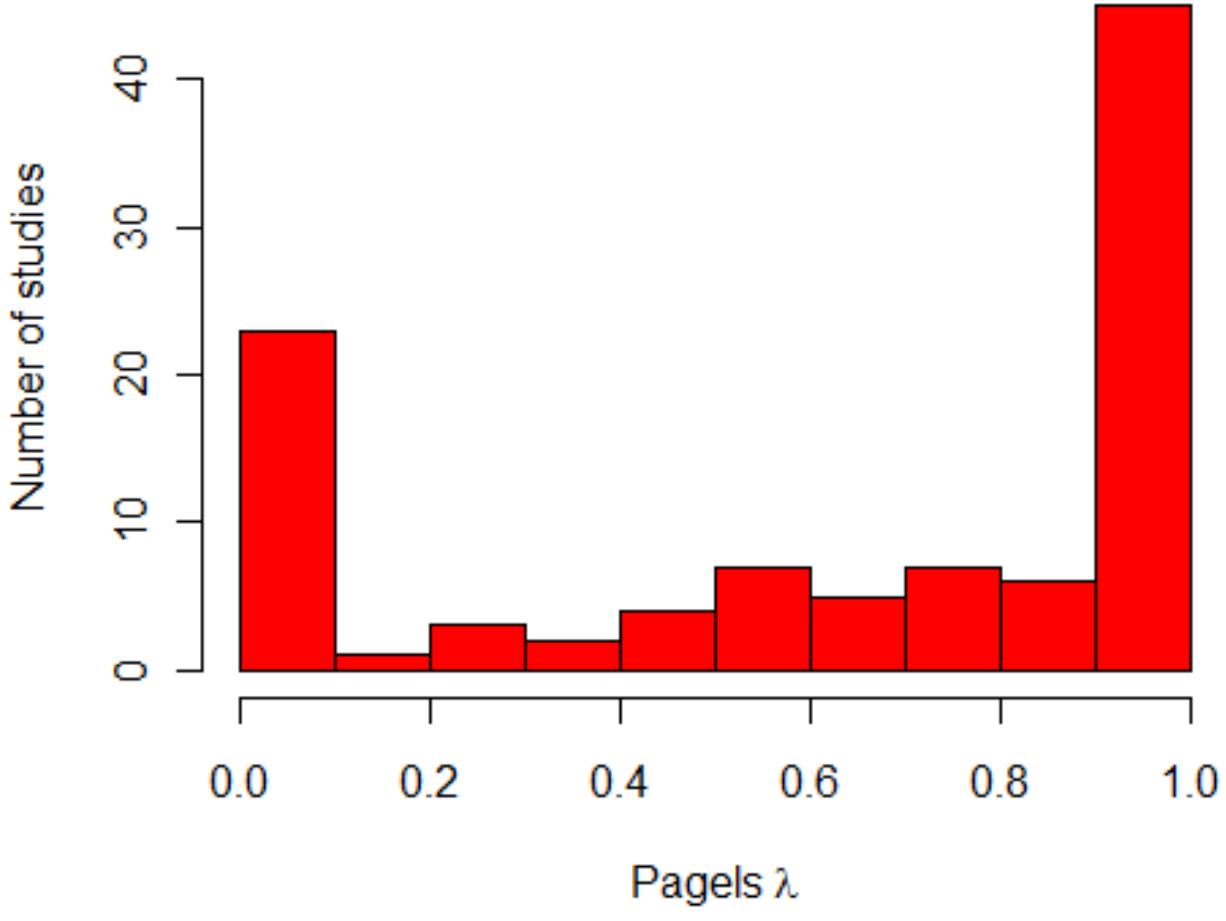
The problem with species data

Non-independence of the species data points violates a major assumption of conventional statistical methods for data analysis.



How prevalent is phylogenetic signal in ecologically relevant traits?

Pagel's λ measures the extent to which closely related species are similar in their trait values (phylogenetic signal). Here is a survey of λ-values from many studies and traits by Freckleton et al (2002)**:**

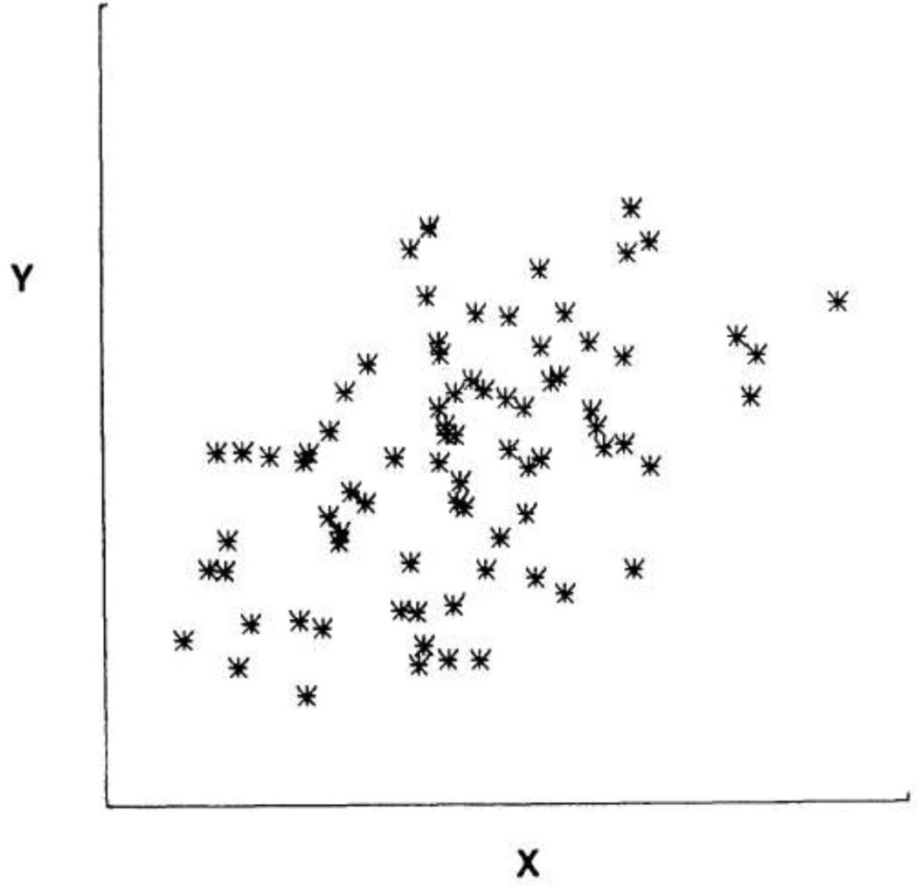


Why is phylogenetic signal a problem?

Non-independence leads to wrong calculations of precision (standard errors, confidence intervals). It leads to wrong Type 1 error rates in null hypothesis significance testing.

Example scenario:
Data on two traits
for 40 species

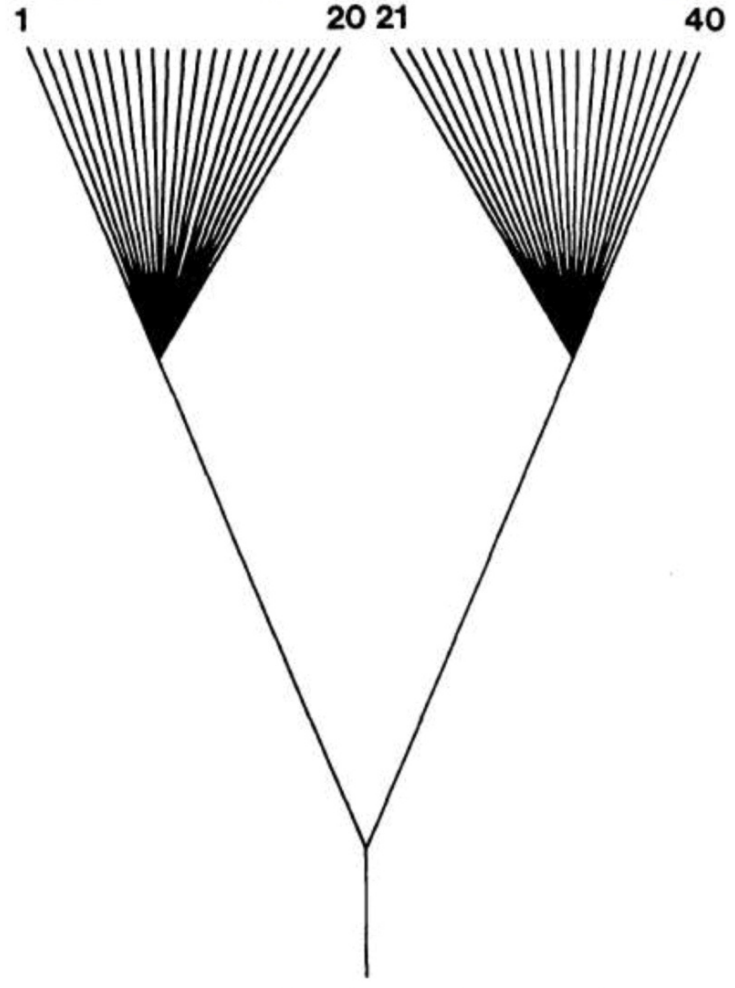
Looks like a strong
correlation between
variables Y and X



Felsenstein (1985) *Am Nat*

Why is phylogenetic signal a problem?

Felsenstein's "worst case scenario" for the phylogeny of the 40 species.

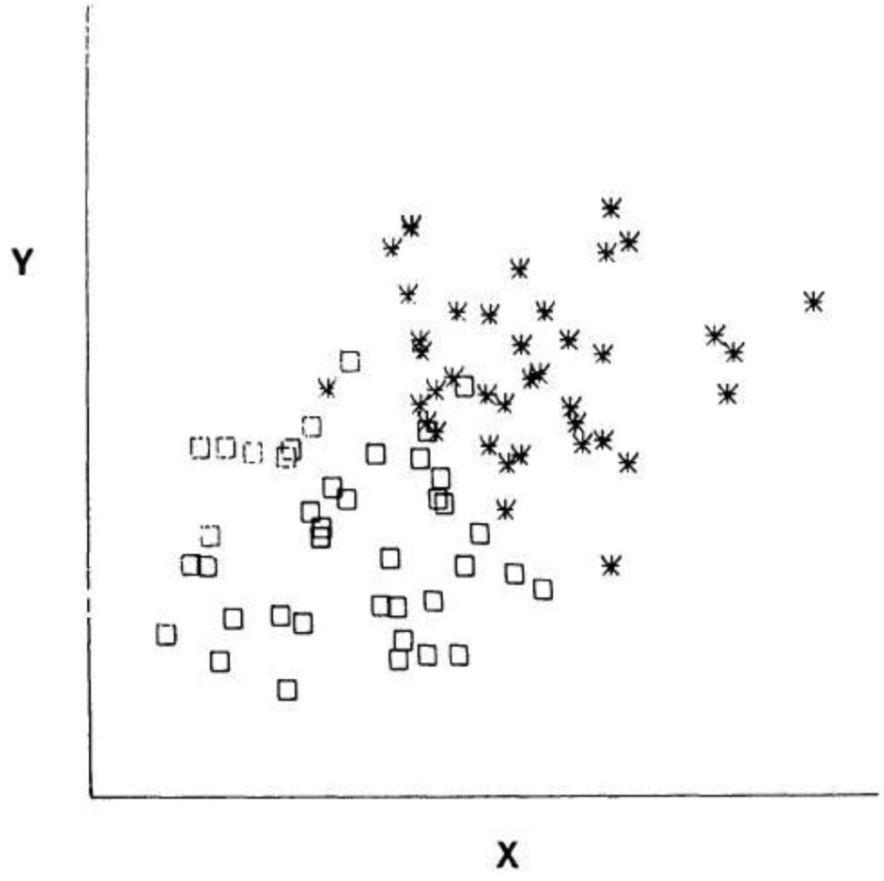


Felsenstein (1985) *Am Nat*

FIG. 5.—A "worst case" phylogeny for 40 species, in which there prove to be 2 groups each of 20 close relatives.

Why is phylogenetic signal a problem?

In this case the non-independence is severe, and creates an apparent association between X and Y where there is none.

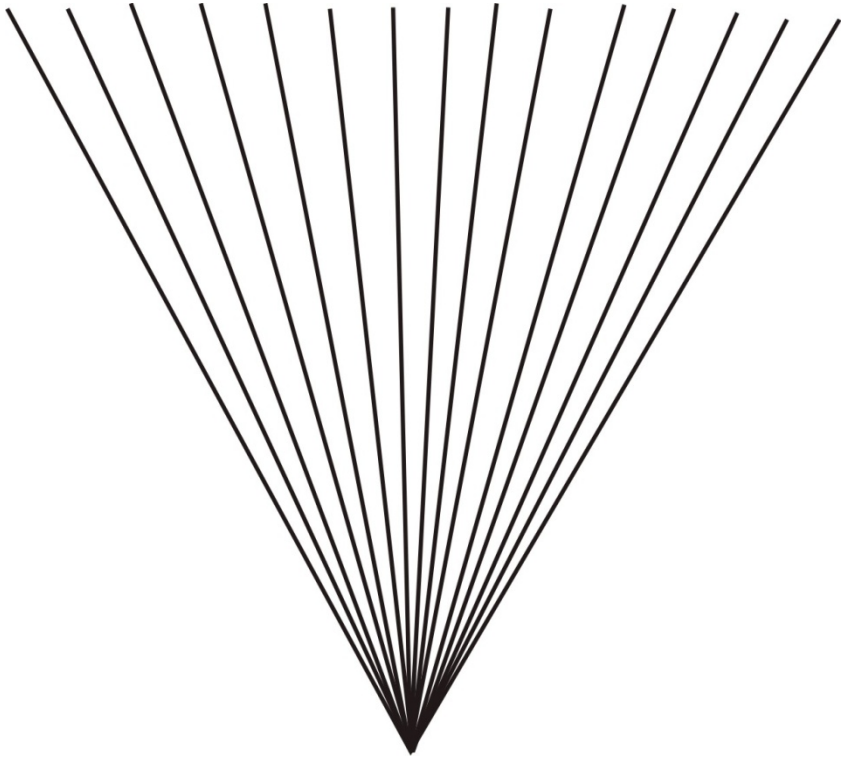


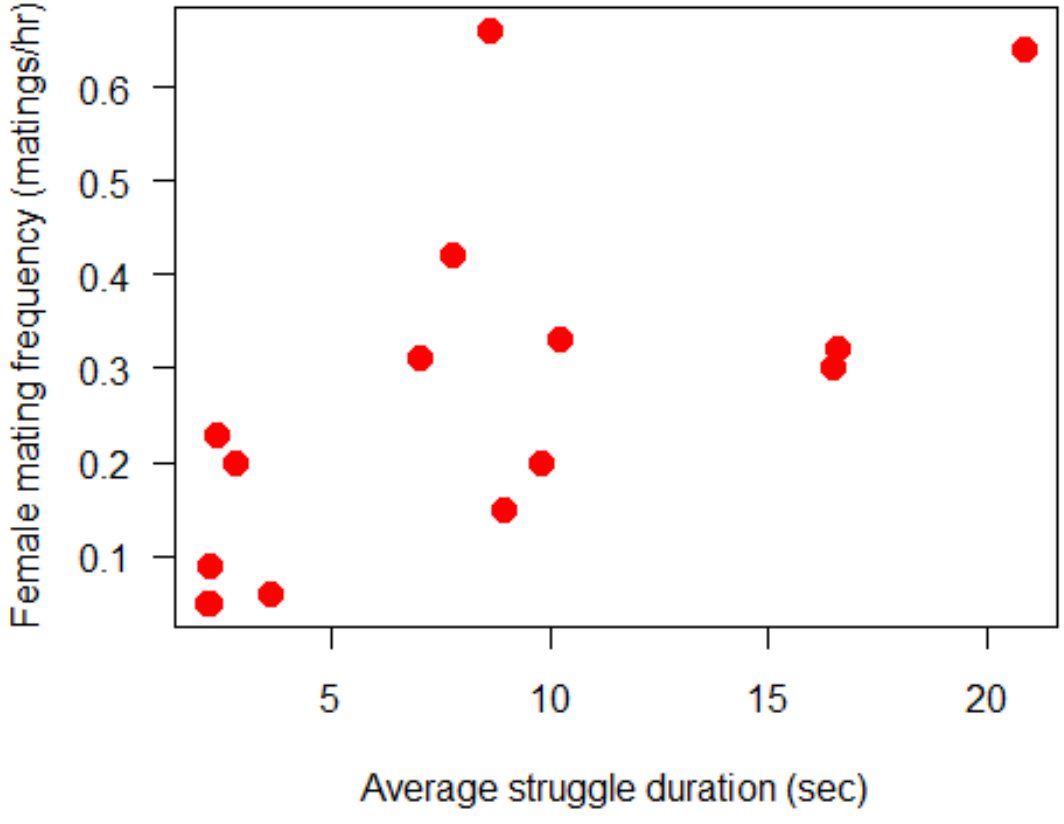
Felsenstein (1985) *Am Nat*

FIG. 7.—The same data set, with the points distinguished to show the members of the 2 monophyletic taxa. It can immediately be seen that the apparently significant relationship of fig. 6 is illusory.

What we are really assuming when we ignore phylogeny

That the species are related as in a “star” phylogeny, which leads to no phylogenetic signal.

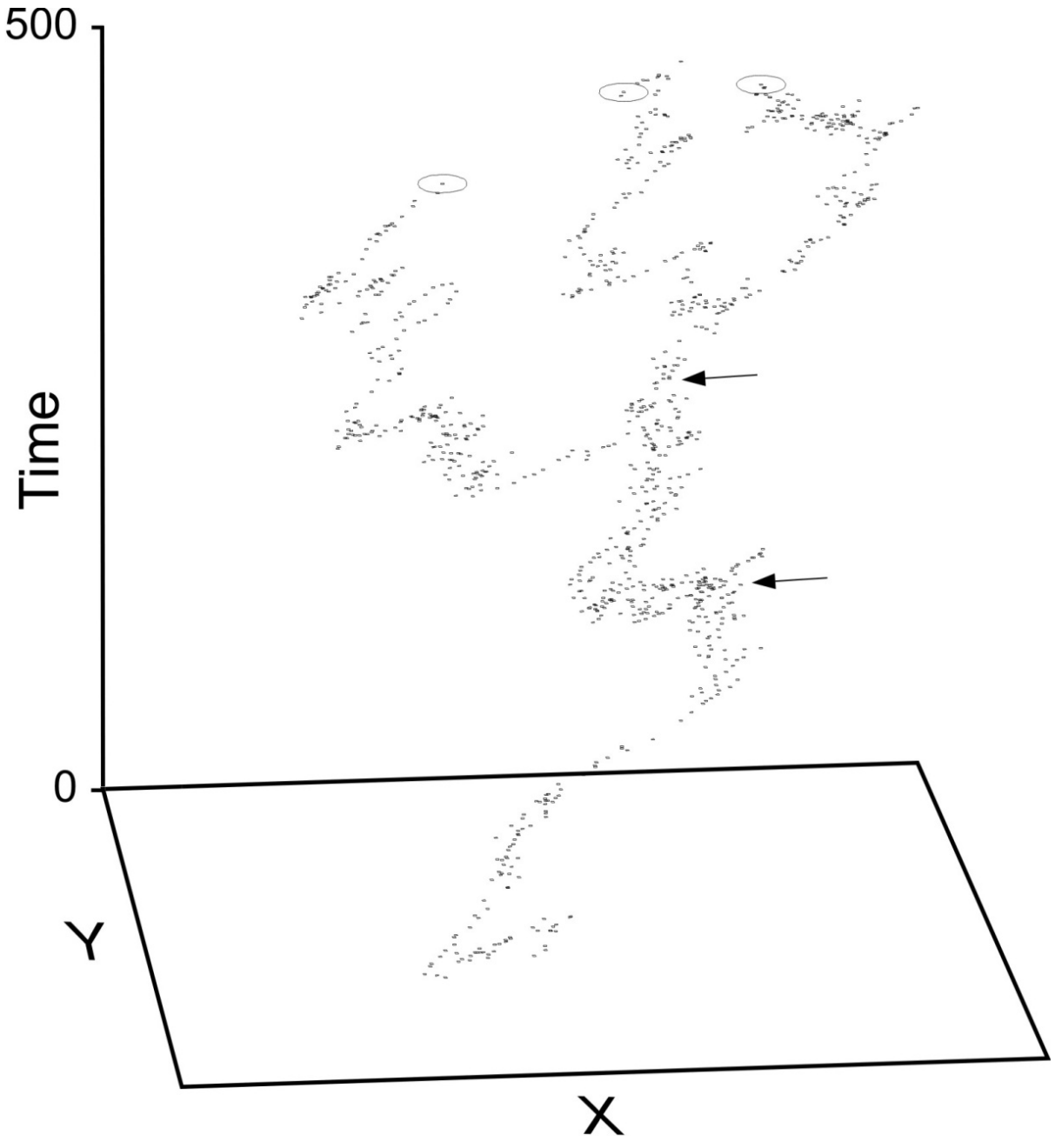




Felsenstein's (1985) solution

Method assumes that the evolution of traits is mimicked by a continuous random walk (Brownian motion).

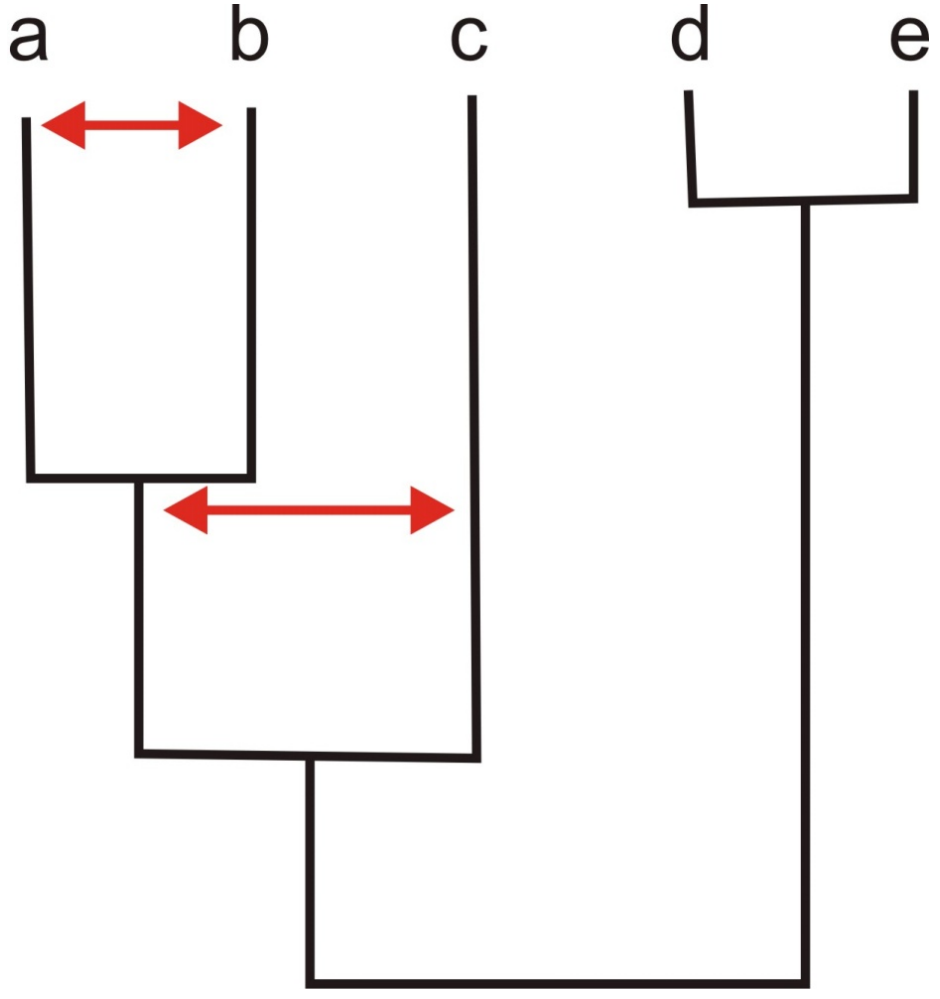
Under Brownian motion, the difference between any two species in a trait has a normal probability distribution with mean 0 and variance proportional to the time since their common ancestor.



Felsenstein (1985) *Am Nat*

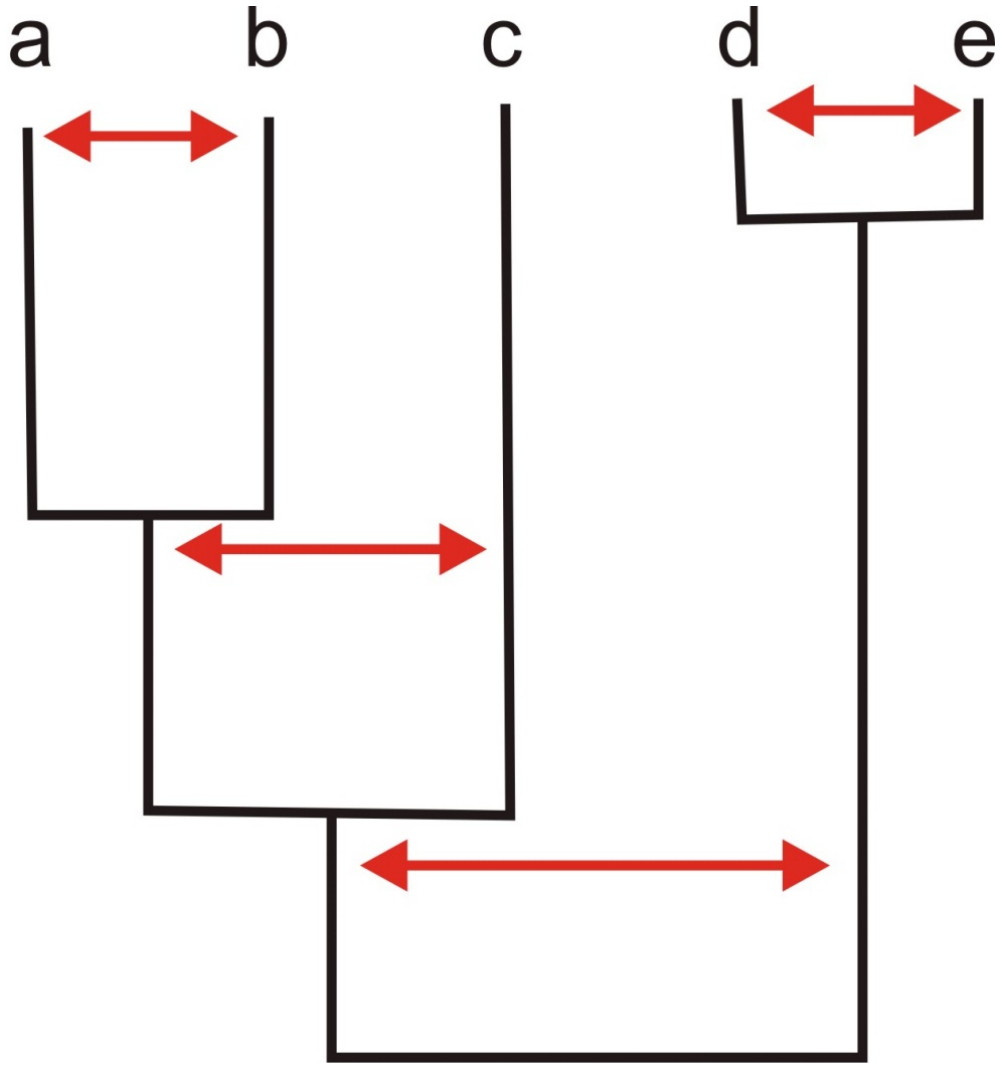
Felsenstein's method of phylogenetically independent contrasts

Under Brownian motion, a, b, and c are not independent, but the difference ("contrast") between a and b is independent of the difference between c and (a+b)/2.



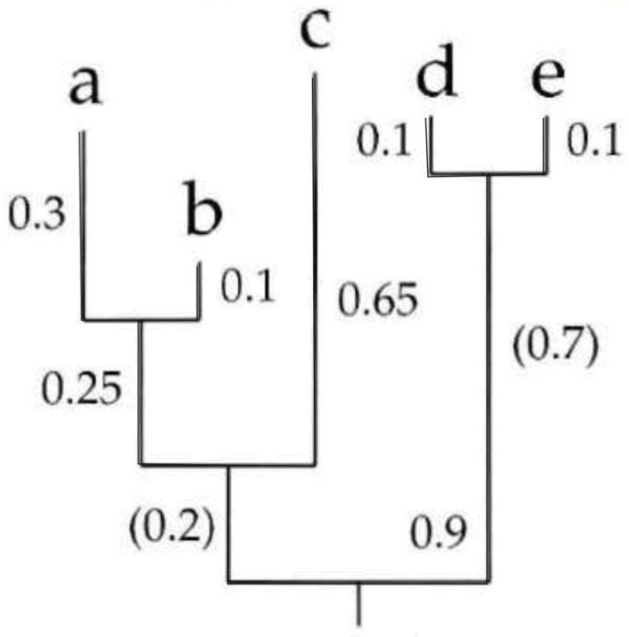
Phylogenetically independent contrasts

There are *n* – 1 independent contrasts for *n* species.



Phylogenetically independent contrasts

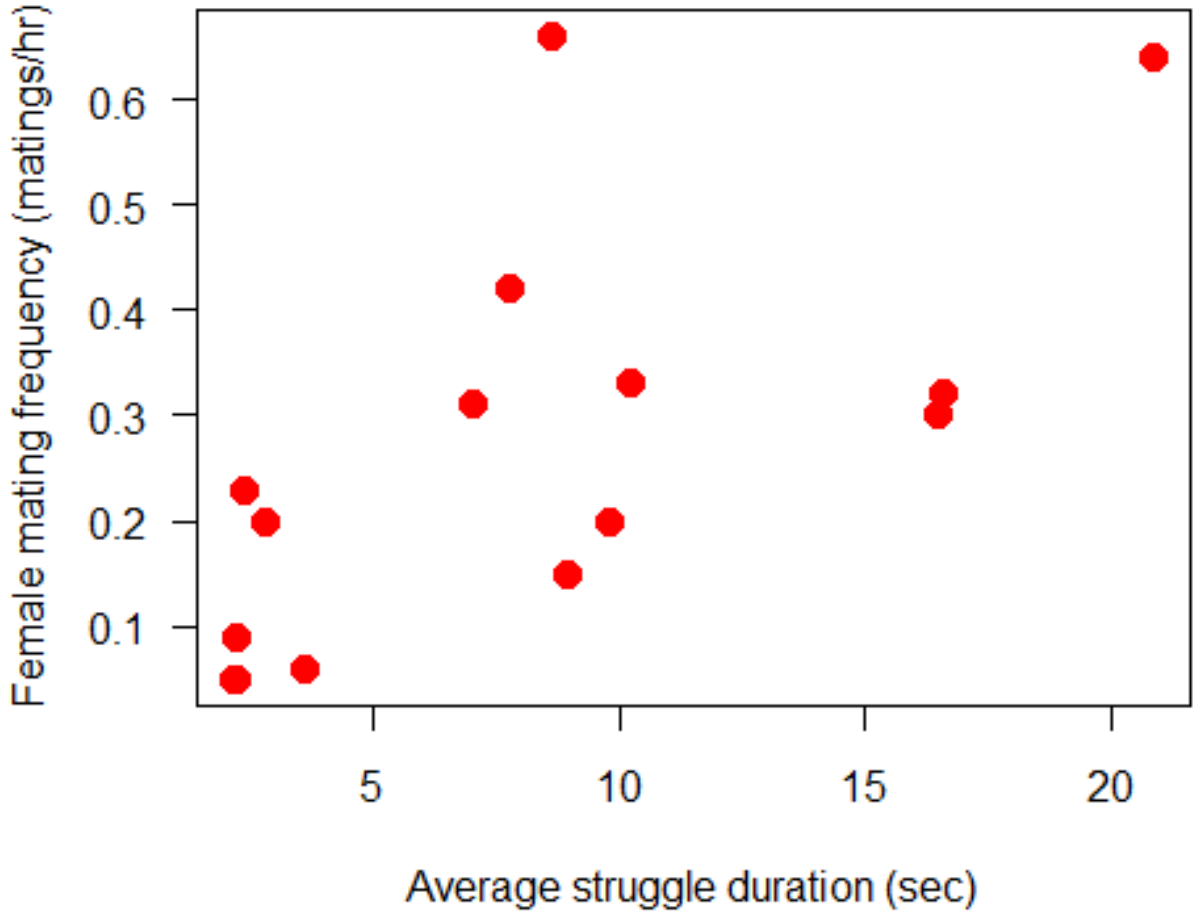
Calculation details. Usually, contrasts are standardized by the square root of the expected variance, which is proportional to branch length.



Contrast									Variance proportional to
y_1	=	x_a	−	x_b					0.4
y_2	=	$\frac{1}{4}\,x_a$	+	$\frac{3}{4}\,x_b$	−	x_c			0.975
y_3	=						x_d	− x_e	0.2
y_4	=	$\frac{1}{6}\,x_a$	+	$\frac{1}{2}\,x_b$	+	$\frac{1}{3}\,x_c$	− $\frac{1}{2}\,x_d$	− $\frac{1}{2}\,x_e$	1.11666

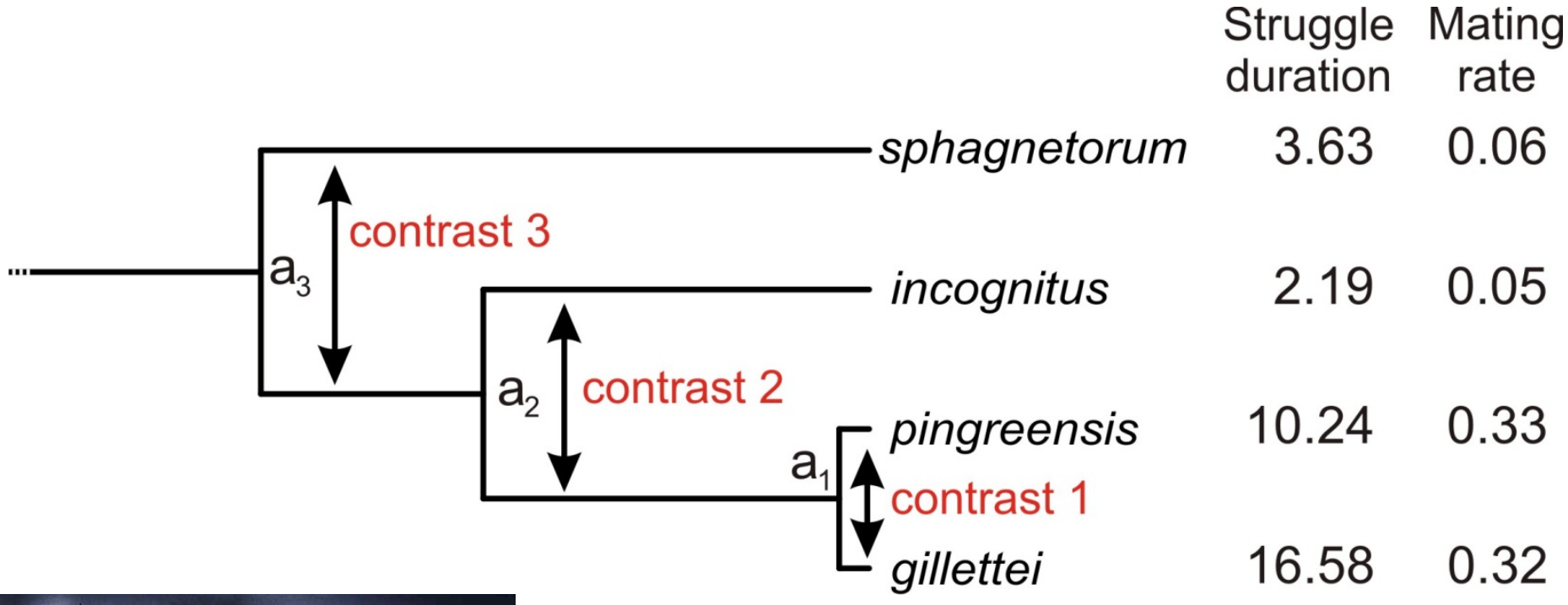
Phylogenetically independent contrasts

The idea is to convert the data on both traits to their independent contrasts using the phylogeny of the species. Then calculate the correlation between the independent contrasts of the two traits.



Phylogenetically independent contrasts

A cutaway of the independent contrasts for the water strider mating behavior data. The direction of each contrast is arbitrary, but the contrast direction must be the same for both variables.

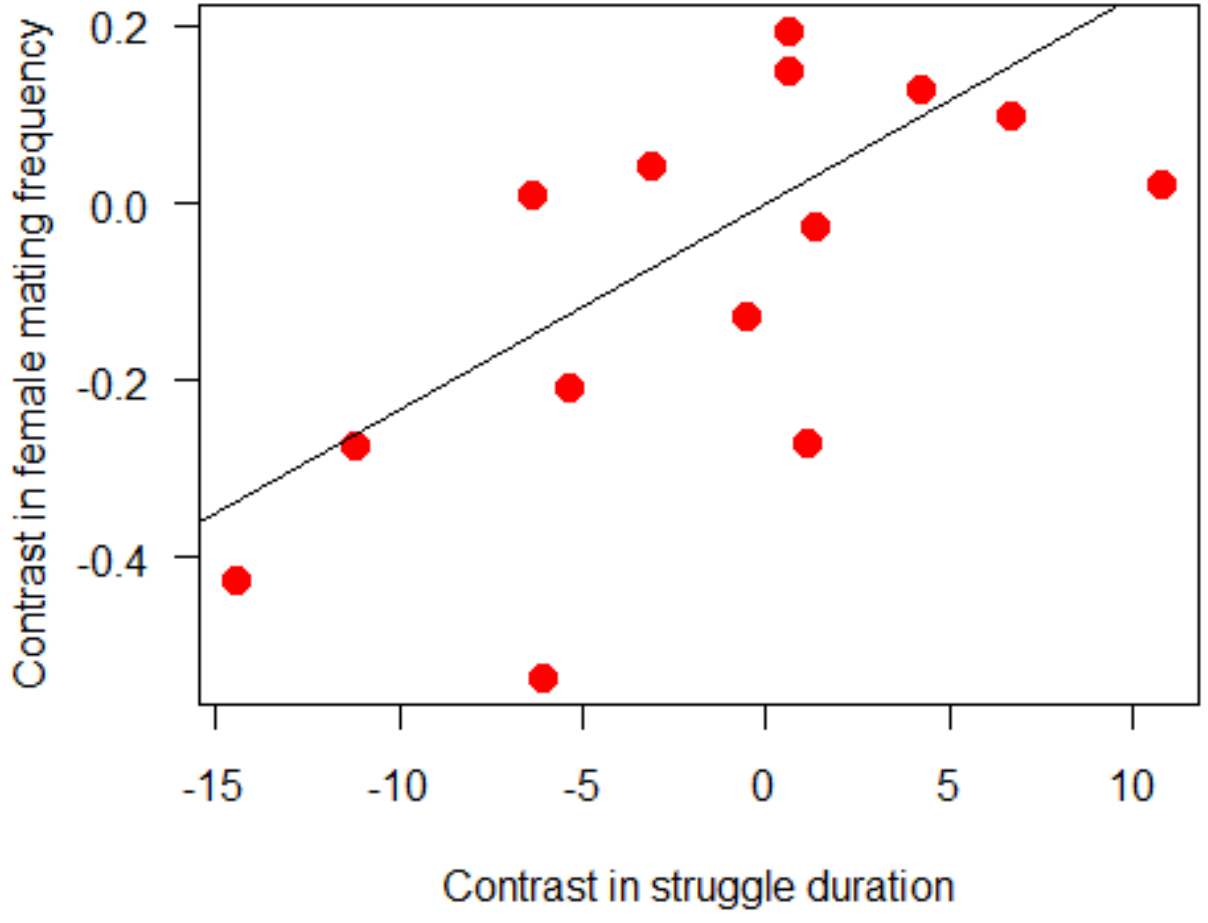


Phylogenetically independent contrasts

Because the direction of the contrast is arbitrary, the correlation or regression using independent contrasts is fitted through the origin (0,0).

The ape package in R implements phylogenetically independent contrasts.

Positive correlation confirmed!



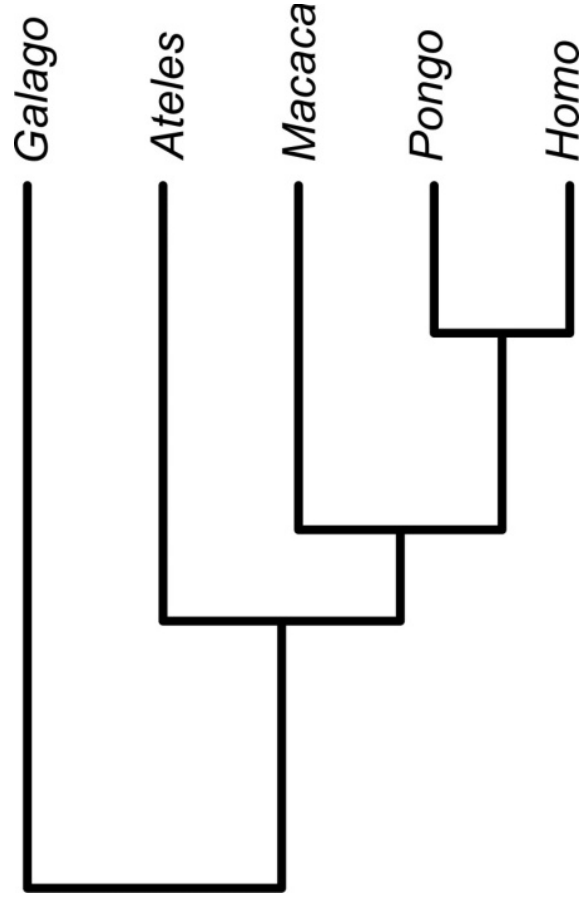
A linear model approach

General least squares (GLS) is a linear model technique mathematically equivalent to phylogenetically independent contrasts.

GLS allows the residuals to be correlated and have unequal variances. The method incorporates them using a “weight” matrix of expected covariances between species traits.

Using GLS gives access to all the tools of linear models, including model selection methods (AIC, etc).

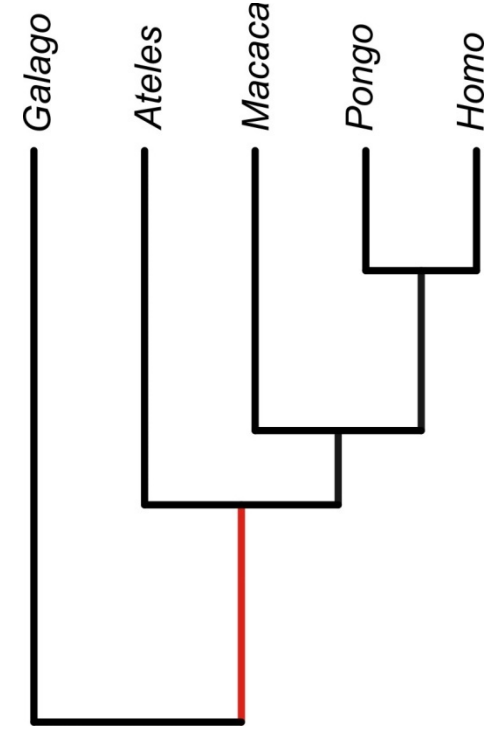
The function `gls()` in the `nlme` package can be used to fit phylogenetic linear models.



Specifying the covariance matrix between data points

	Homo	*Pongo*	*Macaca*	*Ateles*	*Galago*
Homo	1.00	0.79	0.51	0.38	0
Pongo	0.79	1.00	0.51	0.38	0
Macaca	0.51	0.51	1.00	0.38	0
Ateles	0.38	0.38	0.38	1.00	0
Galago	0.00	0.00	0.00	0.00	1

To analyze, we must know what the variances and correlations are between species. Under Brownian motion, the expected covariance between two species is the proportion of total history, from root to tip, that they share.

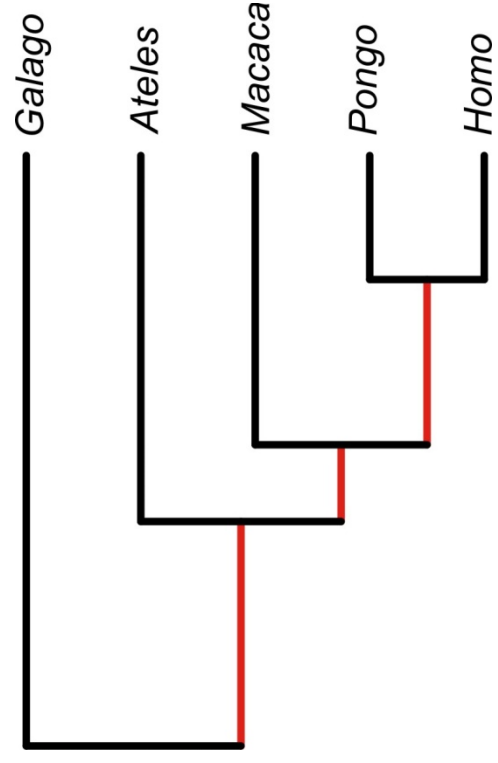


Specifying the covariance matrix between data points

	Homo	*Pongo*	*Macaca*	*Ateles*	*Galago*
Homo	1.00	0.79	0.51	0.38	0
Pongo	0.79	1.00	0.51	0.38	0
Macaca	0.51	0.51	1.00	0.38	0
Ateles	0.38	0.38	0.38	1.00	0
Galago	0.00	0.00	0.00	0.00	1

These expected covariances between pairs of data points (species) are used as "weights" in the linear model fitting.

A pair of data points (species) that share most of their phylogenetic history end up being down-weighted in the analysis. In effect, each of them is counted as only a fraction of a data point.



Assumptions of the method

- Evolution in each trait mimics a continuous random walk in time (Brownian motion).
- The rate of evolution is constant through time and along all branches of the phylogeny.
- Speciation and extinction are unrelated to trait values.

These assumptions are difficult to verify.

Branch lengths of phylogenies can be transformed to improve agreement with Brownian motion assumption.

If the assumptions are not met, then in extreme cases using independent contrasts might be worse than simply treating the species data as though they were independent (Harvey and Rambaut 2000).

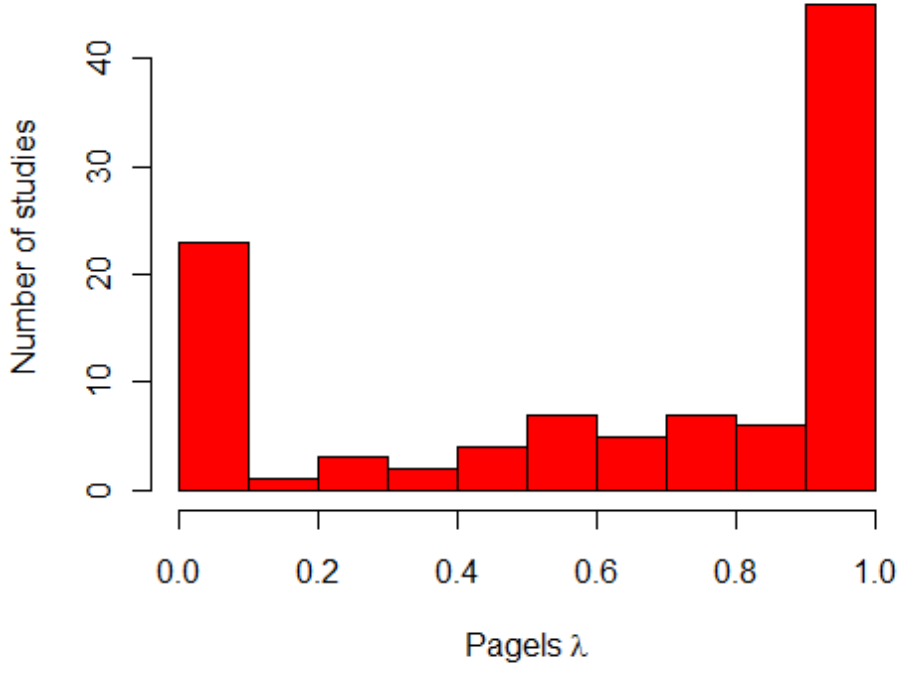
Assumptions of the method

The GLS linear model approach makes it easy to transform branch lengths of the tree to better meet the assumption of Brownian motion.

Under Brownian motion, Pagel's phylogenetic signal $\lambda = 1$.

If phylogenetic signal λ is less than one, each of the non-diagonal elements of the phylogenetic matrix can be multiplied by the estimated λ. This allows us to fit a model in which phylogenetic signal in the data is weaker than expected under simple Brownian motion.

The `ape` package in R can finds the "best" estimate of λ for a given data set using maximum likelihood.



Discrete species data

Patterson and Givnish (2002) found that lily species flowering in the low light environment of the forest understory, such as the blue bead lily (*Clintonia borealis*), tend to have small and inconspicuous flowers whitish or greenish in color.

Lilies that live in sunny, open habitats, or that live in deciduous woods but flower before the tree leaves come out, such as the Turk's-cap lily (*Lilium superbum*), tend to have large, showy flowers.

Discrete species data

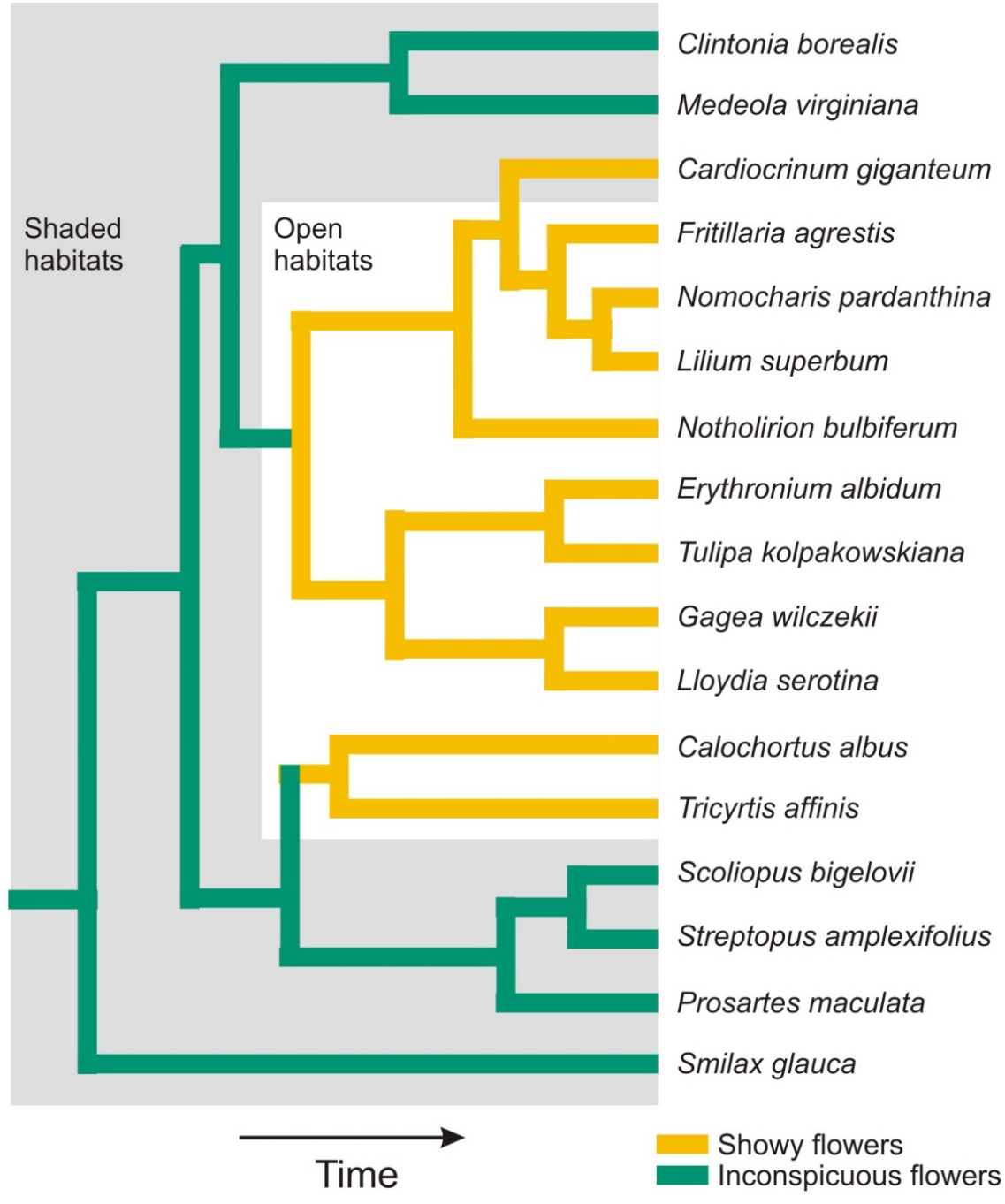
Data from 17 lily species indicated an almost perfect association between habitat and flower type. All ten species flowering in open habitats had large and showy flowers. Six of the seven species flowering in shaded habitats had relatively small and inconspicuous flowers. This seemed like a strong association.

	Open habitat	**Shaded habitat**
Showy flowers	10	0
Inconspicuous flowers	1	6

Discrete species data

But the phylogeny of the group reveals the same problem as in the water strider example: closely related species tend to be similar.

Even though there are 17 species, there might have been as few as three transitions between habitats in the past, leaving fewer effective data points than first assumed.



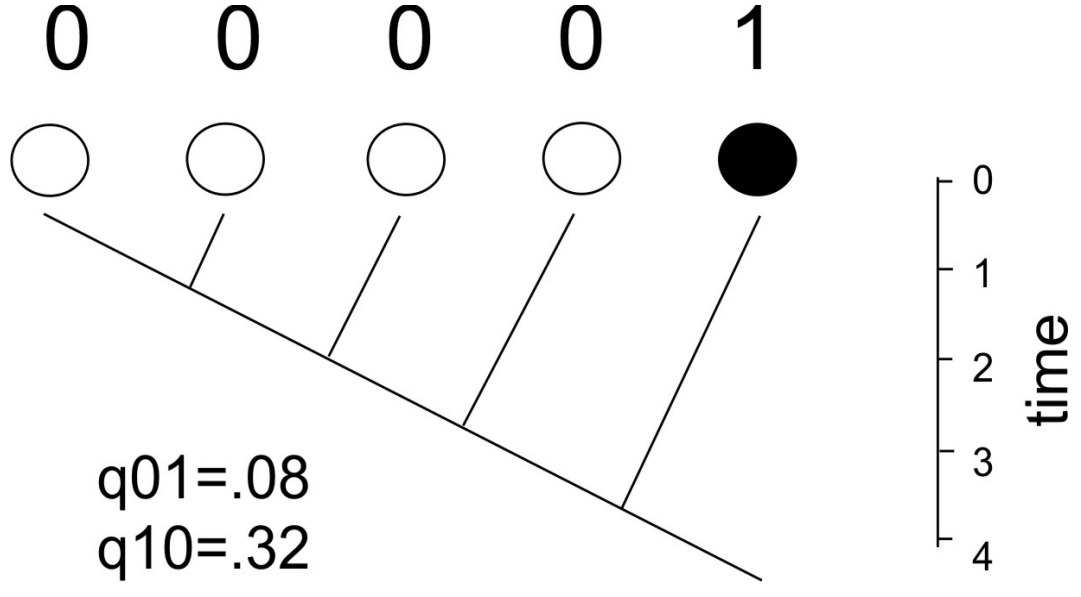
Discrete species data

Pagel (1994) developed a maximum likelihood method for analyzing discrete characters. The method assumes that evolution in each trait mimics a discrete random walk in time (Markov process).

It estimates the transition rates *q* between states through time on a phylogeny.

It uses likelihood to estimate and test how transitions between states in one trait (e.g., flower conspicuousness) depend on the character states of a second trait (e.g., habitat).

The method is implemented in the `corHMM` package in R.



Discrete species data

Maddison, W. and R. Fitzjohn. 2015. *The unsolved challenge to phylogenetic correlation tests for categorical characters*. Syst. Biol. 64:127–136.

"*… Pagel's test is susceptible to yielding significant results from the effects of a single change in one of the characters, …. Other tests suffer the same problem, which we will call "within-clade pseudoreplication*".

"…we suspect that any comparative method that responds to the effect of a state, rather than the effect of a change, will be susceptible to within-clade pseudoreplication."

No solutions yet…

Is phylogenetically independent contrasts/GLS also susceptible?

"...phylogenetically independent contrasts can be misled by a single extraordinary event..."

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell. 2018. *Rethinking phylogenetic comparative methods*. Syst. Biol 67: 1091-1109.

"…biologists likely have a favorite example of a lineage that has evolved something spectacular such as devilishly horned lizards that squirt blood from their eye sockets or marine sloths that grazed ancient seabeds. As macroevolutionary researchers, it is hard to know what to do with these types of events …. Their singular and unreplicated nature seems incompatible with models that we typically use to describe change over time, such as Brownian motion..."

Method development continues apace.

R: an embarrassment of riches

CRAN Task View: Phylogenetics, Especially Comparative Methods

Maintainer: Brian O'Meara
Contact: omeara.brian at gmail.com
Version: 2019-08-13
URL: https://CRAN.R-project.org/view=Phylogenetics

The history of life unfolds within a phylogenetic context. Comparative phylogenetic methods are statistical approaches for analyzing historical patterns along phylogenetic trees. This task view describes R packages that implement a variety of different comparative phylogenetic methods. This is an active research area and much of the information is subject to change. One thing to note is that many important packages are not on CRAN: either they were formerly on CRAN and were later archived (for example, if they failed to incorporate necessary changes as R is updated) or they are developed elsewhere and have not been put on CRAN yet. Such packages may be found on GitHub, R-Forge, or authors' websites.

Getting trees into R : Trees in R are usually stored in the S3 phylo class (implemented in ape), though the S4 phylo4 class (implemented in phylobase) is also available. ape can read trees from external files in newick format (sometimes popularly known as phylip format) or NEXUS format. It can also read trees input by hand as a newick string (i.e., "(human,(chimp,bonobo));"). phylobase and its lighter weight sibling rncl can use the Nexus Class Library to read NEXUS, Newick, and other tree formats. treebase can search for and load trees from the online tree repository TreeBASE, rdryad can pull data from the online data repository Dryad. RNeXML can read, write, and process metadata for the NeXML format. PHYLOCH can load trees from BEAST, MrBayes, and other phylogenetics programs (PHYLOCH is only available from the author's website). phyext2 can read and write various tree formats, including simmap formats. rotl can pull in a synthetic tree and individual study trees from the Open Tree of Life project. The treeio package can read trees in Newick, Nexus, New Hampshire eXtended format (NHX), jplace and Phylip formats and data output from BEAST, EPA, HyPhy, MrBayes, PAML,

PHYLDOG, pplacer, r8s, RAxML and RevBayes. phylogram can convert Newick files into dendrogram objects. brranching can fetch phylogenies from online repositories, including phylomatic .

Utility functions: These packages include functions for manipulating trees or associated data. ape has functions for randomly resolving polytomies, creating branch lengths, getting information about tree size or other properties, pulling in data from GenBank, and many more. phylobase has functions for traversing a tree (i.e., getting all descendants from a particular node specified by just two of its descendants). geiger can prune trees and data to an overlapping set of taxa. tidytree can convert a tree object in to a tidy data frame and has other tidy approaches to manipulate tree data. evobiR can do fuzzy matching of names (to allow some differences). rphast implements an R interface to the PHAST, which can be used for many types of analysis in comparative and evolutionary genomics, such as estimating models of evolution from sequence data, scoring alignments for conservation or acceleration, and predicting elements based on conservation or custom phylogenetic hidden Markov models. SigTree finds branches that are responsive to some treatment, while allowing correction for multiple comparisons. dendextend can manipulate dendrograms, including subdividing trees, adding leaves, and more. apex can handle multiple gene DNA alignments making their use and analysis for tree inference easier in ape and phangorn. aphid can weight sequences based on a phylogeny and can use hidden Markov models (HMMs) for a variety of purposes including multiple sequence alignment.

Ancestral state reconstruction : Continuous characters can be reconstructed using maximum likelihood, generalised least squares or independent contrasts in ape. Root ancestral character states under Brownian motion or Ornstein-Uhlenbeck models can be reconstructed in ouch, though ancestral states at the internal nodes are not. Discrete characters can be reconstructed using a variety of Markovian models that parameterize the transition rates among states using ape. markophylo can fit a broad set of discrete character types with models that can incorporate constrained substitution rates, rate partitioning across sites, branch-specific rates, sampling bias, and non-stationary root probabilities. phytools can do stochastic character mapping of traits on trees.

Diversification Analysis: Lineage through time plots can be done in ape. A simple birth-death model for when you have extant species only (sensu Nee et al. 1994) can be fitted in ape as can survival models and goodness-of-fit tests (as applied to testing of models of diversification). TESS can calculate the likelihood of a tree under a model with time-dependent diversification, including mass extinctions. Net rates of diversification (sensu Magellon and Sanderson) can be calculated in geiger. diversitree implements the BiSSE

method (Maddison et al. 1997) and later improvements (FitzJohn et al. 2009). TreePar estimates speciation and extinction rates with models where rates can change as a function of time (i.e., at mass extinction events) or as a function of the number of species. caper can do the macrocaic test to evaluate the effect of a a trait on diversity. apTreeshape also has tests for differential diversification (see description). iteRates can identify and visualize areas on a tree undergoing differential diversification. DDD can fit density dependent models as well as models with occasional escape from density-dependence. BAMMtools is an interface to the BAMM program to allow visualization of rate shifts, comparison of diversification models, and other functions. DDD implements maximum likelihood methods based on the diversity-dependent birth-death process to test whether speciation or extinction are diversity-dependent, as well as identifies key innovations and simulate a density-dependent process. PBD can calculate the likelihood of a tree under a protracted speciation model. phyloTop has functions for investigating tree shape, with special functions and datasets relating to trees of infectious diseases.

Divergence Times: Non-parametric rate smoothing (NPRS) and penalized likelihood can be implemented in ape. geiger can do congruification to stretch a source tree to match a specified standard tree. treedater implements various clock models, ways to assess confidence, and detecting outliers.

Phylogenetic Inference: UPGMA, neighbour joining, bio-nj and fast ME methods of phylogenetic reconstruction are all implemented in the package ape. phangorn can estimate trees using distance, parsimony, and likelihood. phyclust can cluster sequences. phytools can build trees using MRP supertree estimation and least squares. phylotools can build supermatrices for analyses in other software. pastis can use taxonomic information to make constraints for Bayesian tree searches. outbreaker can infer transmission trees for diseases, as well as other parameters of disease spread. For more information on importing sequence data, see the Genetics task view; pegas may also be of use.

Time series/Paleontology: Paleontological time series data can be analyzed using a likelihood-based framework for fitting and comparing models (using a model testing approach) of phyletic evolution (based on the random walk or stasis model) using paleoTS. strap can do stratigraphic analysis of phylogenetic trees.

Tree Simulations: Trees can be simulated using constant-rate birth-death with various constraints in TreeSim and a birth-death process in geiger. Random trees can be generated in ape by random splitting of edges (for non-parametric trees) or random clustering of tips (for coalescent trees). paleotree can simulate fossil deposition, sampling, and the tree arising from this as well as trees conditioned on observed fossil taxa. TESS can simulate trees with time-dependent speciation and/or extinction rates, including mass extinctions.

Trait evolution: Independent contrasts for continuous characters can be calculated using ape, picante, or caper (which also implements the brunch and crunch algorithms). Analyses of discrete trait evolution, including models of unequal rates or rates changing at a given instant of time, as well as Pagel's transformations, can be performed in geiger. corHMM can look for hidden rates in discrete traits as well as fit correlational models for two or three binary traits (similar to Pagel's old Discrete program) and complex models for multistate traits (similar to Pagel's old Multistate program). Brownian motion models can be fit in geiger, ape, and paleotree. ratematrix can fit univariate or multivariate Brownian motion models with one or more rate regimes. Deviations from Brownian motion can be investigated in geiger and OUwie. mvMORPH can fit Brownian motion, early burst, ACDC, OU, and shift models to univariate or multivariate data. Ornstein-Uhlenbeck (OU) models can be fitted in geiger, ape, ouch (with multiple means), and OUwie (with multiple means, rates, and attraction values). surface wraps ouch to infer shifts in the OU optimum; bayou also allows data-driven selection between different OU models. geiger fits only single-optimum models. Other continuous models, including Pagel's transforms and models with trends, can be fit with geiger. ANOVA's and MANOVA's in a phylogenetic context can also be implemented in geiger. Multiple-rate Brownian motion can be fit in RBrownie. Traditional GLS methods (sensu Grafen or Martins) can be implemented in ape, PHYLOGR, or caper. Phylogenetic autoregression (sensu Cheverud et al) and Phylogenetic autocorrelation (Moran's I) can be implemented in ape or--if you wish the significance test of Moran's I to be calculated via a randomization procedure--in adephylo. Correlation between traits using a GLMM can also be investigated using MCMCglmm. phylolm can fit phylogenetic linear regression and phylogenetic logistic regression models using a fast algorithm, making it suitable for large trees. brms can examine correlations between continuous and discrete traits, and can incorporate multiple measurements per species. phytools can also investigate rates of trait evolution and do stochastic character mapping. metafor can perform meta-analyses accounting for phylogenetic structure. pmc evaluates the model adequacy of several trait models (from geiger and ouch) using Monte Carlo approaches. phyreg implements the Grafen (1989) phyglogenetic regression. geomorph can do geometric morphometric analysis in a phylogenetic context. Disparity through time, and other disparity-related analyses, can be performed with dispRity. MPSEM can predict features of one species based on information from related species

using phylogenetic eigenvector maps. Rphylip wraps PHYLIP which can do independent contrasts, the threshold model, and more. convevol and windex can both test for convergent evolution on a phylogeny.

Trait Simulations : Continuous traits can be simulated using brownian motion in ouch, geiger, ape, picante, OUwie, and caper, the Hansen model (a form of the OU) in ouch and OUwie and a speciational model in geiger. Discrete traits can be simulated using a continuous time Markov model in geiger. phangorn can simulate DNA or amino acids. Both discrete and continuous traits can be simulated under models where rates change through time in geiger. phytools can simulate discrete characters using stochastic character mapping. phylolm can simulate continuous or binary traits along a tree.

Tree Manipulation : Branch length scaling using ACDC; Pagel's (1999) lambda, delta and kappa parameters; and the Ornstein-Uhlenbeck alpha parameter (for ultrametric trees only) are available in geiger. phytools also allows branch length scaling, as well as several tree transformations (adding tips, finding subtrees). Rooting, resolving polytomies, dropping of tips, setting of branch lengths including Grafen's method can all be done using ape. Extinct taxa can be pruned using geiger. phylobase offers numerous functions for querying and using trees (S4). Tree rearrangements (NNI and SPR) can be performed with phangorn. paleotree has functions for manipulating trees based on sampling issues that arise with fossil taxa as well as more universal transformations. dendextend can manipulate dendrograms, including subdividing trees, adding leaves, and more. enveomics.R can prune a tree to keep clade representatives.

Community/Microbial Ecology : picante, vegan, SYNCSA, phylotools, PCPS, caper, DAMOCLES integrate several tools for using phylogenetics with community ecology. HMPTrees and GUniFrac provide tools for comparing microbial communities. betapart allows computing pair-wise dissimilarities (distance matrices) and multiple-site dissimilarities, separating the turnover and nestedness-resultant components of taxonomic (incidence and abundance based), functional and phylogenetic beta diversity. adiv can calculate various indices of biodiversity including species, functional and phylogenetic diversity, as well as alpha, beta, and gamma diversities. entropart can measure and partition diversity based on Tsallis entropy as well as calculate alpha, beta, and gamma diversities. ecospat can also examine phylogenetic diversity. metacoder is an R package for handling large taxonomic data sets, like those generated from modern high-throughput sequencing, like metabarcoding.

Phyloclimatic Modeling : phyloclim integrates several new tools in this area.

Phylogeography / Biogeography : phyloland implements a model of space colonization mapped on a phylogeny, it aims at estimating limited dispersal and competitive exclusion in a statistical phylogeographic framework. jaatha can infer demographic parameters for two species with multiple individuals per species. diversitree implements the GeoSSE method for diversification analyses based on two areas. nodiv can compare sister species distributions at each node to detect major differences in distribution (Borregaard et al., 2014).

Species/Population Delimitation : adhoc can estimate an ad hoc distance threshold for a reference library of DNA barcodes.

Tree Plotting and Visualization: User trees can be plotted using ape, adephylo, phylobase, phytools, ouch, and dendextend; several of these have options for branch or taxon coloring based on some criterion (ancestral state, tree structure, etc.). paleoPhylo and paleotree are specialized for drawing paleobiological phylogenies. Trees can also be examined (zoomed) and viewed as correlograms using ape. Ancestral state reconstructions can be visualized along branches using ape and paleotree. phytools can project a tree into a morphospace. BAMMtools can visualize rate shifts calculated by BAMM on a tree. The popular R visualization package ggplot2 can be extended by ggtree to visualize phylogenies. Trees can also be to interactively explored (as dendrograms) using idendr0. phylocanvas is a widget for "htmlwidgets" that enables embedding of phylogenetic trees using the phylocanvas javascript library. ggmuller allows plotting a phylogeny along with frequency dynamics.

Tree Comparison: Tree-tree distances can be evaluated, and used in additional analyses, in distory and Rphylip. ape can compute tree-tree distances and also create a plot showing two trees with links between associated tips. kdetrees implements a non-parametric method for identifying potential outlying observations in a collection of phylogenetic trees, which could represent inference problems or processes such as horizontal gene transfer. dendextend can evaluate multiple measures comparing dendrograms.

Taxonomy: taxize can interact with a suite of web APIs for taxonomic tasks, such as verifying species names, getting taxonomic hierarchies, and verifying name spelling. evobiR contains functions for making a tree at higher taxonomic levels, downloading a taxonomy tree from NCBI or ITIS, and various other miscellaneous functions (simulations of character evolution, calculating D-statistics, etc.).

Gene tree - species tree: HyPhy can count the duplication and loss cost to reconcile a gene tree to a species tree. It can also sample histories of gene trees from within family trees. rmetasim can simulate loci and individuals across landscapes using the metasim simulation engine.

Interactions with other programs: geiger can call PATHd8 through its congruify function. ips wraps several tree inference and other programs, including MrBayes, Beast, and RAxML, allowing their easy use from within R. Rphylip wraps PHYLIP , a broad variety of programs for tree inference under parsimony, likelihood, and distance, bootstrapping, character evolution, and more. BoSSA can use information from various tools to place a query sequence into a reference tree. pastis can use taxonomic information to make constraints for MrBayes tree searches.

Notes: At least ten packages start as phy* in this domain, including two pairs of similarly named packages (phytools and phylotools, phylobase and phybase). This can easily lead to confusion, and future package authors are encouraged to consider such overlaps when naming packages. For clarification, phytools provides a wide array of functions, especially for comparative methods, and is maintained by Liam Revell; phylotools has functions for building supermatrices and is maintained by Jinlong Zhang. phylobase implements S4 classes for phylogenetic trees and associated data and is maintained by Francois Michonneau; phybase has tree utility functions and many functions for gene tree - species tree questions and is authored by Liang Liu, but no longer appears on CRAN.

Workshop on phylogenetic comparative methods

This Thursday!

Use R!

This course was an introduction to advanced methods in data analysis in ecology and evolution, how they work, and how you can avoid *some* of the most common misinterpretations and perils.

One or more of these methods will likely be useful to your future work, and you will want to review and dig further to understand it better.

The R tips web site and the workshops will remain online and available for the foreseeable future. I'll do my best to keep it up to date. Revisit and refresh your memories as needed.

Lots of people use R for data analysis here, so there is help all around. Start a data analysis group!

Bye!